



UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in
**Scienze Molecolari e Biotecnologie Agrarie, Alimentari e
Ambientali**

Dipartimento di Produzione Vegetale

Dottorato di Ricerca in
Biologia Vegetale e Produttività della Pianta Coltivata
XXIV Ciclo

Characterization of *Camelina sativa* genome, an
oilseed plant, through a combined approach based
on the β -tubulin gene family and microsatellite
molecular markers

Settore disciplinare: AGR/07

Dottorando: Antonella MANCA

Matricola: R08213

Relatori: Dott. Salvatore Roberto PILU
Dott.ssa Incoronata GALASSO
Dott. Diego BREVIARIO

Coordinatore: Prof. Daniele Bassi

Anno Accademico 2010/2011

TABLE OF CONTENT

Publications relevant to this thesis	iv
List of tables.....	v
List of figure	vi
Abstract.....	viii
1.Introduction	1
1.1. <i>Camelina sativa</i> (L.) Crantz, a novel oilseed plants for biodiesel production	2
1.2 <i>Camelina sativa</i> (L.) Crantz.....	4
1.2.1. Plant description.....	4
1.2.2. Seeding and nutrient requirements.....	6
1.2.3. Yield and seed composition	7
1.2.4. Genomic and genetic knowledge in <i>Camelina sativa</i>	8
1.3. Molecular markers.....	9
1.3.1. The TBP (Tubulin Based Polymorphism).....	9
1.3.2. The SSR (Simple Sequence Repeat).....	11
1.4.Objectives.....	14
1.5. Reference.....	15
2. Publication 1	18
2.1. Abstract.....	19
2.2. Introduction	20
2.3. Materials and methods	21
2.3.1. Plant material	21
2.3.2. h-TBP	21
2.3.3. DNA cloning, sequencing and Southern analysis	22
2.3.4. Sequence comparison and phylogenetic analysis.....	23

2.4. Results	23
2.4.1. h-TBP: the rationale	23
2.4.2. Gene structure of the β -tubulin family	27
2.4.3. Genomic organization of the β -tubulin family in camelina.....	30
2.4.4. Phylogenetic analysis	32
2.4.5. Intron sequence analysis	34
2.5. Discussion	34
2.6. Conclusions	37
2.7. Acknowledgments	37
2.8. Appendix	38
2.9. Reference	40
3. Publication 2	42
3.1. Introduction	43
3.2. Material and methods	44
3.2.1. Plant Material and DNA extraction.....	44
3.3. Results and discussion	45
3.4. References	47
4. Manuscript 1	48
4.1. Abstract	49
4.2. Introduction	50
4.3. Material ad methods	51
4.3.1. Plant materials and DNA extraction	51
4.3.2. Analysis of camelina SSR-enriched genomic library	54
4.3.3. Primer design and PCR amplification	54
4.3.4. Genetic data analysis.....	55
4.3.5. Seed quality traits.....	56
4.3.6. ANOVA analysis	57
4.4. Results	57
4.4.1. Genomic library and microsatellites characterization	57
4.4.2. Characteristic of the microsatellite loci and genotypic structure of camelina accessions	59
4.4.3. Genetic diversity and phylogenetic tree	62
4.4.4. Biochemical data.....	65
4.4.5. ANOVA analysis	67

4.5. Discussion.....	68
4.6. Acknowledgments	70
4.7. Reference.....	71

PUBLICATIONS RELEVANT TO THIS THESIS

Galasso I, Manca A, Braglia L, Martinelli T, Morello L, Breviario D (2011) h-TBP: an approach based on intron-length polymorphism for the rapid isolation and characterization of the multiple members of the β -tubulin gene family in *Camelina sativa* (L.) Crantz. *Mol Breeding* 28:635–645.

Manca A, Galasso I (2010) Development of simple sequence repeat (SSR) markers in *Camelina sativa* (L.) Crantz. *Minerva Biotec* 22 Suppl. 1 No. 2: 43-5.

LIST OF TABLES

Table 2.1. List of <i>CsTUBs</i> identified from <i>C. sativa</i> genomic DNA.....	25
---	----

2.8. Appendix:

Table 2.8.1. Percentage of identity of the intron 1 and intron 2 among some of the <i>C. sativa</i> and <i>A. thaliana</i> paralogous and orthologous β -tubulin genes.....	39
--	----

Table 4.1. List of 40 accessions of <i>Camelina sativa</i> used in this study with their accession number, accession name, country of origin and donor.....	53
--	----

Table 4.2. Characteristics of 17 microsatellite loci identified in <i>C. sativa</i> . Locus name, oligonucleotide primer sequences, repeat motif, annealing temperature (Ta), number (No) of alleles and allele size range in base pair (bp).....	58
---	----

Table 4.3. Summary of some genetic diversity parameters.....	60
---	----

Table 4.4. Genetic parameters in 40 genotypes of camelina with 15 microsatellite loci.....	61
---	----

Table 4.5. Genetic diversity indices in 40 genotypes of <i>C. sativa</i> with 15 microsatellite loci.....	63
--	----

Table 4.6. Biochemical parameters.....	66
---	----

Table 4.7. Summary of ANOVA analysis.....	67
--	----

LIST OF FIGURES

Figure 1.1. (A) Trend of the carbon dioxide CO ₂ emissions. (B) World CO ₂ emission by sector.....	2
Figure 1.2. <i>Camelina sativa</i> during flowering.....	5
Figure 1.3. (A) <i>Camelina sativa</i> inflorescence showing open flowers. (B) Dry siliques before harvesting. (C) Open siliques with seeds.....	6
Figure 1.4. Schematic representation of the gene structure of a plant β -tubulin gene.....	10
Figure 1.5. Example of three individuals showing microsatellites of different length.....	11
Figure 1.6. Schematic representation of the microsatellite enrichment method described by Fischer and Bachmann (1998).....	13
Figure 2.1. Polyacrylamide gel electrophoresis of PCR products amplified by h-TBP from genomic DNA.....	26
Figure 2.2. Intron–exon structure of <i>C. sativa</i> (a) and <i>A. thaliana</i> (b) β -tubulin genes.....	29
Figure 2.3. Southern blot analysis of genomic DNA of <i>C. sativa</i> (Cs) and <i>A. thaliana</i> (At).....	31
Figure 2.4. Neighbour-joining tree constructed using Treecon software.....	33

2.8. Appendix:

Figure 2.8.1. Multiple alignment of the *CsTUB* deduced amino acid sequences.....38

Figure 3.1. Some of the SSR markers isolated from *C .sativa* var. Calena.....46

Figure 3.2. Examples of 3 SSR markers (P3GA-C4, P456GA-C11 and P456GA-D11).....46

Figure 4.1. UPGMA dendrogram based on Nei's (1978) genetic distance matrix calculated on the SSR loci analysed.....64

ABSTRACT

The study reported in this thesis was performed in order to increase our knowledge on *Camelina sativa* genome and on the degree of the genetic polymorphism and relatedness present within a collection of camelina accessions. This aim has been pursued through the characterization of the β -tubulin gene family and the use of novel and specific SSR markers.

The β -tubulin gene family of *C. sativa* has been isolated, cloned and characterized using the h-TBP method. This technique allows the rapid cloning of the β -tubulin genomic sequences that encompass the two introns, invariantly present at fixed positions within the coding region of the vast majority of the plant species. We have found that in *C. sativa* this family was composed of at least 20 different β -tubulin isotypes, named *CsTUB1* through *CsTUB20*. This large number of β -tubulin is an indication that *C. sativa* (chromosome number $2n = 40$) might be a polyploid species. The phylogenetic tree obtained from the β -tubulin coding sequences of *C. sativa* and *A. thaliana* has shown a distribution of *CsTUBs* that is spread throughout the clusters without distinction between *A. thaliana* and *C. sativa*.

Then, we reported the isolation and characterization of specific camelina microsatellite markers by using one *C. sativa* (GA)-SSR-enriched library. All SSR markers utilised in this study produced clear and unambiguous amplification fragments permitting the detection of a large number of alleles per locus. A total of 134 alleles were generated at 15 SSR loci in the germplasm analysed with a mean of 8.93 alleles per locus. The high discriminatory capacity of the SSR markers, as observed in other plant species, have also been confirmed in our study, in fact in our camelina collection thirty-eight out of forty accessions are clearly distinguished by the 15 polymorphic microsatellites used. In addition, a certain degree of association among the SSR camelina sub-groups and some of the evaluated agronomic/biochemical traits has been observed.

In conclusion, knowledge of genetic variation and the genetic relationship between genotypes is important for an efficient utilisation of *C. sativa* germplasm resources. Beside of providing a useful tool for germplasm identification and genetic diversity, these 15 newly developed SSR polymorphic markers will prove very useful in genetic mapping and in assisting plant breeders in early progeny selection.

1. Introduction

1.1. *Camelina sativa* (L.) Crantz, a novel oilseed plants for biodiesel production

Since several years many national and international initiatives are underway in order to identify new sources of renewable energy. This growing interest in renewable energy is driven by two main reasons. Firstly, fossil fuels such as oil, coal and natural gas are not unlimited resources on our planet and if the level of our consume will not change the estimation time of depletion of fossil energy sources will be around 50 years for the oil, 70 year for natural gas and 170 years for coal (International Energy Outlook 2006).

Secondly, the combustion of fossil fuels emits large amounts of gas into the atmosphere, increasing the natural greenhouse effect. Carbon dioxide (CO₂) and methane (CH₄) are the main components of greenhouse gases (GHGs). Since 1870, annual CO₂ emission from fuel combustion dramatically increased from near zero to 29 Gt in 2007 and the transportation is second only to the electricity and heat in terms of CO₂ emissions (International Energy Agency 2009) (Fig. 1.1.a and b).

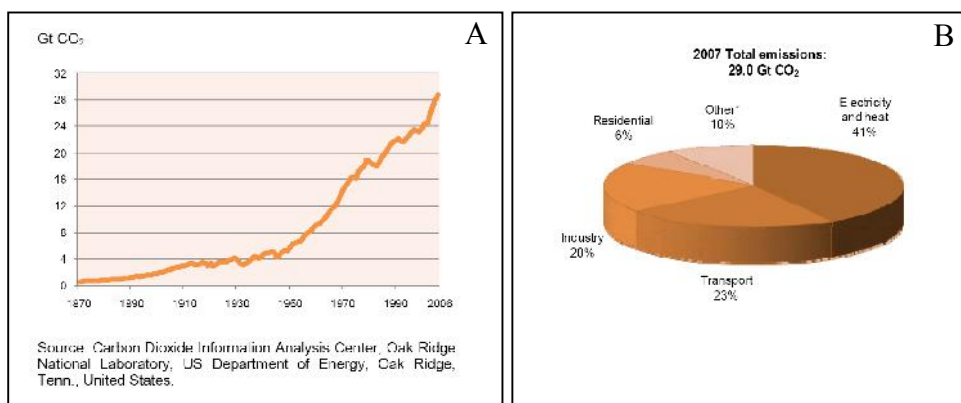


Figure 1.1. (a) Trend of the carbon dioxide CO₂ emissions. (b) World CO₂ emission by sector. Source from: International Energy Agency 2009.

Electricity and heat generation draws from various renewable energy sources (wind, solar, geothermic) while the transport sector relies almost entirely on petrol. Thus over the next 25 years, world demand for liquid fuels is expected to increase more rapidly in the transportation sector than in any other end-use sector. Therefore, the transportation sector has become the immediate target for

renewable energies. The European Commission encourages the use of biofuels in the transport sector. The last directive from the EC states that the Member States of the EU should gradually increase the consumption of biofuels in transport and by 2020, 10% of transport fuels should be replaced by biofuels (Directive 2009/28/EC).

In a context of growing interest for renewable energy sources the esterified vegetable oil i.e. biodiesel is proposed as one of the possible options to reduce greenhouse gas (GHG) emissions in the transportation sector.

The most known oilseed feedstocks for producing biofuel include rapeseed, sunflower (both mainly used in Europe), soybean (most commonly used in USA) and palm oils (used in the tropics).

However, due to their value in the food sector, crop plants yielding edible oilseeds can be used only to a limited extent. To overcome this problem, several scientific studies are underway in order to evaluate alternative or dedicated plants that do not come into competition with food crops and present attractive characteristics for biodiesel production such as: high oil content, reduced cultivation requirements, growth in marginal soils and compatibility with existing agricultural equipments.

One way to overcome the demand for oils and the need for alternative fuel oils is the use of non-edible oil seed plants. The most interesting alternative oilseed plants are: *Camelina sativa* (L.) Crantz, *Jatropha curcas* (L), *Pongamia pinnata* (L.) Pierre, *Brassica carinata* A. Braun and *Vernicia montana* Lour. (Becker et al. 1999; Chen et al. 2010; Azam et al. 2005). Among these non-food plants *Camelina sativa*, well suitable for temperate climates and poor soils, is attracting increased interest as a source of biodiesel due to a number of beneficial agronomic and economic attributes: good yield/ha, reduced fertiliser inputs, low demand for pesticides and economically useful by-products.

Camelina has also been chosen by the International Air Transport Association (IATA) as candidate for the production of biojet-fuel for aircraft (IATA Report 2010).

1.2 *Camelina sativa* (L.) Crantz

1.2.1. Plant description

Camelina sativa, also known as "false flax" or "gold of pleasure" [chromosome number $2n = 40$, genome size 750 Mbp, (Hutcheon et al. 2010), belongs to the Brassicaceae family. The common name "false flax" comes from the fact that often *C. sativa* was found as a weed in flax (*Linum usitatissimum* L.) fields. It is a plant native from northern Europe and Southeast Asia. Archaeological studies date its cultivation back in the Bronze Age (Bouby 1998).

Camelina has been grown in different parts of the world until the beginning of the twentieth century. It was used for both human consumption and several non-food applications such as the production of soap, cosmetics, oil for lamp and safe paints (Knorzer 1978; Zubr 1997).

Later on, its cultivation was abandoned due to low seed production compared to the amount of seed obtainable from other oil-crops such as rapeseed (*Brassica napus* L.) or sunflower (*Helianthus annuus* L.).

Recently, *C. sativa* has attracted renewed interest as an alternative oilseed plant for biofuel production due to some useful and distinguishable agronomic traits such as: very short growth cycle, resistance to drought and low temperatures, capacity to growth in marginal soils, lower demand of fertilizers, herbicides and pesticides (Putnam et al. 1993; Vollmann et al. 2007).

It is currently cultivated in Canada and US. In 2007, 10,000 hectares of the state of Montana (US) were sown with camelina (McVay and Lamb 2008). Sporadic few hectares of cultivated camelina have also been reported in the northeast Europe (Ireland, Germany, Austria, Poland and Slovenia).

The genus *Camelina* is composed of 11 species (Warwick and Al-Shehbaz 2006) but only five species: *C. sativa*, *C. microcarpa* Andr. ex DC, *C. rumelica* Velen., *C. alyssum* (Mill.) Thell. and *C. hispida* Boiss. are present in the Gene Banks of IPK (Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany) and USDA (USDA-ARS NCRPIS germplasm collection, USA).

Among them only *C. sativa* and *C. microcarpa* are cultivated. Within *C. sativa*, three different subspecies, ssp. *pilosa*, ssp. *sativa* and ssp. *foetida*, have been described (Angelini et al. 1997). From an agronomic point of view, *C. s.* ssp. *sativa* and *C. s.* ssp. *pilosa* seem to be the most promising subspecies. *Camelina sativa* ssp. *pilosa* is usually sown in autumn since it requires vernalisation in order to attain stem elongation and subsequent flowering, while *C. s.* ssp. *sativa* does not require vernalization and can be sown in both autumn and spring.

Camelina sativa has a short life cycle, less than 120 days (from sowing to harvest) when sown in spring. After sowing, the emergence takes place within 7-10 days, with the appearance of two cotyledon leaves. Immediately after the cotyledonary stage, true leaves begin to develop, forming a rosette (stage with 6-8 leaves). Later on, the rosette becomes the starting point of the stem with several alternate leaves. The plants usually reach 70 to 100 cm in high (Fig. 1.2).



Figure 1.2. *Camelina sativa* during flowering.

Such as most of the Brassicaceae, camelina also develops lateral branches. In *C. sativa*, the development of lateral branches is extremely variable and depends on the genotype, plant density in the field and also on the environmental conditions. Therefore, the number of branches can vary from zero (with high plant density) to 27-30 (with a low plant density and/or favourable climatic conditions) (Martinelli and Galasso 2011).

In general, the main inflorescence is composed of a number of flowers, ranging between 35 and 70, according to both the variety and environmental conditions.



Figure 1.3. (a) *Camelina sativa* inflorescence showing open flowers. (b) Dry siliques before harvesting. (c) Open siliques with seeds.

Camelina is a highly self-pollinated plant with small, pale and yellow flowers with 4 petals (Fig. 1.3.a). The fruits are called siliques and contain approximately 15-20 seeds of small size (0.7 mm x 1.5 mm) (Fig. 1.3.b and 1.3.c). The weight of 1000 seeds is from 0.8 to 1.8 grams (Zubr 1997).

1.2.2. Seeding and nutrient requirements

Soil preparation is one of the most sensitive aspects of the camelina growing technique. The preparation of a good seedbed is a necessary condition to obtain uniform germination from very small seeds such as that of camelina. Therefore, after sowing at least one harrowing and one rolling are necessary in order to obtain a good seed-soil contact. Given the small size of the seed, the sowing depth should not exceed 2-3 cm with a row spacing of 13-15 cm. About 5-6 kg/ha of seed are required to obtain a population of 450-500 plants per m².

Several studies have shown that fertilization and in particular the nitrogen requirement is relatively low in camelina compared to other oil crops (Putnam et al. 1993; Budin et al. 1995; Zubr 1997). Generally, camelina requires from 80 to 100 kg N/ha to obtain the maximum seed yield (Grant 2008) Recently (Losak et al. 2010), have shown that the increase in nitrogen uptake while enhancing the number of lateral branches per plant, the amount of seed per hectare and the weight of 1000 seeds, does not increase oil content. This negative correlation between the oil and protein content present in seeds, has also been reported by Agegnehu and Honermeier (1997) and Urbaniak et al. (2008).

1.2.3. Yield and seed composition

According to several reports camelina seed yield remains variable. French et al. (2009) reported a seed yield of about 1000 kg/ha in Arizona, USA, while Moser (2010) and Vollmann et al. (2007) reported a seed production between 0.336 t/ha to 2.24 t/ha and from 1.57 t/ha to 2.25 t/ha, respectively. In Montana, where the cultivation of camelina is particularly extensive, yields of about 1.7 t/ha to 2.4 t/ha were recorded (McVay and Lamb 2008). Recently, agronomic trials repeated for 2 consecutive years in Lombardia (Italy) reported an average quantity of 2.1 t/ha and 2.4 t/ha of seeds resulting from spring or autumn sowing, respectively (Galasso et al., in preparation). The observed high variability in seed production is not only attributable to environmental and growing conditions, but also to the lack of selection and genetic improvement for this plant. Few varieties of *C. sativa* are actually cultivated and programs of genetic improvement, aimed at stabilizing the yields per hectare and other features of interest, have just begun.

The oil content in dry weight seeds ranges between 35 and 40% and consists of approximately 54% polyunsaturated fatty acids, 34% monounsaturated and 12% saturated (Putnam et al. 1993). The most abundant polyunsaturated fatty acid is the linolenic acid (C18:3), but significant amount of linoleic acid (C18:2) is also present. Among the monounsaturated fatty acids fraction oleic acid (C18:1) and gondoic acid (C20:1) are the most abundant while minimal amounts of erucic acid (C22:1) are present, unlike to what found in other cruciferous (Putnam et al. 1993).

Thanks to its oil composition, camelina acts as a valuable renewable resource not only for biodiesel production but also for the production of hydraulic oil and lubricants. Moreover, since the oil of camelina is rich in linolenic acid, omega-3, but low in erucic acid, it can also be used for potential applications in the nutraceutical and cosmetics industry (Huber et al. 2006). In addition, the meal that is obtained after mechanical extraction of the oil, must be considered a by-product of high additional value. Meal is composed of a 45% of protein, 13% fiber, 5% minerals in addition to 10% of oil residue and other minor compounds such as vitamins (Acamovic et al. 1999). Since the meal is an excellent source of protein, it can be used in the preparation of feeds for livestock. In fact, the protein component is made up of amino acids such as glutamine, asparagines, arginine, leucine, glycine, valine, serine, lysine, proline with a value comparable to that of rapeseed and soy protein fraction. However, the meal contains also some anti-nutritional compounds such as glucosinolates, tannins, trypsin inhibitors, phytates

and sinapines (Matthäus 1997; Matthäus and Angelini 2005; Matthäus and Zubr 2000; Schuster and Friedt 1998).

In particular the presence of glucosinolates has prevented the utilization of the meal for livestock feed until a few years ago. The concentration of glucosinolates in camelina seed cake varies from 15 to 24 $\mu\text{mol/g}$ (Matthäus 1997). Since these values are not excessively high, in 2009 the U.S. Food and Drug Administration (FDA) has approved the use of 10% of camelina meal from crushed seeds in the daily rations of cattle (Schill 2009). The possibility of using by-products of camelina for animal feed is very important because it turns to positive the overall energy balance, that is the ratio between the energy used in the production process and the energy supplied by biofuel.

1.2.4. Genomic and genetic knowledge in *Camelina sativa*

In these last years, as reported above, many studies at agronomic and biochemical level have been carried out on *C. sativa* (Gugel et al. 2006; Martinelli and Galasso 2011; Vollmann et al. 1996; Vollmann et al. 2007).

In contrast, the number of studies at genetic and genomic level is still limited. In fact the genome of *C. sativa*, is still largely unexplored, and to date there are very few DNA sequences available in the EMBL database (48 DNA sequences for *C. sativa* and 45 for *C. microcarpa*). Another much debated issue concerns the diploid nature of this plant. (Hutcheon et al. 2010), studying two genes in the fatty acid biosynthesis pathway, *fatty acid desaturase (FAD) 2* and *fatty acid elongase (FAE) 1*, argue that *C. sativa* has tripled its genome over the centuries. Hence, they suggest to consider *C. sativa* as an allohexaploid. A polyploid genome has also suggested by Gehringer et al. (2006) in studies performed with the use of AFLP and SSR molecular markers. Also, the knowledge about germplasm diversity and genetic relationships among different camelina accessions is very low. At the moment, only two papers have analysed, the genetic variability occurring among different camelina genotypes, making use of either RAPD (Random Amplified Polymorphic DNA) or AFLP (Amplified fragment length polymorphism) molecular markers. Vollmann et al. (2005), using 24 RAPD primers on 41 camelina accessions, found that only 15 (63%) of them yielded polymorphic bands, producing only one to three loci per marker (a total of 30 marker loci). This low number of loci per marker suggested a low degree of genetic diversity. Ghamkhar et al. (2010) used 8 AFLP markers to study the level of genetic diversity, within and among 53 camelina germplasm lines collected

from different geographic locations. They demonstrated a positive correlation between oil quality traits and the ecogeography collection site. Although sampling was relatively biased towards the Russian-Ukrainian area, this region was suggested by the authors as a genetic diversity hotspot and the possible centre of origin for camelina.

Knowledge about germplasm diversity is an invaluable aid in crop improvement strategies. However, this requires the identification of molecular markers able to characterise the different accessions, to distinguish them, to put them in connection with each other and with the agronomic characteristics of growth and productivity.

In this regard, it is important to identify a number of significant and diverse molecular markers that are able to characterize the available accessions of *C. sativa*, to analyse the genetic variation among them and to evaluate their usefulness in plant breeding programs. To date, very few analyses have been performed to explore the genetic diversity present within the camelina germplasm. This hampers further improvement in the line of making *C. sativa* a very profitable plant.

1.3. Molecular markers

Several types of molecular markers have been described in the literature [see review Semagn et al. (2006)]. Their use has greatly improved our understanding in genetic diversity, with reference to populations or groups of individuals in crop plants. Knowledge of the genetic variation present in different crop accessions is not only important for breeding programs but also for evolution studies and for conservation purposes.

This thesis reports data concerning the genetic variability found in one collection of *C. sativa* by the use of two codominant molecular markers: TBP (Tubulin Based Polymorphism) and SSR (Simple Sequence Repeat), both based on the polymerase chain reaction (PCR).

1.3.1. The TBP (Tubulin Based Polymorphism)

The TBP marker belongs to the intron length polymorphism (ILP) markers. ILPs are sequence-tagged-site (STS) markers targeted to specific gene loci (Einax and Voigt 2003; Wang et al. 2005). Genes encoding for β -tubulins are present in all eukaryotic organisms. In any given eukaryote, with the exception of few

unicellular organisms like yeast, β -tubulin genes give rise to a small gene-family. In plants, a β -tubulin gene family can account for 4 to 23 different members depending on the species. In almost all of the plant species each β -tubulin gene is characterized by the presence of three coding exons, with a highly conserved nucleotide sequence, interrupted by two introns, highly different in both length and DNA sequence, that are invariantly located at fixed positions (Liaud et al. 1992).

TBP marker results from the selective amplification, done with PCR, of all the introns present in the β -tubulin gene family. This is achieved by using degenerated primer mixtures that target the exon sequences at the boundaries of the introns (Bardini et al. 2004; Breviario et al. 2007) (Fig. 1.4). After PCR, the amplified products are resolved on a sequencing polyacrylamide gel giving rise to a clear and consistent fingerprinting pattern, specific for each plant species.

TBP has been successfully used for fingerprinting the genomes of several plant species including crops and wild species for which no genomic information was available (Braglia et al. 2010).

β -Tubulin gene

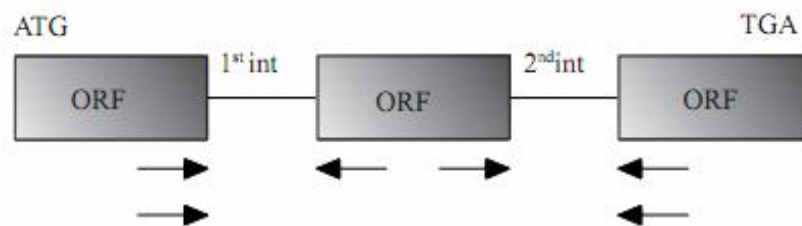


Figure 1.4. Schematic representation of the gene structure of a plant β -tubulin gene. Arrows indicate the primer positions.

Beside its usefulness as molecular marker, TBP also represents a simple and efficient method for the rapid isolation, of the complete β -tubulin gene family from any plant species.

This is done with a variant of the method by which the genomic region that encompassed both introns 1 and 2 of each β -tubulin gene is firstly amplified and

then cloned and sequenced (Fig. 1.4). This will be demonstrated in the first paper of this thesis, reporting the characterization of the *C. sativa* β -tubulin gene family.

1.3.2. The SSR (Simple Sequence Repeat)

SSR markers, commonly known as microsatellites (Tautz 1989; Kalia et al. 2011; Morgante and Olivieri 1993), are tandem repeated motifs of 1-6 base pairs (bp) widespread in all eukaryotes and prokaryotes genomes (Zane et al. 2002). During recent decades, microsatellites have become the most popular source of genetic markers to assess genetic diversity in several plant species due to their high reproducibility, multi-allelic nature, co-dominant inheritance and wide genome coverage. It is estimated that there is, on average, a SSR region every 50-750 kb of DNA, depending on the type of repeating pattern.

SSRs are present in the nuclear, in the plastid and mitochondrial genome (Chung et al. 2006; Rajendrakumar et al. 2007), and are distributed in both coding and noncoding regions. The variability of these repeats, might in some cases, act as a factor of gene regulation. In a work of Bao et al. (2002), it was demonstrated that the amylose content in rice is correlated with the change in the number of repetitions GA or CT in the 5' UTR of the waxy gene. Polymorphisms are highlighted by single specific primers designed on the sequences flanking the SSRs, which are generally highly conserved. The differences do not affect the repeating pattern but the number of times that this repeat is present in the amplified DNA, which results in a variation in length (Fig. 1.5). The products obtained by PCR amplification can be detected either by electrophoresis on polyacrylamide gel or by capillary electrophoresis.

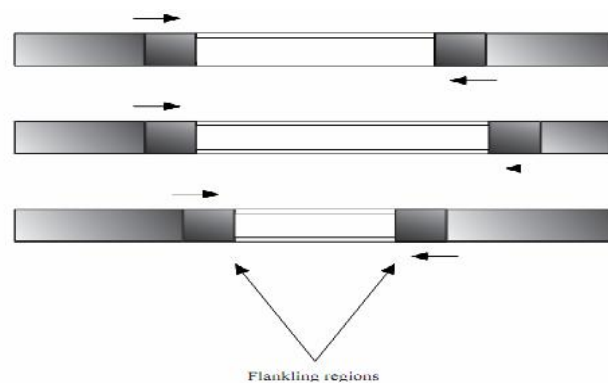


Figure 1.5. Example of three individuals showing microsatellites of different length. Arrows indicate the flanking regions used for primer design.

Microsatellite can be developed using both coding and not coding DNA sequences present in public sequence EMBL databases or by the construction and screening of genomic DNA libraries. Screening of the public database using DNA sequences isolated from the species of interest or from related species provides a rapid, simple and low cost source of microsatellite markers. However, in the case of poorly studied species where sequence information are not available cloning of microsatellite sequences is the alternative.

This is done through the construction of genomic libraries, which are then screened with probes for repeated sequences. Sequencing of positive clones and the design of primers complementary to flanking regions follow up.

However, this procedure is quite time-consuming, laborious and also very expensive. An alternative to identify and isolate microsatellite sequences in short time is to construct a genomic library enriched in microsatellites using the method based on magnetic beads selection (Fig.1.6), described by Fischer and Bachmann (1998). This method has been successfully applied to construct microsatellite-enriched genomic libraries in many plant species with high yields of clones containing microsatellites (Jones et al. 2001; Nunome et al. 2009; Phumichai et al. 2010).

In accordance, the second paper of this thesis reports the construction of the first microsatellite enriched genomic library of *C. sativa*, while the application of the developed SSR markers on one collection of camelina is reported in the third manuscript.

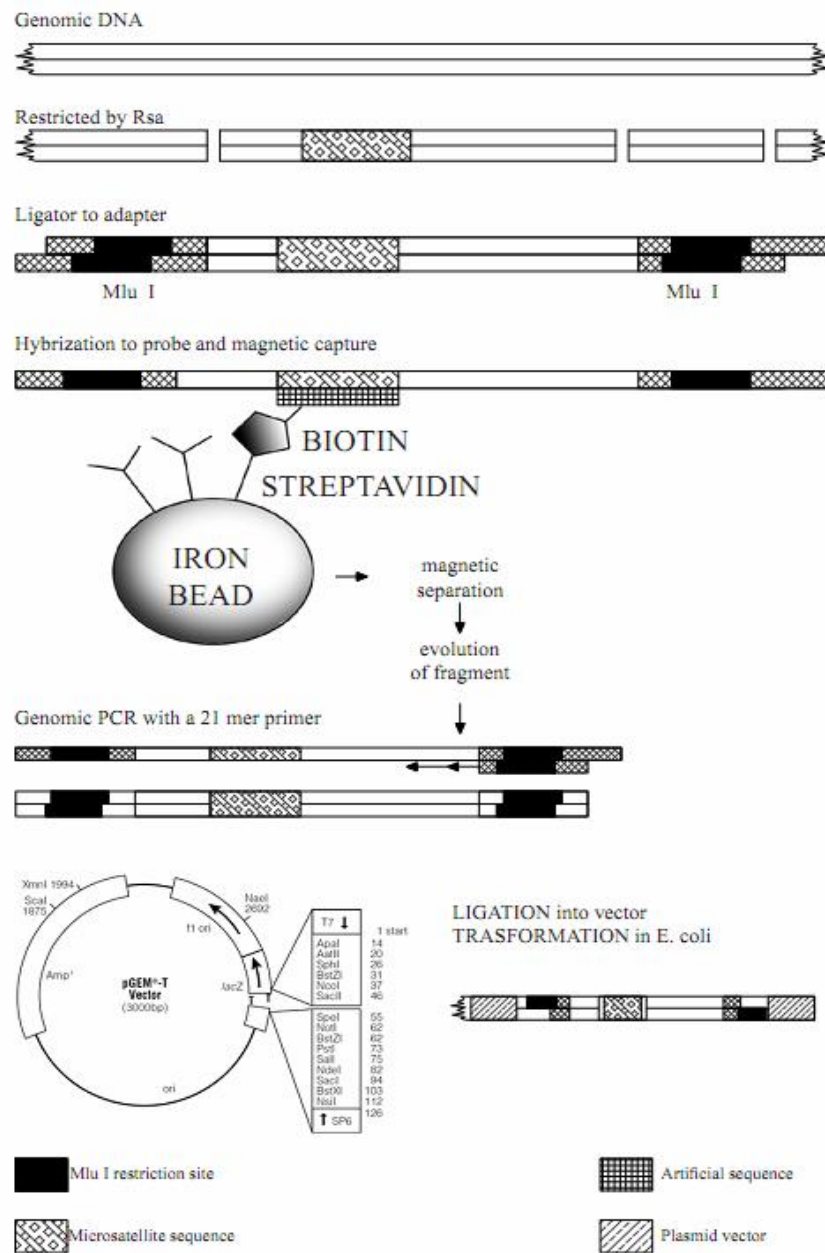


Figure 1.6. Schematic representation of the microsatellite enrichment method described by Fischer and Bachmann (1998), with some modifications.

1.4.Objectives

The studies reported in this thesis were performed in order to increase our knowledge on *C. sativa* genome, on the degree of the genetic polymorphism and relatedness present within a collection of camelina accessions. This aim has been pursued through the characterization of the β -tubulin gene family and the use of novel and specific SSR markers.

The different objectives can be better summarized as follows:

- A. Characterization of the genomic organization of the β -tubulin gene family.
- B. Construction of a microsatellite-enriched genomic library and development of novel SSR markers.
- C. Assessment of the genetic variability in a collection of 40 *C. sativa* accessions using the newly developed SSR markers
- D. Correlation of the genetic analysis with some biochemical and agronomic traits.

1.5. Reference

- Acamovic T, Gilbert C, Lamb K, Walker KC** (1999) Nutritive value of *Camelina sativa* meal for poultry. *British Poultry Science* **40**: 27-27.
- Agegnehu M, Honermeier B** (1997) Effects of seeding rates and nitrogen fertilization on seed yield, seed quality and yield components of false flax (*Camelina sativa* Crtz). *Bodenkultur* **48**: 15-21.
- Angelini LG, Moscheni E, Colonna G, Belloni P, Bonari E** (1997) Variation in agronomic characteristics and seed oil composition of new oilseed crops in central Italy. *Industrial Crops and Products* **6**: 313-323.
- Azam M, Waris A, Nahar NM** (2005) Prospects and potential of fatty acid methyl esters of some non-traditional seed oils for use as biodiesel in India. *Biomass and Bioenergy* **29**: 293-302.
- Bao JB, Corke HC, Sun MS** (2002) Microsatellites in starch-synthesizing genes in relation to starch physicochemical properties in waxy rice (*Oryza sativa*; L.). *TAG Theoretical and Applied Genetics* **105**: 898-905.
- Bardini M, Lee D, Donini P, Mariani A, Giani S, Toschi M, Lowe C, Breviario D** (2004) Tubulin-based polymorphism (TBP): a new tool, based on functionally relevant sequences, to assess genetic diversity in plant species. *Genome* **47**: 281-291.
- Becker HC, Löptien H, Röbbelen G, Gómez-Campo C** (1999) 13 Breeding: An overview. In: *Developments in Plant Genetics and Breeding*. Elsevier, pp 413-460.
- Bouby L** (1998) Two early finds of gold-of-pleasure (*Camelina* sp.) in middle Neolithic and Chalcolithic sites in western France. *Antiquity* **72**: 391-398.
- Braglia L, Manca A, Mastromauro F, Breviario D** (2010) cTBP: A Successful Intron Length Polymorphism (ILP)-Based Genotyping Method Targeted to Well Defined Experimental Needs. *Diversity* **2**: 572-585.
- Breviario D, Baird WV, Sangoi S, Hilu K, Blumetti P, Giani S** (2007) High polymorphism and resolution in targeted fingerprinting with combined β -tubulin introns. *Molecular Breeding* **20**: 249-259.
- Budin JT, Breene WM, Putnam DH** (1995) Some Compositional Properties of *Camelina* (*Camelina sativa* L Crantz) Seeds and Oils. *Journal of the American Oil Chemists Society* **72**: 309-315.
- Chen YH, Chen JH, Chang CY, Chang CC** (2010) Biodiesel production from tung (*Vernicia montana*) oil and its blending properties in different fatty acid compositions. In: *Bioresource Technology*. pp 9521-9526.
- Chung SM, Staub JE, Chen JF** (2006) Molecular phylogeny of *Cucumis* species as revealed by consensus chloroplast SSR marker length and sequence variation. *Genome* **49**: 219-229.
- Directive 2009/28/EC** of the European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC. (http://eurlex.europa.eu/JOIndex.do?year=2009&serie=L&textfield2=140&Submit=Cercare&_submit=Cercare&ihmlang=it).
- Einax E, Voigt K** (2003) Oligonucleotide primers for the universal amplification of β -tubulin genes facilitate phylogenetic analyses in the regnum Fungi. *Organisms Diversity & Evolution* **3**: 185-194.
- Fischer D, Bachmann K** (1998) Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.). *BioTechniques* **24**: 796-800, 802.

- French AN, Hunsaker D, Thorp K, Clarke T** (2009). Evapotranspiration over a camelina crop at Maricopa. Arizona. *Ind. Crops Prod.* **29**: 289-300.
- Grant DJ** (2008) Response of camelina to nitrogen, phosphorus, and sulfur. Fertilizer facts. Montana State University, Ext. Service, Western Triangle. Ag. Research Center, Conrad. <http://landresources.montana.edu/fertilizerfacts/pdf/FF%2049.pdf>
- Gehringer A, Friedt W, Luehs W, Snowdon RJ** (2006) Genetic mapping of agronomic traits in false flax (*Camelina sativa* subsp *sativa*). *Genome* **49**: 1555-1563.
- Ghamkhar K, Croser J, Aryamanesh N, Campbell M, Kon'kova N, Francis C** (2010) Camelina (*Camelina sativa* L. Crantz) as an alternative oilseed: molecular and ecogeographic analyses. *Genome* **53**: 558-567.
- Gugel, RK, Falk, KC** (2006) Agronomic and seed quality evaluation of *Camelina sativa* in western Canada. Agricultural Institute of Canada, Ottawa, ON, CANADA, pp 1047-1058.
- Huber GW, Iborra S, Corma A** (2006) Synthesis of transportation fuels from biomass: Chemistry, catalysts, and engineering. *Chemical Reviews* **106**: 4044-4098.
- Hutcheon C, Ditt RF, Beilstein M, Comai L, Schroeder J, Goldstein E, Shewmaker CK, Nguyen T, De Rocher J, Kiser J** (2010) Polyploid genome of *Camelina sativa* revealed by isolation of fatty acid synthesis genes. *Bmc Plant Biology*. doi: 10.1186/1471-2229-10-233
- IATA Report (2010)** "Report on Alternative Fuels" 5th Edition Ref. No: 9709-03 ISBN 978-92-9233-491-8 Montreal Geneva, <http://www.iata.org/ps/publications/pages/alternative-fuels.aspx>.
- Jones ES, Dupal MP, Kölliker R, Drayton MC, Forster JW** (2001) Development and characterisation of simple sequence repeat (SSR) markers for perennial ryegrass (*Lolium perenne* L.). *TAG Theoretical and Applied Genetics* **102**: 405-415.
- Kalia R, Rai M, Kalia S, Singh R, Dhawan A** (2011) Microsatellite markers: an overview of the recent progress in plants. *Euphytica* **177**: 309-334.
- Knorzer KH** (1978) Evolution and Spreading of Gold of Pleasure (*Camelina sativa*). *Berichte Der Deutschen Botanischen Gesellschaft* **91**: 187-195.
- Liaud MF, Brinkmann H, Cerff R** (1992) The β -Tubulin Gene Family of Pea - Primary Structures, Genomic Organization and Intron-Dependent Evolution of Genes. *Plant Molecular Biology* **18**: 639-651.
- Losak, T, Vollmann, J, Hlusek, J, Peterka, J, Filipeik, R, Praskova, L** (2010) Influence of combined nitrogen and sulphur fertilization on False flax (*Camelina sativa* L. Crtz.). *Acta Alimentaria* **39**: 431-444.
- Martinelli T, Galasso I** (2011) Phenological growth stages of *Camelina sativa* according to the extended BBCH scale. *Annals of Applied Biology* **158**: 87-94.
- Matthäus B** (1997) Antinutritive Compounds in Different Oilseeds. *Lipid / Fett* **99**: 170-174.
- Matthäus B, Angelini LG** (2005) Anti-nutritive constituents in oilseed crops from Italy. *Industrial Crops and Products* **21**: 89-99.
- Matthäus B, Zubr J** (2000) Variability of specific components in *Camelina sativa* oilseed cakes. *Industrial Crops and Products* **12**: 9-18.
- McVay K, Lamb P** (2008) Camelina Production in Montana. <http://www.msuextension.org/park/Images/Ag/pdfs/camelina.pdf>.

- Morgante M, Olivieri AM** (1993) PCR-Amplified Microsatellites as markers in Plant Genetics. *Plant Journal* **3**: 175-182.
- Moser BR** (2010) Camelina (*Camelina sativa* L.) oil as a biofuels feedstock: Golden opportunity or false hope? *Lipid Technology* **22**: 270-273.
- Nunome T, Negoro S, Kono I, Kanamori H, Miyatake K, Yamaguchi H, Ohyama A, Fukuoka H** (2009) Development of SSR markers derived from SSR-enriched genomic library of eggplant (*Solanum melongena* L.). *TAG Theoretical and Applied Genetics* **119**: 1143-1153.
- Phumichai C, Phumichai T, Kongsiri N, Wongkaew A, Sripichit P, Kaveeta R** (2010) Isolation of 55 microsatellite markers for *Jatropha curcas*; and its closely related species. *Biologia Plantarum* **55**: 387-390.
- Putnam D, Budin J, Field L, Breene W** (1993) Camelina: A promising low-input oilseed. *New crops*. , Wiley, New York, pp. 314-322.
- Rajendrakumar P, Biswal AK, Balachandran SM, Srinivasarao K, Sundaram RM** (2007) Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. *Bioinformatics* **23**: 1-4.
- Schill SR** (2009) Camelina meal approved for feedlot cattle. *Biodiesel Mag.* http://www.biodieselmagazine.com/article.jsp?article_id=3837.
- Schuster A, Friedt W** (1998) Glucosinolate content and composition as parameters of quality of camelina seed. *Industrial Crops and Products* **7**: 297-302.
- Semagn K, Bjornstad A, Ndjiondjop MN** (2006) An overview of molecular marker methods for plants. *African Journal of Biotechnology* **5**: 2540-2568.
- Tautz D** (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**: 6463-6471.
- Urbaniak SD, Caldwell CD, Zheljzkov VD, Lada R, Luan L** (2008) The effect of seeding rate, seeding date and seeder type on the performance of *Camelina sativa* L. in the Maritime Provinces of Canada. *Canadian Journal of Plant Science* **88**: 501-508.
- Vollmann J, Damboeck A, Eckl A, Schrems H, Ruckebauer P** (1996) Improvement of *Camelina sativa*, an underexploited oilseed. *Progress in New Crops*, American Society for Horticultural Science. Alexandria, VA, USA, pp 357-362.
- Vollmann J, Grausgruber H, Stift G, Dryzhyruk V, Lelley T** (2005) Genetic diversity in camelina germplasm as revealed by seed quality characteristics and RAPD polymorphism. *Plant Breeding* **124**: 446-453.
- Vollmann J, Moritz T, Kargl C, Baumgartner S, Wagentristl H** (2007) Agronomic evaluation of camelina genotypes selected for seed quality characteristics. *Industrial Crops and Products* **26**: 270-277.
- Wang X, Zhao X, Zhu J, Wu W** (2005) Genome-wide investigation of intron length polymorphisms and their potential as molecular markers in rice (*Oryza sativa* L.). *DNA Research* **12**: 417-427.
- Warwick SI, Al-Shehbaz IA** (2006) Brassicaceae: Chromosome number index and database on CD-Rom. *Plant Systematics and Evolution* **259**: 237-248.
- Zane L, Bargelloni L, Patarnello T** (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology* **11**: 1-16.
- Zubr J** (1997) Oil-seed crop: *Camelina sativa*. *Industrial Crops and Products* **6**: 113-1.

2. Publication 1

h-TBP: an approach based on intron-length polymorphism for the rapid isolation and characterization of the multiple members of the β -tubulin gene family in *Camelina sativa* (L.) Crantz

Incoronata Galasso, Antonella Manca, Luca Braglia, Tommaso Martinelli, Laura Morello, Diego Breviario.

Published 2011 in *Molecular Breeding*. 28:635–645

2.1. Abstract.

We have developed a new version of the cTBP (combinatorial tubulin-based polymorphism) method, a previously described approach based on intron-length polymorphism (ILP), to rapidly characterize the β -tubulin gene family of *Camelina sativa* (L.) Crantz, a plant species of importance for oil production but still largely unexplored at genomic level.

The method, named h-TBP, allows the rapid cloning of the β -tubulin genomic sequences that encompass the two introns, invariantly present at fixed positions within the coding region of the vast majority of the plant species. The β -tubulin sequences cloned by h-TBP also comprise part of exon1 and exon3 and the whole sequence of exon2. The h-TBP method has then been used to isolate, clone and characterize the β -tubulin gene family of *C. sativa*, composed of at least 20 different β -tubulin isotypes, named *CsTUB1* through *CsTUB20*. The relatively high number of β -tubulin genes has been further substantiated by Southern-blot analysis.

Comparison of the β -tubulin exon sequences of *C. sativa* with those of *Arabidopsis thaliana*, the closest relative among crucifers, defines distinct groups of putative orthologous genes, identified by a UPMGA cluster analysis. Analysis of the *C. sativa* β -tubulin intron sequences reveals some molecular features that can provide the first hints for the understanding of intron plasticity and evolution. From a more immediate perspective, these data provide the first substantial contribution to the characterization of the largely unexplored genome of *C. sativa*, and the tools for assisting programmes of breeding and selection of the most productive plants.

Keywords: ILP, β -tubulin, Introns, Oilseed plant, False flax, Brassicaceae.

2.2. Introduction

Camelina sativa (L.) Crantz or false flax, a member of the mustard family (Brassicaceae), is an oilseed plant native to Central Asia and the Mediterranean with both winter and spring annual biotypes. Camelina is an ancient crop plant used extensively in the past for cooking, cosmetics and as a lamp oil (Knorz 1978). Camelina cultivation subsequently declined due to low yield when compared to other oilseed crops such as rapeseed or sunflower.

Recently, *C. sativa* has attracted renewed attention as an alternative oilseed plant for biofuel transformation because of its low input requirements such as fertilizers, pesticides and energy (Putnam et al. 1993; Vollmann et al. 2007). In addition, *C. sativa* meal is rich in protein of high biological value (Moloney et al. 1998) and recently the US Food and Drug Administration (FDA) has approved the use of meal from crushed seeds in cattle feed supplement (Schill 2009). Despite this renewed interest, camelina is still an underexploited crop species. As such it has received less attention from plant breeders than other oilseed plants such as rapeseed. Its genome has also been poorly characterized. Relatively few gene sequences have been isolated and few reports concerning the use of molecular markers for the genetic characterization of different *C. sativa* genotypes have been published (Vollmann et al. 2005; Gehringer et al. 2006; Ghamkhar et al. 2010).

Both limitations can be overcome with the use of suitable intron length polymorphism (ILP) markers. In fact, ILPs are sequence-tagged-site (STS) markers targeted to specific gene loci, thus providing readily usable molecular tools for the assessment of genetic variability in unexplored, still largely uncharacterized plant genomes (Einax and Voigt 2003; Wang et al. 2005). ILP markers are codominant, neutral, and stable and can be easily transferred among different plant species. Typically, ILP markers result from the amplification of specific gene loci by exon-primed intron-crossing PCR (EPIC-PCR) (Palumbi 1995). cTBP (combinatorial tubulin-based polymorphism) describes a successful and widely applicable ILP based approach that allows for the amplification of either of the two introns commonly present in conservative positions within the coding sequences of plant β -tubulin genes (Breviario et al. 2007; Braglia et al. 2010). cTBP has been successfully used for fingerprinting the genomes of several plant species including crops and plant species for which no genomic information was available (Braglia et al. 2010). Here we report a further development of the method that has allowed the rapid isolation of 20 different members of the *C. sativa* β -tubulin gene family.

This approach, named h-TBP (horse-TBP), applicable in the field of food traceability like other TBP-based methods (Casazza et al. 2011), found a unique usage in the rapid cloning and sequencing of the genomic region that encompassed both intron1 and intron2 of each β -tubulin gene. *C. sativa* β -tubulin coding sequences were also compared to their counterparts in the *Arabidopsis thaliana* genome (Snustad et al. 1992), the closest species among the Brassicaceae family (Flannery et al. 2006), to define different groups of associations. Analysis of the *C. sativa* β -tubulin intron sequences identifies single isotype molecular traits that are useful for marker assisted selection (MAS) and genetic relationship studies.

2.3. Materials and methods

2.3.1. Plant material

Seeds of *C. sativa* varieties Calena accession number CAM134, Lindo CAM40, Bavaria CAM180, Soledo CAM46 and camelina species *C. microcarpa* Andr. ex DC., PI 650134, *C. rumelica* subsp. *transcaspica* (Fritsch) Hedge, PI 650139 and *C. hispida* var. *grandiflora* (Boiss.) Hedge, PI 650133 used in the present study were kindly provided by the Institute of Plant Genetics and Crop Plant Research (IPK), Germany, and the USDA-ARS-NCRPIS germplasm collection, USA, while the *C. sativa* var. Ligena was kindly provided by Dr. Rod Snowdon, Department of Plant Breeding, Justus-Liebig-University of Giessen, Germany.

2.3.2. h-TBP

Genomic DNA was extracted from fresh leaf tissue, according to the method reported by Doyle and Doyle (1987) modified as follows: 3 g of young leaves were ground in liquid nitrogen to a fine powder and extracted with cetyltrimethylammonium bromide (CTAB) extraction buffer (100 mM Tris-HCl, pH 8.0; 20 mM EDTA; 1.4 M NaCl; 2% [w/v] CTAB; and 0.2% [v/v] β -mercaptoethanol). The mixture was incubated for 30 min at 65°C, followed by one extraction with chloroform/isoamyl alcohol (24:1). Isopropanol was used to precipitate nucleic acids and the pellet obtained was dissolved in Tris-EDTA (TE) buffer (10 mM Tris-HCl, pH 8.0 and 1 mM EDTA, pH 8.0).

Extracted DNA was treated with RNase and remaining impurities were extracted with a second treatment of chloroform/isoamyl alcohol (24:1) and

chloroform. Total DNA was precipitated using cold ethanol and the pellet was dissolved in TE buffer. Conditions for the amplification of the genomic regions that encompass both intron1 and intron2 of the multiple members of the *C. sativa* β -tubulin gene family were as follows: 100 ng of genomic DNA extracted from the leaf tissue of all samples was used for each PCR reaction. Degenerated oligonucleotides TBP-F: 5'-AACTGGGCBAARGGNCAYTAYAC-3' and TBP-R: 5' CRAAVCCBACCATGAARAARTG-3' were used as the forward and the reverse primer, respectively.

The target sequence for TBP-F is within exon1 in close proximity to the first intron, whereas that for TBP-R is located within exon3. The following touchdown PCR protocol for the h-TBP method was used: 94°C for 4 min, followed by 14 cycles of 30 s at 94°C, 45 s at 65°C (decreasing by 0.7°C every cycle from initial to final value), and 2.5 min at 72°C, then 15 cycles of 30 s at 94°C, 30 s at 55°C and 2.5 min at 72°C. The reaction was held at 10°C after a final extension at 72°C for 8 min. h-TBP amplified products were separated on a 6% acrylamide gel and band visualized by silver nitrate staining as in Bardini et al. (2004).

2.3.3. DNA cloning, sequencing and Southern analysis

Genomic DNA was extracted from leaf tissue of *C. sativa* var. Calena as reported above. Once purified, the PCR products resulting from the amplification with the h-TBP method were cloned in the pGEM-T Easy Vector Systems (Promega). About 100 clones were screened according to the insert size, and to the presence/absence of restriction sites for *EcoRI*, *EcoRV* and *HindIII*. Inserts of selected clones were sequenced, in both directions, with an automated sequencer by Macrogen, (Seoul, South Korea).

A search of nucleotide sequence homologies was carried out in the National Center Biotechnology Information (NCBI) database using the BLAST algorithm (<http://www.ncbi.nlm.nih.gov>). Searching for tandem repeat motifs and microsatellite sequences was performed using the Tandem Repeats Finder and Sputnik programs, which are publicly available at <http://tandem.bu.edu/trf/trf.html> (Benson 1999) and <http://www.cbib.u-bordeaux2.fr/pise/sputnik.html> (Abajian 1994).

The nucleotide sequences obtained are available in the EMBL (European Molecular Biology Laboratory) Nucleotide Database with the accession numbers listed in Table 2.1. For Southern hybridization, 5 μ g of *C. sativa* CAM134 and *A. thaliana* (var. Columbia) genomic DNA were digested with the restriction

enzyme *Eco*RI, size-separated and transferred onto a nylon membrane. Exon2 of *CsTUB3* was PCR amplified using two specific primers Tubexon2F: 5'-GATCCAAGTGTGTCACCTCGTTG-3' and Tubexon2R: 5'-TTACAGCTAGGAGTGGTGAGCTT-3' and labelled with α -[³²P]-dCTP using a random primer DNA labelling kit (Fermentas, Life Sciences) before filter hybridization. The filter was washed in 1.5 mM sodium citrate pH 7.0, 15 mM NaCl, 0.5% SDS (0.1 9 SSC), at 65°C.

2.3.4. Sequence comparison and phylogenetic analysis

Nucleotide and deduced amino acid sequences were multialigned with the ClustalW programme using default parameters. In order to compare our partial camelina *CsTUB* sequences with the *Arabidopsis* β -tubulin genes the nucleotide sequences corresponding to the N- and C-terminal amino acids regions of *Arabidopsis* members were excluded from the sequence alignment and phylogenetic analyses. Genomic sequences corresponding to the nine *ArathTUB* genes (*ArathTUB1*: M20405, *ArathTUB2*: M84700; *ArathTUB3*: M84701; *ArathTUB4*: M21415; *ArathTUB5*: M84702; *ArathTUB6*: M84703; *ArathTUB7*: M84704; *ArathTUB8*: M84705; *ArathTUB9*: M84706) were retrieved from the EMBL Database. The genetic distance was determined using the Treecon program for Windows (Van de Peer and DeWachter 1994). The similarity values were used to perform the cluster analyses according to the unweighted pair group method with arithmetic average (UPGMA, Sneath and Sokal 1973). The statistical confidence of a particular group of sequences in the tree was evaluated by a bootstrap test with 1000 replicates (Hillis and Bull 1993).

2.4. Results

2.4.1. h-TBP: the rationale

h-TBP amplifies the whole genomic region that encompasses the two introns found at conservative positions within the coding region of the vast majority of the plant β -tubulin genes characterized to date. Amplification of these multiple gene-specific loci is achieved with the combinatorial use of a forward

primer targeting a sequence upstream of the 5' end of the first intron and a reverse primer annealing downstream from the 3' end of the second intron.

As shown at the top of Fig. 2.1.a, this leads to the amplification of genomic regions that contains the first and the second intron of the different members of the β -tubulin gene family, together with the second coding exon and a partial sequence of both coding exons 1 and 3.

The effectiveness of such a strategy is first demonstrated by the genomic fingerprinting of the five different varieties of *C. sativa* shown in Fig. 2.1.a. Different ILP markers are generated that can be of use for a preliminary characterization and grouping of the different varieties. Data in Fig. 2.1.a demonstrate that h-TBP can be successfully applied to previously uncharacterized plant genomes for which no molecular information was available.

Also, the number of the amplified β -tubulin bands is unexpectedly high when compared to the number of β -tubulin genes reported for the closely related model plant *A. thaliana*. The consistency of the banding pattern observed across the different *C. sativa* varieties indicates a high degree of genetic similarity that is lost when h-TBP is applied to different species of the genus *Camelina*.

In this latter case the pattern produced by h-TBP is highly dissimilar among the different species, distinguishing each of them clearly, rapidly and consistently (Fig. 2.1.b). These findings prompted us to isolate 20 putatively different β -tubulin genes of *C. sativa* generated by h-TBP mediated amplification.

2. Publication 1. *h-TBP: an approach based on intron-length polymorphism for the rapid isolation and characterization of the multiple members of the β -tubulin gene family in *Camelina sativa* (L.) Crantz*

<i>C. sativa</i>	EMBL accession numbers	Intron1 length (bp)	Intron1 AT (%)	Intron2 length (bp)	Intron2 AT (%)
<i>CsTUB1</i>	FN811148	82	63	97	71
<i>CsTUB2</i>	FN811149	99	64	92	70
<i>CsTUB3</i>	FN811150	115	63	94	69
<i>CsTUB4</i>	FN811151	136	69	105	66
<i>CsTUB5</i>	FN811152	144	67	109	64
<i>CsTUB6</i>	FN811153	97	69	178	62
<i>CsTUB7</i>	FN811154	180	68	100	70
<i>CsTUB8</i>	FN811155	183	73	99	68
<i>CsTUB9</i>	FN811156	104	69	179	63
<i>CsTUB10</i>	FN811157	101	63	277	66
<i>CsTUB11</i>	FN811158	111	64	489	65
<i>CsTUB12</i>	FN811159	91	64	511	65
<i>CsTUB13</i>	FN811160	210	64	393	66
<i>CsTUB14</i>	FN811161	100	65	530	65
<i>CsTUB15</i>	FN811162	159	66	504	66
<i>CsTUB16</i>	FN811163	161	62	519	65
<i>CsTUB17</i>	FN811164	174	65	528	67
<i>CsTUB18</i>	FN811165	217	63	503	66
<i>CsTUB19</i>	FN811166	707	66	92	68
<i>CsTUB20</i>	FN811167	917	69	83	68
Average mean		204	65.8	279	66.5

Table 2.1. List of *CsTUBs* identified from *C. sativa* genomic DNA with respective EMBL accession number, intron length and percentage of AT nucleotides.

2. Publication 1. *h-TBP*: an approach based on intron-length polymorphism for the rapid isolation and characterization of the multiple members of the β -tubulin gene family in *Camelina sativa* (L.) Crantz

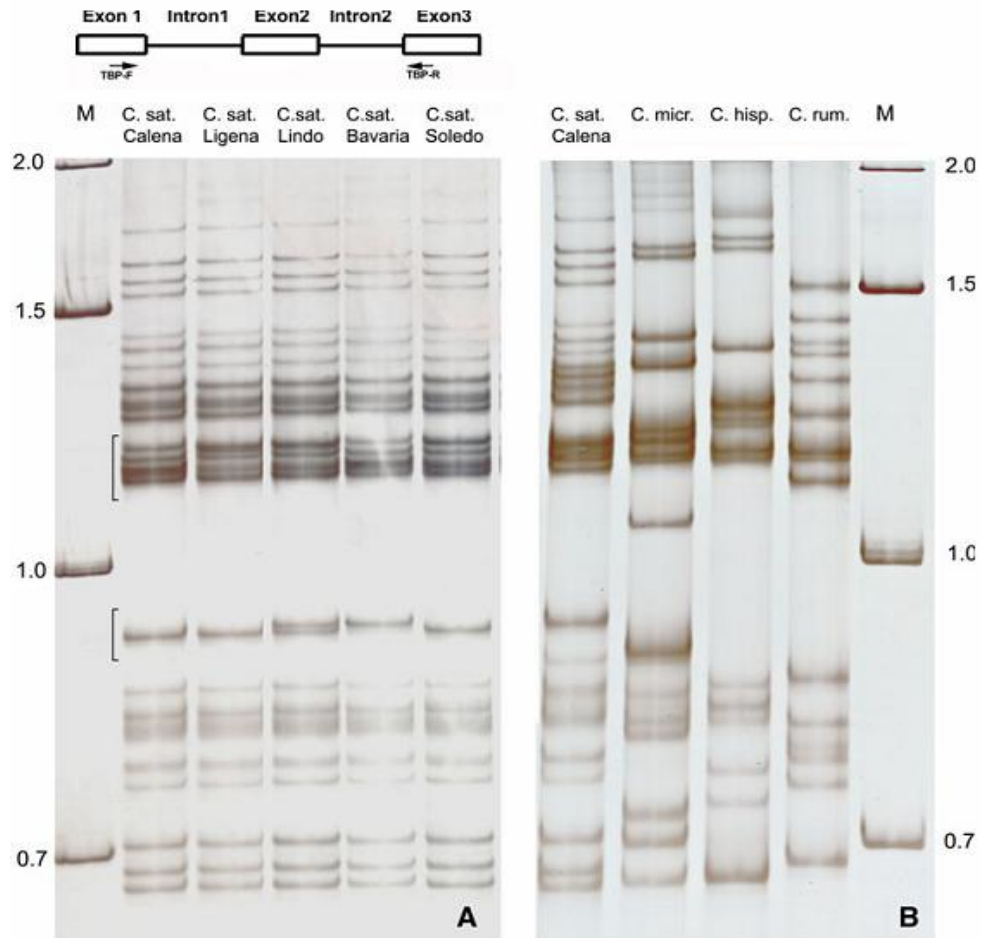


Figure 2.1. Polyacrylamide gel electrophoresis of PCR products amplified by *h-TBP* from genomic DNA of five *C. sativa* varieties (Calena, Ligena, Lindo, Bavaria, Soledo) (a) and four different species of camelina: *C. sat.*, *C. sativa* var. Calena; *C. micr.*, *C. microcarpa*; *C. hisp.*, *C. hispida* var. *grandiflora*; *C. rum.*, *C. rumelica* subsp. *transcaspica* (b). M, molecular size marker in kbp. Two major regions for DNA polymorphism among the five camelina varieties are indicated by brackets on the left side of the gel A. At the top is shown a schematic representation of the intron–exon structure of a typical plant β -tubulin gene. Arrows indicate position of the primers TBP-F and TBP-R on the exons 1 and 3.

2.4.2. Gene structure of the β -tubulin family

Twenty partial β -tubulin gene sequences, named *CsTUB1* through *CsTUB20*, were isolated from the genomic DNA of the commercially available Calena variety by applying the h-TBP method. Isolated bands were then cloned by routine molecular genetics techniques. DNA sequencing revealed that all of the h-TBP-derived genomic bands exclusively contained β -tubulin sequences.

As predicted by the cloning strategy, each of the *CsTUB* sequences contains 97 bp of coding exon1, 270 bp of exon2 (complete) and 147 bp of exon3 together with the two full-length introns, 1 and 2 (Fig. 2.2.a).

Introns were conserved in both number and position with respect to *A. thaliana* (*ArathTUB*) and to the β -tubulin genes of most other plant species (Snustad et al. 1992; Liaud et al. 1992; Gianì' and Breviaro 1996). Intron length varied among the 20 different *C. sativa* β -tubulin isotypes. Most intron sequences were rather short (30 introns ranged from 82 to 277 bp), while 10 were relatively long (from 393 to 917 bp). As observed for most dicot introns, *C. sativa* β -tubulin introns were highly rich in AT nucleotides with an average of 66% over the total (Table 2.1.).

This is similar to the mean content (68%) observed in introns of *A. thaliana*. Overall, the average length of intron2 was slightly higher (279 bp) than that of intron1 (204 bp). Intron/exon borders of the *CsTUB* sequences contained the canonical 5' and 3' splicing consensus motifs [GT...AG] (Lorkovic et al. 2000). A search for tandem repeat motifs and microsatellite sequences revealed the presence of two direct head-to-tail arranged tandem repeats of 34 bp and 46 bp within intron2 of *CsTUB10* (EMBL acc. number FN811157) and intron1 of *CsTUB13* (EMBL acc. number FN811160), respectively.

In addition, two mononucleotide stretches (T)₂₆ and (T)₃₆ were detected in introns 1 and 2 of *CsTUB7* and *CsTUB8*, respectively, and two trinucleotides (CTT)₇ and (CTT)₆ were found in intron1 of *CsTUB10* and *CsTUB14*, respectively.

Analyses of the *CsTUBs* coding sequences revealed a high level of homology across the different paralogs (97–78% identity) that typically drops within the intron sequences. This is fully consistent with the organization of the β -tubulin genes observed in many other plant species. Although not complete, since it accounts for an open reading frame of 514 bp corresponding to a total of 171 residues, the deduced aminoacid sequence of the different *C. sativa* β -tubulin typically contains specific functional domains such as the predicted guanidinotriphosphate hydrolase (GTPase) site and the GTP-exchange domain that

is strongly conserved with the exception of one single conservative change in *CsTUB2* (Gly instead of Glu at position 109) (2.8. Appendix—Fig. 2.8.1).

On the other hand, intron DNA sequences are highly variable across the different members of the β -tubulin gene family, facilitating easy recognition of each individual isotype. Such specificity represents a useful tool for monitoring molecular diversity among varieties and for assisting breeding programmes.

2. Publication 1. *h-TBP: an approach based on intron-length polymorphism for the rapid isolation and characterization of the multiple members of the β -tubulin gene family in *Camelina sativa* (L.) Crantz*

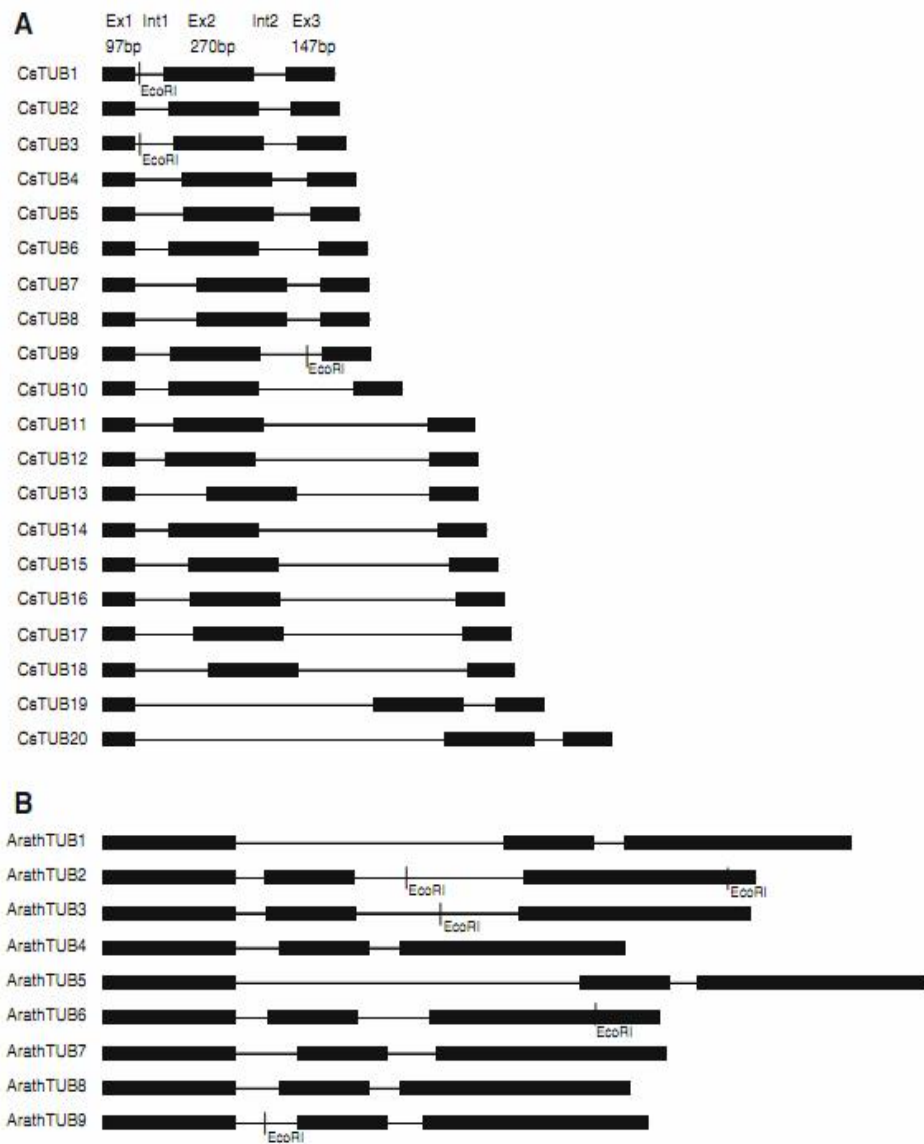


Figure 2.2. Intron-exon structure of *C. sativa* (a) and *A. thaliana* (b) β -tubulin genes. Exons are represented as black boxes while lines represent introns. For *C. sativa* the sizes of partial exons 1 (97 bp) and 3 (147 bp) and those of full-length exon2 (270 bp) are shown at the top. *EcoRI* restriction sites are also indicated.

2.4.3. Genomic organization of the β -tubulin family in camelina

The high number of β -tubulin genes present in *C. sativa* was unexpected, since the genome of its closest relative *A. thaliana* contains only nine different β -tubulin isotypes.

Therefore, to obtain additional information on the genome organization and number of β -tubulin genes present in the *C. sativa* genome, we performed a Southern blot analysis on the *EcoRI* restricted genomic DNA of Calena using, as a probe, a radioactively-labelled exon2 fragment isolated from *CsTUB3* (Fig. 2.3). *EcoRI* endonuclease was chosen since no restriction sites for this enzyme are present in any of the exon2 sequences of *C. sativa* β -tubulin isotypes (see Fig. 2.2.a). This implies that each of the detected hybridization fragments should correspond to one gene. As shown in Fig. 2.3, no less than 17 clearly defined bands, ranging in size from 1.6 to 4.9 kbp, were detected by Southern blot hybridization analysis.

This indicates the presence of multiple camelina β -tubulin genes up to a total number that closely matches that of the fragments cloned through h-TBP. Because of its high homology with the corresponding sequence of *A. thaliana*, the same *C. sativa* exon2-specific probe was used on *A. thaliana* genomic DNA restricted with *EcoRI*.

The number and size of the hybridizing bands detected on the *EcoRI* digested genomic DNA of *A. thaliana* are consistent with those previously reported by Snustad et al. (1992), taking into account that some fragments are missing due to the presence, in some of the β -*ArathTUB* genes, of additional *EcoRI* sites located outside of the region recognized by the *CsTUB3* exon2 probe (Fig. 2.2.b).

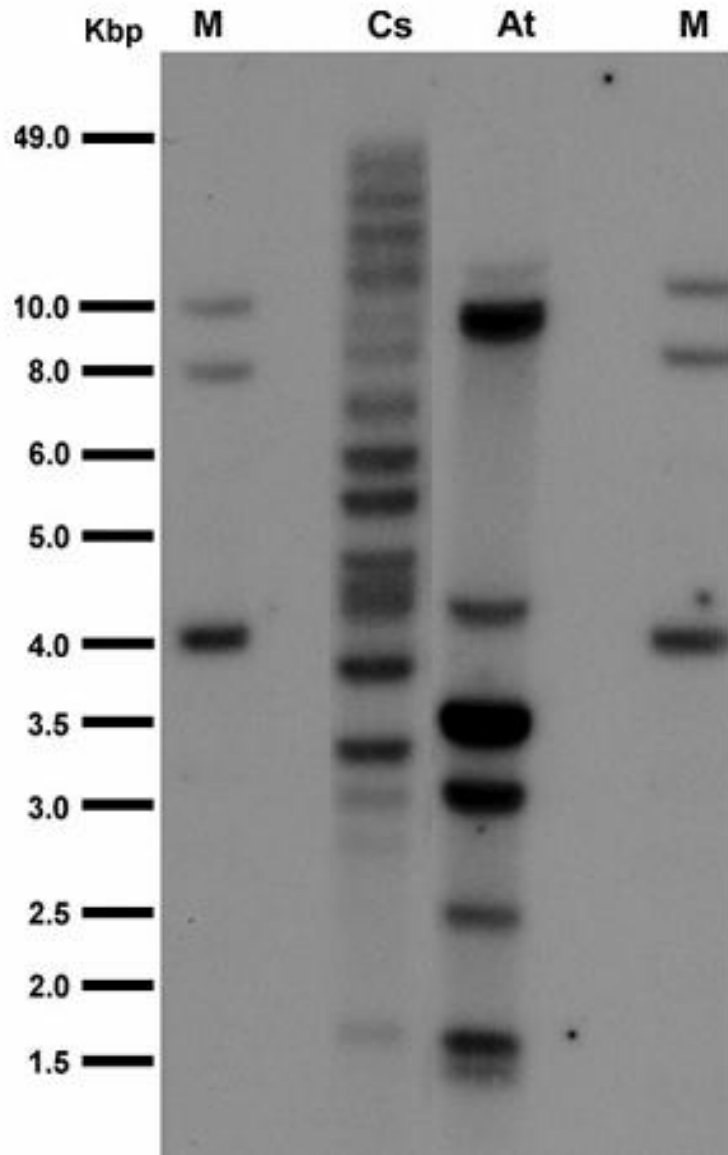


Figure 2.3. Southern blot analysis of genomic DNA of *C. sativa* (Cs) and *A. thaliana* (At). DNAs were digested with *EcoRI* and probed with the labelled α -[32 P]-dCTP exon2 of *CsTUB3*. Molecular size markers (kbp) are indicated.

2.4.4. Phylogenetic analysis

Since each of the *A. thaliana* β -tubulin genes has been fully characterized at both sequence and expression level, this latter defining protein isotypes that are specific for developmental stages rather than in response to external stimuli, we compared the *A. thaliana* β -tubulin coding sequences with those of *C. sativa* in order to identify putative orthologous genes.

Such a comparison has been reciprocally confined to the corresponding 514-bp-long portion of the coding sequence. Estimated genetic similarities among the 20 *C. sativa* and the nine *Arabidopsis* β -tubulin genes were used to infer a cluster analysis by the neighbour-joining method (Fig. 2.4). The reliability of this cross-species similarity study was supported by a preliminary analysis on the *A. thaliana* β -tubulin gene family that demonstrates a strict correspondence between the clusters identified with the partial nucleotide sequences and those obtained from the corresponding full length amino acid sequences (data not shown).

As shown in the neighbour-joining tree of Fig. 2.4, the 20 different *CsTUBs* spread throughout the four β -tubulin subfamilies previously defined by Oakley et al. (2007), with nine members in class II and only two in classes I and III. One to three (two in most cases) members of the *C. sativa* β -tubulin gene family grouped with a closely related isotype of *A. thaliana*, thus defining small clades of higher similarity (93–99% of nucleotide identity) in which strictly related paralogous genes (defined as homeologs) were clustered with their putative *Arabidopsis* orthologs.

The only exception is represented by the homeologs *CsTUB15*, *CsTUB17* and *CsTUB18* that do not appear to have any β -tubulin counterpart in the *Arabidopsis* genome (Fig. 2.4).

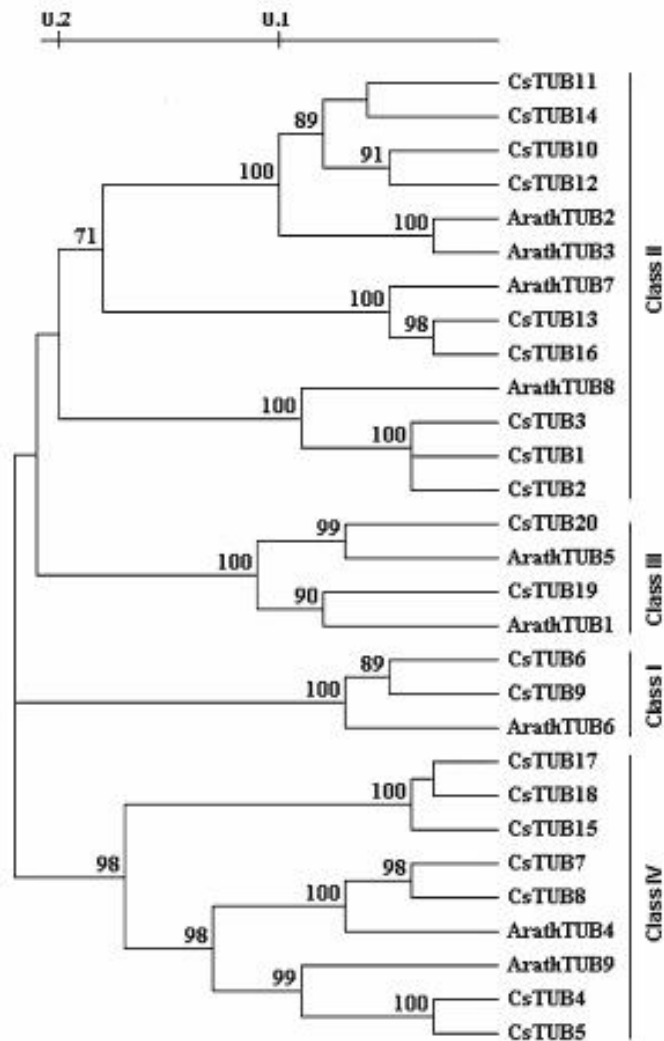


Figure 2.4. Neighbour-joining tree constructed using Treecon software. *CsTUB* and *ArathTUB* abbreviation indicate the 20 camelina and the nine *Arabidopsis* β -tubulin genes, respectively. On each node the bootstrap values out of 1000 replicates are indicated. The scale bar of genetic distance is shown at the top. Classes I-IV are assigned according to Oakley et al. (2007).

2.4.5. Intron sequence analysis

Pairwise alignment of the intron sequences of the *C. sativa* homeologous genes, as defined by clusters, revealed a large range of variability in nucleotide conservation, ranging from 25 to 97% (2.8. Appendix—Table 2.8.1).

Most differences were due to short sequence duplications, insertions or deletions, or to single base or short sequence repeats, interspersed within more conserved blocks, rather than to single nucleotide substitutions. Such modifications resulted in a mean length variation among homeologs of 7.8% for the first intron and 6.5% for the second intron.

Larger fluctuations, which can even lead to a two fold variation in the length of the intron, were observed in some cases, as for the second intron of *CsTUB10* which is 277 bp long, while the length of the other members of the cluster (*CsTUB11*, *CsTUB14* and *CsTUB12*) varied from 489 to 530 bp. Similarly, the second intron of *CsTUB16* (EMBL acc. number FN811163) shows an extensive internal duplication of 76 bp (regions 627–702 bp and 807–884 bp) that increases its size to 519 bp, more than 100 bp longer than its homeolog *CsTUB13* (393 bp long) (Table 2.1.).

Similar observations are also valid when comparing camelina β -tubulin genes with their putative *Arabidopsis* orthologs. On average, length variability was 11.2% for the first and 13.7% for the second intron, respectively, with some bigger differences due to segmental duplications or deletions. With regards to sequence conservation among orthologs, cases of both strong homology (up to 86% of sequence identity) and great divergence were found, in the first as well as in the second introns (2.8. Appendix—Table 2.8.1).

Whether higher conservation of particular introns or sequence blocks among orthologs occurs by chance or relates to intron function is not known at present, but a comparison with tubulin intron sequences from other closely related species could provide some information.

2.5. Discussion

Here we have shown that one of the major features that characterizes a good ILP-based marker, namely sequence site-specific tagging, is indeed useful not only for providing plant genomic fingerprinting information but also for allowing the rapid cloning and characterization of specific multiple gene loci, in this case β -tubulin of *C. sativa*.

This general approach is particularly useful when dealing with plant species for which no or poor information is available at the genomic level. To achieve such a goal we have modified pre-existing versions of a tubulin- based polymorphism method (Bardini et al. 2004; Breviario et al. 2007) in order to amplify longer traits of the β -tubulin loci, thus recovering enough information for gene characterization, phylogenesis and genetic similarity studies.

The newly devised h-TBP method has allowed the identification in *C. sativa* of a very high number of β -tubulin genes. Although we do not know whether all 20 *CsTUBs* are actually expressed, since they were isolated from genomic DNA, such a large number of β -tubulin might be an indication that the herbaceous annual *C. sativa* (chromosome number: $2n = 40$) is a polyploid species, not a diploid as previously thought. This is in accordance with a recent suggestion made by Gehringer et al. (2006) who, on the basis of simple sequence repeat and amplified fragment length polymorphism data, reached the conclusion that *C. sativa* is either an auto- or an allopolyploid.

This hypothesis is further supported by the phylogenetic tree shown in Fig. 2.4, which clearly shows a general, although not absolute, 1:2 correspondence between *Arabidopsis* and *Camelina* β -tubulin genes. Such a distribution is highly reminiscent of that observed when comparing the five α -tubulin sequences of barley (*Hordeum vulgare* L.) with the 15 members of the closely related, hexaploid species *Triticum aestivum* L. (Farajalla and Gulick 2007). There, each barley tubulin sequence clusters together with three wheat tubulins, one from each of the A, B and D genome sets.

In camelina, subsequent tandem gene duplication may have occurred, producing in some cases three highly related genes, instead of two. The ease with which we have been able to clone out many if not all of the β -tubulin genes is quite remarkable and may help in designing similar strategies for other as yet unexplored plant genomes. The unusually large size of the β -tubulin gene family of *C. sativa* has been further confirmed by Southern hybridization analyses, although only 17 fragments, not 20, could be clearly identified. This discrepancy is likely due to co-migration of some *EcoRI* fragments of similar size.

Similar experiments performed on *A. thaliana* genomic DNA, using *CsTUB3* exon2 as probe, detected a number and size of hybridizing fragments fully consistent with those reported in the literature (Snustad et al. 1992). This excludes the occurrence of any cross-hybridization with aspecific DNA sequences and further confirms that the number of β -tubulin genes in *C. sativa* is double that of *A. thaliana*, the most closely related species. Expansion of the β -tubulin gene

family in plants, although rare, is not a new event. In fact 20 and 19 expressed β -tubulin have been found in the woody perennial *Populus tremuloides* Michx. and in the shrub *Gossypium hirsutum* L., respectively (Oakley et al. 2007; He et al. 2008). This expansion likely resulted from genome-wide duplication events that occurred early in the genetic lineage. The authors suggest that the selective conservation of such a high number of β -tubulin genes, compared to the limited number of α -tubulin genes, may be due to an adaptive functional role in relation to wood (aspen) and fiber (cotton) formation.

Why this is also occurs in *C. sativa* is a matter of further investigation, particularly so when considering the data of the phylogenetic tree obtained from the β -tubulin coding sequences of *C. sativa* and *A. thaliana*. In fact, the expression profile and putative roles of the β -tubulin isoforms of the latter have been well documented. Since two or more members of the β -tubulin family of *C. sativa* can be associated with each member of the *A. thaliana* family, their putative role and mode of expression could be postulated in accordance. *C. sativa* – *A. thaliana* interspecies comparisons also indicate putative orthologs for future molecular evolution studies (see below). Among the *C. sativa* paralogs, a cluster of β -tubulin isoforms with no correspondence with *A. thaliana* genes can be easily recognized (*CsTUB15*, *CsTUB17* and *CsTUB18*). Future studies should address the reason for this exclusiveness. The phylogenetic tree has also shown a distribution of *CsTUBs* that is spread throughout the clusters without distinction between *A. thaliana* and *C. sativa*. A similar distribution of the β -tubulin genes, irrespective of the analysed species, was also reported by others (Jost et al. 2004; Oakley et al. 2007) comparing several β -tubulin isolated from monocotyledonous and dicotyledonous plants. In their phylogenetic trees the β -tubulins from rice and maize intermingled with those from dicotyledonous plants and vice versa, while the five β -tubulin genes of the moss *Physcomitrella patens* grouped together in a separate cluster. Our results provide further evidence that the angiosperm β -tubulin genes have likely originated from a single ancestral gene. At the structural genomic level, the organization of *C. sativa* genes is similar to that found for the vast majority of plant β -tubulin.

Each member of the family is characterized by the presence of three coding exons separated by the two introns found at conservative positions in all the plant species investigated thus far, with the minor exceptions of maize *ZeamaTUB1* and rice *OryzaTUB2* where only the first intron is present. Intron conservative positioning represents a useful base for the development of ILPs when

surrounding exon sequences are not too variable and allow the designing of versatile primers, as is the case for β -tubulin.

Intron length and sequence variation not only represent useful information for the genetic characterization of different *C. sativa* varieties, but also provide a readily exploitable tool for assisting breeding programmes designed to improve the agronomic and biochemical traits of such an important crop.

Ultimately, a careful study of the molecular events that have contributed to sequence variation (InDel, SNPs, repeats etc.) in individual introns of closely related varieties (and species) is likely to shed new light on intron molecular evolution, a still exciting and yet unresolved molecular genetics issue.

2.6. Conclusions

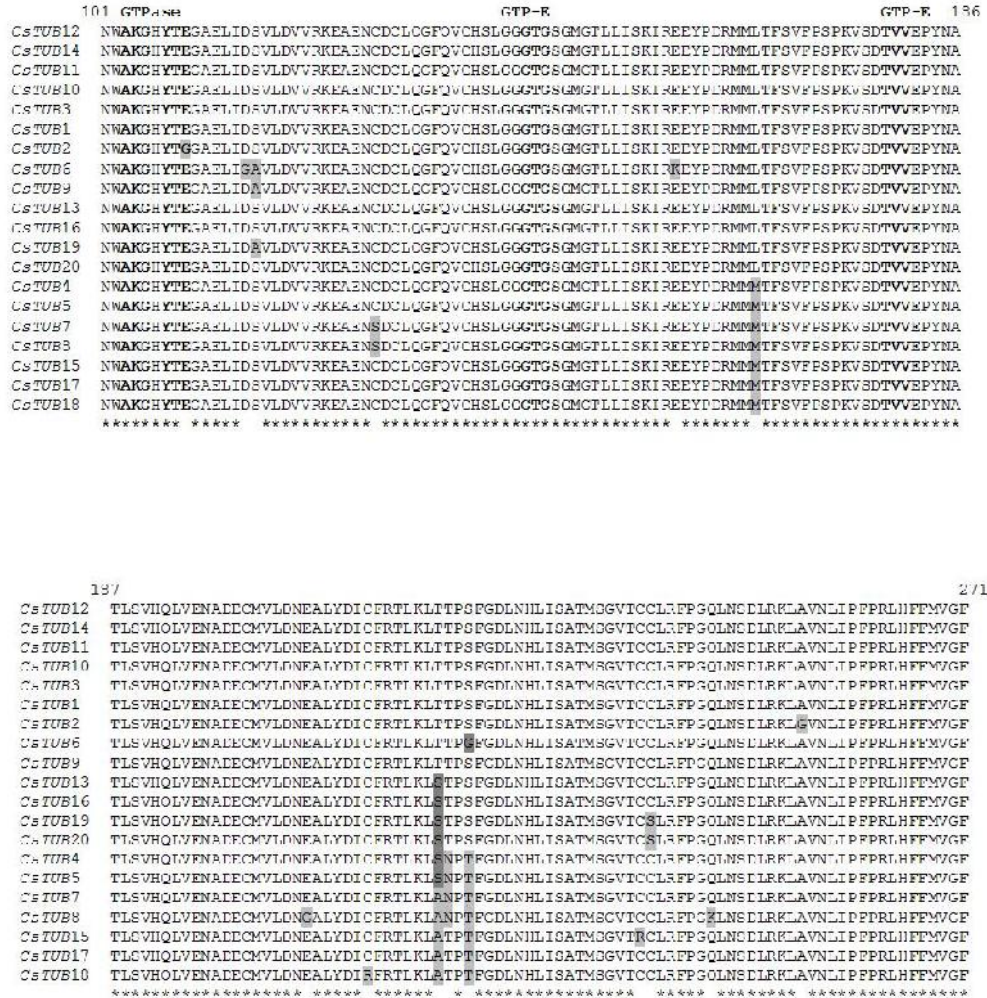
Few laboratories have used the β -tubulin gene family to assess genetic diversity and to carry out evolution studies among eukaryotic organisms.

Here we provide new information concerning the β -tubulin gene family isolated by the largely unexplored yet important oil plant *C. sativa*. This was achieved by the use of h-TBP, a novel ILP-based method that allowed the rapid isolation of 20 different β -tubulin gene sequences. New tools for assisting genetic characterization and breeding are now available.

2.7. Acknowledgments

We thank Mrs. Floriana Gavazzi and Gloria M. Daminati for technical support. This work was partially supported by Regione Lombardia, agreement Regione/CNR, project 2 “Risorse biologiche e tecnologie innovative per lo sviluppo sostenibile del sistema agro-alimentare”.

2.8. Appendix



2.8. Appendix—Figure 2.8.1. Multiple alignment of the CsTUB deduced amino acid sequences. Differences are shown in grey. The predicted GTP-exchange domains (in bold) and the GTPase domain are included. Asterisks indicate identical amino acids.

2. Publication 1. *h-TBP: an approach based on intron-length polymorphism for the rapid isolation and characterization of the multiple members of the β -tubulin gene family in Camelina sativa (L.) Crantz*

	<i>CsTUB14-1</i> <i>CsTUB14-2</i>	<i>CsTUB10-1</i> <i>CsTUB10-2</i>	<i>CsTUB12-1</i> <i>CsTUB12-2</i>	<i>ArathTUB2-1</i> <i>ArathTUB 2-2</i>	<i>ArathTUB3-1</i> <i>ArathTUB3-2</i>
<i>CsTUB11-1</i>	25	41	46	44	42
<i>CsTUB11-2</i>	86	28	36	23	50
<i>CsTUB14-1</i>		83	91	85	78
<i>CsTUB14-2</i>		31	38	24	49
<i>CsTUB10-1</i>			81	81	77
<i>CsTUB10-2</i>			61	46	51
<i>CsTUB12-1</i>				78	67
<i>CsTUB12-2</i>				31	64
<i>AraTUB2-1</i>					81
<i>AraTUB2-2</i>					66

	<i>CsTUB2-1</i> <i>CsTUB2-2</i>	<i>CsTUB3-1</i> <i>CsTUB3-2</i>	<i>ArathTUB8-1</i> <i>ArathTUB8-2</i>
<i>CsTUB1-1</i>	89	97	26
<i>CsTUB1-2</i>	92	90	31
<i>CsTUB2-1</i>		89	34
<i>CsTUB2-2</i>		95	11
<i>CsTUB3-1</i>			42
<i>CsTUB3-2</i>			25

	<i>CsTUB9-1</i> <i>CsTUB9-2</i>	<i>ArathTUB6-1</i> <i>ArathTUB6-2</i>
<i>CsTUB6-1</i>	92	45
<i>CsTUB6-2</i>	90	85
<i>CsTUB9-1</i>		45
<i>CsTUB9-2</i>		86

2.8. Appendix—Table 2.8.1. Percentage of identity of the intron 1 and intron 2 among some of the *C. sativa* and *A. thaliana* paralogous and orthologous β -tubulin genes. The first line indicates the percentage of identity among intron 1 sequences, while the second line indicates the percentage of identity among intron 2 sequences. The grey and white cells indicate camelina and *Arabidopsis* intron groups, respectively.

2.9. Reference

- Abajian C (1994)** SPUTNIK, <http://www.abajian.com/sputnik>.
- Bardini M, Lee D, Donini P, Mariani A, Gianì S, Toschi M, Lowe C, Breviario D (2004)** Tubulin-based polymorphism (TBP): a new tool, based on functionally relevant sequences, to assess genetic diversity in plant species. *Genome* **47**: 281–291.
- Benson G (1999)** Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**:573–580.
- Braglia L, Manca A, Mastromauro F, Breviario D (2010)** cTBP: a successful ILP-based genotyping method targeted to well defined experimental needs. *Diversity* **2**:572–585.
- Breviario D, Vance Baird Wm, Sangoi S, Hilu K, Blumetti P, Gianì S (2007)** High polymorphism and resolution in targeted fingerprinting with combined β -tubulin introns. *Mol Breed* **20**:249–259.
- Casazza AP, Gavazzi F, Mastromauro M, Gianì S, Breviario D (2011)** Certifying the feed to guarantee the quality of traditional food: an easy way to trace plant species in complex mixtures. *Food Chem* **124**:685–691.
- Doyle JJ, Doyle JL (1987)** A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytoch Bull* **19**:11–15.
- Einax E, Voigt K (2003)** Oligonucleotide primers for the universal amplification of β -tubulin genes facilitate phylogenetic analyses in the regnum Fungi. *Org Divers Evol* **3**:185–194.
- Farajalla MR, Gulick P (2007)** The α -tubulin gene family in wheat (*Triticum aestivum* L.) and differential gene expression during cold acclimation. *Genome* **50**:502–510.
- Flannery ML, Mitchell FJ, Coyne S, Kavanagh TA, Burke JI, Salamini N, Dowding P, Hodkinson TR (2006)** Plastid genome characterisation in Brassica and Brassicaceae using a new set of nine SSRs. *Theor Appl Genet* **113**:1221–1231.
- Gehringer A, Friedt W, Luhs W, Snowdon RJ (2006)** Genetic mapping of agronomic traits in false flax (*Camelina sativa* subsp. *sativa*). *Genome* **49**:1555–1563.
- Ghamkhar K, Croser J, Aryamanesh N, Campbell M, Kon'kova N, Francis C (2010)** Camelina (*Camelina sativa* (L.) Crantz) as an alternative oilseed: molecular and ecogeographic analyses. *Genome* **53**:558–567.
- Gianì S, Breviario D (1996)** Rice β -tubulin mRNA levels are modulated during flower development and in response to external stimuli. *Plant Sci* **116**:147–157.
- He X-C, Qin Y-M, Xu Y, Hu C-Y, Zhu Y-X (2008)** Molecular cloning, expression profiling, and yeast complementation of 19 β -tubulin cDNAs from developing cotton ovules. *J Exp Bot* **59**:2687–2695.
- Hillis DM, Bull JJ (1993)** An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* **42**:182–192.
- Jost W, Baur A, Nick P, Reski R, Gorr G (2004)** A large plant β -tubulin family with minimal C-terminal variation but differences in expression. *Gene* **340**:151–160.
- Knorz KH (1978)** Evolution and spread of gold of pleasure (*Camelina sativa* S.L.). *Ber Dtsch Bot Ges* **91**:187–195.
- Liaud MF, Brinkmann H, Cerff R (1992)** The β -tubulin gene family of pea: primary structures, genomic organization and intron-dependent evolution of genes. *Plant Mol Biol* **18**:639–651.

2. Publication 1. *h-TBP: an approach based on intron-length polymorphism for the rapid isolation and characterization of the multiple members of the β -tubulin gene family in Camelina sativa (L.) Crantz*

- Lorkovic ZJ, Wieczorek Kirk DA, Lambermon MHL, Filipowicz W (2000)** Pre-mRNA splicing in higher plants. *Trends Plant Sci* **5**:160–167.
- Moloney AP, Woods VB, Crowley JG (1998)** A note on the nutritive value of camelina meal for beef cattle. *Irish J Agric and Food Res* **37**:243–247.
- Oakley RV, Wang YS, Ramakrishna W, Harding SA, Tsai CJ (2007)** Differential expansion and expression of α - and β -tubulin gene families in *Populus*. *Plant Physiol* **145**:961–973.
- Palumbi SR (1995)** Nucleic acids II: the polymerase chain reaction. In: Hillis D, Moritz C (eds) *Molecular systematics*, 2nd edn. Sinauer Associates Inc, Sunderland, MA, pp 205–247.
- Putnam DH, Budin JT, Field LA, Breene WM (1993)** Camelina: a promising low-input oilseed. In: Janick J, Simon JE (eds) *New crops*. Wiley, New York, pp 314–322.
- Schill SR (2009)** Camelina meal approved for feedlot cattle. *Biodiesel Mag*. http://www.biodieselmagazine.com/article.jsp?article_id=3837.
- Sneath PHA, Sokal RR (1973)** *Numerical taxonomy*. Freeman, W. H. and Co., San Francisco, CA, USA, p 573.
- Snustad DP, Haas NA, Kopczak SD, Silflow CD (1992)** The small genome of *Arabidopsis* contains at least nine expressed β -tubulin genes. *Plant Cell* **4**:549–556.
- Van de Peer Y, De Wachter R (1994)** TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput Appl Biosci* **10**:569–570.
- Vollmann J, Grausgruber H, Stift G, Dryzhyruk V, Lelley T (2005)** Genetic diversity in camelina germplasm as revealed by seed quality characteristics and RAPDs polymorphism. *Plant Breed* **124**:446–453.
- Vollmann J, Moritz T, Kargl C, Baumgartner S, Wagentristl H (2007)** Agronomic evaluation of camelina genotypes selected for seed quality characteristics. *Indust Crops Prod* **26**:270–277.
- Wang X, Zhao X, Zhu J, Wu W (2005)** Genome-wide investigation of intron length polymorphisms and their potential as molecular markers in rice (*Oryza sativa* L.). *DNA Res* **12**:417–42

3. Publication 2

Development of simple sequence repeat (SSR) markers in *Camelina sativa* (L.) Crantz

Antonella Manca, Incoronata Galasso.

Published 2010 in MINERVA BIOTEC; 22 Suppl. 1 No. 2:43-5.

3.1. Introduction

Camelina sativa is an oilseed plant that belongs to the *Brassicaceae* family. Camelina was widely grown in Europe and Russia until the 1940s for the production of vegetable oil but later on higher-yielding crops displaced camelina.

Recently, interest in *C. sativa* has been renewed in some parts of North America and Europe due to its high content of unsaturated fatty acids suitable for the production of soap, varnish, and biofuel. In fact, the oil contained in its seeds is very rich in polyunsaturated fatty acids such as linolenic acid (35% - 45%), linoleic acid (15% - 20%), and a low proportion of erucic acid (1% - 3%). It is also rich in antioxidants such as tocopherols, which confers stability towards oxidation. Camelina in comparison to other oilseed plants, has a very short life cycle (85-100 days) and is one of the most economical plant to grow thanks to minimal input requirements¹. For this reason *C. sativa* can grow on marginal lands avoiding competition with food crops.

These properties, together with its low nutrient demand, potentially make the production of *C. sativa* oil cheaper than oils from traditional crops such as soybean, sunflower and rapeseed. Despite its potential, the main limitation in camelina use is the lack of agronomic knowledge as well as the limited information about the genetic diversity of available germplasm. For this reason, a better knowledge of camelina genetics becomes incumbent.

Microsatellites, also known as simple sequence repeats, are short repeated nucleotide motifs usually one to four base pairs in length that are flanked by conserved sequences.

They are considered one of the most useful classes of markers to assess genetic diversity within a species. So far, no microsatellite sequences have been isolated from camelina, and very few nucleotide sequences have been deposited in the EMBL nucleotide database. For the above reasons, the development of SSR markers in camelina is of great importance.

3.2. Material and methods

3.2.1. Plant Material and DNA extraction

Seeds of different accessions of *C. sativa* were kindly provided by the Institute of Plant Genetics and Crop Plant Research (IPK), Germany. DNA was extracted from plant leaf material using either the GenElute™ Plant Genomic DNA Miniprep Kit (Sigma-Aldrich) or according to De Miera and de la Vega². DNA from *C. sativa* var. Calena was used for the genomic library development. Construction of an enriched microsatellite library *Camelina sativa* var. Calena genomic DNA (10 µg/ul) was digested with *RsaI* restriction enzyme.

Blunt-end DNA fragments were ligated by T4 DNA ligase to a *MluI* adaptor consisting of a 21- and 25-mer primer.

Fragments containing microsatellites were selected by hybridisation with the biotinylated synthetic oligonucleotide (GA)₁₅ and recovered by magnetic beads (MagneSphere Magnetic, Promega Madison, USA) linked to streptavidin³. Beads were then washed several times at high-stringency conditions and the captured DNAs were eluted from the beads. Enriched fractions were amplified by PCR, using primers complementary to the adapters. Amplification products were purified, cloned into pGEM-T vector and then transformed into *Escherichia coli* cells.

Recombinant plasmids were extracted from the white colonies and sequenced. Searching for microsatellite repeats was carried out using the program Sputnik (<http://cbi.labri.fr/outils/Pise/sputnik.html>). Primers flanking each unique microsatellite were designed using Primer3 (<http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi>). All PCR reactions using the SSR markers were performed in a total volume of 15 µl with the following program: 94°C for 4 min, followed by 36 cycles (30s at 94°C, 30s at optimal annealing temperature, and 30s at 72°C), and ending with 30 min extension at 72°C. The amplified PCR products were separated on 4% agarose gels and visualized using ethidium bromide staining.

3.3. Results and discussion

Microsatellite marker has become one of the most widely used molecular markers for genetic analysis in recent years. The enrichment of DNA fragments through the binding of the SSR probe is simple and efficient approach for microsatellite isolation and has been successfully applied to a number of plant genomes^{3,4}. Based on the methodologies applied for onion and various *Brassicaceae* species, we constructed the first camelina SSR enriched genomic library and isolated microsatellite markers. The microsatellite enriched library, consisting in approximately 800 clones, was constructed using the enrichment procedure using streptavidin-coated magnetic beads.

From this library, we sequenced about 100 clones and 70% of them showed to contain microsatellites. Based on sequence data, all of them contained GA/CT repeats with a minimum and maximum repeat length of 13 and 50, respectively. Although, all this nucleotide sequences contain the microsatellite (GA)_n of interest, not all were suitable for primer design due to the too close distance between the microsatellite and the cloning site within the vector.

As a result, we succeeded in designing specific primers for about 40 sequences. This first analysis allows us to point out that about 15 out of 40 primer pairs amplified a single fragment of expected molecular size on *C. sativa* var. Calena, while the remaining primers amplified two or more fragments (Figure 3.1).

This preliminary result indicates that *C. sativa* (chromosome number: 2n=40) may be a polyploid and not a diploid species as previously thought. This is in accordance with the recently suggestion made by Gehringer et al.⁵, (2006) that, on the basis of SSR and ALFP data, reached the same conclusion. In order to test whether the isolated SSRs were polymorphic seven of them were used to examine the amplification of microsatellite loci using genomic DNA of 10 camelina genotypes as templates.

The result of this analysis demonstrated that five SSR markers out of seven showed polymorphisms among the ten genotypes (Figure 3.2). All the primers designed in this work will be analysed using an automated fragment analyzer ABI Prism 310 (Applied Biosystems).

In conclusion, in this paper we reported the isolation and characterization of specific camelina microsatellite markers by using a *C. sativa* SSR-enriched library and our preliminary results indicate that the polymorphic microsatellite markers identified will undoubtedly be important for germplasm characterization and assessment of genetic diversity in our camelina collection.

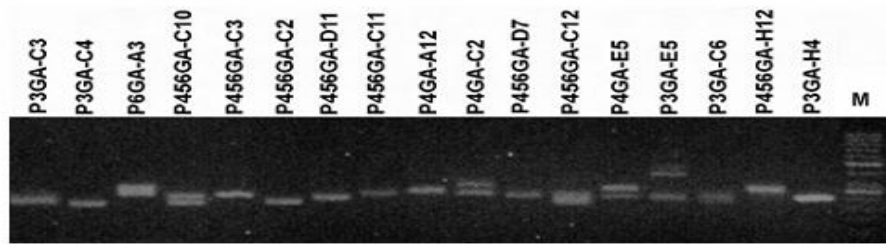


Figure 3.1. Some of the SSR markers isolated from *C. sativa* var. Calena. M = 50 bp DNA ladder.

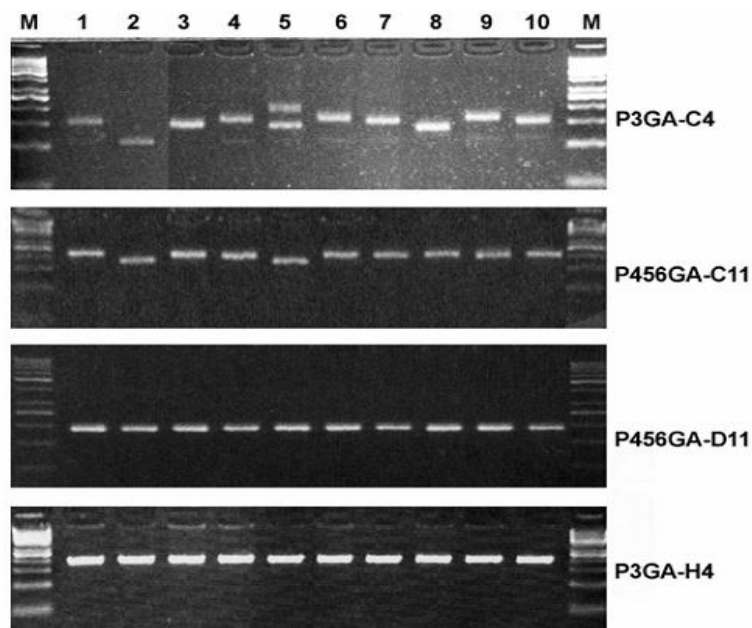


Figure 3.2. Examples of 3 SSR markers (P3GA-C4, P456GA-C11 and P456GA-D11) showing polymorphisms among the 10 analysed genotypes and one example of one monomorphic SSR marker (P3GA-H4). Numbers on top indicate: (1) var. Calena; (2) var. Ligena; (3) CAM45; (4) CAM40; (5) CAM172; (6) CAM180; (7) CAM46; (8) CAM120; (9) FF084; (10) FF06. M = 50 bp DNA ladder.

3.4. References

1. **Putnam DH, Budin JT, Field LA, Breene WM.** Camelina: A promising low-input oilseed. In: J. Janick J, Simon JE, editors. New crops. Wiley, New York; 1993. p. 314-2.
2. **De Miera LES, de la Vega MP.** A comparative studies of vicilin genes in Lens: negative evidences of concerted evolution. *Mol Biol Evol* 1998; **15**:303-1.
3. **Fischer D and Bachmann K.** Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.). *BioTechniques* 1998; **24**:796-2.
4. **Nunome T, Negoro S, Kono I, Kanamori H, Miyatake K, Yamaguchi H, et al.** Development of SSR markers derived from SSR-enriched genomic library of eggplant (*Solanum melongena* L.). *Theor Appl Genet* 2009; **119**:1143-3.
5. **Gehring A, Friedt W, Luhs W, Snowdon RJ.** Genetic mapping of agronomic traits in false flax (*Camelina sativa* subsp. *sativa*). *Genome* 2006; **49**:1555-3.

4. Manuscript 1

Evaluation of genetic diversity in one *Camelina sativa* (L) Crantz collection using SSR markers and biochemical traits.

Antonella Manca, Paola Pecchia, Sergio Mapelli, Piernicola Masella, Incoronata Galasso

Istituto di Biologia e Biotecnologia Agraria, Via E. Bassini 15,
20133 Milano

4.1. Abstract

A genomic DNA library enriched with GA/TC repeats from *Camelina sativa* var. Calena has been analysed. After sequencing of about 200 randomly selected clones, approximately 60% of them showed to contain simple or compound microsatellites with a high number of repeats. Among all SSRs analysed 15 primer pairs amplified polymorphic fragments. Fourteen camelina accessions of different origin were genotyped with 15 SSR markers that generated 134 alleles with an average of 8.93 alleles per locus. The observed heterozygosity (H_o) among the accessions ranged from 0.000 to 0.1500 with an average of 0.0370 whereas the average expected heterozygosity (H_e) among accessions was 0.2769. The analysis of the total heterozygosity ($H_T = 0.651$), the intra-population genetic diversity ($H_S = 0.260$) and the inter-population genetic diversity ($D_{ST} = 0.391$) demonstrated that 60% of the genetic differentiation is among the accessions, while 40% resides within them.

Phylogenetic tree of the 40 camelina accessions was constructed based on Nei's genetic distance. The UPGM (Unweighted Pair Group Method) dendrogram shows, except for G1-CAM108 and G9-CAM170, a clear discrimination among camelina accessions grouping them in five sub-groups.

ANOVA analysis indicates significant differences in some biochemical/agronomic parameters among the *C. sativa* accessions grouped according to Nei's genetic distance.

Keywords: SSR, Microsatellite, *Camelina sativa*, Genetic diversity, Biochemical traits

4.2. Introduction

Camelina sativa (L.) Crantz, is an annual oilseed plant belonging to the Brassicaceae family with chromosome number $2n=40$ and genome size close to 750 Mbp (Hutcheon et al. 2010). Very recently, this ancient non-food plant has been widely recognised as a potential alternative resource for biofuel production (Fröhlich et al. 2003; Patil et al. 2009; IATA report 2010). The identification of new energy plants that do not compete with food crops, such as *Glycine max* (L.) Merr *Helianthus annuus* (L) and *Elaeis guineensis* Jacq., can provide an effective solution to the food versus fuel issue. Several papers reported that *C. sativa* is a suitable plant for its low input requirements, with reduced fertilizations and pesticides, and because is capable of growing in marginal lands (Zurb, 1997; Vollmann et al. 2007; Sèguin-Swartz et al. 2009). Given these features, *C. sativa* will not displace or compete with other food-crops and will not impact on food supply.

Despite all this potential, few are the camelina varieties that have been genetically improved and genetic and genomic characterization of camelina is still lagging out. In fact, the first preliminary genetic map of *C. sativa* has been published in 2006 using 157 Amplified Fragment Length Polymorphism (AFLP) markers and three *Brassica* Simple Sequence Repeats (SSRs) (Gehring et al. 2006). Also the number of papers studying the genetic diversity in camelina germplasm is limited. The first work on the use of molecular markers for genetic characterization in camelina has been published by Vollmann et al. (2005). Vollmann and co-authors using 24 RAPD primers on 41 camelina accessions found a low degree of genetic diversity. Later on Ghamkhar et al. (2010), using 8 AFLP markers on 53 camelina accessions collected from different geographic locations, found a link between oil quality traits and the ecogeography collection site. Although sampling was relatively biased, the Russian-Ukrainian area was suggested to be a genetic diversity hotspot and the possible centre of origin of camelina.

Molecular markers are useful tools for germplasm characterization, cultivar fingerprinting and biodiversity study in plants. Several types of molecular markers have been described in the literature (Semagn et al. 2006). Their use has greatly improved our understanding of crop plants genetic diversity that may occur between or within accessions or groups of individuals. Knowledge of the genetic variation present in different crop accessions is not only important for germplasm characterization but also for evolution studies and for conservation purposes. Among the many molecular techniques available, microsatellites, or

simple sequence repeats (SSRs), are becoming increasingly widespread since they are highly polymorphic, codominant, reproducible and relatively easy to interpret. Belaj et al. (2003), comparing SSRs, RAPDs and AFLPs for their discriminating power/effectiveness in olive tree (*Olea europaea* L.) cultivars, reported the highest level of polymorphisms for SSR molecular marker. Similar results were also reported in soybean and in others species when SSRs were compared to other molecular markers (Powell et al. 1996; Russell et al. 1997; Pejic et al. 1998).

Microsatellite markers (Tautz, 1988, Morgante and Olivieri, 1993; Bell and Ecker, 1994), are tandem arrays of 1 to 6 base repeat units that are widely distributed in all eukaryotes and prokaryotes genomes (Zane et al. 2002). Microsatellites are present in the nuclear, in the plastid and mitochondrial genome (Chung et al. 2006; Rajendrakumar et al., 2007), and are distributed in both coding and noncoding regions. The rate of mutation in microsatellites is very high and they are thought to play a significant role in genome evolution by creating and maintaining quantitative genetic variation (Kashi et al. 1997). However, since microsatellite sequences differ in different organisms and even in closely related species, the level of transferability of SSR markers across related species may be quite moderate (Thiel et al. 2003). This means that microsatellite sequence must be developed for each species.

Recently, a genomic library enriched for GA-microsatellite sequences has been developed by Manca and Galasso (2010) using genomic DNA from *C. sativa* var. Calena. This library, consisting of approximately 800 clones, showed an enrichment in microsatellites close to 60%. It was used in this work for the following purposes: (1) to develop specific *C. sativa* SSR markers (2) to analyse the genetic variability in a collection of 40 *C. sativa* accessions of different origin using the newly developed SSR markers (3) to analyse the level of correlation between identified SSR camelina-groups and some biochemical and agronomic traits

4.3. Material ad methods

4.3.1. Plant materials and DNA extraction

Fourteen different *C. sativa* accessions were analysed in this study. The list of all accessions with their origin, donor code and relative accession number is reported in Table 4.1. Plants were grown in an open field during autumn 2008 and spring 2009. In particular one autumn sowing (1st October 2008) was carried out only for the accessions G16-CAM31, G19-CAM37 and G33-CAM76 since they

require vernalization while for the remaining accessions (37 accessions) the sowing was performed in spring (the 2nd of April 2009). The camelina accessions were sown in small plots of 1.4 m x 0.5 m size with an inter-row of 16 cm, using a sowing rate of 600 seeds/m². Nitrogen fertilization was applied at a rate of 30 kg/ha both after sowing and at the beginning of internodes elongation. Single camelina plots were harvested at full maturity in July (the 4th of July) and then the seeds were air dried at 35°C in a ventilated chamber until 8% seeds humidity was reached. Five weeks after sowing, fresh leaf material was collected from 8 single plants from each accession for a total of 320 individuals, frozen in liquid nitrogen and stored at -80°C up until use. Genomic DNA was extracted from leaf tissue of eight individuals of each accession using the “GenElute Plant Genomic DNA Miniprep Kit” (SIGMA) according to the manufacturer’s instructions.

4. Manuscript 1- Evaluation of genetic diversity in one *Camelina sativa* (L) Crantz collection using SSR markers and biochemical traits.

Code Number	Accession number	Scientific name	Accession name	Country of origin	Donor
G1	CAM108	<i>Camelina sativa</i>	Bronowska	Poland	IPK ⁺
G2	CAM110	<i>Camelina sativa</i>	Volynskaja	Poland	IPK
G3	CAM111	<i>Camelina sativa</i>	Omskaja	URS [∞]	IPK
G4	CAM116	<i>Camelina sativa</i>		Belgium	IPK
G5	CAM123	<i>Camelina sativa</i>	Czenstochwska	Poland	IPK
G7	CAM136	<i>Camelina sativa</i>	Sortadinskij	Poland	IPK
G8	CAM137	<i>Camelina sativa</i>		Denmark	IPK
G9	CAM170	<i>Camelina sativa</i>		Poland	IPK
G11	CAM173	<i>Camelina sativa</i>	Irkutskja Mestnaja	URS	IPK
G13	CAM175	<i>Camelina sativa</i>		Sweden	IPK
G14	CAM187	<i>Camelina sativa</i>		Spain	IPK
G15	CAM25	<i>Camelina sativa</i>	Omskij Mestnyj	URS	IPK
G16	CAM31	<i>Camelina sativa</i>		Poland	IPK
G17	CAM34	<i>Camelina sativa</i>	Omskij	URS	IPK
G18	CAM35	<i>Camelina sativa</i>	Ukrajinskij	URS	IPK
G19	CAM37	<i>Camelina sativa</i>	Zavolzkij	URS	IPK
G20	CAM38	<i>Camelina sativa</i>		Austria	IPK
G21	CAM39	<i>Camelina sativa</i>		Austria	IPK
G22	CAM58	<i>Camelina sativa</i>		Germany	IPK
G23	CAM7	<i>Camelina sativa</i>	Kingizskij	Kyrgyzstan	IPK
G24	D11852	<i>Camelina sativa</i>		URS	IPK
G25	D11854	<i>Camelina sativa</i>		Bulgaria	IPK
G26	D11855	<i>Camelina sativa</i>		United Kingdom	USDA ^o
G27	D11857	<i>Camelina sativa</i>		Switzerland	IPK
G30	FF004	<i>Camelina sativa</i>	St.Pernitzen	Austria	Arche Noah ^x
G31	AMES26669	<i>Camelina sativa</i>	Cs•er1672	Sweden	USDA
G32	AMES26665	<i>Camelina sativa</i>	Cs•163•207372	Denmark	USDA
G33	CAM76	<i>Camelina sativa</i>	Zavolzkaja, Auslese 2	URS	IPK
G35	CAM29	<i>Camelina sativa</i>	Volyn'skaja	Ukraine	IPK
G36	CAM8	<i>Camelina sativa</i>	Sortadinskij	URS	IPK
C1	—	<i>Camelina sativa</i>	Calena	Germany	BSV Saaten [±]
C2	—	<i>Camelina sativa</i>	Ligena	Germany	Dott. Snowdon [‡]
C3	CAM45	<i>Camelina sativa</i>	Ukrajinskaja	URS	IPK
C4	CAM40	<i>Camelina sativa</i>	Lindo	Unknown	IPK
C5	CAM172	<i>Camelina sativa</i>	Zarja Socialisma	URS	IPK
C6	CAM180	<i>Camelina sativa</i>	Bavaria	Germany	IPK
C7	CAM46	<i>Camelina sativa</i>	Soledo	Unknown	IPK
C8	CAM120	<i>Camelina sativa</i>	Wroclawska	Poland	IPK
C9	FF084	<i>Camelina sativa</i>	Morgesonne	Austria	Arche Noah
C10	FF006	<i>Camelina sativa</i>	Kärntner	Austria	Arche Noah

— Samples without accession number.

[∞] URS Union of Republic Soviet.

^o USDA United States Department of Agriculture, USA.

⁺ IPK Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany.

^x Arche Noah The Austrian Seed Savers Association, Austria - Europe private seed bank.

[±] BSV Saaten Bayerische Futtersaatenbau GmbH- Germany-Seed company.

[‡] Dott. Snowdon Rod Justus-Liebig-University Giessen.

Table 4.1. List of 40 accessions of *Camelina sativa* used in this study with code number and their accession number, accession name, country of origin and donor

4.3.2. Analysis of camelina SSR-enriched genomic library

In order to find microsatellite markers useful for camelina characterization the GA-enriched library developed by Manca and Galasso (2010), was used. The GA-enriched-library, consisting of about 800 clones, was obtained by hybrid capture method as described in Manca and Galasso (2010). Plasmid DNAs from about 200 randomly selected clones were isolated using the Wizard[®] SV 96 Plasmid DNA purification System (Promega) and sequenced with an automated sequencer by Macrogen, (Seoul, South Korea) using the universal T7 Sequencing Primer.

The identification of microsatellites within the cloned sequences was carried out using the Sputnik program, which is publicly available <http://www.cbib.u-bordeaux2.fr/pise/> (Abajian, 1994). The nucleotide sequences obtained are available in the EMBL (European Molecular Biology Laboratory) Nucleotide Database.

4.3.3. Primer design and PCR amplification

Primer pairs, flanking each unique microsatellite, were designed using the web-based computer program Primer3 version 0.4.0 (<http://frodo.wi.mit.edu/primer3/>) (Rozen and Skaletsky 1998) Primers were designed following the program's default parameters, with the following exceptions: product size range from 150 to 300 bp and melting temperature (T_m) differences in forward and reverse primers not more than 5°C - 6°C. Following the method described by Schuelke (2000) all PCR reactions were performed with three primers: a forward primer with M13(-21) tail (5'-TGT AAA ACG ACG GCC AGT-3') to its 5' end, a reverse primer, and the universal M13(-21) primer fluorescent labelled with one of the following fluorescent dyes: 6-FAM (6-carboxyfluorescein), VIC (2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein) or NED (7',8'-benzo 5'-fluoro-2',4,7 trichloro-3 carboxyfluorescein) (Applied Biosystems). Loci were amplified in a total volume of 14 μ l using 12.5 ng of genomic DNA, 7ul of GoTaq[®] Green Master Mix (Promega), 0.2 μ M of reverse primer, 0.2 μ M of fluorescence labelled universal M13(-21) primer and 0,07 μ M of forward primer with M13(-21) tail.

All PCR reactions were performed in a 2720 Thermal Cycler (Applied Biosystems) using conditions optimised for each primer pairs. The amplification condition was 5 min initial denaturation at 94°C, followed by 40 cycles of denaturation at 94°C for 30 s, primer annealing at optimal annealing temperature

(see Ta, Table 4.2) for 30 s, and extension at 72°C for 30 s, with a final extension at 72°C for 30 min. In some cases, two lightly different touchdown PCR protocols (called TD1 and TD2) were used. TD1: consisted of an initial denaturation step of 5 min at 94°C followed by 10 cycles of 30 s at 94°C, 30 s annealing temperature at 65°C reduced by 1°C every cycle, and 30 s at 72°C. This was followed by 30 cycles of 30 s at 94°C, 30 s at 55°C, 30 s at 72°C and one final extension step at 72°C for 30 min. TD2 program differed from TD1 profile in annealing temperature and number of cycles. The annealing temperature was reduced in steps of 0.5°C every cycle (8 cycles), from 54°C to 50°C, followed by 32 cycles at 50°C.

An amount of PCR product ranging from 0.5 to 2µl, depending on the performance of the amplification of each primer pairs, was mixed with 13µl of Hi-Di formamide and 0.25 µl of LIZ-500 size standard and denatured at 95°C for 5 min. Up to three PCR products labelled with FAM, VIC, and NED respectively, were pooled together before separation in the ABI 310 Genetic Analyzer (Applied Biosystems Inc., USA). All the SSRs allele sizes were estimated by using GeneMapper software v4.0 (Applied Biosystems Inc., USA).

4.3.4. Genetic data analysis

Genetic diversity parameters such as percentage of polymorphic loci (P_p), observed number of alleles (N_a), effective number of alleles (N_e) (Kimura and Crow 1964), observed (H_o) and expected (H_e) heterozygosity (Levene, 1949) and genetic distance (Nei, 1978) among populations/accessions were analysed using POPGENE 1.32 software package (Yeh et al. 1999). Finally, MEGA 3.1 software was combined with POPGENE 1.32 to generate the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) dendrogram based on genetic distance of Nei's (1978). While, the quantification of genetic diversity within and among populations, based on Nei's (1973) genetic diversity statistics, was calculated using DISPAN program (Ota, 1993). In particular, were calculated for each population the total gene diversity (H_T), the intra-population genetic diversity (H_S) and the inter-population diversity (D_{ST}). Finally, the coefficient of differentiation among populations $G_{ST}=D_{ST}/H_T$ was also determined for each accessions.

4.3.5. Seed quality traits

From each camelina accession, three samples of 0.5 g of dry seeds were ground in a homogenizer and mixed with hexane (1.5 mL/g tissue). The solution was energetically shaken for 30 min. The extract was then centrifuged (Sorvall rotor HB4) at 7000 rpm for 20 min and the upper-liquid collected. The extraction procedure was repeated on the sediment for three times. The hexane was evaporated in nitrogen flow until constant volume and weight, to have the pure oil fraction.

The extracted oils were hydrolyzed to obtain fatty acids (FAs). Essentially, an aliquot of oil (20 μ l) was hydrolyzed in 1 ml of 1% NaOH in MeOH, at 80°C for 60 min. The solution was dried under vacuum and the residue was dissolved in 2 ml of H₂O plus 0.3 ml of 1 N H₂SO₄, and then energetically shaken. The fatty acids, recovered from water sulfuric acid solution with 1 ml of hexane, were analyzed by the qualitative and quantitative HPLC system Jasco PU2089 pump equipped with Alltech 3300 ELSD detector using a Phenomenex Luna 2u C8 column (150 \times 4.6 mm), in thermostat at 18°C, eluted with acetonitrile: isopropanol: water (50: 30: 20 v/v, 1 mL/min) for 20 min. Between each chromatographic run the column was washed from triglycerids for 10 min with one elution of acetonitrile: isopropanol: water (50: 45: 5 v/v, 1 mL/min). Data from detector were recorded, integrated and elaborated by Borwin software program. The concentrations of individual fatty acids were expressed as percentage of total FA content.

Tocopherols have been analyzed in HPLC by direct injection of oil. The tocopherol isomers were separated using HPLC Jasco Tritotar III pump and Jasco Uvidec detector. Five-10 μ l of pure oil were loaded into a Merck Chromolith RP-18e column (100 \times 4.6 mm), and eluted with 1.5 ml/min MeOH 95%. The data at 280 nm were acquired and elaborated by the Borwin software system. The tocopherol amounts in oil samples have been determined by comparison with standard tocopherols.

The weight of 1000 seeds was determined by measuring 10 replicates of 100 seeds each. Seeds were counted manually and weighed in grams.

4.3.6. ANOVA analysis

In order to identify associations between the SSR camelina sub-groups obtained by the UPGMA and the biochemical and agronomic data, the significance of differences in biochemical/agronomic parameters was tested among accessions grouped by SSR markers. At this purpose an analysis of variance (ANOVA) and Tukey honestly significant difference (HSD) multiple comparison post hoc test (p at 0.05) were performed using SYSTAT Software v. 12.

4.4. Results

4.4.1. Genomic library and microsatellites characterization

The library generated by the hybrid capture method was highly enriched in repetitive DNA. In fact, after sequencing of 192 clones, about 60% of the inserts (115 sequences) showed to contain simple or compound microsatellites with a very high number of repeats. However, analysis carried out on these 115 clones has shown that 35% of the sequenced clones did not meet the selection criteria for primer design. This group includes clones with insufficient flanking sequence (31), clones whose repeats were too short (5), duplicate clones (2) or clones with more than two microsatellites very close to each other (2). Therefore primers were designed only on the remaining 65% (74) clones containing a repeated motif of adequate size and position within the cloned insert. Since the library was enriched for (GA) repeats this was the most common repeated motif observed in almost all the clones, nevertheless some (CA) and (GT) repeats, as well as (GAA) motifs were found. The number of (GA) repeats ranged from 13 to 45 in the perfect microsatellites (Table 4.2).

4. Manuscript 1- Evaluation of genetic diversity in one *Camelina sativa* (L) Crantz collection using SSR markers and biochemical traits.

Locus name	Primer pair sequence (5' - 3')	Repeat motif	Ta (°C)	No alleles	Size Range (bp)
P4B3	F:ATGGACGACGAGAGGTGTGA R:CGATCTGAGCAGGCTCCATT	(GA) ₂₃	55	1	210
P4D11	F:TGATTGAGCGGGAAGTGGAG R:CAACGCATACGTGGCGAGTT	(GA) ₂₃	58	1	190
P3H4	F:CCAGACAAACACACGAAGCCGAAT R:CTGGCGTAGGGTTTTAGTCCATT	(GA) ₁₃	58	3	200 • 204
P3C3	F:TTGGCCATTGCTTGGGTGTA R:ACTGGATTCCGTGCCTTGGT	(GA) ₃₁	55	7	178 • 222
P4C2	F:CGTCTCCCGTTGTCCAAGT R:TCACAAATCACCAATCCCAAG	(GA) ₁₅	58	6	174 • 186
P4H3	F:TCTGTGTCCGCATGGGAAGT R:CACGTAAGCAATCAAATCCCTCTC	(GA) ₃₁	TD1	4	219 • 225
P6E4	F:CCGTGGAGGAAGATGATTGAGA R:TGCCCCAAATCCAAGTTCC	CT(CTT) ₂ (CT) ₂₀ CTT	54	10	303 • 327
P6C2	F:CATGCTCTACTACCAAACCAACA R:TTGATCCTCTCTCAAGGTATCTCTTC	(GA) ₂₆	TD1	13	134 • 170
P4C7	F:GCCGTCTCCAAAGGCAGAAG R:GCAGCTCTCATCCCTGGTTTTG	(GA) ₂₁	58	9	257 • 276
P4E6	F:CACATCCAAAAGCTCTCTTCTCTT R:GACAGCGATGGTTTTAAGAAAGTTGA	(CT) ₂₆	58	8	189 • 226
LIB19	F:AACCGTGAACCTACTACAGAGAAAGA R:TCTGGGATTACCGAACCAA	(GA) ₂₅	54	8	115 • 151
P7A4	F:CCCAGGAAGCCACAGACA R:GCATAAATCGCAAATGAAAGAGTAG	(GA) ₂₆	TD 2	16	146 • 179
P7D9	F:CGGATAGATGACCCAGATCCA R:GTGGAGCATTCCATTCAGTCAG	(GA) ₂₃	TD 2	6	170 • 182
P7D4	F:TGGTCCCCCTGACGAAAGAA R:CCTCTGCCACAACCTTCACA	(GA) ₂₄	TD 2	9	129 • 173
P7C2	F:ATCCAACATGCTCACAAGG R:GGACATCAGACAACACAGAGA	(GA) ₂₀	TD 2	13	174 • 210
P7A8	F:GCGGTGGAGGAAAGGTTTTA R:CCGGCCCAATATCTCTGATG	(GA) ₁₁	TD 2	9	176 • 206
P4C11	F:TGAGCCTTTATAGAAGTTTCGGAACAA R:TGCTGATGTCACACGGAGGA	(GA) ₂₆	58	13	181 • 229
Total				134	

TD1 and TD2 = Touchdown PCR program 1 and 2.

Table 4.2. Characteristics of 17 microsatellite loci identified in *C. sativa*. Locus name, oligonucleotide primer sequences, repeat motif, annealing temperature (Ta), number (No) of alleles and allele size range in base pair (bp).

4.4.2. Characteristic of the microsatellite loci and genotypic structure of camelina accessions

In total, seventy-four microsatellite sequences were retrieved from the GA-enriched genomic library and used as candidates to design specific primers using the software Primer3. The amplification quality of these developed markers was previously evaluated using the genomic DNA isolated from *C. sativa* var. *Calena*, which was used for making the genomic library. Among the 74 available SSR primers tested, fifty-seven were immediately discarded because when products were resolved on agarose (3%) gel, poor amplification or multiple and complex bands were observed. On the contrary, the remaining 17 couple of primers amplified clear fragments consisting of one strong band of the expected size. Therefore only 17 SSR primer pairs of those originally designed were chosen to genotype a collection of *C. sativa* consisting of 40 accessions of different provenience (Table 4.1). The analysis was performed on 320 individuals of *C. sativa* (8 individuals for each accession).

Among the 17 primer pairs used 15 showed polymorphisms while 2 (P4B3 and P4B11) were monomorphic (Table 4.2). The expected mean heterozygosity (H_e) was higher than the observed mean heterozygosity (H_o) for all populations except for G9-CAM170. Average mean expected heterozygosity within the 40 camelina populations across all 15 loci was 0.2769, ranging from 0.0000 (G33-CAM76) to 0.6126 (G2-CAM110), while the observed average mean heterozygosity was 0.370, with G17-CAM34 scoring the highest (0.1500) value. Thirteen out of the 40 accessions exhibited homozygosity across all loci ($H_o = 0.000$) (Table 4.3.) and among them G33-CAM76 showed both H_o and H_e values equal to 0.0000, in all the 8 individuals analyzed with all the SSR markers used. Observed heterozygosity (H_o) and expected heterozygosity (H_e) for each locus are reported in Table 4.4. Observed heterozygosity values were lower than expected heterozygosity values for almost all loci. For the 15 microsatellite loci examined, the total number of alleles was 134 across all populations, with an average of 8.93. The observed number of alleles per locus ranged from 3 (P3H4) to 16 (P7A4). In particular, among the 15 loci the P7A4 locus showed the highest H_e value or genetic diversity ($H_e = 0.8644$) and P3H4 the lowest ($H_e = 0.2296$). Locus P7A8 showed the highest observed heterozygosity ($H_o = 0.0656$) and P4H3 the lowest ($H_o = 0.0063$).

4. Manuscript 1- Evaluation of genetic diversity in one *Camelina sativa* (L) Crantz collection using SSR markers and biochemical traits.

Code number	Accession number	Origin*	Genetic diversity indices					
			P_p	N_a	N_e	H_o	H_e	No P_L
G1	CAM108	PLN	13.33	1.1	1.0	0.0000	0.0311	2
G2	CAM110	PLN	100.00	2.9	2.6	0.0345	0.6126	15
G3	CAM111	URS	86.67	2.7	2.1	0.0881	0.4865	13
G4	CAM116	BBB	13.33	1.2	1.1	0.0000	0.0578	2
G5	CAM123	PLN	100.00	2.9	2.2	0.0595	0.5566	15
G7	CAM136	PLN	46.67	1.7	1.3	0.0083	0.2039	7
G8	CAM137	DKK	60.00	1.6	1.2	0.0083	0.1794	9
G9	CAM170	PLN	6.67	1.1	1.0	0.0167	0.0156	1
G11	CAM173	URS	80.00	2.3	1.7	0.0167	0.3667	12
G13	CAM175	SSS	40.00	1.4	1.2	0.0000	0.1600	6
G14	CAM187	EEE	60.00	1.7	1.4	0.0000	0.2422	9
G15	CAM25	URS	100.00	2.9	2.0	0.0833	0.4833	15
G16	CAM31	PLN	53.33	1.7	1.4	0.0417	0.1806	8
G17	CAM34	URS	100.00	2.9	2.3	0.1500	0.5733	15
G18	CAM35	URS	66.67	1.8	1.4	0.0250	0.2550	10
G19	CAM37	URS	80.00	2.4	1.8	0.1083	0.3756	12
G20	CAM38	AAA	66.67	2.2	1.7	0.0083	0.3228	10
G21	CAM39	AAA	73.33	2.6	2.1	0.0000	0.4556	11
G22	CAM58	DDD	33.33	1.3	1.2	0.0000	0.1067	5
G23	CAM7	KSS	26.67	1.3	1.1	0.0000	0.0622	4
G24	D11852	URS	66.67	1.9	1.5	0.0167	0.3011	10
G25	D11854	BGA	73.33	1.7	1.4	0.0083	0.2661	11
G26	D11855	GBR	26.67	1.3	1.2	0.0000	0.1089	4
G27	D11857	CHZ	46.67	1.5	1.3	0.0167	0.2100	7
G30	FF004	AAA	86.67	2.3	1.7	0.0607	0.3781	13
G31	AMES26669	SSS	73.33	2.1	1.7	0.0833	0.3389	11
G32	AMES26665	DKK	46.67	1.5	1.1	0.0000	0.1089	7
G33	CAM76	URS	0.00	1.0	1.0	0.0000	0.0000	0
G35	CAM29	UKR	93.33	3.5	2.6	0.0929	0.5878	14
G36	CAM8	URS	53.33	1.7	1.4	0.0250	0.2417	8
C1	— Calena	DDD	33.33	1.3	1.2	0.0250	0.1283	5
C2	— Ligena	DDD	66.67	1.9	1.4	0.0595	0.2156	10
C3	CAM45	URS	73.33	2.3	1.5	0.0512	0.3151	11
C4	CAM40	UNKW	66.67	1.7	1.1	0.0833	0.0983	10
C5	CAM172	URS	86.67	2.6	2.1	0.1000	0.4956	13
C6	CAM180	DDD	6.67	1.1	1.0	0.0000	0.0156	1
C7	CAM46	UNKW	6.67	1.1	1.0	0.0000	0.0156	1
C8	CAM120	PLN	86.67	3.2	2.7	0.0000	0.5311	13
C9	FF084	AAA	86.67	2.7	1.9	0.1095	0.4473	13
C10	FF006	AAA	93.75	2.8	2.3	0.1000	0.5467	14
	Mean		59.51	2.0	1.6	0.0370	0.2769	8.925

*AAA Austria; BBB Belgium; BGA Bulgaria; CHZ Switzerland; DDD Germany; DKK Denmark; EEE Spain; GBR United Kingdom; KSS Kyrgyzstan; PLN Poland; SSS Sweden; UKR Ukraine; URS Union of Republic Soviet; UNKW Unknown.

Table 4.3. Summary of some genetic diversity parameters: P_p (percent polymorphic loci); N_a (mean number of alleles in each accession or population); N_e (effective number of alleles), H_o (observed heterozygosity); H_e (expected heterozygosity); No P_L (number of polymorphic loci).

4. Manuscript 1- Evaluation of genetic diversity in one *Camelina sativa* (L) Crantz collection using SSR markers and biochemical traits.

Locus	Na	Ne	He	Ho
P3H4	3	1.30	0.2296	0.0125
P3C3	7	3.51	0.7159	0.0478
P4C2	6	2.61	0.6173	0.0345
P4H3	4	1.07	0.0673	0.0063
P6E4	10	2.26	0.5590	0.0219
P6C2	13	5.66	0.8249	0.0406
P4C7	9	4.50	0.7790	0.0375
P4E6	8	2.96	0.6632	0.0377
L1B19	8	3.27	0.6952	0.0533
P7A4	16	7.30	0.8644	0.0533
P7D9	6	3.40	0.7072	0.0281
P7D4	9	4.18	0.7618	0.0252
P7C2	13	6.12	0.8379	0.0377
P7A8	9	2.56	0.6108	0.0656
P4C11	13	6.34	0.8436	0.0505
Mean	8.93	3.80	0.6518	0.0368

Table 4.4. Genetic parameters in 40 genotypes of camelina with 15 microsatellite loci. **Na** = observed number of alleles; **Ne** = effective number of alleles; **Ho** = observed heterozygosity; **He** = expected heterozygosity.

4.4.3. Genetic diversity and phylogenetic tree

The indices of genetic diversity (H_T , H_S , D_{ST} , G_{ST}), genetic distance and genetic identity were calculate to investigate the genetic diversity between the 40 genotypes of *C. sativa*. The genetic diversity is reported in Table 4.5. The mean of inter-population diversity ($D_{ST}= 0.391$; 60%) was higher than the mean of intra-population genetic diversity ($H_S = 0.260$; 40%) indicating that the diversity among populations is greater than that observed within populations (Table 4.5). The total genetic mean diversity (H_T) ranged from 0.067, in the P4H3 locus, to 0.863, in locus P7A4, with an average value of 0.651, while the inter-population (G_{ST}) coefficient of gene differentiation resulted between 0.273, in locus P4H3, and 0.656 in P7A4, with an average value of 0.575. The smallest genetic distance was observed between G1-CAM108 and G9-CAM170 (0.023). Phylogenetic tree of the 40 camelina accessions was constructed based on Nei's genetic distance (Fig. 4.1). The UPGM (Unweighted Pair Group Method) dendrogram shows a clear discrimination among camelina accessions with the exception of G1-CAM108 and G9-CAM170. The 40 accessions distribute in 2 main groups. Group A includes 35 and group B includes only 5 accessions. Group A can be further subdivided into sub-groups A1 trough A4 (Fig. 4.1).

Locus	Nei's genetic diversity indices			
	H _T	H _s	D _{ST}	G _{ST}
P3H4	0.229	0.123	0.107	0.465
P3C3	0.712	0.281	0.431	0.605
P4C2	0.617	0.228	0.389	0.630
P4H3	0.067	0.049	0.018	0.273
P6E4	0.560	0.239	0.321	0.574
P6C2	0.824	0.366	0.457	0.555
P4C7	0.778	0.285	0.493	0.633
P4E6	0.664	0.263	0.400	0.603
LIB19	0.695	0.268	0.427	0.614
P7A4	0.863	0.297	0.566	0.656
P7D9	0.706	0.296	0.411	0.582
P7D4	0.761	0.310	0.451	0.592
P7C2	0.837	0.308	0.529	0.632
P7A8	0.610	0.233	0.376	0.617
P4C11	0.843	0.347	0.496	0.588
Mean	0.651	0.260	0.391	0.575

Table 4.5. Genetic diversity indices in 40 genotypes of *C.sativa* with 15 microsatellite loci. H_T = total genetic diversity, H_s = intra-population genetic diversity, D_{ST} = inter-population diversity and G_{ST} = coefficient of gene differentiation inter-population.

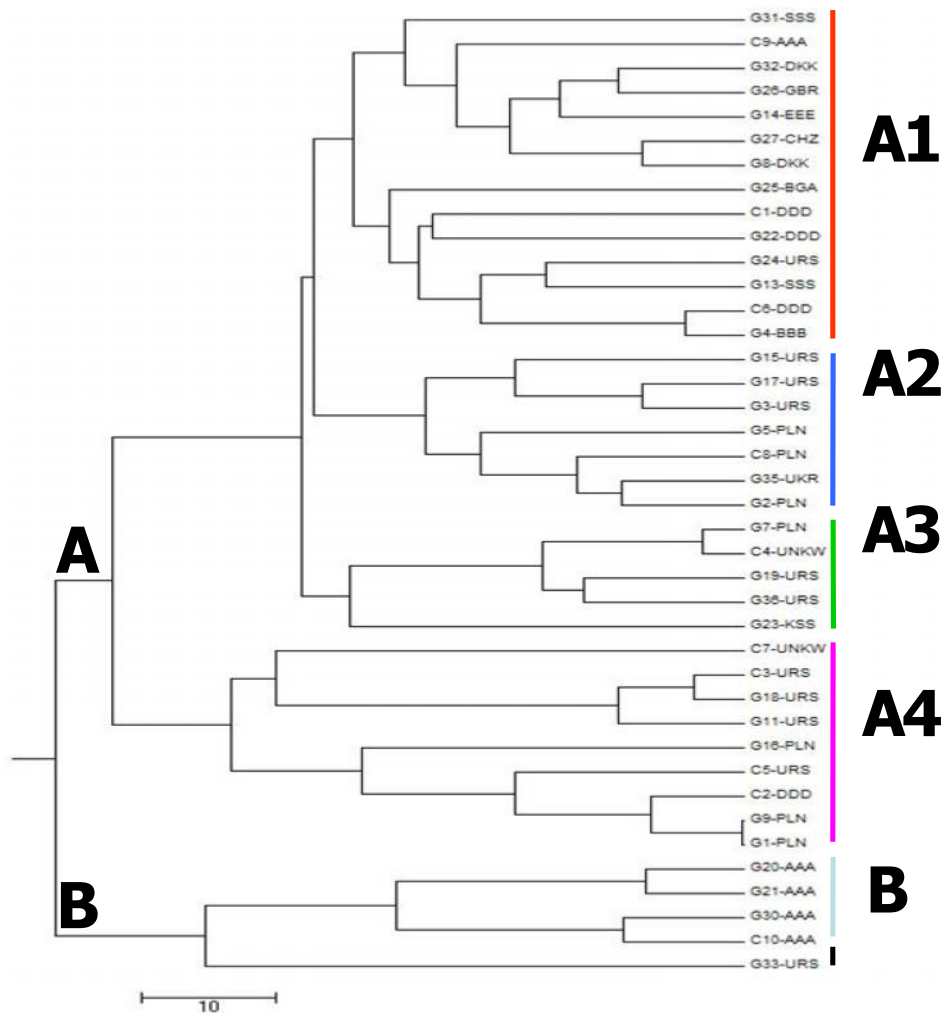


Figure 4.1. UPGMA dendrogram based on Nei's (1978) genetic distance matrix calculated on the SSR loci analysed. Code number as reported in Table 4.1. and provenience code: **AAA** Austria; **BBB** Belgium; **BGA** Bulgaria; **CHZ** Switzerland; **DDD** Germany; **DKK** Denmark; **EEE** Spain; **GBR** United Kingdom; **KSS** Kyrgyzstan; **PLN** Poland; **SSS** Sweden; **UKR** Ukraine; **URS** Union of Republic Soviet; **UNKW** Unknown.

4.4.4. Biochemical data

Oil content analysis shows the occurrence of relevant differences among the 40 genotype of *C. sativa* (Table 4.6.). In fact oil content (OIL) ranged from 27.28% in C3-CAM45 to 34.70% in the G25-D11856 accession. The fatty acid (FA) composition was primarily made up by poly- and mono-unsaturated fatty acids, as expected. The most abundant polyunsaturated FA is the linolenic acid (C18:3), that ranges from 45.85% to 69.03%, but a significant amount of linoleic acid (C18:2), from 12.64% to 28.28%, is also present. Among the monounsaturated FAs the most abundant are oleic acid (C18:1), that ranges from 10.94% to 19.51% and gondoic acid (C20:1) that may be present from 5.67% to 8.80%. Saturated FAs represent a smaller quantity of the total FAs that is mainly represented by palmitic acid (from 3.97 % to 5.02 %) (Table 4.6). With regard to tocopherols, the highest content was measured in the accession G11-CAM173 (2.07 µg/mg) and the lowest in C4-CAM40 (0.68 µg/mg).

The weight of the 1000 seeds (TWS) varied from 0.87g in G8-CAM137 to 1.41g C2-Ligena (Table 4.6).

4. Manuscript 1- Evaluation of genetic diversity in one *Camelina sativa* (L) Crantz collection using SSR markers and biochemical traits.

SSR Sub-group	Code number	Accession number	OIL %	PA %	OA %	LA %	ALA %	GA %	TOC µg/mg	TWS g/DW
A1	G4	CAM116	30.98	3.48	15.02	13.97	61.50	6.04	1.43	0.96
A1	G8	CAM137	29.07	2.64	12.71	16.81	62.15	5.70	1.37	0.87
A1	G13	CAM175	32.10	2.43	11.92	17.85	61.35	6.45	1.50	1.07
A1	G14	CAM187	31.10	2.83	11.09	12.88	67.23	5.98	1.68	0.90
A1	G22	CAM158	29.85	3.32	12.47	16.70	60.56	6.95	1.30	0.97
A1	G24	D11852	32.69	2.05	12.54	16.18	63.54	5.70	1.11	0.95
A1	G25	D11854	34.70	2.37	14.05	18.15	59.20	6.23	1.21	0.92
A1	G26	D11855	31.14	2.23	19.51	12.64	58.08	7.54	1.06	0.91
A1	G27	D11857	31.74	2.67	14.35	15.67	60.95	6.36	1.19	0.92
A1	G31	AMES26669	33.34	2.13	16.49	14.08	60.05	7.25	0.81	1.25
A1	G32	AMES26665	30.38	2.13	11.95	14.59	65.12	6.21	1.22	0.93
A1	C1	-- Calena	31.18	4.60	11.41	16.77	58.66	8.56	1.08	1.11
A1	C6	CAM180	33.47	4.70	13.64	13.23	60.53	7.89	1.02	0.90
A1	C9	FF084	30.66	4.07	11.20	15.41	62.58	6.74	0.96	0.97
Mean			31.60	2.98	13.45	15.35	61.54	6.68	1.21	0.97
A2	G2	CAM110	29.36	4.05	13.12	18.02	58.00	6.81	1.23	0.97
A2	G3	CAM111	28.83	3.75	11.62	19.29	58.42	6.92	1.56	1.14
A2	G5	CAM123	31.46	3.97	12.38	18.16	59.15	6.35	1.66	1.01
A2	G15	CAM125	31.80	3.07	13.55	18.50	59.05	5.84	1.61	1.00
A2	G17	CAM134	31.99	3.14	13.64	19.38	58.17	5.67	1.25	1.03
A2	G35	CAM129	32.53	2.16	12.90	18.17	60.98	5.79	1.97	0.98
A2	C8	CAM120	29.82	4.08	10.94	17.10	61.00	6.88	1.19	1.23
Mean			30.83	3.46	12.59	18.37	59.25	6.32	1.50	1.05
A3	G7	CAM136	31.42	4.22	14.84	17.86	57.13	5.95	1.41	1.02
A3	G19	CAM37	33.31	2.72	12.70	14.79	64.09	5.70	1.29	1.14
A3	G23	CAM7	30.88	2.32	14.66	20.47	55.72	6.83	1.40	1.24
A3	G36	CAM8	33.12	2.55	15.83	19.56	56.22	5.84	1.02	1.01
A3	C4	CAM40	30.97	5.02	16.68	22.63	47.53	8.15	0.68	1.09
Mean			31.94	3.37	14.94	19.06	56.14	6.49	1.16	1.10
A4	G1	CAM108	31.39	3.22	13.08	20.90	56.31	6.50	1.00	1.29
A4	G9	CAM170	32.48	3.04	12.15	23.57	55.15	6.08	1.42	1.30
A4	G11	CAM173	29.34	3.02	12.92	22.00	56.25	5.82	2.07	1.23
A4	G16	CAM31	31.53	2.60	10.04	11.93	69.03	6.39	1.39	1.30
A4	G18	CAM35	29.43	2.80	14.43	28.28	48.05	6.44	1.62	1.25
A4	C2	-- Ligena	31.37	6.45	14.38	20.84	50.82	7.50	0.76	1.41
A4	C3	CAM45	27.28	4.79	11.38	29.61	45.85	8.37	1.20	1.05
A4	C5	CAM172	32.08	4.56	12.70	19.49	55.14	8.11	1.05	1.09
A4	C7	CAM46	32.05	4.75	13.61	17.29	55.55	8.80	1.16	1.26
Mean			30.77	3.91	12.74	21.55	54.68	7.11	1.29	1.24
B	G20	CAM38	33.22	3.12	15.70	17.18	57.88	6.12	1.34	0.98
B	G21	CAM39	30.08	3.56	16.29	16.27	57.61	6.26	0.92	1.00
B	G30	FF004	33.85	2.72	15.75	15.77	58.56	7.20	0.82	0.96
B	C10	FF006	32.91	3.97	12.96	16.72	59.89	6.46	1.08	1.01
Mean			32.52	3.34	15.17	16.49	58.49	6.51	1.04	0.99
—	G33	CAM76	32.50	1.76	13.15	18.77	59.59	6.72	1.06	0.89

Table 4.6. Biochemical parameters. **OIL**= % of oil content, **PA** = palmitic acid; **OA** = oleic acid; **LA**= linoleic acid; **ALA** = linolenic acid; **GA** = gondoic acid, **TOC** = tocopherol; **TWS** = 1000 seed weight.

4.4.5. ANOVA analysis

Univariate ANOVA analysis indicates that significant differences in some biochemical/agronomic parameters may exist among the *C. sativa* accessions grouped according to Nei's (1978) genetic distance (Fig. 4.1). In particular, according to Tukey HSD test, the A4 SSR sub-group showed a significant higher TWS (equivalent to 1.24g) and linoleic acid (LA) content, than the remaining SSR sub-groups (Table 4.7). Sub-group A1 showed a significant higher linolenic acid level (61.54%) but lower linoleic acid (15.35%) in comparison with the other groups. No significant differences were observed for the remaining biochemical parameters i.e. total oil content, oleic acid, palmitic acid, gondoic acid and tocopherols.

SSR Sub-group	TWS g/DW	OIL %	LA %	ALA %	PA %	OA %	GA %	TOC µg/mg
A1	0.97b	31.60a	15.35a	61.54b	2.98a	13.45a	6.68a	1.21a
A2	1.05b	30.83a	18.37ab	59.25ab	3.46a	12.59a	6.32a	1.50a
A3	1.10ab	31.94a	19.06ab	56.14ab	3.37a	14.94a	6.49a	1.16a
A4	1.24c	30.77a	21.55b	54.68a	3.91a	12.74a	7.11a	1.29a
B	0.99b	32.52a	16.49ab	58.49ab	3.34a	15.17a	6.51a	1.04a

Table 4.7. Summary of ANOVA analysis. TWS= weight of 1000 seeds, LA = linoleic acid, ALA = linolenic acid; PA = palmitic acid; OA = oleic acid; GA = gondoic acid, TOC = tocopherol. Different letters identify statistically significant differences between means, according to Tukey's multiple comparisons (p = 0.05).

4.5. Discussion

The present work reports data concerning the development of 15 specific *C. sativa* microsatellite markers used to study the genetic variability present in one collection of 40 camelina accessions, of different geographical provenience. All SSR markers have been isolated from one GA-enriched genomic DNA library, previously prepared by Manca and Galasso (2010). Sequencing of about 200 genomic clones, randomly isolated from this library, allowed the recovery of about 60% of clones positive for the presence of DNA sequences with GA/TC repeats. This value, similar to that of many other microsatellite enriched-libraries developed for other species (Aranzana et al. 2002; Métais et al. 2002), confirms the reliability of the library, a tool particularly useful for species like *C. sativa*, that is characterized by few available DNA sequences in the EMBL database (only 48 DNA nucleotide sequences) and a moderate level of SSR transferability across related species (Gehring et al. 2006).

Although primers on 70 microsatellites containing nucleotide sequences (see material and methods) were designed, only 17 (15 polymorphic and 2 monomorphic) amplified clearly discernable fragments consisting of one single strong band of the expected size. Almost all of the remaining couple of primers amplified multiple fragments suggesting the presence of multiple priming sites across the genome. Since multiple-locus amplification is common in species with an allopolyploid origin (Pairon and Jacquemart 2008) this high percentage of microsatellites that yields multiple PCR bands, strongly supports the hypothesis of a polyploid origin of camelina genome. Very unlikely, this is the result of unspecific priming. Genome polyploidization in *C. sativa* has already been suggested by previous studies reporting about the use of AFLP molecular markers (Gehring et al. 2006), the analysis of copy number, gene structure and genome organization of two genes in the fatty acid biosynthesis pathway: *fatty acid desaturase (FAD) 2* and *fatty acid elongase (FAE) 1* (Hutcheon et al. 2010) and the isolation and characterization of the beta-tubulin multigene family (Galasso et al. 2011).

All SSR markers employed in this study produced clear and unambiguous amplification fragments allowing the detection of a large number of alleles per locus (from 3 to 16). The total number of alleles and allelic richness observed in this study were quite high. A total of 134 alleles were generated at 15 SSR loci in the germplasm analysed with a mean of 8.9 alleles per locus.

However a low number of heterozygotes were detected in all genotypes. This finding is most likely attributable to the self-pollinating nature of camelina

species rather than resulting from selective pressure since most of the accessions analysed are neither improved nor cultivated. A further confirmation of the inbred nature of camelina comes from the Nei's genetic diversity indices. The analysis, conducted on the 40 camelina accessions, demonstrated that 60% of the genetic differentiation occurs among accessions, while 40% resides within them ($D_{st} = 0.391$).

The high discriminatory capacity of the SSR markers, as observed in other plant species, has also been confirmed in our study. In fact, thirty-eight out of forty accessions are clearly distinguished in our camelina collection by the 15 polymorphic microsatellites used. G1-CAM108 and G9-CAM170 were the only two accessions that remained undistinguishable from one another. Since we analysed 8 individuals from each accessions and slight differences were observed between them, one explanation might be that these two were actually the same accession as they both originate from Poland. Beside these two accessions, the UPGM dendrogram, based on Nei's genetic distances, shows that all camelina accessions can be clearly separated in two main groups including 35 and 5 accessions respectively. The largest A group could be further divided into 4 sub-groups (A1 to A4). Although our SSR markers were effective in differentiating camelina accessions, the phylogenetic tree largely fails to show clusters of camelina accessions strictly based on the provenience. A similar results was obtained by Ghamkhar et al. (2010) analysing 53 camelina accessions from different locations using 8 AFLP markers.

A number of studies have demonstrated that genetic relationships based on molecular markers do not always agree with morphological/agronomic traits (Van Tienderen et al. 2002). This is particular true when polymorphisms are detected in non-functional regions of the genome. However, molecular and morphological/biochemical analyses could be complementary tools for varieties characterization and for distinguishing accessions, once a certain degree of association between the two approaches has been established. In this contest, results shown in Table 4.7 represent a promising result for the assessment of correlation between molecular and agronomic/biochemical data. To this regard, significant variation was observed among the SSR camelina sub-groups when compared for some biochemical traits. In particular, the significant relationships between the A4 group and the TWS and AL content, appear quite interesting for plant breeders who aim to improve seed size and/or the fatty acid profile.

In conclusion, knowledge of genetic variation and the genetic relationship between genotypes is important for an efficient utilisation of *C. sativa* germplasm

resources. Beside of providing a useful tool for germplasm identification and genetic diversity, these 15 newly developed SSR polymorphic markers will prove very useful in genetic mapping and in assisting plant breeders in early progeny selection.

4.6. Acknowledgments

This work was supported by Regione Lombardia, agreement Regione/CNR, project 2: 'Risorse biologiche e tecnologie innovative per lo sviluppo del sistema agro-alimentare'. We thank Mrs. Floriana Gavazzi for technical assistance. We are also very grateful to Dr. Tommaso Martinelli who grown and followed the plants in the field

4.7. Reference

- Abajian C** (1994) SPUTNIK, <http://www.abajian.com/sputnik>.
- Aranzana MJ, Garcia-Mas J, Carbó J, Arús P** (2002) Development and variability analysis of microsatellite markers in peach. *Plant Breeding* **121**: 87-92.
- Belaj A, Satovic Z, Cipriani G, Baldoni L, Testolin R, Rallo L, Trujillo I** (2003) Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. *TAG Theoretical and Applied Genetics* **107**: 736-744.
- Bell CJ, Ecker JR** (1994). Assignment of thirty microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19**(1): 137-144.
- Chung SM, Staub JE, Chen JF** (2006) Molecular phylogeny of *Cucumis* species as revealed by consensus chloroplast SSR marker length and sequence variation. *Genome* **49**: 219-229.
- Fröhlich A, Rice B** (2005) Evaluation of *Camelina sativa* oil as a feedstock for biodiesel production. *Industrial Crops and Products* **21**: 25-31.
- Galasso I, Manca A, Braglia L, Martinelli T, Morello L, Breviario D** (2011) h-TBP: an approach based on intron-length polymorphism for the rapid isolation and characterization of the multiple members of the β -tubulin gene family in *Camelina sativa* (L.) Crantz. *Molecular Breeding* **28**: 635-645.
- Gehringer A, Friedt W, Luehs W, Snowdon RJ** (2006) Genetic mapping of agronomic traits in false flax (*Camelina sativa* subsp *sativa*). *Genome* **49**: 1555-1563.
- Ghamkhar K, Croser J, Aryamanesh N, Campbell M, Kon'kova N, Francis C** (2010) *Camelina* (*Camelina sativa* (L.) Crantz) as an alternative oilseed: molecular and ecogeographic analyses. *Genome* **53**: 558-567.
- Hutcheon C, Ditt RF, Beilstein M, Comai L, Schroeder J, Goldstein E, Shewmaker CK, Nguyen T, De Rocher J, Kiser J** (2010) Polyploid genome of *Camelina sativa* revealed by isolation of fatty acid synthesis genes. *Bmc Plant Biology* doi:10.1186/1471-2229-10-233
- IATA Report (2010)** "Report on Alternative Fuels" 5th Edition Ref. No: 9709-03 ISBN 978-92-9233-491-8 Montreal Geneva, <http://www.iata.org/ps/publications/pages/alternative-fuels.aspx>.
- Kashi Y, King D, Soller M** (1997) Simple sequence repeats as a source of quantitative genetic variation. *Trends in Genetics* **13**: 74-78.
- Kimura M, Crow JF** (1964) The number of alleles that can be maintained in a finite population. *Genetics* **49**:725-738.
- Levene H** (1949) On a matching problem arising in genetics. *The Annals of Mathematical Statistics*, Oxford **20**:91-94.
- Manca A, Galasso I** (2010) Development of simple sequence repeat (SSR) markers in *Camelina sativa* (L.) Crantz. *Minerva Biotec.* **22**(2): 43-5.
- Morgante M, Olivieri AM** (1993) PCR-amplified microsatellites as markers in plant genetics. *The Plant Journal* **3**: 175-182.
- Nei M** (1973) Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences* **70**: 3321-3323.

4. Manuscript 1- Evaluation of genetic diversity in one *Camelina sativa* (L) Crantz collection using SSR markers and biochemical traits.

- Nei M** (1978) Estimation of a average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583-590.
- Ota T** (1993). Dispan: Genetic distance and phylogenetic analysis. Pennsylvania State University, University Park, PA, USA. <http://www.bio.psu.edu/People/Faculty/Nei/Lab/Programs.html>.
- Patil PD, Gude VG, Deng S** (2009) Biodiesel production from *Jatropha Curcas*, waste cooking, and *Camelina sativa* Oils. *Industrial & Engineering Chemistry Research* **48**: 10850-10856.
- Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M** (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. *TAG Theoretical and Applied Genetics* **97**: 1248-1255.
- Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A** (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* **2**: 225-238.
- Rajendrakumar P, Biswal AK, Balachandran SM, Srinivasarao K, Sundaram RM** (2007) Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. *Bioinformatics* **23**: 1-4.
- Rozen S, Skaletsky HJ** (1998) Primer 3, http://www.genome.wi.mit.edu/genome_software/other/primer3.html.
- Russell JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R** (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. *TAG Theoretical and Applied Genetics* **95**: 714-722.
- Schuelke M** (2000) An economic method for the fluorescent labelling of PCR fragments. *Nature Biotechnology* **18**:233-234
- Semagn K, Bjornstad A, Ndjiondjop MN** (2006) An overview of molecular marker methods for plants. *African Journal of Biotechnology* **5**: 2540-2568.
- Séguin-Swartz G, Eynck C, Gugel RK, Strelkov SE, Olivier CY, Li JL, Klein Gebbinck H, Borhan H, Caldwell C, Falk KC** (2009). Diseases of *Camelina sativa* (L.) Crantz (false flax) *Can. J. Pl. Path* **31**:375-386.
- Tautz D** (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**: 6463-6471.
- Vollmann J, Grausgruber H, Stift G, Dryzhyruk V, Lelley T** (2005) Genetic diversity in camelina germplasm as revealed by seed quality characteristics and RAPD polymorphism. *Plant Breeding* **124**: 446-453.
- Vollmann J, Moritz T, Kargl C, Baumgartner S, Wagentristl H** (2007) Agronomic evaluation of camelina genotypes selected for seed quality characteristics. *Industrial Crops and Products* **26**: 270-277.
- Yeh FC, Yang RC, Boyle T** (1999). POPGENE version 1.3.1. Microsoft Window-bases Freeware for Population Genetic Analysis. Available: www.ualberta.ca/~fyeh/. University of Alberta and the Centre for International Forestry Research.
- Zane L, Bargelloni L, Patarnello T** (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology* **11**: 1-16.
- Zubr J** (1997) Oil-seed crop: *Camelina sativa*. *Industrial Crops and Products* **6**: 113-119.