UNIVERSITE CATHOLIQUE DE LOUVAIN
EARTH AND LIFE INSTITUTE
FACULTE D'INGENIERIE BIOLOGIQUE, AGRONOMIQUE ET ENVIRONNEMENTALE

QUANTITATIVE APPROACH OF THE GENETIC DETERMINISM OF SEX EXPRESSION IN MONOECIOUS HEMP (*CANNABIS SATIVA* L.), AND ITS RELATIONSHIP WITH FLOWERING PHENOLOGY AND STEM AND SEED YIELDS

ANNE-MICHELLE FAUX

FEVRIER 2014

THESE PRESENTEE EN VUE DE L'OBTENTION DU GRADE DE DOCTEUR EN SCIENCES AGRONOMIQUES ET INGENIERIE BIOLOGIQUE

THESIS EXAMINING BOARD:

SUPERVISOR: PROF. PIERRE BERTIN

PRESIDENT OF THE JURY: PROF. BRUNO HENRY DE FRAHAN

MEMBERS OF THE JURY:

PROF. XAVIER DRAYE

DR. JEAN-PIERRE GOFFART

PROF. RICHARD LAMBERT

DR. GIUSEPPE MANDOLINO

Cette fleur il est vrai avait une déformation professionnelle,
elle s'acharnait à sentir bon.

Julos Beaucarne

ACKNOWLEDGMENTS

A l'heure de clore ce travail, il me tient particulièrement à cœur d'exprimer ma profonde reconnaissance envers les nombreuses personnes qui en ont permis la réalisation.

Mes remerciements les plus chaleureux s'adressent à mon promoteur, le Professeur Pierre Bertin. Merci, Pierre, de m'avoir accueillie au sein du laboratoire ECAV il y a sept ans et demi. Merci pour la confiance que tu m'as accordée, tes précieux conseils et les préceptes de bonne rédaction scientifique que tu m'as enseignés. Merci pour tes nombreuses corrections, ta disponibilité et ta patience, en particulier lorsque le rythme de correction dépassait celui de rédaction. Merci aussi pour ton humour qui ponctuait agréablement nos discussions ainsi que pour le partage de ta vision du monde et de tes voyages, en Bolivie, à Cuba ou ailleurs.

Je remercie vivement les Professeurs Xavier Draye et Richard Lambert et Dr Jean-Pierre Goffart et Giuseppe Mandolino d'avoir accepté de faire partie de mon Comité d'accompagnement et/ou Jury de thèse. A tous, merci pour vos conseils et critiques constructives.

J'adresse un remerciement particulier au Professeur Xavier Draye pour sa précieuse contribution tout au long de la réalisation de ce travail. Merci, Xavier, pour tout le temps que tu m'as généreusement consacré, que ce soit pour une initiation à SAS, une dissection de modèles mixtes ou une discussion de cartographie génétique. Merci pour les connaissances et ton intérêt pour les statistiques, en particulier, que tu m'as partagés.

Je remercie également chaleureusement le Professeur Richard Lambert. Merci, Richard, d'avoir accepté d'accompagner cette thèse alors que le sujet incluait une composante qui s'écarte de ton domaine d'expertise. Merci d'avoir permis la mise en place d'essais à Michamps. Merci pour ton soutien et ta disponibilité pour toute question.

Que les Dr Jean-Pierre Goffart et Giuseppe Mandolino voient ici l'expression de mon entière gratitude pour leur contribution à l'aboutissement du présent document. Jean-Pierre, après vous avoir rencontré dans le cadre des séminaires organisés pour le cours de fertilisation, c'est un plaisir de vous revoir pour discuter du chanvre. Merci pour vos remarques judicieuses sur les aspects agronomiques du travail.

Dr Mandolino, we met at the Annual Congress of the Italian Society of Agricultural Genetics in September 2009 in Torino where we had the opportunity to discuss about hemp. It is a great pleasure to see you again and to count you in my thesis Jury. I gratefully thank you for your valuable remarks, the interesting conversations we have had and for sharing your deep knowledge and passion for hemp.

Mes remerciements s'adressent également au Professeur Jean-François Ledent pour ses précieux conseils en matière d'écophysiologie des grandes cultures, ses encouragements au cours de la réalisation de ce travail et le partage de sa passion pour l'agronomie et l'agriculture.

Le travail sur le chanvre n'aurait été possible sans une connaissance fine des processus de création et multiplication variétale appliqués chez l'espèce. Je remercie vivement Olivier Beherec, améliorateur à la Fédération Nationale des Producteurs de Chanvre (Le Mans), pour son accueil chaleureux lors des Journées Portes ouvertes de la FNPC en août 2007, ses promptes réponses à mes demandes d'information et son intérêt porté à nos recherches.

La tâche aurait par ailleurs été gargantuesque sans la collaboration de nombreuses personnes.

Merci à Gérard Collignon, Philippe Hamiet (Centre A. de Marbaix), Christian Decamps (asbl Fourrages-Mieux), Mickaël Mary, Jean-Pol Mazy, Frédéric Vandeputte (CIPF), Daniel Van der Vennet et Raymond Blaise (Centre de Michamps), en particulier, pour l'aide à la mise en place et au suivi des essais en champs à Corroy-le-Grand et à Michamps. Merci à Guy Foucart (CIPF), Pascal Colomb et David Becker (ECOSEM) pour la mise à disposition d'étuves et trieuse de graines, respectivement. Je remercie également tous les volontaires et jobistes, trop nombreux pour les citer, qui ont activement participé aux récoltes des essais en champ.

A Louvain-la-Neuve, j'adresse un tout grand merci à Viviane Van Hese, Pierre Van Thorre et Willy Dussart pour leur préciseuse collaboration à la culture des plantes en serres et aux manipulations de biologie moléculaire. Je remercie également le Dr Marie-Christine Flamand et Marie-Christine Eloy (ISV, UCL) pour la révélation des produits d'amplification AFLP par séquençage capillaire. Merci aussi à Nicolas Dauguet (de Duve Institute, UCL) pour la mise au point du protocole de cytométrie de flux.

Un merci tout particulier à Isabelle Vanacker, véritable chef d'orchestre spécialisée en symphonie administrative, pour sa disponibilité, son accueil toujours souriant et son efficacité.

Je remercie également les mémorants qui ont contribué à ce travail. Merci à Alice Berhin, Frédéric Collard, Raphaël d'Andrimont, Adrien Occre et Pierre Raulier pour leur très précieuse et efficace collaboration. Un merci tout spécial à Pierre et Raphaël, alias 'HempTeam', pour les moments inoubliables passés ensemble lors des récoltes des essais en champ. Chef, merci d'avoir même songé à sceller notre amitié dans ma main!

Leur compagnie a rendu le quotidien si agréable. Je remercie chaleureusement tous les membres du groupe de vie ECAV cotoyés au cours de ces sept années et demi. Merci à Adrien, Alek, Audrey, Aurélie, Beata, Benjamin, Charles, Claire, Dieudonné, Gauthier, Gisèle, Guillaume, Gustave, Guy, Hanh, Igor, Inès, Isabelle, Jean-Elysée, Khalid, Laure-Anne, Léopold, Many, Maxime, Nathalie, Pierre B, Pierre R., Pierre VT, René, Sophie, Taha, Teresa, Thomas, Tristan, Viviane, Walid, Willy, Xavier, Yang Min, Yurelkys, … pour l'ambiance familiale qui règne en ECAV et qui contribue à un enrichissement humain et culturel permanent.

Au niveau de la Faculté, je tiens à remercier les Professeurs et assistants avec qui j'ai eu l'occasion de collaborer. Merci aux Professeurs Anne-Laure Jacquemart, Cathy Debier, Joseph Dufey, Charles Bielders, Michel Focant, Bernard Toussaint, Jean-François Ledent, Richard Lambert, Xavier Draye et Pierre Bertin pour le partage de leurs passions respectives au cours des excursions. Merci au Professeur Jacques Mahillon, Doyen jusqu'il y a peu, pour son écoute à l'égard des préoccupations des assistants. Merci aux nombreux collègues assistants de l'A-Cube pour tous les bons moments passés ensemble,

que ce soit lors des réunions, d'activités académiques ou de sorties extra-académiques.

Grâce à eux, les heures de travail et de détente étaient allègrement rythmées. Merci, chers amis, de m'avoir sortie de la matrice enchanvrée le temps de nombreux soupers, sorties ou escapades. Un merci particulier à Alice, Yas, Emi, Caro, Séverine, Mathieu, Delphine, Seb, Lolo, Dédé, Adeline, Arslane, Charles, Julien et Lio pour leurs encouragements et bouffées d'oxygène durant les derniers mois. Un merci particulier à Adeline pour sa précieuse aide dans le dernier coup de rush, à Lio et Arslane pour les impressions livrées à domicile, à Charles pour la mécanique vélo et à Julien pour ses talents culinaires!

Enfin, je tiens à remercier ma famille pour son soutien inconditionnel. Merci à mes grand-parents pour le regard admiratif qu'ils portent ou ont porté sur mes projets. Merci à mes oncles, tantes, cousins et cousines, pour leurs nombreuses attentions.

Mes remerciements les plus chers vont à ma fratrie et à mes parents. Merci Pierre, Gilles et Emilie, Philippe-François et Hélène pour tout ce que vous m'apportez. Votre complicité est une source de bonheur intarissable. Un merci particulier à Pierre pour ses conseils avisés à tout point de vue et le partage de son expertise en sélection génomique, calcul matriciel ou autre. Merci à Philippe-François et Pierre pour leur aide salvatrice lors de l'un ou l'autre pic de travail. Enfin, merci du fond du cœur à mes parents de m'avoir toujours soutenue et encouragée tout au long de la réalisation de ce travail. Merci pour votre écoute, vos conseils et votre confiance. Merci pour le goût pour la recherche et l'envie d'aller jusqu'au bout des projets que vous m'avez transmis.

Louvain-la-Neuve, le 13 février 2014.

ABSTRACT

The sex expression is a major concern for the genetic improvement of hemp. Although naturally dioecious and bearing heteromorphic sex chromosomes, the species displays a highly variable sexual phenotype. The monoecious cultivars allow the mechanical harvest of both stem and seed; however, their seed production requires a strict control of the sex expression. This thesis investigated the genetic determinism of sex expression in monoecious hemp and established its relationship with flowering phenology and stem and seed yields.

The existence of genotypic variability for the sex expression in monoecious hemp was evidenced despite the high environmental sensitivity of the trait. Sex expression and earliness varied mostly consistently among cultivars. Higher seed yields were associated with (mid-) early feminised cultivars. Monoecious hemp was demonstrated to bear XX chromosomes, which showed homologies with both the X and Y ones of dioecious hemp. Three F1 segregating populations were created, and genetic maps were constructed. The distribution of male and female flowers in monoecious plants was modelled. A relatively high number of genomic regions (QTLs) showed variation associated with the model parameters. Some of them were likely located on the sex chromosomes. Closely linked QTLs for sex expression, earliness and yields were found. Their putative implications for breeding were discussed.

In conclusion, the present quantitative approach successfully provided fundamental insights to improve our understanding of the genetic determinism of sex expression in monoecious hemp and support its cultivation for a dual production.

SHORT CONTENTS

TABLE OF CONTENTS

CHAPTER III

INVESTIGATING THE GENETIC DETERMINISM OF A QUANTITATIVE TRAIT 61

PART II – EXPERIMENTATIONS AND RESULTS 97

CHAPTER IV

THE RELATIONSHIP OF STEM AND SEED YIELDS TO FLOWERING PHENOLOGY AND SEX EXPRESSION IN MONOECIOUS HEMP 99

CHAPTER V

CHAPTER VI

CHAPTER VII

CHAPTER VIII

CHAPTER IX

INTRODUCTION AND OBJECTIVES

INTRODUCTION

Hemp (Cannabis sativa L.) was domesticated at least 6000 years ago (Schultes 1970 cited by van der Werf et al. 1996). In Europe, the species has been cultivated over a period of many centuries as raw material for the production of rope, canvas, textiles, paper and oil products (Bocsa and Karus 1998). After a heyday during the golden age of sailing ships in the 17th century, the cultivation of hemp progressively declined, firstly as a result of the increasing concurrence of the cotton fibre. However, since about the last twenty years, there has been a renewed interest for hemp. This recent comeback of hemp in the European fields can be explained by the ecological characteristics of its cultivation and the diversity and properties of renewable resources that it produces (van der Werf et al. 1996; Bocsa and Karus 1998; Struik et al. 2000; Ranalli and Venturi 2004). At the local level, the creation of the ASBL 'Chanvre Wallon' in 2006 formalised the increasing interest for hemp from scientific, industrial and political actors in the Region wallonne in Belgium (Chanvre Wallon 2013). The present thesis started in this context considering that hemp might constitute an interesting alternative crop for farmers.

From an agronomic point of view, the production of hemp is greatly affected by two major features of its flowering system: its sensitivity to the photoperiod and reproductive morphology (Struik et al. 2000). Firstly, hemp is a quantitative short-day plant. By modulating the flowering time, the photoperiod has a key influence on the stem yield since the maximum stem yield is reached shortly after flowering (Lisson et al. 2000b). Secondly, hemp is a naturally dioecious plant with highly variable sexual phenotype. Both dioecious and monoecious cultivars exist. However, in dioecious cultivars, the male plants flower and senesce earlier than the female ones, which makes the mechanical harvest of the seed difficult (Bocsa and Karus 1998). Monoecious cultivars have been bred from plants bearing flowers of both sexes in dioecious stands (de Meijer 1995; Ranalli 2004; Moliterni et al. 2004). Compared to the dioecious ones, the monoecious cultivars display higher seed yields and allow the mechanical harvest of both stems and seeds due to the synchronized maturity of the plants (Mandolino and Carboni 2004). Nevertheless, breeding monoecious

cultivars is not an easy task. Firstly, the multiplication of monoecious cultivars requires the manual elimination of the residual male plants before flowering in order to avoid the reversion to the dioecious form (Beherec 2000). Secondly, the sex expression, defined as the ratio between the female and male flowers, varies among plants from the predominantly male to predominantly female extreme phenotypes. The strongly masculinised plants must be eliminated during the process of seed production of monoecious cultivars because they are less fruiting and more prone to develop fungal infections (Beherec 2000). From these considerations, it results that the use of monoecious cultivars might be advantageous to reduce the crop heterogeneity and to produce both stems and seeds simultaneously, while the sex expression of the monoecious hemp plants constitutes a major concern for breeders.

The production of stem and fibre in hemp has been investigated in several studies (de Meijer and Keizer 1994; Meijer et al. 1995; van der Werf et al. 1995a, 1995b, 1995c, 1996; Sankari and Mela 1998; Struik et al. 2000; Lisson and Mendham 2000; Mediavilla et al. 2001; Amaducci et al. 2008c). Stem yields can be improved by breeding late-flowering cultivars (van der Werf et al. 1994; Meijer et al. 1995). The duration from emergence to flowering has been modelled in response to temperature and photoperiod (Lisson et al. 2000b; Amaducci et al. 2008a). Although the harvest of both stems and seeds constitutes a common practice in different countries like France (Bocsa and Karus 1998; Bouloc 2006) and Canada (Vera and Hanks 2004), studies related to the simultaneous production of stems and seeds in hemp are scarce.

The mechanism of sex determination in hemp has been considered to be probably the most complicated among dioecious species (Westergaard 1958). Hemp is a diploid species that displays sexual dimorphism and includes sex chromosomes (Hirata 1924). In dioecious hemp, the inflorescences consist in drooping axillary panicles in the male plants and crowded racemes in the female ones (Mohan Ram and Nath 1964). The chromosomes XX are found in female plants, and XY in male plants. The sex determinism would be based on an X-to-autosomes equilibrium and not on a Y-active mechanism (Westergaard 1958; Ainsworth 2000). In monoecious hemp, the inflorescences are similar to those of dioecious female plants with both male and female flowers (Bocsa and Karus 1998). However, the determinism of their sex expression remains widely unknown. Though presumably XX (Menzel 1964), the composition in sex chromosomes in monoecious hemp has never been stated (Mandolino et al.

2002; Moliterni et al. 2004). In addition, the sex expression in hemp, both dioecious and monoecious, shows a high plasticity. External factors such as photoperiod, nitrogen status or hormonal treatments (Freeman et al. 1980), are able to induce the production of flowers of the opposite sex on dioecious hemp plants and modify the sex expression of monoecious plants (Schaffner 1921; Borthwick and Scully 1954; Arnoux 1966a, b; Fournier and Beherec 2006).

Recent molecular studies on the sex expression in hemp led to the identification of DNA markers associated with the male phenotype (Sakamoto et al. 1995; Mandolino et al. 1999; Flachowsky et al. 2001; Mandolino et al. 2002; Torjek et al. 2002; Sakamoto et al. 2005) and DNA markers located on the X and/or Y chromosomes (Peil et al. 2003; Rode et al. 2005). These studies provided insights on the structure of the Y chromosome of hemp – higher size compared to the X chromosome, presence of retrotransposon-like sequences and existence of a pseudoautosomal region, in particular. However, the sex expression in monoecious hemp has been little investigated. The absence of male-associated DNA markers has been observed in 26 plants of monoecious cultivars (Mandolino et al. 1999; Torjek et al. 2002). Moreover, these studies including monoecious cultivars considered the monoecious state as a qualitative trait.

OBJECTIVES

The present thesis focuses on the study of the sex expression in monoecious hemp. However, on the one hand, hemp is a short-day plant and, on the other hand, the use of monoecious cultivars has significant implications on both hemp breeding and cultivation. Therefore, two main objectives were addressed: (i) investigating the genetic determinism of the sex expression in monoecious hemp and (ii) establishing the relationships between sex expression, flowering phenology and yields in stem and seed in monoecious hemp. With respect to the first main objective, the originality of the present thesis relied in the quantitative approach of the sex expression, ranging from strongly masculinised to strongly feminised phenotypes. Therefore, we assumed that the genetic determinism of the sex expression in monoecious hemp could be investigated through the identification of quantitative trait loci (QTLs).

Five specific objectives were defined in order to achieve the main objectives:

1. Assessing the genotypic variability of the sex expression in monoecious hemp;

2. Establishing the relationships between sex expression, flowering phenology and stem and seed yields in monoecious hemp;

3. Establishing the composition in sex chromosomes of monoecious hemp;

4. Constructing linkage maps of the sex expression in hemp;

5. Identifying QTLs involved in the determinism of the sex expression, earliness and yields in stem and seed in hemp.

The specific objectives were addressed by the following experimental designs.

(1) Assessing the genotypic variability of the sex expression in monoecious hemp (Chapters IV & V)

The genotypic variability of the sex expression expressed as a quantitative trait was assessed among five monoecious hemp cultivars in the field (Chapter IV) and in controlled conditions (Chapter V). The cultivars covered a range of earliness varying from very early ('Uso 31'), early ('Fedora 17'), mid-early ('Santhica 27' and 'Felina 32') to late ('Epsilon 68'). In the field, environmental variation was created by the selection of two contrasting sites in Belgium, five sowing dates ranging between Mid-April and Mid-July and the completion of two trial years. In controlled conditions, three trials were performed under distinct photoperiodic conditions. The sex expression of the monoecious plants was assessed by recording the *degree of monoecy*, an ordinal scale established by Sengbush (1952) that classifies the monoecious hemp forms according to their ratio between female and male flowers. In controlled conditions, the scale of Sengbusch (1952) was modified by including seven levels instead of five. In the field, the monoecy degree was determined at the plot level, while in controlled conditions, it was determined for each plant node. The observation of the sex expression was repeated in time.

(2) Establishing the relationships between sex expression, flowering phenology and yields in both stem and seed in monoecious hemp (Chapters IV & V)

The relationship between sex expression and flowering phenology was assessed among five monoecious hemp cultivars in both field (Chapter IV) and controlled conditions (Chapter V), while the relationships between stem and seed yields, flowering phenology and sex expression was determined in the field only (Chapter IV). The experimental designs used to this purpose were described for the first specific objective. In the field, the flowering phenology was determined by seven floral developmental stages based on the phenological scale of Mediavilla et al. (1998). In controlled conditions, the flowering phenology was determined by the time of first flower appearance and the duration of flowering only. In the field, the plant density was measured, and the dry matter yields in stem and seed were determined. The relationships between sex expression, flowering phenology and yields were discussed with an eye

towards supporting the crop management and thus the yields in hemp cultivated for the simultaneous production of stem and seed.

(3) Establishing the composition in sex chromosomes of monoecious hemp (Chapter VI)

The composition in sex chromosomes of monoecious hemp was assessed by flow cytometry and a male-associated DNA marker. The use of flow cytometry relied on the presence of size differences between the X and Y chromosomes of hemp (Yamada 1943 cited by Sakamoto et al. 1995; Sakamoto et al. 1998) and the ability of the technique to highlight very small genome-size differences (Costich et al. 1991; Vagera et al. 1994; Dolezel and Göhde 1995) and analyze large populations of cells (Dolezel and Bartos 2005). According to Mandolino et al. (2002), the male-associated DNA marker MADC2 is completely associated to the male phenotype and should therefore be located on the Y chromosome in a region excluded from recombination during meiosis. The amplification of this marker by monoecious plants would indicate the presence of a male-specific region in monoecious hemp. The cytometric and molecular analyses were performed through the evaluation of 55 and 115 plants, respectively, belonging to five monoecious hemp cultivars.

(4) Constructing linkage maps of the sex expression in hemp (Chapter VII)

Three linkage maps were constructed in order to provide a framework to perform QTL analyses. To this purpose, three F_1 segregating populations were used: two derived from a cross between male and female plants of the dioecious cultivar ‘Carmagnola’, and one from a cross between two monoecious cultivars, ‘Uso 31’ and ‘Fedora 17’. The creation of dioecious populations aimed at the identification of markers linked to the sex and thus located on the sex chromosomes, while the creation of a monoecious population aimed at the segregation of the sex expression in the offspring. Two dioecious populations instead of one were used because of the relatively low number of seeds obtained from the ‘Carmagnola’ crosses. DNA polymorphism was provided by AFLP molecular markers. The selection of the AFLP technique relied on its ease of

implementation, its ability to amplify many loci (Vos et al. 1995; Mueller and Wolfenbarger 1999; Meudt and Clarke 2007) and the availability of AFLP primer pairs providing sex-linked markers in hemp (Flachowsky et al. 2001).

(5) Identifying QTLs involved in the determinism of the sex expression, earliness and yields in stem and seed in hemp (Chapters VIII and IX)

The sex expression, flowering time, stem and seed yields were recorded in the three F_1 segregating populations as described for the fourth specific objective. The sex expression of both dioecious and monoecious plants was recorded at distinct times, while all of the remaining three traits were recorded once. In addition, the sex expression of the monoecious plants was noted at each plant node. Due to the large number of data points per plant, distinct variables were constructed in order to summarise the sex expression of the monoecious plants. A method based on the modelling of the distribution of the male and female flowers as a function of the plant node along the stem was established for characterizing the sex expression of the monoecious plants. The development and discussion of this approach were reported in Chapter VIII.

The QTL analysis was conducted in each segregating population independently by expressing each phenotypic variable in terms of QTLs in the structure of a mixed model (Chapter IX). The linkage maps that were constructed according to the fourth specific objective provided the framework for the location of QTLs. The use of a mixed model for QTL mapping (van Eeuwijk et al. 2010) allowed to consider the existence of genetic correlations between the observations of the sex expression made on each plant at distinct times.

THESIS OUTLINES

The present thesis is structured in three parts.

Part I provides the context of the present research. The first three chapters address the questions '*why*', '*what*' and '*how*', respectively: why hemp is an interesting crop today (Chapter I), what features are significant for the genetic improvement of the species (Chapter II), and how to investigate the genetic determinism of a quantitative trait (Chapter III).

Part II presents the experimentations and results as a succession of five chapters, either published or prepared for submission to peer review journals (Chapters IV to IX). Each chapter can be easily understood independently; however, I would like to apologize for the discomfort that the redundancies among chapters may provide during a complete and continuous reading of the present document.

Part III discusses the main advances of the present work and provides perspectives for further studies aiming to improve our understanding of the genetic determinism of the sex expression in monoecious hemp and/or support the production of stem and seed in monoecious hemp.

PART I – CONTEXT

CHAPTER I

HEMP, AN ANCIENT CROP WITH HIGH INNOVATIVE POTENTIAL

Hemp refers to cultivars of *Cannabis sativa* L. grown for the production of fibre, straw, cannabinoids or seed and characterized by low levels of Δ^9-tetra-hydrocannabinol (van der Werf et al. 1996; Lisson et al. 2000a). It is one of the oldest cultivated crops: the species was domesticated at least 6000 years ago (Schultes 1970 cited by van der Werf et al. 1996) and perhaps 8500 years ago (Schultes and Hofmann 1973 cited by Small and Cronquist 1976). Hemp originates from Central Asia but has been cultivated from the Equator to the Polar Circle (Vavilov 1926 cited by van der Werf et al. 1996; Bocsa and Karus 1998).

From the 16th to the 18th century, hemp and flax (*Linum usitatissimum* L.) were the major fibre crops in Russia, Europe and North America. Hemp fibres has been used in the textile, sail and paper industries (Ranalli and Venturi 2004). However, the large-scale cultivation of cotton, jute and other tropical fibres, the development of new technologies to process wood into paper pulp, and the presence of psychoactive components in hemp resulted in the decline of its cultivation in the 19th century. The hemp production declined further in the 20th century due to the advent of synthetic fibres (van der Werf 2004). From the second world war until the 1980's, hemp has been a largely forgotten crop, although the breeding work continued in Eastern and Central Europe and in France (de Meijer 1995; van der Werf et al. 1996). For about the last twenty years, there has been a renewed interest for hemp as a source of cellulose fibre and seed oil in Western European countries, Australia, the US and Canada (de Meijer 1995; van der Werf et al. 1996; Schumann et al. 1999; Lisson et al. 2000a; Struik et al. 2000; Ranalli and Venturi 2004). Besides, legal measures against the use of *Cannabis* drug in Western countries can no more discourage the cultivation and transformation of hemp (de Meijer 1995). The recent revival of interest for hemp can be explained by features of both its cultivation and uses.

(1) Hemp is an annual crop that has the right profile to fit into a sustainable agricultural system in a temperate climate.

Hemp is able to provide high yields under a wide range of agro-ecological conditions (Struik et al. 2000). As an annual crop, it does not require a long-term commitment in land use and can be easily integrated in the rotation (Ranalli 2004). While harvesting hemp straw and seed requires suitable equipment (Chanvre Wallon 2013), the management of a hemp crop is similar to other conventional crops.

The cultivation of hemp has little pesticide and modest fertilizer needs and provides several advantages in the crop rotation. Its capacity to grow fast constitutes an opportunity for weed control. The plant develops a long tap root and thereby has an improving effect on the soil structure (van der Werf et al. 1996; Ranalli and Venturi 2004; Piotrowski and Carus 2011). Hemp could increase the yield of the succeeding crop in the rotation cycle up to 10-15% (Ranalli and Venturi 2004). According to van der Werf et al. (1996), hemp is an excellent candidate to fill the niche of an annual crop in a temperate climate.

Compared to other crops such as potato or sugar beet, hemp is a low-input crop with low environmental impacts (van der Werf 2004). The low-input trait of hemp refers to the low use of fertilizers NPK (75-38-113, 170-80-293 and 220-101-180 kg ha^{-1} of $N\text{-}P_2O_5\text{-}K_2O$ for hemp, potato and sugar beet, respectively, in France), pesticides (0, 5.5 and 3.7 kg ha^{-1} of active ingredient for each crop, respectively) and diesel (65, 165 and 137 kg ha^{-1}, respectively) (van der Werf 2004). Hemp is characterized by low environmental impacts in terms of eutrophication (20.5, 23.8, 24.1 and 24.4 kg ha^{-1} of PO_4-equivalents with hemp, potato, sugarbeet and pea, respectively), emission of greenhouse gases (2300, 4120 and 4900 kg ha^{-1} of CO_2-equivalents with hemp, potato and sugarbeet, respectively) and acidifying pollutant (9.8, 22.4 and 240.5 kg ha^{-1} of SO_2-equivalents with hemp, potato and sugarbeet, respectively) (van der Werf 2004). According to Ranalli and Venturi (2004), hemp would be a suitable crop for both conventional and organic farming systems, able to provide high biomass production with low inputs. Forapani et al. (2001) attributed the renewed interest for hemp to its comparatively low chemical fertilizer requirements and its exceptional disease resistance, in addition to an increased public perception of the value of natural fibres compared with synthetic ones (Montford and Small 1999).

(2) Hemp is a multi-purpose crop able to provide renewable raw materials and end-products with valuable properties.

Hemp is grown for a multitude of end products derived from the seed, fibre, wooden core and cannabinoids (Struik et al. 2000). In the European Union, hemp fibres are mainly intended for the paper and automotive industries (Karus and Vogt 2004). They are used as biocomposites for thermoplastic and thermoset press-moulded parts in the vehicles. The current use of hemp fibres in the textile industry would be limited in EU (Karus and Vogt 2004). However, a project aiming at the development of a competitive hemp fibre production chain for textile destination has been conducted in Italy (Amaducci 2003). The hurds, or wooden core of the stems, are used for animal bedding and in the insulation sector. Most hemp seeds are sold for animal feed, mainly for birds. The remaining markets for hemp seed include the cosmetic and human food sectors (Karus and Vogt 2004). In addition, hemp produces terpenophenolic substances known as cannabinoids, which accumulate mainly in the glandular trichomes of the plant and are unique to the species (de Meijer et al. 2003). At least 66 different cannabinoids were identified in hemp (Pacifico et al. 2008). The therapeutical potential of these substances has been investigated, and *Cannabis*-based medicines have been developed (Pertwee 2004; Stott and Guy 2004). According to Struik et al. (2000), the multi-purpose feature of hemp uses distinguishes it from other non-food crops, which usually produce only one type of material, and makes it an interesting model crop for the development of multi-output systems.

In addition, valuable properties have been recognised for specific hemp products. Correia et al. (2001) showed that hemp fibre allows the production of paper pulp at lower energy cost than traditional wood pulp species. In the automotive industry, hemp fibre reinforced plastics show favourable mechanical properties such as rigidity and strength in combination with low density (Haufe and Karus 2011). Lime–hemp composites are innovative building materials with hygrothermal properties such that they can significantly improve the indoor comfort (Evrard and De Herde 2009). The potential of the hempseed oil in human nutrition has been pointed out: the hempseed oil presents an optimal omega-6 to omega-3 ratio for human health and is a rich source of two essential fatty acids (linoleic and alpha-linolenic acids) (Callaway 2004). Essential oils can be produced from the inflorescences of fibre hemp plants. They present an interesting composition in terpenes with legal and safe cannabinoid content and

therefore could be used as natural flavour and fragrance additives (Bertoli et al. 2010).

Furthermore, hemp can contribute to the sustainability of distinct industries by providing substitutes with lower environmental impact. The use of hemp fibre for textile constitutes an alternative to the high-input cotton fibre (van der Werf 2004). The use of hemp in the paper industry allows to relieve the pressure on forest reserves (Lisson et al. 2000b; Ranalli and Venturi 2004). In the automotive industry, the replacement of fossil energy-based composites by hemp fibre reinforced materials results in considerable savings in greenhouse gases emissions, which are further decreased when the storage of carbon in the hemp-based materials is considered (Haufe and Karus 2011). However, as biomass species for the production of biofuels, the interest of hemp is limited compared to fibre corn or miscanthus (Godin et al. 2013).

The total area of hemp cultivation in EU was close to 12 000 ha in 2011 (GC lin et chanvre 2010). Two thirds of the EU production originate from France. In 2004, the largest European hemp producers after France were Germany, United Kingdom, Czech republic, Romania, Russia and Ukraine, with a hemp surface larger than 1000 ha each (Bouloc 2006). However, the total hemp cultivation area in EU was higher during the period ranging from 1997 to 2003, with an average hemp area close to 15 000 ha (Hobson and Karus 2008). This situation could be explained by two main reasons (EIHA 2011). Firstly, hemp is suffering from the absence of commercial support, such as subsidies or tax incentives, resulting in an under investment from the private sectors in bio-based materials. This situation contrasts with flax fibre, which benefits from aids for processing long fibres (Hobson and Karus 2008). Secondly, hemp fibre has to compete with imported exotic fibres like jute, kenaf or sisal, which are, according to EIHA (2011), produced under low social and environmental standards. In addition, the investment in alternative crops such as hemp is affected by the cereal prices (Hobson and Karus 2008), which remain relatively high since 2007 (Index Mundi 2013).

In Wallonia, the hemp surface is limited and highly variable from one year to another. There were 135 ha in 2010, only 27 in 2011 and 80 in 2012 (Chanvre Wallon 2012). Four actors of production and transformation of hemp

have been identified: *ChanvrEco*, hurds producer for the construction sector and animal and horticultural bedding, *Pur Chanvre*, producer of organic hempseed oil, *Belchanvre*, agricultural cooperative aiming the development of a defibering line, and *IsoHemp*, producer of hemp blocks for the construction sector (Chanvre Wallon 2013). According to Hobson and Karus (2008), the production of hemp has a relatively high regional added value because storage, processing and, in most cases, further steps of the process chain take place close to the field, resulting in a higher employment rate per hectare of hemp compared to wheat used as reference crop.

In conclusion, hemp is an interesting crop from both agronomical and environmental points of view, and its production can enhance the regional development (Hobson and Karus 2008). Besides, given its interesting technical properties (Evrard and De Herde 2010; Hobson and Karus 2008; Haufe and Carus 2011), hemp can play an important role in the production of innovative biomaterials – *e.g.*, natural fibres reinforced plastics, insulation and construction materials –, while the demand for natural fibres from bio-based plastics and insulation industries is rising (EIHA 2011). However, the evolution of the European hemp production appears highly dependent on the support of both its cultivation and transformation by the EU (Hobson and Karus 2008). The European Industrial Hemp Association (EIHA) is claiming for a sustainability certification for all natural fibres to guarantee a fair competition (EIHA 2011). Last but not least, the Common Agricultural Policy (CAP) is currently being deeply reformed. In this context, the consideration of hemp in the list of crops available for the 'greening' of the first CAP pillar could offer an opportunity for the hemp production.

CHAPTER II

PHENOTYPIC PLASTICITY AND GENOTYPIC VARIABILITY IN HEMP

2.1. *CANNABIS SATIVA* L. IS A HIGHLY VARIABLE SPECIES

2.1.1. Taxonomy

Cannabis and *Humulus* constitute the only two genera of the *Cannabaceae* family (Lambinon et al. 2004), which belongs to the Rosales order (APGIII 2009). However, there is no general agreement on the infra-generic taxonomic classification of *Cannabis* (Meijer and Keizer 1996). All strains within the genus *Cannabis* intercross readily and present a continuous variation pattern for all observed morphological and agronomical traits, suggesting that all plants of *Cannabis* belongs to one single species, *C. sativa* (Small et al. 2008). According to Small and Cronquist (1976), *C. sativa* includes two subspecies, *sativa* and *indica* (Lam.) Small & Cronq, each of these comprising two botanical varieties: *C. sativa* L. subsp. *sativa* var. *sativa* and var. *spontanea*, on the one hand, and *C. sativa* L. subsp. *indica* var. *indica* and var. *kafiristanica*, on the other hand. The discrimination between both subspecies and varieties has been based on their psychoactivity and domestication status, respectively (Small and Cronquist 1976). On the opposite, other taxonomists recognize three species within the genus *Cannabis*: *C. sativa* L., *C. indica* Lam. and *C. ruderalis* Janischevsky (Schultes et al. 1974 cited by de Meijer and Keizer 1996; Emboden 1974). This classification relies on the plant morphology (height, degree of branching and achene morphology and leaf morphology) rather than on the psychoactivity. However, the classification of Small and Cronquist (1976) seems to be the most accepted today (de Meijer 1995; Forapani et al. 2001; Alghanim and Almirall 2003; van Bakel et al. 2011). The currently

cultivated varieties of hemp were domesticated from the botanical variety *Cannabis sativa* L. subsp. *sativa* var. *sativa* (Small and Cronquist 1976).

Fig. 2.1 Inflorescences in the genus *Cannabis*: (a) drug strain (Barney's farm 2014) and (b) hemp cultivar.

2.1.2. Botanical description

Hemp is an annual species. It is naturally dioecious and characterized by sexual dimorphism (Fig. 2.2a, b). The male plants have a shorter life cycle, are generally taller and have thinner stalks than the female ones, which are shorter but thicker (Bocsa and Karus 1998; Schumann et al. 1999; Struik et al. 2000). In addition to the dioecious hemp, monoecious forms of hemp exist. Monoecious cultivars have been developed from plants bearing hermaphrodite flowers or bisexual inflorescences (Moliterni et al. 2004).

(i) Vegetative system

Hemp has a well developed primary root, which can reach a depth of 2-2.5 meters depending on the soil properties (Bocsa and Karus 1998). The stem length varies from 0.75 to 3 m or more depending on the daylength, cultivar, soil and available nutrients, water supply, spacing of the plants and sex (Mohan Ram and Nath 1964; Bocsa and Karus 1998; Ranalli 2004).

The stalk is constituted of the bark and the wooden core, which are found outside and inside the vascular cambium, respectively. The bark fraction accounts for 30-40% of the stem dry matter yield (van der Werf et al. 1996; Ranalli 2004). It contains primary bast fibres (3-55 mm long) and may also contain secondary bast fibres (about 2 mm long). Primary fibres run longitudinally along the stem from bottom to top and can reach almost the full length of the plants (Van Dam and Gorshkova 2003 cited by Westerhuis et al. 2009b). Secondary fibres derive from tangential division of cambium cells when a stem part has reached its maximum length (Westerhuis et al. 2009b). Much more shorter fibres (0.5-0.6 mm long) are found in the core fraction (Lisson and Mendham 2000; Mediavilla et al. 2001).

The single fibre cells are organised in bundles by binding substances that consist mainly of pectins. The most important component of the fibre is cellulose (55% in weight). In addition, the fibre also contains hemicellulose (16%), pectins (18%) and lignins (4%) (Bocsa and Karus 1998; Chabbert et al. 2006). The cellulose is less easily decomposed by the process of stalk retting than the other fibre components, so that the woody part of the stems and cellulose-filled fibre bundles survive to retting. These can be separated from each other by mechanical breaking and scutching (Westerhuis et al. 2009a).

Hemp has palmately compound leaves which are supported by 3 to 15 cm-long petioles. The number of leaflets increases from one for the first true leaves to a number varying between 3 and 13 and is affected by the variety, age of the plant and position on the stalk (Bocsa and Karus 1998). The phyllotaxis is opposite in young plants and changes to alternate when flowering begins. However, alternate leaves have been observed in non-flowering plants (van der Werf et al. 1994).

The aerial epidermal surface of hemp is covered by numerous glandular and non-glandular trichomes (Bocsa and Karus 1998; Pate 1994). The glandular trichomes are concentrated on the young, usually sessile, leaves located in the inflorescence and on the bract that envelops the fruit (Small and Cronquist 1976; Pate 1994; Pacifico et al. 2006). They secrete a resin that contains terpenoid chemicals called the cannabinoids. The Δ^9-tetra-hydrocannabinol (Δ^9-THC) and cannabinol (CBD) are the most common among the intoxicant and non-intoxicant cannabinoids, respectively (Small and Cronquist 1976). The maximum THC level currently admitted in EU for hemp cultivars is 0.2% of the dry matter (Beherec 2000; Pacifico et al. 2006).

(ii) Reproductive system

Hemp flowers are unisexual. They arise in inflorescences that are terminal at an early stage, and terminal or lateral at a later stage (Moliterni et al. 2004). The male inflorescence consists of drooping axillary panicles with generally few or no leaves (Mohan Ram and Nath 1964; Fig. 2.2a). The female inflorescences are more compact than the male ones (Peil et al. 2003; Fig. 2.2b). They consist of crowded racemes with leafy bracts that develop at the apex of the plant or at the axils of leaves or lateral branches (Moliterni et al. 2004). Monoecious hemp plants bear inflorescences similar to that of female plants with both male and female flowers arising in variable amounts (Fig. 2.3).

The male flower has five stamens with short filaments and large and longitudinally dehiscing anthers (Fig. 2.2c). Pollen grains are copiously produced and mostly wind dispersed (Mohan Ram and Nath 1964; Small and Cronquist 1976). The female flower has a very simple structure (Fig. 2.2d). Its perianth consists of a single, green, spathe-like structure that is covered over by short brownish glands. It encloses an unilocular ovary, which includes an anatropous ovule and is surmounted by a reduced style and a bifid stigma. The fruit is bivalved and consists of a single seed (achene), which is greyish, smooth and enclosed in the persistent perianth (Mohan Ram and Nath 1964; Moliterni et al. 2004). The hempseed typically contains over 30% oil and about 25% protein, with considerable amounts of dietary fibre, vitamins and minerals (Callaway 2004).

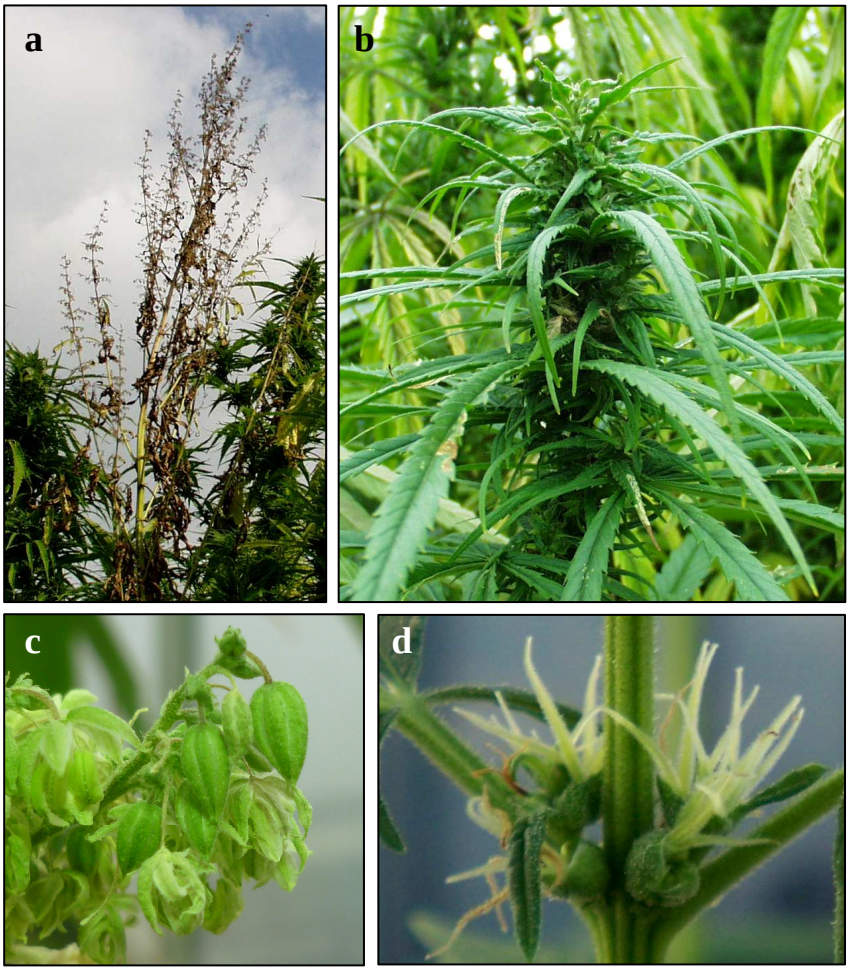

Fig. 2.2 Sexual dimorphism in dioecious hemp: male (a) and female (b) inflorescences, and male (c) and female (d) flowers.

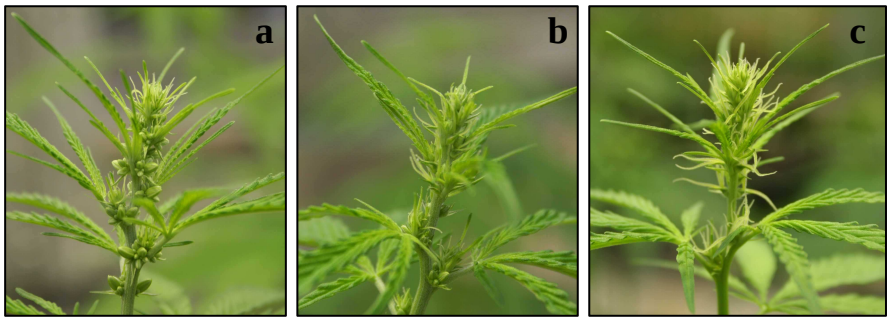

Fig. 2.3 Inflorescences of monoecious hemp displaying distinct degrees of monoecy: masculinised (a), feminised (b) and strongly feminised (c) phenotypes.

2.1.3. Growth and development

The life cycle of a hemp plant has been divided into four main stages: germination and emergence, vegetative stage, flowering and seed formation, and senescence (Mediavilla et al. 1998). The period from the first leaf pair appearance to the first flower buds at the leaf axis includes a phase of slow growth, until the fifth pair of leaves, followed by a phase of rapid growth (Bocsa and Karus 1998). As a short-day plant (Fleishmann 1938 cited by de Meijer and Keizer 1994; Borthwick and Scully 1954), flowering in hemp is affected by the photoperiod. The stem elongation is affected by the time of flowering since the stem growth decreases after flowering. Therefore, the duration from sowing to flowering and therefore the photoperiod is a factor of paramount importance in the stem yield determination (Lisson et al. 2000b; van der Werf et al. 1994).

In addition, the phenological development is affected by the sex of the plant. In a dioecious hemp cultivar, male plants flower and senesce generally earlier than the female ones (Struik et al. 2000). The phenological scale of Mediavilla et al. (1998) distinguishes the floral development (stage 3) of male, female and monoecious plants, with developmental stages specific to each flower sex in monoecious hemp (Table 2.1).

Table 2.1 Definitions and codes of growth stages of *Cannabis sativa* L. plants (Mediavilla et al. 1998)

Code	Definition	Remarks
GERMINATION AND EMERGENCE		
0000	Dry seed	
0001	Radicle apparent	
0002	Emergence of hypocotyl	
0003	Cotyledons unfolded	
VEGETATIVE STAGE refers to main stem. Leaves are considered as unfolded when leaflets are at least one cm long.		
1002	1st leaf pair	1 leaflet
1004	2nd leaf pair	3 leaflets
1006	3rd leaf pair	5 leaflets
1008	4th leaf pair	7 leaflets
1010	5th leaf pair	
: : :	: : :	
10xx	nth leaf pair	xx = 2(nth leaf pair)
FLOWERING AND SEED formation refers to the main stem including branches.		
2000	GV point	Change of phyllotaxis on the main stem from opposite to alternate. Distance between petioles of alternate leaves at least 0.5 cm
2001	Flower primordia	Sex nearly indistinguishable
Dioecious male plant		
2100	Flower formation	First closed staminate flowers
2101	Beginning of flowering	First opened staminate flowers
2102	Flowering	50% opened staminate flowers
2103	End of flowering	95% of staminate flowers open or withered
Dioecious female plant		
2200	Flower formation	First pistillate flowers. Bract with no styles.
2201	Beginning of flowering	Styles of first female flowers
2202	Flowering	50% of bracts formed
2203	Beginning of seed maturity	First seeds hard
2204	Seed maturity	50% of seeds hard
2205	End of seed maturity	95% of seeds hard or shattered
Monoecious plant		
2300	Female flower formation	First pistillate flowers. Perigonal bract with no pistils.
2301	Beginning of female flowering	First pistils visible
2302	Female flowering	50% of bracts formed
2303	Male flower formation	First closed staminate flowers
2304	Male flowering	Most staminate flowers open
2305	Beginning of seed maturity	First seeds hard
2306	Seed maturity	50% of seeds hard
2307	End of seed maturity	95% of seeds hard or shattered
SENESCENCE		
3001	Leaf desiccation	Leaves dry
3002	Stem desiccation	Leaves dropped
3003	Stem decomposition	Bast fibres free

2.1.4. Current hemp cultivars

The great majority of the current hemp cultivars consists of populations, obligately outbred in the case of dioecious hemp and reproduced by open-field pollination (Bocsa and Karus 1998; Forapani et al. 2001; Fournier and Beherec 2006). Some inbred monoecious lines are developed to breed monoecious cultivars (de Meijer 1995). The current hemp cultivars can be characterized according to their sexual type and earliness.

(i) Sexual type

Three sexual types (dioecious, 'unisexual' and monoecious) closely linked to the breeding strategy can be distinguished (de Meijer 1995).

Dioecious cultivars

The dioecious cultivars derive directly from dioecious landraces (*e.g.*, the Italian 'Carmagnola') or from crosses between distinct dioecious strains (*e.g.*, the Italian 'Fibranova') (de Meijer 1995). Compared to the monoecious cultivars, the dioecious cultivars allow the application of the method of Bredemann (1924 cited by Ranalli 2004) to increase the fibre content. This strategy consists in determining *in vivo* the fibre content of the male plants before flowering and restricting the pollination to the best male plants. Its application resulted in an increase of the fibre content from approximately 15 to 25% in thirty years (Bocsa and Karus 1998). However, the use of dioecious hemp cultivars is limited by the seed production. Indeed, waiting for seed maturity implies the dead of the male plants as a result of their earlier senescence compared to the female ones, and the presence of such dead plants makes the mechanical harvest of the seed not feasible (Bocsa and Karus 1998; Fournier and Beherec 2006).

'Unisexual' hybrids

The 'unisexual' type derives from a cross between female plants of a dioecious cultivar and a monoecious cultivar, and mainly contains female and monoecious plants (Bocsa and Karus 1998). Breeding strategies including the

creation of such 'unisexual' F_1 have been developed to solve the seed production shortage that occurs with dioecious cultivars.

'Uniko-B', 'Kompolti Hybrid TC' and the 'dioecious-monoecious hybrid' cultivars (next section) are examples of cultivars derived from unisexual hybrids. 'Uniko-B' results from the intercross of an F_1 derived from a cross between a dioecious cultivar and a monoecious one. This F_1 mostly contains female and monoecious plants, while the commercial F_2 has ca. 30% of male plants (de Meijer et al. 1995; Bocsa and Karus 1998). 'Kompolti Hybrid TC' is a dioecious cultivar obtained from a three-way cross including an unisexual F_1. This cultivar is known for its high stalk yield resulting from heterosis (de Meijer 1995; Bocsa and Karus 1998).

Monoecious cultivars

The monoecious type includes the 'truly' monoecious and 'dioecious-monoecious hybrid' cultivars. The truly monoecious cultivars (*e.g.*, 'Fibrimon') derive from crosses including monoecious inbred lines. The monoecious trait was first bred from material from Central Russia by Grisko (Ranalli 2004). The dioecious-monoecious hybrid cultivars (*e.g.*, the French 'Fedora 17', 'Felina 32' and 'Epsilon 68') are created following a scheme in two steps. A first cross is performed between female plants of a dioecious cultivar and a monoecious cultivar. The F_1 unisexual hybrid is then backcrossed with the monoecious cultivar. The F_2 accounts for the commercial seeds (Fournier and Beherec 2006).

The production of truly and hybrid monoecious cultivars is highly demanding in manpower because it requires the manual elimination of the sporadically occurring male plants before their flowering in order to avoid the reversion to the dioecious form. In the case of hybrid cultivars, this means eliminating approximately 50% of the F_1 (Fournier and Beherec 2006). In addition, the sexual phenotype is highly variable in monoecious hemp. It ranges from the strongly masculinised to strongly feminised extreme phenotypes, and, besides, is sensitive to environmental factors, among which the photoperiod and nitrogen status (Freeman et al. 1980; Fournier and Beherec 2006; Truta et al. 2007). The strongly masculinised plants are less fruiting and more prone to develop fungal infections from *Botrytis cinerea* and are therefore eliminated during the seed production process of the dioecious-monoecious hybrid cultivars in addition to the male plants. As a result of the selection processes, the 'truly'

monoecious cultivars contrast from the dioecious-monoecious hybrids by a higher variability of sexual phenotype and lower seed yields. Therefore, they are little used (Fournier and Beherec 2006). The dioecious-monoecious hybrid cultivars are usually referred to as monoecious without distinction from the 'truly' monoecious cultivars.

Compared to the dioecious ones, the monoecious cultivars display higher seed yields and crop homogeneity, and their more synchronised maturity makes the mechanical harvest of both stem and seed possible (Bocsa and Karus 1998; Mandolino and Carboni 2004; Fournier and Beherec 2006).

(ii) Earliness

The earliness of hemp cultivars is commonly defined as the time at which 50% of the plants have visible flowers (Lisson et al. 2000b; Amaducci et al. 2008a). According to de Meijer and Keizer (1994), the earliness of a given accession is primarily determined by its latitude of origin: lower latitude of origin is associated with later flowering in the growing conditions of our latitudes.

de Meijer and Keizer (1994) stated that the earliness is a highly heritable trait that varies widely among hemp accessions, the heritability being expressed at its broad sense, assessed from the variability over years of the flowering time of a wide range of hemp accessions (from 30 to 90 depending on the trial year). The existing variability in earliness is exploited by breeders to provide varieties covering a large range of earliness (Beherec 2000).

The currently available fibre hemp cultivars have been distributed along three groups according to their origin: the central Russian ecotypes (Russia, Ukraine, Poland), the Southern European ecotypes (Spain, Italy, Balkan region, Hungary, Romania), and the Far East Asian ecotypes (China, Japan, Korea) (de Meijer 1995; Bocsa and Karus 1998). The vast majority of the Southern varieties are dioecious. The central Russian types are mostly monoecious. Ukrainian varieties and most of the French varieties were developed by crosses between the southern and central Russian types. Many of these are monoecious or dioecious-monoecious hybrid populations (Bocsa and Karus 1998). From the

analysis of breeding histories, de Meijer (1995) concluded to the presence of a considerable genetic relatedness among the modern European and West Asian cultivars. On the opposite, the Far Eastern Asian hemp includes landraces rather than cultivated varieties, and would be somewhat different from the European and West Asian cultivars (de Meijer 1995; Bocsa and Karus 1998).

2.1.5. Genetic diversity in *Cannabis*

Recent molecular characterizations of the genetic diversity revealed the presence of a high degree of polymorphism in hemp (Faeti et al. 1996; Forapani et al. 2001). Four fibre hemp cultivars – three dioecious and one truly monoecious – and two drug strains – one dioecious and one female inbred line – were studied by Forapani et al. (2001). The structure of the diversity in similar amounts of within- and among-cultivar variation suggested the existence of a single and widely shared gene pool with the proportion of among-cultivar variance strongly depending upon the compared cultivars and has been described as typical of dioecious, outbreeding species (Forapani et al. 2001).

The genetic diversity among six fibre hemp varieties and nine drug strains has been analysed from five STR loci including a total of 79 distinct alleles (Gilmore et al. 2003). Significant genetic variation among accessions was found, with an average of 25% genetic differentiation, while only 6% of the variation was attributable to the chemotype. Therefore, the authors concluded to the absence of clear split between the drug and non-drug materials (Gilmore et al. 2003).

The structures of the genetic diversity observed in the genus *Cannabis* by Forapani et al. (2001) and Gilmore et al. (2003) agree with the conclusion of Small and Cronquist (1976) who assumed that *Cannabis* is characterized by a massively panmictic situation, as a result of its extensive ability to form spontaneous populations throughout the world, its wind pollination that facilitates long-distance genetic interchange and the absence of sterility barriers observed between all *Cannabis* strains.

Recently, the sequencing of the *Cannabis* genome revealed that genomic analyses might provide insights to elucidate the still controversial taxonomic treatment of the genus. A draft genome and transcriptome sequence

of the *Cannabis* drug strain 'Purple Kush' has been compared with those of the hemp cultivars 'Finola' and 'Uso 31' (van Bakel et al. 2011). An analysis of single-nucleotide variants allowed the distinction of two distinct groups including the marijuana and hemp forms, respectively. In addition, a comparison of gene expression in female *Cannabis* flowers of 'Purple Kush' and 'Finola' revealed similar overall tanscript profiles, while the entire cannabinoid pathway is expressed at higher levels in 'Purple Kush' than in 'Finola' (van Bakel et al. 2011). According to van Bakel et al. (2011), the use of sequence-based genotyping might be promising to trace the relationships in *Cannabis*, including wild germoplasm, landraces, cultivars and drug strains.

Among the hemp cultivars, the level of heterozygosity is lower in the monoecious cultivar than in the dioecious ones due to its ability to self-pollinate and the strong selective pressure needed to maintain its monoecious trait (Forapani et al. 2001; Mandolino and Carboni 2004).

2.1.6. Implications for hemp cultivation

Hemp can be grown for a large diversity of end products (Struik et al. 2000; Ranalli and Venturi 2004) and under a wide range of agro-ecological conditions (Sankari and Mela 1998; Struik et al. 2000; Cosentino et al. 2012, 2013). Given that most current hemp products derive from the stem, fibre or seed (Chapter I), we distinguished three production purposes: maximizing the stem, fibre or seed yield. However, a successful cultivation requires a special attention in the selection of a cultivar that is adapted to the production purpose and environmental cropping conditions. As previously described, the cultivar can be characterized by its sexual type and earliness (section 2.1.4), and the photoperiod appears to be a significant environmental factor in hemp cultivation since it affects both the flowering time and sexual phenotype (section 2.1.3). The three following sections review the literature on the genetic determinism of the sex expression, flowering response to the photoperiod and yield formation, respectively, in hemp.

2.2. HEMP DISPLAYS A HIGHLY VARIABLE SEXUAL PHENOTYPE

Most flowering plants are hermaphrodites, and only approximately 4 and 7% of them are strictly dioecious and monoecious, respectively (Dellaporta and Calderon-Urrea 1993). In addition, the association of dioecy with chromosomal heteromorphism is extremely rare (Parker and Clarke 1991): the presence of sex chromosomes in dioecious angiosperms has only been reported in a limited number of genera, including *Rumex*, *Cannabis*, *Humulus* and *Silene*.

2.2.1. Sex chromosomes of hemp

Hemp is diploid with 2n = 20 including sex chromosomes (Hirata 1924; Fig. 2.4). The chromosomes XX are found in female plants, and XY in male plants, with the Y chromosome larger than the X one and autosomes (Yamada 1943 cited by Sakamoto et al. 1995). The size difference between the X and Y chromosomes has been attributed to the large long arm of the Y chromosome and could be responsible for the larger genome size of diploid male plants compared to female ones (97.2% of the genome size of male plants) (Sakamoto et al. 1998).

The X chromosome in *Cannabis* is submetacentric and has a short satellite at the end of its short arm. The Y chromosome is subtelocentric. It bears the largest long arm and is also characterized by a satellite at the extremity of its short arm (Sakamoto et al. 1998). Pairing of the X and Y chromosomes at meiotic prophase I of pollen mother cells has been observed at the short arm of the Y chromosome, but not at the end of the Y long arm (Sakamoto et al. 2000).

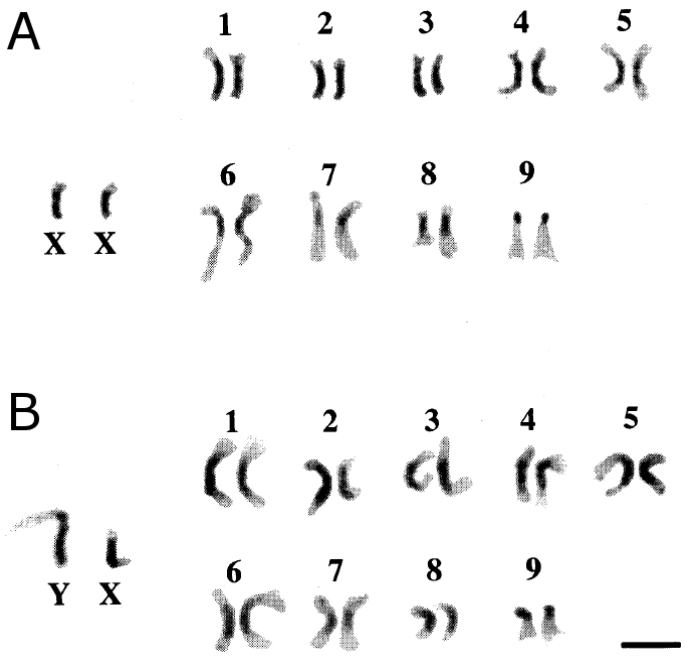

Fig. 2.4 Karyotype of female (A) and male (B) plants of *Cannabis sativa* at the pro-metaphase stage. Bars indicate 5 μm. From Sakamoto et al. (1998).

Several RAPD markers closely linked to the male phenotype have been identified. All these markers, referred to as MADC, for male-associated DNA sequence in *C. sativa*, showed similarities with retrotransposon-like sequences. The sequence of MADC1 (Sakamoto et al. 1995) and its flanking regions includes a region coding for a reverse transcriptase that is strongly homologous to those of long interspersed element-like (LINE-like) retrotransposons from various organisms (Sakamoto et al. 2000). This LINE-like retrotransposon found in *C. sativa* has been named LINE-CS. MADC3 and MADC4 include a region homologous to open reading frames that encode for polyproteins of *copia*-like retrotransposons (Sakamoto et al. 2005). The highest homologies found for MADC5 and MADC6 (50-60%) were observed with plant sequences belonging

to repetitive genome regions or retrotransposon-like sequences of different plant species (Mandolino et al. 1999; Torjek et al. 2002).

The LINE-CS retrotransposons – which include the MADC1 sequence – are present in high concentration at the terminal region of the long arm of the Y chromosome, where no pairing with the X chromosome has been observed (Sakamoto et al. 2000). Such fragments are found in the X chromosome and autosomes too; however, their accumulation at an extremity of the Y chromosome is particularly high. The insertion of such sequences might be one cause of the larger size of the Y chromosome (Sakamoto et al. 1998, 2000).

However, in addition to their male-specific portion, all of the MADC sequences include fragments that are not found exclusively in the male plants. Hybridization of DNA from male and female plants by southern blotting analysis revealed either the absence of no sex-linked polymorphism (Mandolino et al. 1999), either multiple bands among which only few are male-specific (Sakamoto et al. 2000; Torjek et al. 2002; Sakamoto et al. 2005). Similarly, the *in situ* hybridization of chromosomes with the MADC3 and MADC4 sequences produced signals dispersed on all chromosomes and thus not specific of the Y chromosome. Thus multiple sequences encoding retrotransposable elements should exist ubiquitously in the genome of *C. sativa* (Sakamoto et al. 2005). This appears in agreement with the view of Clark et al. (1993), who observed no major difference in the distribution of repeated DNA sequences between X, Y and autosomes in *Rumex acetosa*, a dioecious species with heteromorphic sex chromosomes. According to Charlesworth (2002), the abundances in repeated DNA in plants bearing sex chromosomes would be mostly similar on both sex chromosomes and autosomes.

The genetic structure of the sex chromosomes of hemp was investigated by linkage analysis between AFLP markers showing distinct types of segregating patterns with the male phenotype, referred to as ‘sex’ marker (Peil et al. 2003). The sex chromosomes of hemp include a pseudoautosomal region (PAR), *i.e.*, a region which is assumed to be homologous between X and Y (Peil et al. 2003). This region has been defined by two DNA fragments that were mapped on the Y chromosome but recombined with the sex locus – *i.e.*, the locus at which the phenotypic marker ‘sex’ was mapped – so that they were transferred to the X of the male parent and thereby were present in a significant number of female progenies. The presence of partial homology between the X

and Y chromosomes of hemp has also been indicated by markers detected on both one X of the female parent and the Y of the male parent. In addition, polymorphism was observed between different X chromosomes (Peil et al. 2003; Rode et al. 2005). Very few recombinations were observed among all of the sex-linked markers, suggesting that the sex chromosomes of hemp present typical sex behaviour (Peil et al. 2003). According to Mandolino et al. (2002), the complete linkage of markers to the male phenotype supports their location in a region of the Y chromosome excluded from recombination during meiosis. This region could belong to the extra-portion of the Y chromosome, which accounts for 2.8% of the male genome size, though the possibility of a location very close to male-determining loci cannot be ruled out. However, the position of male-determining loci in the sex chromosomes of hemp is unknown.

2.2.2. Evolution of sex chromosomes

The presence of heteromorphic sex chromosomes in plants may represent relatively advanced forms of dioecy. Such chromosomes would have occurred through a process of degeneration of the Y chromosome (Dellaporta and Calderon-Urrea 1993; Charlesworth et al. 2005). The restriction of recombination between male and female fertility factors would be the first step in establishing heteromorphic sex chromosomes. Then, more extensive restrictions of recombination between the sex chromosomes would have evolved by selective advantage if sex-linked alleles have opposite effects on the fitness in the two sexes. This would result in the accumulation of deleterious mutations linked to sex determination factors on the Y chromosome because of the absence of opportunity for such mutations to be eliminated by recombination. The accumulation of such mutations would lead to the degeneracy of the Y chromosome (Dellaporta and Calderon-Urrea 1993).

The evolution of sex chromosomes has been characterized by six successive stages, each of them being represented by a plant species (Ming et al. 2011; Fig. 2.5). In stage 1 (*Fragaria virginiana*), homologous chromosomes carry sex-determining loci for male and female sterility on proto-Y (W) and proto-X (Z) chromosomes, respectively, and recombination between the sex-determining loci is allowed. In stage 2 (*Asparagus officinalis*), recombination is suppressed between both sex-determining loci and their immediate neighboring

region, while the YY genotype is still viable. The stage 3 (*Carica papaya*) is characterized by the extension of the suppression of recombination and formation of a small male-specific region of the Y chromosome (MSY). This region would expand through the accumulation of retrotransposons and translocation and duplication of genomic fragments. The YY genotype is no more viable, while the X and Y chromosomes have the same size. In stage 4 (*Silene latifolia*), the MSY spreads to the majority of the Y chromosome. Heteromorphic sex chromosomes result from the accumulation of transposable elements and duplications in the Y chromosome. The stage 5 (*Cycas revoluta*) is characterized by the loss of non-functioning Y chromosome sequences, which can result in a shrinking of the Y chromosome. Meiotic pairing is allowed in a small portion of the sex chromosomes. At stage 6, the suppression of recombination is spread to the entire Y chromosome, leading to the loss of the pseudoautosomal region and, finally, the loss of the Y chromosome (*Rumex acetosa*). A new sex determination system based on X-to-autosome ratio evolves, whereas a new Y chromosome could be formed but plays no role in sex determination (Ming et al. 2011).

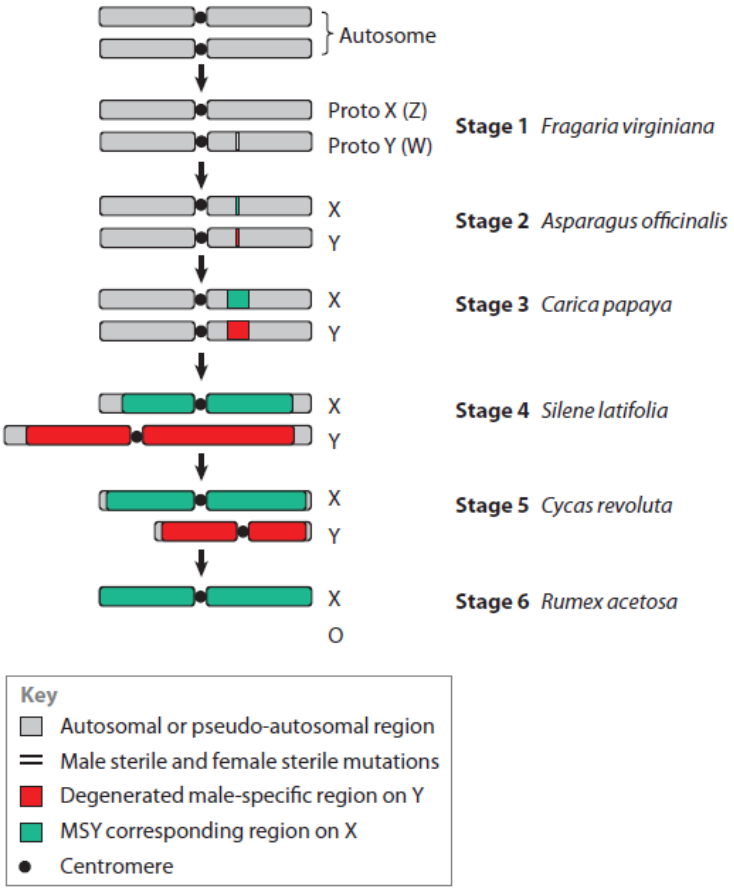

Fig. 2.5 The six stages of sex chromosomes evolution (from Ming et al. 2011).

The increased length of the Y chromosome in hemp (Yamada 1943 cited by Sakamoto et al. 1995) suggests some similarity with the sex chromosomes of *Silene latifolia* (Sakamoto et al. 1998; Fig. 2.5). In this latter species, males have XY and females XX sex chromosomes (Dellaporta and Calderon-Urrea 1993). The structure of the Y chromosome is characterized by a non-pairing region including sex-determining loci and a pairing, pseudoautosomal, region which undergoes recombination with the X chromosome during meiosis. The non-pairing region of the Y chromosome

contains female suppressing, male promoting and male fertility regions (Westergaard 1953 cited by Di Stilio et al. 1998; Charlesworth 2002). The sex determinism in *Silene* is based on an Y-active system, with dominant male factors and female suppressing factors mapping to the Y chromosome (Dellaporta and Calderon-Urrea 1993). In contrast, in hemp, the sex determinism is based on a X-to-autosomes system. In addition, the genetic basis of sex determinism in *Silene* is strong and shows little evidence for lability or environmental effects (Ainsworth 2000), while hemp displays a high plasticity of sexual phenotype. Thus, despite the presence of heteromorphic sex chromosomes similarly to *Silene*, major differences between *Silene* and hemp are observed in the sex determinism, and the stage of sex-chromosomes evolution in hemp is unclear. The genetic determinism of the sex expression in hemp is described in the next section.

2.2.3. Genetic determinism of the sex expression

In the present study, the *sex expression* of a hemp plant was defined as the ratio between the female and male flowers. The *sex* refers to as the binary characterization of the sex expression in dioecious hemp, the plants being male or female.

The sex determinism in dioecious hemp would be based on an X-to-autosomes equilibrium and not on a Y-active mechanism (Westergaard 1958; Ainsworth 2000). This assumption agrees with the experiences carried out on polyploid hemp individuals (Warmke and Davidson 1944).

The X-to-autosomes system of sex determination is based on the ratio of the number of X chromosomes to the ploidy level. This system can explain why hemp polyploid individuals with XXY and XXXY formulae are female or female-hermaphrodite, and XY, XXYY and XYY individuals are male (Warmke and Davidson 1944). An example of species with X-to-autosomes system of sex determination is *Rumex acetosa*. Females are XX, and males XY_1Y_2 (2n = 14 and 2n = 15, respectively). Both Y chromosomes are shorter than the X one (85 and 75% in length, respectively) (Parker and Clark 1991). Diploid plants with XXY and XXY_1Y_2 genotypes are fertile females. In polyploids, plants with an X-to-autosome ratio of 1 or higher are female, plants with a ratio of 0.5 or lower are male, and plants with a ratio between 0.5 and 1

bear both female and male flowers or are hermaphrodite. The Y chromosomes are required for pollen fertility but do not seem to be required for stamen development. Conversely, their presence does not inhibit the development of the female gynoecium (Dellaporta and Calderon-Urrea 1993).

Unlike in dioecious hemp, the genetic determinism of monoecious forms of hemp remains widely unknown. Out of the 13 species that Westergaard (1958) listed as having well-differentiated sex chromosomes(Westergaard 1958), the ability to exhibit sex change and/or to produce hermaphrodite offspring is known in only three species: *Rumex hastatulus*, *Humulus japonicus* and hemp (Freeman et al. 1980). Both the North-Carolina race of *R. hastatulus* and *H. japonicus* are characterized by XX and XY_1Y_2 sex chromosomes. In the former species, XX plants are strictly female while 5 to 10% of the XY_1Y_2 plants exhibit varying degrees of pistillate development (Smith 1963), and the sex determinism has elements of both Y-active and X-to-autosomes systems (Parker and Clark 1991). In *H. japonicus*, both male and female plants are able to produce flowers of the opposite sex (Schaffner 1923). The sex is determined by an X-to-autosomes system (Jacobsen 1957). The Y chromosomes are not necessary for the production of male flowers; however, the pollen cannot mature in the absence of Y chromosome (Parker and Clark 1991). According to Neve (1961 cited by Shephard et al. 1999), the sex expression in *Humulus* may be governed by genes located on the autosomes and X chromosomes. In monoecious hemp, although Hoffman (1961 cited by Truta et al. 2007) assumed the existence of XX, XY and YY forms, Menzel (1964) observed presumably XX chromosomes, and male-associated DNA markers have not been detected. However, Mandolino et al. (2002) and Moliterni et al. (2004) stated that there is no specific report on the chromosome set in monoecious hemp.

Furthermore, the sex expression in hemp is strongly affected by environmental factors, such as the photoperiod, light intensity and nitrogen status. These factors are able to induce the production of flowers of the opposite sex on dioecious hemp plants (Schaffner 1921; Borthwick and Scully 1954).

Photoperiod

On the one hand, a masculinising effect of long days is reported (Heslop-Harrison 1957, 1972 cited by Freeman et al., 1980). Arnoux (1966a, b) observed that long days increase the proportion of monoecious to female plants

that were obtained from a cross between a dioecious female parent and a monoecious plant, both in the field and controlled conditions. On the other hand, Borthwick and Scully (1954) noted that the transfer of plants from a long photoperiod (18h) to appreciably shorter ones (8 or 11h) resulted in the production of male flowers on a greater percentage of female plants than when they were transferred to photoperiods only slightly shorter than the critical photoperiod for flowering (14h).

Light intensity

According to Borthwick and Scully (1954), a high light intensity would induce the production of male flowers on a greater percentage of hemp female plants. In constrast to hemp, a feminising effect of high light intensity on the sex expression of many other dioecious species has been observed (Freeman et al. 1980).

Nitrogen status

The effect of nitrogen on the sex expression in hemp is not clear. No significant difference in the production of male flowers by hemp female plants was observed between low and high nitrogen treatments by Borthwick and Scully (1954). According to Arnoux (1966a, b), rich nitrogenous nutrition has a masculinising effect on hemp plants. On the opposite, van der Werf and van der Berg (1995) observed a higher ratio of female to male plants under increased nitrogen fertilization during one of two trial years. This observation agreed with the feminising effect of nitrogen that was reported in hemp and many other dioecious species (Freeman et al. 1980). However, the shift of sex-ratio under high nitrogen fertilization was observed only within a sub-population of relatively short plants, suggesting that it might result from the effect of nitrogen fertilization on size variability. The absence of nitrogen effect on the sex ratio during the second trial year remained unexplained (van der Werf and van der Berg 1995).

According to Freeman et al. (1980), long day conditions and low nitrogen status could be considered as stress factors in hemp that are able to reduce the ratio of female to male plants. In this sense, it is possible that the transfer of plants from long to short photoperiods as mentioned above (Borthwick and Scully 1954) acted as a stress factor, resulting in the production

of male flowers by a higher number of female plants. In contrast to the photoperiod, it appears that the effects of light intensity and nitrogen status on the sex expression of hemp have been little described and remain widely unclear.

In addition to environmental factors, effects of exogenous hormones on the sex expression in hemp have been widely reported. Gibberellin induces predominantly male plants, while auxin, cytokinin, abscisic acid and ethylene induce the feminisation of the plants (Heslop-Harrison 1956; Mohan Ram and Jaiswal 1972; Chailakyan and Khryanin 1978, 1979; Freeman et al. 1980).

The ability to undergo sexual reversion might have a genetic base. Borthwick and Scully (1954) noted a variable tendency to form male flowers among different dioecious female lines and concluded that the occurrence of monoecy in dioecious hemp is, to some extent, hereditary. Similarly, differences in the resistance to sex reversion treatments were noted between two dioecious hemp cultivars (Grassi, de Meijer, pers. comm. cited by Moliterni et al. 2004).

The diversity of intersexual forms, the bipotency of sexually predetermined plants and the occurrence of fertile male flowers on female plants of dioecious hemp suggested that the sex in hemp would be determined by the activity of genes that are located not only on the sex chromosomes but also on the autosomes (Grisko 1937 cited by Truta et al., 2007; Sengbusch, 1952 cited by Westergaard, 1958; Rath, 1968 cited by Truta et al., 2007; Migail, 1986 cited by Mandolino and Ranalli, 2002). Grişko (1937 cited by Truta et al. 2007) considers that the determinants of phenotype and those of sexualisation are independent since female plants are able to produce male flowers under given environmental conditions. The ability to reverse the sex determination mechanism by hormonal treatment suggested that sex determination genes could regulate alternative programs of sexuality through a signal transduction mechanism that modifies endogenous hormonal levels (Dellaporta and Calderon-Urrea 1993).

2.2.4. Sexual differentiation

The male and female flowers of hemp differ radically in general morphology and size, and show no evidence of the missing sex: the female flowers result from the direct formation of carpel initials from perianth initials, they never form any vestiges of stamen initials (Mohan Ram and Nath 1964).

According to Galoch (1980), the transition to the phase of generative development in hemp is associated with increased content of an ABA-like inhibitor in both female and male individuals. Female plants exhibit higher ABA and auxin endogenous levels than male plants, which conversely have a higher gibberellin level. The formation of male generative organs may be associated with an increased demand of the differentiating floral meristem for gibberellin, and that of the female organs for auxins (Galoch 1980).

A morphological and molecular characterization of apices at the leaf axil in male and female hemp plants was conducted by Moliterni et al. (2004). Gene expression was analysed on apices that were at the earliest step of sexual commitment, when floral buds were not yet visible, so that differences between both sexes could be related to the onset of apex sexual differentiation. An amount of 900 cDNA-AFLP fragments out of 4800 was polymorphic between the male and female plants. The sexual polymorphism of 22 of them was confirmed by northern hybridization, and five fragments, all more expressed in female apices, were sequenced. This higher expression in female apices suggested that, at this development stage, some metabolic processes are more active in the females than in males, either due to their specific activation in females, or to their repression in males (Moliterni et al. 2004). Four of the five fragments presented a high degree of similarity with known sequences. A similarity was found with a Rac-GPT binding protein, which plays a signalling role in auxin-regulated gene expression in *Arabidpopsis*.

2.3. THE FLOWERING TIME IS MODULATED BY THE PHOTOPERIOD

Hemp is a quantitative short-day plant (SDP) with flowering delayed by photoperiods greater than approximately 14 hours (Borthwick and Scully 1954; Lisson et al. 2000b; Amaducci et al. 2008a). Provided that the supply of energy, nutrients and water for growth is sufficient, the rate of development of a crop towards flowering is controlled by temperature and daylength. The temperature affects the development in two ways: firstly, as a general promoter of development through activation of enzymatic processes; secondly, as a preconditioner of photoperiodism (Horie 1994). The duration from emergence to flowering in hemp has been modelled in response to the temperature and photoperiod according to three distinct approaches (Lisson et al. 2000b; Amaducci et al. 2008a; Cosentino et al. 2012).

The first approach (Major 1980) divides the duration from emergence to flowering into a vegetative phase of development, which lasts from emergence to floral initiation, and a flower development phase (*FDP*), from floral initiation to first flower appearance (Lisson et al. 2000b; Fig. 2.6). The vegetative phase of development itself is split into a temperature-dependent basic vegetative phase (*BVP*) and a daylength-dependent photoperiod induced phase (*PIP*). The duration of the *PIP* is assumed to be null under photoperiods shorter than the critical photoperiod for flowering (P_c). Lisson et al. (2000b) observed that hemp plants flowered in a minimum and constant thermal duration under photoperiods shorter than approximately 14h. On the opposite, under photoperiods longer than P_c, the duration of the *PIP* increases linearly as a function of the photoperiod and genotypic sensitivity to the photoperiod (*PS* for photoperiod sensitivity). The *PS* is expressed in units of thermal time delay per hour of increase in daylength (Lisson et al. 2000b). According to Summerfield et al. (1991) and Yan and Wallace (1998), the genes conferring the photoperiod sensitivity act as inhibitors of the plant development to flowering, *i.e.*, they can delay but not promote flowering, and their activity occurs only under photoperiods that delay flowering ($P > P_c$ for SDP). Hemp flowering under 24-h daylength has been reported after significant delays by Schaffner (1926), suggesting the absence of plateau for the flowering response of hemp to the photoperiod. Similarly to the *BVP* duration, the duration of the *FDP* did not appear affected by the photoperiod (Lisson et al. 2000b). This was also true for the duration from

flowering to harvest, which did not show any clear response to the photoperiod (Lisson et al. 2000b).

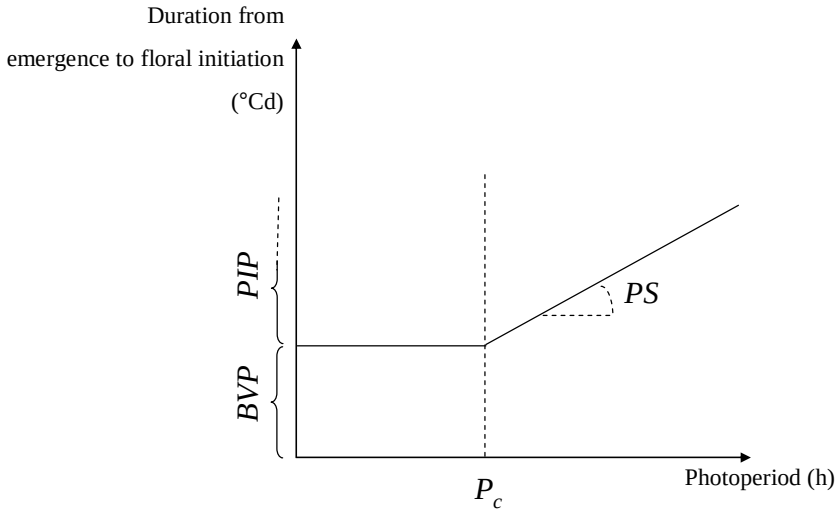

Fig. 2.6 Flowering response of hemp to the photoperiod. Illustration of the model described by Major (1980) and applied in hemp by Lisson et al. (2000b), where P_c is the critical photoperiod for flowering, *BVP* and *PIP* account for the basic vegetative and photoperiod induced phases of development, respectively, and *PS* is the photoperiod sensitivity, expressed in °Cd h^{-1}.

The model of Amaducci et al. (2008a) was implemented using the framework of the model of Lisson et al. (2000b), *i.e.*, the division of the duration from emergence to floral initiation in three successive phases (*BVP*, *PIP* and *FDP*). However, their modelling approach was based on data from field trials and integrated the cumulative effect of hourly variations of the temperature and daily variations of the photoperiod. Non-linear developmental responses to the temperature and photoperiod were used. The large variability observed in the thermal duration of the *PIP* supported the significant modulating role of the photoperiod in the duration of this phase. Besides, the differences in flowering time among cultivars were mainly related to differences in the duration of the

PIP, indicating that a substantial part of the genotypic variation in flowering time is due to the sensitivity to sub-optimal photoperiods (Amaducci et al. 2008a). The cardinal temperatures found for hemp cultivation by Lisson et al. (2000c) and Amaducci et al. (2008a) were consistent: 1 and 1.9°C for the base temperature for the *BVP*, 29 and 26.4°C for the optimum temperature and 41 and 40°C for the ceiling temperature, respectively.

A third model predicting the time of 50% flowering has been proposed by Cosentino et al. (2012). Their approach differed from both previous ones by the assumption that the effect of the temperature is the highest under optimal photoperiods for flowering and decreases under both shorter and longer photoperiods than the optimal ones. In addition, the duration from sunrise to sunset plus the civil twilight rather than the daylength was used to account for the effect of the photoperiod. The obtained optimal daily light length was 14.7 h regardless of the cultivar.

The flowering duration, computed as the duration from the first to the last flower appearance, and flowering dynamics were investigated in both dioecious and monoecious hemp cultivars (Amaducci et al. 2008b). The plants flowered earlier and their flowering duration was shorter in the late sown treatments, which experienced shorter daylengths during their photoperiod-induced phase of development. The similarity of the patterns of flowering duration and flowering time in response to sowing time suggested that a same underlying mechanism controls both parameters. The percentage of flowering plants was described as a function of time by a bi-logistic curve. This indicated the presence of two underlying processes, corresponding to the successive male and female flowerings in the dioecious cultivars. The goodness of the bilogistic curve for fitting the flowering dynamics in monoecious hemp suggested the presence of two groups with distinct flowering characteristics. Amaducci et al. (2008b) assumed that early plants in monoecious hemp cultivars would bring male, protandric characteristics.

2.4. INCREASING THE YIELD IN STEM, FIBRE OR SEED

Monteith (1972, 1977) defined the efficiency of a crop production in thermodynamic terms, *i.e.*, as the ratio between the energy output (carbohydrates) and input (radiation). The dry matter accumulated by a non-stressed crop varies linearly as a function of the amount of photosynthetically active radiation (PAR) intercepted by the canopy, and the slope of this relationship determines the radiation-use efficiency (RUE). Greater yields rise from increasing the proportion of dry matter consisting of the harvested organs of the plant (van Dobben 1962). This proportion, which accounts for the economic part of the accumulated biomass, has been referred to the harvest index (HI) (Donald 1962). Considering the above, the crop physiological characteristics of hemp were described by van der Werf et al. (1996) as follows:

$$Y = PAR_{intercepted} \times RUE \times HI \qquad (2.1)$$

where Y is the yield in harvested dry matter of a crop that achieves its full potential, *i.e.*, growing in non-limiting conditions of water and nutrients supply and in the absence of pest, disease attack and any other stress. The factors affecting the amount of intercepted PAR, RUE and dry matter partitioning are reviewed here with the aim of increasing the yield in stem, fibre and/or seed.

2.4.1. Light interception

The amount of PAR intercepted by a crop depends on its emergence date, the rate of canopy establishment, the proportion of light intercepted by the fully established canopy, the date of onset of canopy senescence and the rate at which the light interception decreases during the senescence. These factors can be affected by environmental parameters, including temperature, radiation and daylength, and by management decisions (van der Werf et al. 1996).

Assuming that the canopy is a homogenous environment with leaves uniformly distributed over the land, the transmission of the light in the canopy can be described on the basis of Beer's law:

$$F = 1 - e^{-k\,LAI} \tag{2.2}$$

where F is the fraction of PAR intercepted by the canopy, LAI is the leaf area index and k is the coefficient of light extinction in the canopy. The value of k is commonly 0.7 – 1.0 in a canopy with horizontal leaves and 0.3 – 0.5 in a canopy with vertically inclined leaves (Monsi and Saeki 1953 cited by Hirose 2005). The high extinction coefficient estimated for hemp (0.96) was attributed to the planophile character of its canopy (Meijer et al. 1995).

A base temperature of 2.5°C was estimated for the full canopy closure (90% of light intercepted), and this stage was reached at 340°Cd (base 2.5°C) or 465°Cd (base 0°C) after emergence at 64 plants m^{-2}. The course of proportional light interception by the canopy from plant emergence to full canopy closure has been described as a logistic function of thermal time accumulation (Spitters 1990):

$$F_t = \frac{1}{1 + ((1/Nf_0) - 1)e^{-R_0 t}} \tag{2.3}$$

where F_t is a daily value of F, R_0 the initial relative leaf area growth rate, N the plant density, f_0 the initial fraction of light intercepted per plant and t the temperature sum calculated with the appropriate base temperature (Meijer et al. 1995). Cumulative light interception until canopy closure has been calculated using daily values of F and PAR (Meijer et al. 1995; van der Werf et al. 1995a, 1995b). The decrease of F during senescence has been described by a straight line (van der Werf et al. 1995b).

2.4.2. Radiation-use efficiency

The RUE of a non-stressed crop species depends on the crop gross photosynthesis, the respiration for maintenance and growth (Charles-Edwards 1982) and dry matter losses during the growing season, which may induce an apparent decrease in RUE (van der Werf et al. 1996). In hemp, Meijer et al. (1995) found an average RUE of 1.94 g MJ^{-1} over the entire growing season and RUE values of 2.1 and 1.2 g MJ^{-1} for the periods prior and posterior to flowering, respectively. These values were higher than those found in *Festuca arundinacea* (from 1.2 to 1.8 g MJ^{-1} depending on the growing season; Belanger et al. 1994) and *Dactylis glomerata* (1.2 g MJ^{-1}; Mills et al. 2009). In contrast, the pre-flowering RUE of hemp was relatively lower than the pre-flowering RUEs of 2.2 to 2.9 g MJ^{-1} reported for other C_3 species such as sunflower, rice, wheat and potato (Kiniry et al. 1989; Haverkort et al. 1992). The relatively low RUE in hemp compared to these latter species has been explained by several factors (Meijer et al. 1995). Firstly, the photosynthesis rate of the hemp canopy is negatively affected by the high extinction coefficient of the hemp canopy during most of the growing season. Indeed, at small LAI, horizontal foliage (k ~ 0.9) intercepts more light than erect foliage (k ~ 0.3), which enhances the canopy photosynthesis rate. However, at high LAI – which is the case during most of the growing season –, a high extinction coefficient reduces the canopy photosynthesis (Trenbath and Angus 1975; Monteith 1977). According to Meijer et al. (1995), an extinction coefficient of 0.75 instead of 0.96 would have increased the RUE and above-ground dry matter production by approximately 2.75%. Secondly, the cost for converting primary assimilates in lignin is higher than that associated with the synthesis of starch or cellulose (Penning de Vries et al. 1974 cited by Meijer et al. 1995). The hemp stem is relatively rich in lignin compared to the storage organs of carbohydrate producing crops like wheat and sugarbeet. Therefore the growth respiration due to the synthesis of lignin would be relatively larger in hemp than in wheat or sugar beet. The synthesis of lignin in the stems would account for a decrease of 6-7% of the above-ground dry matter production and RUE (Meijer et al. 1995). Thirdly, dry matter losses as a result of leaf senescence or self-thinning were not considered in the total dry matter production and could have reduced the measured total above-ground dry matter by 2.5 t ha^{-1}, resulting in an underestimation of the RUE (Meijer et al. 1995; van der Werf et al. 1996).

(i) Radiation-use efficiency and flowering

According to van der Werf et al. (1994, 1996), only a minor part of the decrease of the RUE of hemp following the flowering would be due to larger losses in dry matter (shed leaves and dead plants) and increased growth respiration due to oil and protein synthesis. Rather, the low post-flowering RUE would be primarily due to a significant decrease in crop photosynthesis likely as a result of the senescence of leaves. This explanation arose from field trials in which two daylength treatments – ambient daylength and 24-h long daylength – were applied before plants started to flower (van der Werf et al. 1994). The total dry matter production and the allocation of dry matter to the leaves and stem were higher while the allocation of dry matter to the inflorescence was lower under long days than under ambient daylength. However, the high energy costs for oil and protein synthesis in the seed and the loss of leaves under ambient daylength were estimated to account for only a minor part of the difference in total dry matter production between both treatments. On the opposite, the LAI of plants at final harvest was strongly lower under ambient daylength than under long days, while the PAR intercepted by the canopy did not significantly differ from one treatment to the other. The low RUE found under ambient daylength would therefore likely be due to a lower rate of canopy photosynthesis after than before flowering (van der Werf et al. 1994).

(ii) Radiation-use efficiency and plant density

Lower RUE values were obtained in high dense hemp crops (270 plants m^{-2}) although both the living and dead dry matters were included in the computation of the RUE (van der Werf et al. 1995b). Similar amounts of PAR were intercepted by the canopy between plant densities of 30, 90 and 270 plants m^{-2}, while the LAI was higher at 270 m^{-2} (4.8) than at 90 and 30 plants m^{-2} (4.16 and 3.63, respectively). In contrast, the crop growth rate was lower at 270 plants m^{-2} (120.7, 125.8 and 90.8 kg ha^{-1} d^{-1} at 30, 90 and 270 plants m^{-2}). This observation has been attributed to small RUE, likely as a result of higher self-thinning at high plant density considering that the dry weight of the dead plants was underestimated because of dry matter decomposition (van der Werf et al. 1995b, 1996). In addition, according to Whaley et al. (2000) in wheat, it is possible that the lower plant densities (30 and 90 plants m^{-2}) allowed a more even distribution of the light in the canopy, such that more leaves, lower in the

canopy, could photosynthesise at a high rate, increasing thereby the photosynthesis capacity of the canopy.

2.4.3. Seed yield formation

The formation of the seed yield has been little investigated in hemp. The present paragraph provides general information on the constitution of the seed yield in grain crops in response to environmental and genetic factors. It is based on a chapter book and a review of Egli (1994, 2004).

In grain crops, the seed yield is the weight of seeds produced per unit of area. Though the compensatory effects can be such that no relation can be observed between the yield at the crop and seed levels, a better understanding of the processes affecting the seed yield formation can be achieved by studying the growth of a single seed. The weight of a single seed is determined by its rate and duration of dry matter accumulation.

Genetic variation and environmental changes can affect the availability of assimilates or other growth factors to the seed and thereby induce differences in the seed growth rate or seed-fill duration. These changes occur through mechanisms that can operate at two levels: the plant or the seed. Seed growth requires a photosynthetically active plant to provide raw materials. Therefore, the progressions of leaf senescence and seed-fill duration tend to change in concert: a slow decline in canopy photosynthesis, associated with delayed plant senescence, can result in an increased seed-fill duration and seed size. On the opposite, having a source of raw materials is not a guarantee that seed growth continues: the seed itself can limit the accumulation of dry matter and the increase in seed size, so that the termination of seed growth can occur before the completion of senescence and death of the vegetative plant.

Genetic differences in the seed growth rate and seed-fill duration were observed in soybean (*Glycine max* L.) (Egli 1994). The seed growth rate was positively related to the number of cells in the cotyledons. The seed-fill duration can be affected by the timing of anthesis: assuming that the maturation of seeds on a plant is usually more uniform than the beginning of seed filling of individual seeds, late appearing flowers result in shorter seed-fill durations. The seed-fill duration linearly decreases with temperatures below 30°C in many

crops including maize, wheat, sunflower, lentil and oat. However, the negative effect of short seed-fill duration at high temperature may be reduced by an increased seed growth rate, though differences among species would be important. Water stress during the seed development shortens the filling period, probably primarily because of accelerated leaf senescence, and reduces yield in many crop species (Egli 2004). In soybean, a short-day species, the duration of flowering includes a photoperiod-sensitive phase and is longer under long photoperiods (Asumadu et al. 1998), and a positive effect of long days on the seed-fill duration has been reported (Kantolic and Slafer 2001). However, the seed-fill duration in soybean was not sensitive to the planting date in the field (Egli et al. 1987), suggesting that the photoperiod would not have a significant effect on the seed-fill duration in production fields (Egli 2004).

In hemp, positive correlations were observed between the seed yield, on the one hand, and plant dry weight and height, on the other hand. The highest seed yields were obtained with the earliest flowering and ripening varieties; however, relatively high seed yields were also obtained with late-flowering accessions (Schumann et al. 1999). According to Picault (2006), higher seed yields are obtained with early sowings and low seeding rates. However, excessively early sowing must be avoided because they are associated with long flowering duration, which increases the probability of heterogeneous seed maturity, and thereby the number of green seeds at harvest (Picault 2006).

2.4.4. Dry matter partitioning

(i) Total dry matter yield and stem *vs.* inflorescence partitioning

Delaying flowering can efficiently increase the total above-ground dry matter yield (van der Werf et al. 1994). Indeed, preventing flowering by long days decreased the proportion of the inflorescence in the above-ground dry matter and conversely increased the proportion of stem, which resulted in a higher total above-ground dry matter yield (van der Werf et al. 1994). This effect was more significant in a monoecious cultivar than in a dioecious cultivar which was characterized by a smaller allocation of dry matter to the inflorescence, likely as a result of its late flowering and the presence of approximately 50% of male plants with smaller inflorescence fraction compared

to the female plants (van der Werf et al. 1996). Therefore, the use of late flowering and dioecious **cultivars** can positively affect the proportion of stem in the above-ground dry matter.

Postponing the **sowing time** decreases the plant weight at harvest and stem dry matter yield, while postponing the **harvest time** increases both of them (Westerhuis et al. 2009b). According to van der Werf et al. (1996), delaying the harvest date – from 15/08 to 15/09 or 15/10 at De Bilt, Netherlands – could improve the crop yield by increasing the amount of intercepted PAR.

The **plant density** can positively affect the total dry matter production of a hemp crop only within a certain range. The number of hemp living plants decreases during the growing season, and this decrease is higher in densely sown crops as a result of higher inter-plant competition (Meijer et al. 1995; Lisson and Mendham 2000). Stem and total dry matter yields respond similarly to the plant density (van der Werf et al. 1995b; Lisson and Mendham 2000). In hemp, the highest dry matter yields were observed at plant densities comprised between 80 and 120 plants m^{-2} (van der Werf et al. 1995b; Cromack 1998; Lisson and Mendham 2000; Amaducci et al. 2008c). Furthermore, the proportion of stem in the total dry matter increases with the plant density. This higher proportion of stem in dense crops was associated with a delayed flowering and consequent decline in dry weight of the inflorescence, and to an increased initial stem growth caused by higher inter-plant competition (van der Werf et al. 1995b). The plants are slender (increased ratio height/weight) and shorter at final harvest in dense populations (270 plants m^{-2}).

Nitrogen fertilization increases the final total dry matter production and stem dry matter yield of hemp. Indeed, higher stem yields have been observed at 200 than at 80 kg N ha^{-1} throughout the entire growing season (van der Werf et al. 1995c). However, higher self-thinning rate and lower plant slenderness were obtained at high N level. According to van der Werf et al. (1995c), this was due to the higher LAI and PAR interception at high N level, resulting in an enhanced competition for light. The proportion of stem in the total dry matter yield was barely affected by the N fertilization (van der Werf et al. 1995c).

(ii) Fibre yield and content

Within the stem, a high proportion of bark is generally sought because of its higher economic value compared to the stem core (van der Werf et al. 1996). For long fibre uses (*e.g.*, textiles), high bark yield with high primary bast fibre content and low secondary bast fibre content are advantageous (Mediavilla et al. 2001; Westerhuis et al. 2009a). In dioecious hemp, peaks of stem, bark and fibre yield would be reached at the peak of female flowering and end of male flowering, which was described as the time of 'technical maturity' (Mediavilla et al. 2001). With respect to the stem quality, postponing the harvest after the time of technical maturity results in a decrease in bark quality due to an increased production of secondary fibres (Mediavilla et al. 2001).

According to Westerhuis et al. (2009a, b), the total fibre weight per stem would be mostly determined by the stem weight, cultivar and position on the stem. The total fibre weight per stem is positively correlated with the **stem weight**, while the fibre content in the stem is negatively correlated with the stem weight (Westerhuis et al. 2009b). This observation appears in agreement with that of van der Werf et al. (1994), who noted a negative correlation between stem dry weight and bark content in the stem: an increase in stem dry weight by postponing the harvest was associated with a decrease in bark content.

With respect to the **position on the stem**, the fibre content is higher in the middle part of the stem and lower towards both bottom and top (Westerhuis et al. 2009a).

The fibre content in the stem is significantly affected by the **cultivar** (van der Werf et al. 1995b; Cromack 1998; Sankari 2000; Amaducci et al. 2008c; Westerhuis et al. 2009b). The application of the Bredemann method (1924 cited by Ranalli 2004) and the high narrow-sense heritability of the fibre content (Hennink 1994) led to a considerable improvement of the trait (Ranalli 2004).

In addition, the bark content in the stem would be positively affected by the **plant density**, and thus by the plant slenderness, until a certain level of plant density (van der Werf et al. 1995b). Finer fibres and less secondary growth were obtained at high plant densities (Amaducci et al. 2008c). However, at a certain plant density, the percentage of bark in the stem would be lower as a result of

the elimination of particularly slender plants with relatively higher bark content by self-thinning (van der Werf et al. 1995b). According to Westerhuis et al. (2009a), the plant density would affect the fibre content through its effect on the stem weight only.

Lower bark content was observed under high **nitrogen fertilization**. This effect likely resulted from a difference in plant density between both nitrogen treatments, difference which had arisen as a result of self-thinning under high nitrogen (van der Werf et al. 1995c). According to Westerhuis et al. (2009a), it is possible that, similarly to the plant density, the effect of the nitrogen fertilization on the bark content occurs through an effect on the stem weight.

2.4.5. Yield as affected by the crop management

Increasing the dry matter yield in stem, seed or fibre can be achieved by acting on distinct factors. The effects of cultivar, sowing date, seeding rate, N fertilization and harvest date on the stem, fibre and seed yields are summarised here.

(i) Cultivar selection

Late flowering cultivars increase the total production of dry matter by delaying the flowering and associated reduction of radiation-use efficiency (van der Werf et al. 1994). This increase in total dry matter production has been linked to higher stem yields and a lower allocation of assimilates to the floral parts. On the opposite, highest seed yields were obtained in the earliest flowering accessions but earlier accessions not always had higher seed yields (Schumann et al. 1999).

(ii) Sowing date

Advancing the sowing date – from 15/04 to 31/03 or 16/03 at De Bilt, Netherlands – can be effective to get an earlier canopy closure and thus higher total amount of intercepted PAR, but must be balanced against the increased probability of frost damage (van der Werf et al. 1994, 1996). Besides, sowing

before April in France must be avoided because it induces a long flowering duration and, thereby, a high heterogeneity of seed maturity and quality (Picault 2006).

(iii) Seeding rate

A high seeding rate accelerates the canopy closure, delays flowering, increases the proportion of stem in the total dry matter and the proportion of bark content in the stem and, to a certain extent, the total dry matter production (van der Werf et al. 1995b, 1996). Maximizing the total dry matter production by increasing the seeding rate can be achieved under the highest possible plant density not causing self-thinning (van der Werf et al. 1996).

According to Picault (2006), a seeding rate of 20 kg ha^{-1} would allow achieving 95 and 90% of the potential seed and stem yields of a hemp crop, respectively. However, the height of the stems increases while their slenderness decreases with plant density (van der Werf et al. 1995b), so that, at low seeding rate, the stems are more difficult to harvest (Picault 2006). Seeding rates comprised between 50 and 60 kg ha^{-1} would be recommended for fibre production. At such rates, the seed yield would reach approximately 90% of its potential (Picault 2006).

(iv) Nitrogen fertilization

N fertilization of 200 rather than 80 kg N ha^{-1} increases both total and stem dry matter yields (van der Werf et al. 1995c). The N fertilization does not seem to affect the proportion of stem in the total dry matter yield. On the opposite, a negative effect on the bark content has been reported (van der Werf et al. 1995c). According to Picault (2006), a late N fertilization would positively affect the seed yield in hemp, similarly to wheat (Ellen and Spiertz 1980).

(v) Harvest date

Delaying the harvest date can increase the stem weight and stem dry matter yield (Westerhuis et al. 2009b). However, this makes the field drying of hemp stems more difficult. Moreover, increased stem dry weights are related to

decreased fibre contents in the stem (Westerhuis et al. 2009a). Furthermore, postponing the harvest after the peak of flowering affects the fibre quality since it increases the amount of secondary fibres in the stem (Mediavilla et al. 2001). Thus, when harvesting the seed in addition to the stem, waiting for seed maturity can negatively affect the fibre yield and quality.

The existence of a market for hemp seed led to the adaptation of monoecious fibre cultivars for the simultaneous production of fibre and seed (Fournier and Beherec 2006). These cultivars are characterized by a relatively early flowering in order to reach the seed maturity. However, as reported in the present section, sowing date, seeding rate and harvest date may have opposite effects on the yields in stem and seed. Therefore, managing a dual-purpose hemp crop will rely on compromises for most agronomic factors.

2.5. SYNOPSIS

This chapter firstly presented the variability existing in the species *Cannabis sativa* (section 2.1) before reviewing the literature on the genetic determinism of the sex expression (section 2.2), flowering response to the photoperiod (section 2.3) and yield formation (section 2.4) in hemp. Considering this literature review, the present section aims to highlight the sources of variation that may affect each of the four traits investigated in the present work: the sex expression, flowering phenology, and yields in stem and seed. To this purpose, we used the model of the quantitative genetic theory, which describes the phenotype (**P**) as a function of the genotype (**G**), environment (**E**) and their interaction (Lynch and Walsh 1998). The situation in the present study is as follows (Fig. 2.7).

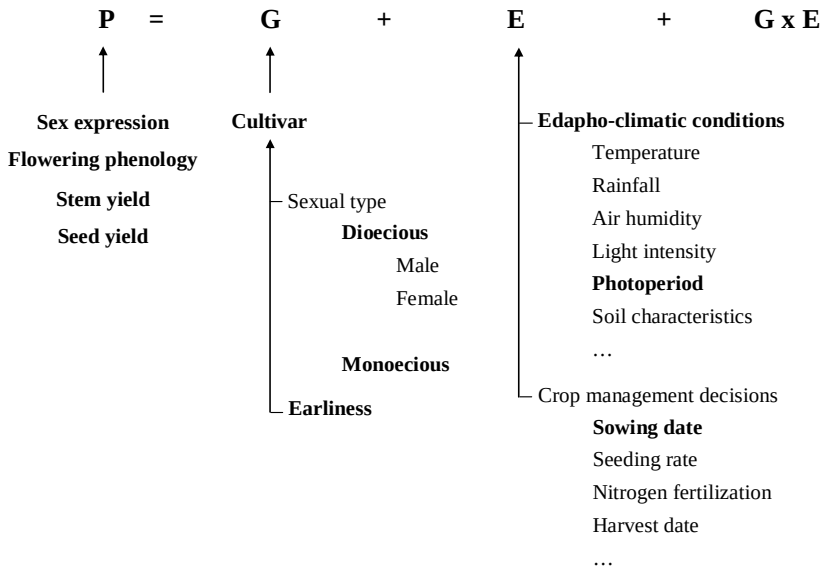

Fig. 2.7 The four traits investigated in the present thesis (sex expression, flowering phenology and yields in stem and seed) and their potential sources of variation expressed on the basis of the model of the quantitative genetic theory. P, G and E account for phenotype, genotype and environment, respectively. The experimental factors that were tested in this work are in bold face.

P is the observed trait value, *i.e.*, the sex expression, expressed as the ratio between the female and male flowers, the flowering phenology, expressed as the duration of a given phenophase, and the yields in stem and seed.

G accounts for the mean phenotypic value observed for a given genotype. The ability of a genotype to produce distinct phenotypes when exposed to different environments is referred to as the phenotypic plasticity (Pigliucci 2005). The genotype depends on the cultivar, which can be characterized by its sexual type and earliness (section 2.1.4). The sexual type is either dioecious or monoecious. The genotype of dioecious hemp plants varies with sex given the presence of heteromorphic sex chromosomes (section 2.2.1). However, the sex chromosomes of monoecious hemp have not been stated to date (Mandolino et al. 2002; Moliterni et al. 2004). In the present work, the four studied traits were characterized in five monoecious cultivars covering a range of earliness. An additional dioecious cultivar was used to investigate the genetic determinism of the sex expression in hemp.

E is the deviation due to the growing conditions. The effect of the environment can be explained by the edapho-climatic conditions – *e.g.*, temperature, rainfall, air humidity, light intensity, photoperiod, soil characteristics – and the crop management – *e.g.*, sowing date, seeding rate, nitrogen fertilization, harvest date. In the present study, trials were carried out in both field and controlled conditions. In the field, two trial locations were investigated, each of them being characterized by its own edapho-climatic conditions. In addition, due to its effect on both the sex expression (section 2.2.3) and the flowering time (section 2.3), the photoperiod was selected as an experimental factor to create environmental variation. In the field, the effect of distinct photoperiodic conditions was included in the 'sowing date' factor, while, in controlled conditions, distinct photoperiodic treatments were applied.

Non-parallelism of the phenotypic responses of a set of genotypes to the environment refers to as **G x E interaction**. The occurrence of these interactions affects plant breeding strategies when the phenotypic superiority of genotypes changes in relation to the environment (Malosetti et al. 2004).

Before addressing the experimental part, the next chapter investigates the dissection of the genotypic effect on the value of quantitative traits by the use of genetic molecular markers and appropriate statistical methods. This

technical presentation aims to support the understanding of chapters VII and IX, which are dedicated to the construction of linkage maps of the sex expression and investigation of its genetic determinism by the search of quantitative trait loci linked to the sexual phenotype, respectively.

CHAPTER III

INVESTIGATING THE GENETIC DETERMINISM OF A QUANTITATIVE TRAIT

Most traits in nature and of importance to agriculture (*e.g.*, yield, quality, drought resistance) are quantitatively distributed. Their expression is regulated by the cumulative action of many genes and, in most cases, is influenced by environmental factors (Lynch and Walsh 1998). These traits are called quantitative, polygenic or complex traits. While monogenic traits follow patterns of simple Mendelian inheritance, polygenic traits typically vary along a continuous gradient of phenotypes. Quantitative traits complicate the works of breeders because performance only partially reflects the genetic values of the individuals (Semagn et al. 2010).

A *quantitative trait locus (QTL)* is a region of the genome that is associated with variation in a quantitative trait of interest. Conceptually, a QTL can be a single gene or a cluster of linked genes that affect the trait directly (a gene coding for a structural enzyme) or indirectly (a gene coding for regulators of the expression of other genes). It may also be non-coding regions that affect gene expression (Vinod 2006).

A *QTL analysis* is a statistical method that links two types of information – phenotypic data (trait measurements) and genotypic data (usually molecular markers) – in an attempt to explain the genetic basis of variation in complex traits (Miles and Wayne 2008). In particular, performing a QTL analysis is relevant in three areas. Firstly, it provides an efficient tool to dissect the genetic architecture of complex traits by investigating the following questions (Vinod 2006):

(1) How many regions of the genome affect the trait of interest? Where these regions are located?

(2) What is the effect of the QTL?

- How much variation of the trait is caused by a specific region?
- What is the gene action associated with the QTL: additive, dominance effect?
- Which allele is associated with the favorable effect?

Secondly, the QTL methodology can be associated with a functional analysis by the identification of candidate genes, which connect QTLs to proteins and regulatory elements of known functions (Asins 2002; Cooper et al. 2009). Thirdly, the markers that have been identified as associated with major genes or QTL can be employed to practical plant breeding purposes (Vinod 2006).

Basically, conducting a QTL analysis requires, firstly, two strains of individuals that differ genetically with regard to the trait of interest and, secondly, genetic markers that distinguish between these parental lines (Miles and Wayne 2008). The localization of the quantitative trait loci is carried out by using a *genetic map*, which shows the ordering of markers along a chromosome and the relative distances between them (Lynch and Walsh 1998).

So far, two non-saturated molecular maps have been published in hemp (Mandolino and Ranalli 2002; Peil et al. 2001 cited by Mandolino and Carboni 2004), while no QTL analysis has been reported in the species under study. However, no segregating population and associated linkage map are currently available in hemp. Therefore, the investigation of the genetic determinism of the sex expression by QTL analysis requires the creation of a segregating population and construction of a linkage map. Both of these steps were carried out in the present work. F_1 segregating populations were created, and AFLP markers were generated in order to construct linkage maps. The use of F_1 segregating populations relied on the relatively high degree of heterozygosity pointed out in the species (Forapani et al. 2001), while the AFLP technique was selected because of its advantages in terms of ease of implementation and ability to amplify many loci (Vos et al. 1995; Mueller and Wolfenbarger 1999; Meudt and Clarke 2007) and the availability of AFLP primer pairs providing sex-linked markers in hemp (Flachowsky et al. 2001). The present chapter aims to present the basic principles of linkage and QTL mapping starting from the generation of molecular genetic markers. The specificities of linkage and QTL mapping in F_1

segregating populations by using dominant markers are developed because such populations and markers were used in the present study.

3.1. DETECTING THE GENETIC VARIABILITY

3.1.1. Molecular markers

A molecular genetic marker corresponds to a DNA sequence derived from a specific location in the genome. The molecular genetic markers provide therefore a direct access to the genetic variation present at a given locus. Generally, they do not represent the target genes themselves but act as 'signs' or 'flags' that allow differentiating individuals from each other (Collard et al. 2005). Since they are inherited, their transmission from one generation to the next one can be followed. Besides, the assay of a genetic marker is not affected by environmental factors. The ideal genetic marker will be highly variable, or polymorphic, so that the polymorphism revealed by the marker can be used to study the genetic variation at a specific locus (Lynch and Walsh 1998; Wu et al. 2007).

A wide variety of techniques exists to detect DNA variation. A set of techniques commonly used in genetic mapping relies on the hybridisation of denatured single-strand DNA fragments by a DNA probe or primer. These techniques generate molecular markers that differ in number, type of polymorphism (*e.g.*, fragment length, number of simple repeated sequences) and ability to differentiate the homozygous state (*AA*) from the heterozygous one (*Aa*), in particular. This latter characteristic results in the distinction of *codominant* from *dominant* markers and is of primary importance in genetic mapping since the ability to distinguish the heterozygous and homozygous states positively affects the probability of detecting linkage between pairs of markers (Maliepaard et al. 1997).

The codominant markers are generated by DNA primers that score a single-marker locus. The distinct DNA fragments, or markers, amplified by such primer consist of distinct alleles present at the probed marker locus. Thus the codominant markers are *homologous, i.e.*, they derive from the same location in the genome, and the heterozygous individuals can be distinguished from the

homozygous ones. Examples of codominant markers are the RFLP and microsatellite markers.

Unlike the codominant markers, the dominant markers provide a multi-locus fingerprinting (Meudt and Clarke 2007). Each of the amplified markers accounts for an allele derived from a distinct locus. At each locus, the individuals differ from each other by the presence or absence of amplification, and the heterozygous individuals (*Aa*) can thus not be distinguished from the homozygous dominant ones (*AA*). Examples of dominants markers are the RAPD and AFLP markers (Lynch and Walsh 1998). The AFLP technique was used in the present study and is further described below.

3.1.2. The AFLP technique

AFLP markers (Vos et al. 1995) are generated by complete restriction endonuclease digestion of total genomic DNA, followed by selective PCR amplification and electrophoresis of a subset of the fragments, resulting in a unique, reproducible fingerprint (or profile) for each individual. The generation of AFLP markers is performed in four steps (Fig. 3.1).

Step I – Restriction. Genomic DNA is digested with a pair of restriction endonucleases, *e.g.*, *HindIII* and *MseI*, used as rare and frequent cutters, respectively Each of them cut the DNA at a specific sequence (*restriction site*) into restriction fragments.

Step II – Ligation. *Adapters* are ligated to each end of the restriction fragments. The adapters are double-stranded DNA sequences consisting of a core sequence and an enzyme-specific sequence that allows the ligation to the restriction fragments. The sequence of the adapter and restriction sites will be used as target sites for primer annealing during the PCR amplification of the restriction fragments (steps III and IV).

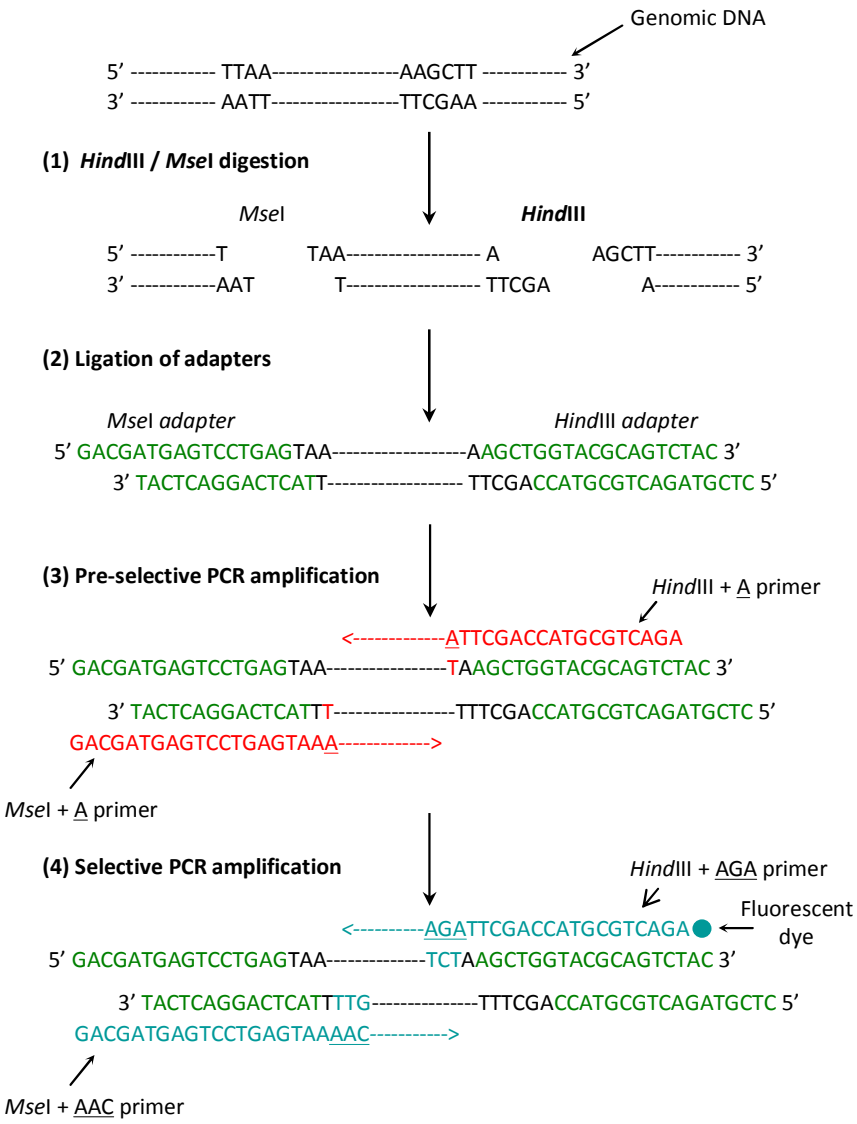

Fig. 3.1 Schematic representation of the AFLP technique with *Hind*III and *Mse*I restriction enzymes. Adapted from Vos et al. (1995) and Meudt and Clarke (2007).

Step III – Pre-selective amplification. A subset of all restricted-ligated fragments is amplified using primers that are complementary to the adapter with an additional nucleotide (*e.g.*, 'A') at the 3' end of the primer. These primers will only prime the DNA synthesis of fragments flanked by a nucleotide that is complementary to the selective nucleotide of each primer. The amount of amplified fragments is thus ~1/16 of the initial amount.

Step IV – Selective amplification. A second round of PCR is performed, in which the primers have two additional flanking nucleotides. The number of amplified fragments is thus further reduced, being now ~$1/16^3$ of the initial amount. The *HindIII*+3 nucleotides primer is labelled with a fluorescent dye so that all strands synthesized from this primer are fluorescently labelled. A range of primers combinations with different selective nucleotides is used in order to amplify alternative subsets of loci.

In the present study, the amplified AFLP fragments were revealed by capillary electrophoresis of fluorescently labelled AFLP fragments. This technique was preferred to gel-based systems because of its high throughput and the high data quality it provides. The capillary instrument detects fragments present in the spectrum of the fluorophore, producing an electronic profile of relative fluorescence units *vs.* fragment size (Fig. 3.2). Polymorphisms, which are observed as peaks present in some samples and absent in others, are caused by the gain or loss of a restriction site, a change in the selective primer binding site or a length difference between restriction sites (Meudt and Clarke 2007). Each AFLP fragment is characterized by its size and primers used for amplification. The number of fragments revealed by primer combination varies generally between 50 and 100 (Vos et al. 1995). The profile of each individual is then described according to the presence (1) or absence (0) of each peak, producing a binary data matrix (individuals x AFLP markers).

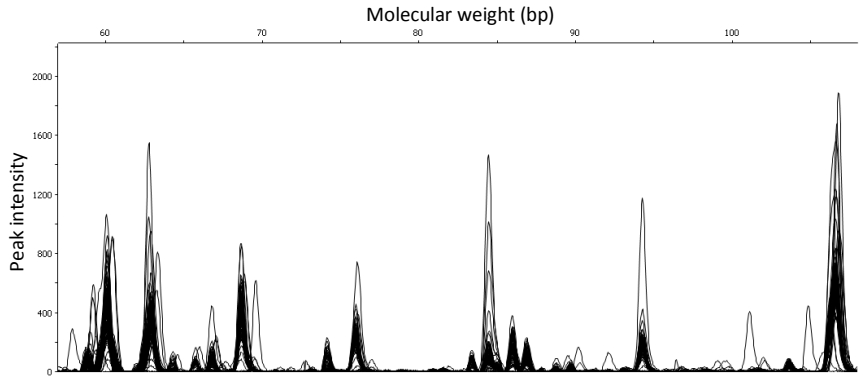

Fig. 3.2 Fluorescently labelled AFLP fragments revealed by capillary electrophoresis. The *x*-axis indicates the molecular weight of the fragments in base pairs, and the *y*-axis, their intensity. Each line corresponds to a single individual.

3.1.3. Advantages and limitations of the AFLP technique

The AFLP technique is highly reliable (Vos et al. 1995; Mueller and Wolfenbarger 1999; Meudt and Clarke 2007). This reliability relies on the combination of the reproducibility of restriction fragment analysis and the amplification of DNA fragments under stringent PCR conditions, *i.e.*, under high annealing temperatures. In addition, similarly to other techniques of multilocus fingerprinting, the amplification of AFLP products can be performed on organisms for which there is no *a priori* sequence information available (Meudt and Clarke 2007). However, the use of a multilocus fingerprinting method raises the question of the homology of the amplification products: does each same-sized DNA fragment originate from one single location in the genome? According to Vos et al. (1995), most AFLP fragments correspond to unique positions on the genome and can therefore been exploited for the construction of linkage maps. Rouppe van der Voort et al. (1997) compared the sequence of co-migrating bands and concluded that the homology of AFLP products is nearly always valid. However, the locus-specificity of AFLP markers

is limited to populations within species or to very closely related species (Qi et al. 1998).

The main disadvantage of the AFLP technique is the dominant nature of the generated markers, which makes the method less useful for studies that require precise assignment of allelic states such as heterozygosity analysis (Mueller and Wolfenbarger 1999). The dominant marker systems generate numerous, genome-wide loci that are individually less informative than codominant marker sytems. However, they derive their statistical power from their high number. Both systems – codominant and dominant markers – are commonly used for linkage analysis (Meudt and Clarke 2007).

In conclusion, the AFLP technique can be ideal when there is no *a priori* sequence information and no suitable established markers, for intra-specific studies, when it is necessary to amplify many loci to ascertain an accurate measure of the genetic diversity – which is the case in outcrossing species such as hemp –, for the rapid generation of data and when the revelation of amplified fragments by capillary electrophoresis is possible (Meudt and Clarke 2007). The AFLP technique has found applications in a wide range of genetic studies, including linkage mapping, population genetics or phylogenetic studies. Genetic mapping with AFLP markers has been performed in potato (Rouppe van der Voort et al. 1997), barley (Qi and Lindhout 1997; Waugh et al. 1997; Qi et al. 1998), rice (Nandi et al. 1997), maize (Vuylsteke et al. 1999) and sugarcane (Hoarau et al. 2001), among others.

3.2. CONSTRUCTION OF A LINKAGE MAP

3.2.1. Principles of linkage analysis and map construction

Linkage is the tendency for genes to be inherited together because they are located close to each other on the same chromosome (Wu et al. 2007). The construction of a linkage map relies on the principle of linkage and recombination between molecular markers (Fig. 3.3): markers that are located close to each other on a same chromosome (*e.g.*, M_2 and M_3) are more likely to be transmitted together to the next generation than markers that are located further apart (*e.g.*, M_1 and M_2).

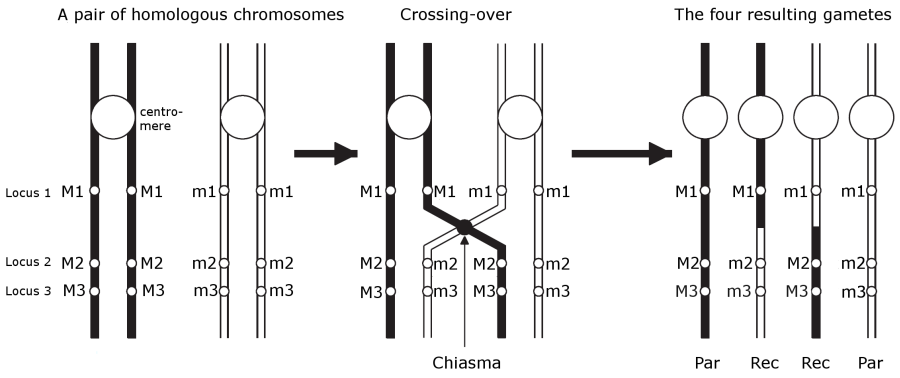

Fig. 3.3 Diagram for crossing-over between two paired homologous chromosomes at an early stage (prophase I) during meiosis. M_1-m_1, M_2-m_2 and M_3-m_3 account for alleles present at three successive loci, respectively. *Par* and *Rec* indicate the parental and recombined gametes, respectively. Adapted from Wu et al. (2007).

In Fig. 3.3, the existence of linkage between the three loci means that there will be more gametes of the parental type $M_1M_2M_3$ and $m_1m_2m_3$ than recombined gametes like $M_1m_2m_3$ and $m_1M_2M_3$. These two latter gametes are

called *recombinants* for loci 1 and 2 (as well as for loci 1 and 3). The proportion of recombinant gametes between two loci, or *recombination fraction*, is denoted *r*. Thus the proportion of parental, or non recombinant gametes, is *1-r*. From the process of meiosis, it follows that a recombination fraction of 0.5 between two markers indicates that they segregate independently, *i.e.*, are located either on distinct chromosomes, either on the same chromosome but relatively far away from each other (Wu et al. 2007).

The linkage analysis consists of estimating the actual recombination fraction, *r*, on the basis of the expected number of recombinants in a segregating progeny. Let $\hat{r}$ denote the estimated recombination fraction between them. The test of linkage between the two loci consists of testing the null hypothesis H_0: $r = 0.5$ (markers are unlinked) against the alternative hypothesis H_1: $r < 0.5$ (markers are linked). The test statistic used to this purpose is the LOD (logarithm of the odds) score, where *L* is the likelihood function:

$$LOD = \log_{10} \frac{L(r = \hat{r})}{L(r = 0.5)} \tag{3.1}$$

Two markers are considered to be linked and thus assigned to a same co-segregation group if the LOD score is higher than a threshold value. A threshold value of 3.0 is commonly used (Morton 1955; Risch 1992 cited by Maliepaard et al. 1997).

Markers assigned to a same co-segregation group are then ordered. The methods for linearly ordering the molecular markers rely on minimizing the recombination between pairs of markers. However, the recombination fractions are not additive because they are probabilities and because of genetic interference. Therefore, they are translated into genetic distance by using a map function (Doerge 2002). Two map functions are commonly used: the Haldane function, which assumes that recombinations occur independently from each other, and the Kosambi function, which integrates the potential effect of interferences between loci (Wu et al. 2007; Doerge and Zeng 2011). The unit for expressing the genetic distance between markers on a chromosome is the Morgan (or, more usually, the centiMorgan, cM), and is defined as the distance along which one recombination event is expected to occur per gamete per generation (Doerge 2002). Finally, the *genetic map* provides a representation of the chromosomes on which the markers and QTL reside (Doerge 2002).

Considering the above, the development of a linkage analysis critically relies on the estimation of the recombination fraction between markers. The accuracy of the recombination fraction estimator depends on the number of recombination events that occurred in the parents and the ability with which these events can be detected. The first component is determined by the recombination fraction itself and the size of the progeny, and the latter component, by the configuration of the markers along the parental chromosomes (Maliepaard et al. 1997). Thus the type of segregating population and marker system used for mapping will strongly affect the power of the linkage analysis in addition to the size of the population. This is developed in the next two sections.

3.2.2. Classical *vs.* F_1 segregating populations

Consider two diploid parental lines denoted as *P* and *Q*, each containing two homologous chromosomes. The *linkage phase* designs the allelic configuration at a pair of loci on these homologous chromosomes in a single parent (Maliepaard et al. 1997).

Classical experimental designs used for linkage and QTL mapping, such as F_2 or backcross populations, derive from an initial cross between two complementary inbred lines (Fig. 3.4a). In such populations, only two alleles segregate at each locus, and the parental linkage phases are known since the parents are homozygous. The distinction of recombinants and non-recombinants in the progeny and the linkage analysis are thus straightforward (Wu et al. 2002a).

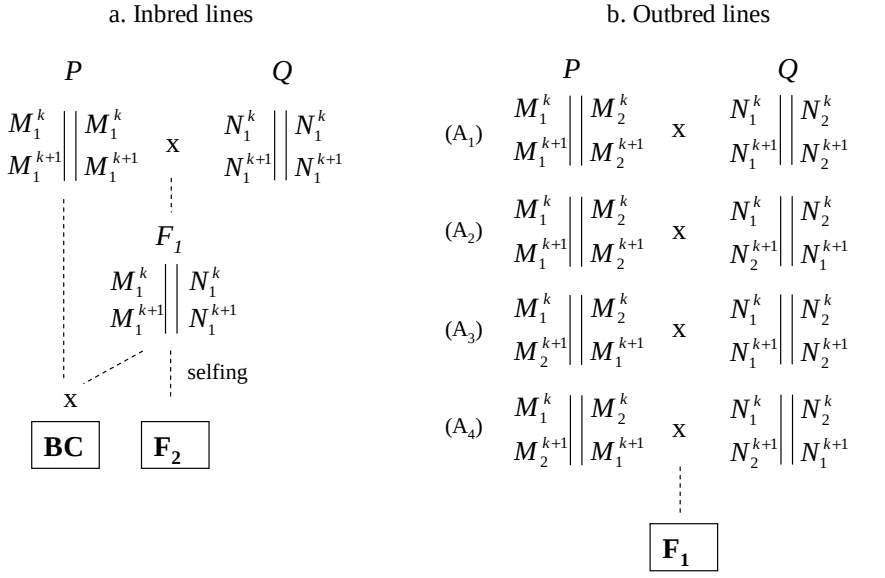

Fig. 3.4 Cross between (a) two diploid inbred lines and (b) two outbred lines, *P* and *Q*, characterized by two marker loci, *k* and *k+1*. M_1^k, M_2^k, N_1^k and N_2^k are alleles found at locus *k*, and M_1^{k+1}, M_2^{k+1}, N_1^{k+1} and N_2^{k+1}, alleles found at locus *k+1*. The segregating populations are indicated in boxes (BC = backcross). In a cross between two diploid outbred lines, four assignments can be distinguished for the two parents depending on the allelic configuration at the two marker loci. For each line, the two vertical bars stand for two homologous chromosomes. The notation is based on Lin et al. (2003).

However, a large group of plant species is recalcitrant to genomic research because of their high genetic heterozygosity, resulting from outcrossing and long generation times (Wu et al. 2002a). In these species, inbred lines are generally not available. Linkage analyses and QTL mapping can be carried out directly from full-sib family derived from a cross between two outbred parents, here referred to as F_1 segregating populations. The efficiency of linkage mapping in such populations is proportional to the level of genetic heterozygosity in the species under study and the genetic divergence between the individuals crossed (Grattapaglia and Sederoff 1994). However, the use of an F_1 segregating population makes the construction of linkage map much more challenging than the use of segregating populations derived from inbred lines. Indeed, in an F_1 segregating population, both the number of alleles segregating at each locus and parental linkage phases are unknown, and the number of segregating alleles can vary from one locus to another (Maliepaard et al. 1997; Wu et al. 2002a). Up to four alleles can segregate at each locus. For each pair of loci in each parent, two linkage phases can be distinguished, resulting in a total of four distinct assignments of alleles (Fig. 3.4b). In parent *P*, the alleles M_1^k and M_1^{k+1} are linked in *coupling* phase in the assignments 1 and 2 and in *repulsion* phase in the assignments 3 and 4.

3.2.3. Segregation of AFLP markers

Classification of all possible cross types between two diploid outbred parents for a single marker locus

As stated in the previous section, up to four alleles may segregate for each marker in a population derived from two diploid outbred parents. Let *a*, *b*, *c* and *d* denote four distinct alleles, and *o* be the null allele, *i.e.*, not-*a*, not-*b*, not-*c* and/or not-*d*. The alleles *a*, *b*, *c* and *d* are codominant with respect to each other, but dominant with respect to the null allele (*e.g.*, the genotypes *aa* and *ao* can not be distinguished from each other). In total, 18 distinct cross types can be distinguished for a single marker locus depending on the allelic configuration in each parent (Maliepaard et al. 1997; Wu et al. 2002a). A classification of these cross types was proposed by Wu et al. (2002a) (Table 3.1).

Table 3.1 Classification of the 18 cross types that may be disinguished for a single marker locus in an F_1 segregating population derived from two diploid lines. The alelles *a*, *b*, *c* and *d* are codominant with respect to each other and dominant with respect to the null allele, *o*. Adapted from Wu et al. (2002a)

Cross type [1]			Parents		Offspring		
			Genotypes	Pheno-types [2]	Distinct phenotypes [2,3]	Segregating pattern	Nb. of pheno-types
A		1	*ab x cd*	*ab x cd*	*ac, ad, bc, bd*	1:1:1:1	4
		2	*ab x ac*	*ab x ac*	*a, ac, ba, bc*	1:1:1:1	4
		3	*ab x co*	*ab x cd*	*ac, a, bc, b*	1:1:1:1	4
		4	*ao x bo*	*a x b*	*ab, a, b, o*	1:1:1:1	4
B	B1	4	*ab x ao*	*ab x a*	*ab, a, b*	1:2:1	3
	B2	6	*ao x ab*	*a x ab*	*ab, a, b*	1:2:1	3
	B3	7	*ab x ab*	*ab x ab*	*a, ab, b*	1:2:1	3
C		8	*ao x ao*	*a x a*	*a, o*	3:1	2
D	D1	9	*ab x cc*	*ab x c*	*ac, bc*	1:1	2
		10	*ab x aa*	*ab x a*	*a, ab*	1:1	2
		11	*ab x oo*	*ab x o*	*a, b*	1:1	2
		12	*bo x aa*	*b x a*	*ab, a*	1:1	2
		13	*ao x oo*	*a x o*	*a, o*	1:1	2
	D2	14	*cc x ab*	*c x ab*	*ac, bc*	1:1	2
		15	*aa x ab*	*a x ab*	*a, ab*	1:1	2
		16	*oo x ab*	*o x ab*	*a, b*	1:1	2
		17	*bo x aa*	*a x b*	*ab, a*	1:1	2
		18	*oo x ao*	*o x a*	*a, o*	1:1	2

[1] The reciprocal cross types (B1 *vs* B2, D1 vs D2) are presented in two distinct groups because they supply different information for the characterization of linkage phase when paired (Wu et al. 2002a).

[2] The phenotype accounts for the observed pattern of bands.

[3] For each cross type, the distinct phenotypes that can be identified in the offspring are given.

Cross type A: Markers heterozygous in both parents and segregating in a 1:1:1:1 ratio.

Cross type B: Markers heterozygous in both parents and segregating in a 1:2:1 ratio. Three groups are distinguished according to the presence of a null allele in one (B1), the other (B2, which is the reciprocal of B1) or none of both parents (B3).

Cross type C: Markers heterozygous in both parents and segregating in a 3:1 ratio.

Cross type D: Markers heterozygous in one parent and homozygous in the other one (testcross configuration) and segregating in a 1:1 ratio. Two reciprocal groups (D1 and D2) are distinguished depending on which of both parents is heterozygous.

In addition, the cross types A, B, C and D show different degrees of informativeness with respect to the differentiation of the genotypes in the offspring. For a given marker, regardless of its cross type, a cross between two diploid parents results in four possible pairings of parental chromosomes. The markers of cross type A allow the distinction of four phenotypes, each of them corresponding to one given pairing of parental chromosomes, and thus are fully informative (Table 3.1). In contrast, the markers of cross types B, C and D allow the distinction of 2 or 3 phenotypes and thus are only partially informative.

Cross types revealed by AFLP markers

The AFLP technique reveals a polymorphism of presence/absence of amplification – which is specific to dominant marker systems. According to Wu et al. (2002a), each AFLP marker may receive one of three distinct cross types: C.8 (*ao* x *ao*), D1.13 (*ao* x *oo*) or D2.18 (*oo* x *ao*), the alleles *a* and *o* corresponding to the presence and absence of amplification, respectively, and their location being interchangeable between both homologous chromosomes in each parent (Table 3.1). The C.8 markers are heterozygous in both parents and segregate 3:1 in the offspring, while the D1.13 and D2.18 markers are heterozygous in only one parent, reciprocally, and segregate 1:1 in the offpring. Because the C and D cross types allow the distinction of only two phenotypes in

the offspring (Table 3.1), the AFLP markers only partially characterize the polymorphism found between individuals of an F_1 segregating population.

3.2.4. Linkage mapping in an F_1 segregating population with AFLP markers

As stated in the section 3.2.2, when mapping in an F_1 segregation population, not only the recombination fractions between pairs of markers but also the linkage phases in the parents are unknown. Therefore, distinct statistical strategies have been developed for estimating the parental linkage phases and recombination fraction in F_1 segregating populations.

The traditional construction of genetic maps with F_1 segregating populations is performed by using dominant markers with 1:1 segregating pattern (D1.13 and D2.18 cross types; Table 3.1). This strategy results in two independent linkage maps, each map having markers segregating in one parent but fixed in the other, as performed by the software MapMaker (Lander et al. 1987; Grattapaglia and Sederoff 1994). It has been applied to construct linkage maps in hemp by Mandolino et al. (2002) (see section 3.2.5). The integration of both parental maps in an unique one requires the use of markers segregating in both parents, with a 3:1 segregating pattern (C.8 cross type). This integration was performed by the first version of the software JoinMap using weighted least squares for the estimation of map distances (Stam 1993).

Maliepaard et al. (1997) presented the equations for estimating the recombination fraction between pairs of markers displaying distinct allelic configurations and the respective LOD score by using a maximum likelihood approach rather than least-square approximation. Starting from this work, Wu et al. (2002a) developed a methodology based on a maximum-likelihood approach to construct genetic maps by simultaneously estimating the parental linkage phases and recombination fraction between markers in a set of markers with various cross types. Their methodology has been implemented in the R package OneMap (Margarido et al. 2007), which was used to construct genetic maps in the present study. The anchoring algorithm of the methodology of Wu et al. (2002a) for estimating the recombination fraction and parental linkage phases

between two markers – or two-point analysis – in an F_1 segregating population is presented here.

Two-point analysis in an F_1 segregating population according to Wu et al. (2002a)

Consider two diploid outbred parents denoted as *P* and *Q*, each containing two homologous chromosomes. For a given marker locus *k*, M_1^k and M_2^k denote the two alleles found in parent *P*, and N_1^k and N_2^k, those found in parent *Q* (Fig. 3.4b). The cross of *P* and *Q* results in four possible genotypes in the offspring at locus *k*: $M_1^k N_1^k, M_1^k N_2^k, M_2^k N_1^k$ and $M_2^k N_2^k$. Similarly, at locus *k+1*, M_1^{k+1}, M_2^{k+1}, N_1^{k+1} and N_2^{k+1} are alleles found in parents *P* and *Q*, respectively, and $M_1^{k+1}N_1^{k+1}, M_1^{k+1}N_2^{k+1}, M_2^{k+1}N_1^{k+1}$ and $M_2^{k+1}N_2^{k+1}$ are the four possible genotypes. Four assignments are distinguished depending on the allelic configuration, or linkage phase, in each parent (Fig. 3.4b).

Let *r* – assumed to be the same in both parents – be the recombination fraction between the marker loci *k* and *k+1*, A_ω be the assignment of the alleles at the second marker locus, *k+1*, given a fixed allelic configuration at locus *k*, and $\mathbf{M} = (m^k, m^{k+1})^T$ be the data for the two marker loci.

The conditional probability of having one of the four possible genotypes in a given individual of the offspring at locus *k+1* given the genotype at locus *k* is referred to as the transition probability of recombination between loci *k* and *k+1*. The transition probabilities under a given assignment of alleles at loci *k* and *k+1* (Fig. 3.4b) can be expressed in a (4 x 4) matrix (Table 3.2).

Table 3.2 Transition probabilities of recombination between two marker loci, *k* and *k+1*, under four assignments of alleles in an F_1 segregating population derived from two diploid outbred lines, *P* and *Q* (Fig. 3.4b). The alleles M_1 and M_2 derive from parent *P*, and N_1 and N_2, from parent *Q*. *r* is the recombination fraction between the two marker loci. From Wu et al. (2002a)

		$M_1^{k+1}N_1^{k+1}$	$M_1^{k+1}N_2^{k+1}$	$M_2^{k+1}N_1^{k+1}$	$M_2^{k+1}N_2^{k+1}$
A_1	$M_1^kN_1^k$	$(1-r)^2$	$r(1-r)$	$r(1-r)$	r^2
	$M_1^kN_2^k$	$r(1-r)$	$(1-r)^2$	r^2	$r(1-r)$
	$M_2^kN_1^k$	$r(1-r)$	r^2	$(1-r)^2$	$r(1-r)$
	$M_2^kN_2^k$	r^2	$r(1-r)$	$r(1-r)$	$(1-r)^2$
A_2	$M_1^kN_1^k$	$r(1-r)$	$(1-r)^2$	r^2	$r(1-r)$
	$M_1^kN_2^k$	$(1-r)^2$	$r(1-r)$	$r(1-r)$	r^2
	$M_2^kN_1^k$	r^2	$r(1-r)$	$r(1-r)$	$(1-r)^2$
	$M_2^kN_2^k$	$r(1-r)$	r^2	$(1-r)^2$	$r(1-r)$
A_3	$M_1^kN_1^k$	$r(1-r)$	r^2	$(1-r)^2$	$r(1-r)$
	$M_1^kN_2^k$	r^2	$r(1-r)$	$r(1-r)$	$(1-r)^2$
	$M_2^kN_1^k$	$(1-r)^2$	$r(1-r)$	$r(1-r)$	r^2
	$M_2^kN_2^k$	$r(1-r)$	$(1-r)^2$	r^2	$r(1-r)$
A_4	$M_1^kN_1^k$	r^2	$r(1-r)$	$r(1-r)$	$(1-r)^2$
	$M_1^kN_2^k$	$r(1-r)$	r^2	$(1-r)^2$	$r(1-r)$
	$M_2^kN_1^k$	$r(1-r)$	$(1-r)^2$	r^2	$r(1-r)$
	$M_2^kN_2^k$	$(1-r)^2$	$r(1-r)$	$r(1-r)$	r^2

According to Bayes's theorem, the posterior probability of A_ω given **M** is expressed as:

$$P(A_\omega|\mathbf{M}) = \frac{P(A_\omega)P(\mathbf{M}|A_\omega)}{\sum_{\omega=1}^{4} P(A_\omega)P(\mathbf{M}|A_\omega)} = \frac{P(\mathbf{M}|A_\omega)}{\sum_{\omega=1}^{4} P(\mathbf{M}|A_\omega)} \tag{3.2}$$

where $P(A_\omega)$ is the unconditional or prior probability of A_ω and is assumed to be uniform, and $P(\mathbf{M}|A_\omega)$ is the likelihood of **M** given A_ω. Assuming that the marker data are independent among *N* individuals in the population, we have:

$$\begin{aligned} P(\mathbf{M}|A_\omega) &= \prod_{i=1}^{N} P(m_i|A_\omega) \\ &= \prod_{i=1}^{N} \sum_{j_1=1}^{p_1} \sum_{j_2=1}^{p_2} P(m_i^k|(MN)_{j_1}) P((MN)_{j_2}|(MN)_{j_1}) P(m_i^{k+1}|(MN)_{j_2}) \end{aligned} \tag{3.3}$$

where p_1 and p_2 are the numbers of distinct phenotypes observed in the offspring at loci *k* and *k+1*, respectively, $P(m_i^k|(MN)_{j_1})$ and $P(m_i^{k+1}|(MN)_{j_2})$ are the indicator variables describing the j_1th phenotype at locus *k* and j_2th phenotype at locus *k+1*, respectively, for individual *i*, and $P((MN)_{j_2}|(MN)_{j_1})$ is the transition probability of having the j_2th phenotype at locus *k+1* given the j_1th phenotype at locus *k*. The indicator variables take a value of 1 or 0: *e.g.*, at locus *k*, the indicator variable is one if the phenotype of individual *i* (m_i^k) is consistent with the $(MN)_{j_1}$ phenotype, zero otherwise. In the present study, the use of AFLP markers, which are characterized by cross types C and D (Table 3.1), results in two possible distinct phenotypes at each marker locus, and thus $p_1 = p_2 = 2$.

For a given assignment at locus *k+1*, the recombination fraction between loci *k* and *k+1* is estimated by maximizing the log-likelihood of equation (3.3) for *r* and solving the following equation:

$$\frac{\delta \ln(P(\mathbf{M}|A_\omega))}{\delta r} = 0 \tag{3.4}$$

The computation of the maximum likelihood estimator of *r* is performed by an iterative procedure based on expectation (E) and maximization (M) algorithms (Dempster 1977). At each $\tau+1$ step, the E step calculates the expected number of recombination events between M^k and M^{k+1} for individual *i* under assignment A_ω based on the r^τ estimate, while the M step calculates $r^{\tau+1}$. These iterative procedures are repeated until *r* converges to a stable value. This value represents the maximum likelihood estimator of the recombination fraction between the marker loci *k* and *k+1* under assignment A_ω. The most likely parental linkage phases between M^k and M^{k+1} are inferred by calculating the posterior probabilities of the four distinct assignments A_ω based on equation (3.2).

Two-point analysis in an F_1 segregating population with AFLP markers

As a result of the presence of three distinct cross types (C.8, D1.13 and D2.18), the linkage of two AFLP markers involves six possible combinations of cross types: C.8-C.8, C.8-D1.13, C.8-D2.18, D1.13-D1.13, D2.18-D2.18 and D1.13-D2.18 (Table 3.3).

However, the use of C.8, D1.13 and D2.18 cross types for mapping in F_1 segregating populations implies some uncertainty in the estimation of the parental linkage phases and recombination fraction between markers (Maliepaard et al. 1997; Wu et al. 2002a).

None of the present six cross-types combinations allows the distinction of the four possible assignments of alleles (Fig. 3.4b; Table 3.3). For example, assignments A_2 and A_3 between two C.8 markers provide the same posterior probabilities (Wu et al. 2002a). Similarly, assignments A_1 and A_2, on the one hand, and A_3 and A_4, on the other hand, are not distinguishable for the linkage analysis involving D1.13 markers. The same holds true for assignments A_1 and A_3, on the one hand, and A_2 and A_4, on the other hand, with D2.18 markers.

The power, or probability, of detecting linkage between two dominant markers has been characterized as a function of the cross type and population size by Maliepaard et al. (1997) (Fig. 3.5).

The least informative combination is found between D1.13 and D2.18 markers, for which no recombinant can be identified (Table 3.3). Therefore, the linkage between D1.13 and D2.18 markers is established through C.8 markers

(Maliepaard et al. 1997; Wu et al. 2002a). Thereafter, the power of detecting linkage is the lowest with C.8-C.8 markers in coupling phase in parent *P* and repulsion phase in parent *Q* (C x R), or conversely in R x C linkage phases, then with C.8-C.8 markers in R x R linkage phases, C.8-D1.13 (or C.8-D2.18) markers and C.8-C.8 markers in coupling phase in each parent, and the highest with D1.13-D1.13 (or D2.18- D2.18) markers, for which eight recombinants can be detected in the offspring (Fig. 3.5; Table 3.3).

In addition, the power of detecting linkage is positively correlated with the population size: the maximum detectable recombination fraction is approximately 0.2, 0.25 and 0.35 with population sizes of 50, 100 and 200, respectively (Maliepaard et al. 1997; Fig. 3.5).

Table 3.3 Linkage analysis between two marker loci, *k* and *k+1*, characterized by a dominant marker system in an F_1 segregating population *(this table is split into two parts; the second one is on the next page)*

Marker locus	Cross type [1]	Parents [2]					Offspring [2,5]																
		Genotypes		LP [3]		Assign-ment [4]	No recombination					Recombination in parent *Q*					Recombination in parent *P*					Rececombination in both parents				
		P	*Q*	*P*	*Q*																				
k	C.8	a ‖ o	a ‖ o	C	C	A_1	a ‖ a	a ‖ o	o ‖ a	o ‖ o		a ‖ a	a ‖ o	o ‖ a	o ‖ o		a ‖ a	a ‖ o	o ‖ a	o ‖ o		a ‖ a	a ‖ o	o ‖ a	o ‖ o	
k+1	C.8	a ‖ o	a ‖ o				a ‖ a	a ‖ o	o ‖ a	o ‖ o		a ‖ o	a ‖ a	o ‖ o	o ‖ a		o ‖ a	o ‖ o	a ‖ a	a ‖ o		o ‖ o	o ‖ a	a ‖ o	a ‖ a	
k	C.8	a ‖ o	a ‖ o	C	R	A_2	a ‖ a	a ‖ o	o ‖ a	o ‖ o		a ‖ a	a ‖ o	o ‖ a	o ‖ o		a ‖ a	a ‖ o	o ‖ a	o ‖ o		a ‖ a	a ‖ o	o ‖ a	o ‖ o	
k+1	C.8	a ‖ o	o ‖ a				a ‖ o	a ‖ a	o ‖ o	o ‖ a		a ‖ a	a ‖ o	o ‖ a	o ‖ o		o ‖ o	o ‖ a	a ‖ o	a ‖ a		o ‖ a	o ‖ o	a ‖ a	a ‖ o	
k	C.8	a ‖ o	a ‖ o	R	C	A_3	a ‖ a	a ‖ o	o ‖ a	o ‖ o		a ‖ a	a ‖ o	o ‖ a	o ‖ o		a ‖ a	a ‖ o	o ‖ a	o ‖ o		a ‖ a	a ‖ o	o ‖ a	o ‖ o	
k+1	C.8	o ‖ a	a ‖ o				o ‖ a	o ‖ o	a ‖ a	a ‖ o		o ‖ o	o ‖ a	a ‖ o	a ‖ a		a ‖ a	a ‖ o	o ‖ a	o ‖ o		a ‖ o	a ‖ a	o ‖ o	o ‖ a	
k	C.8	a ‖ o	a ‖ o	R	R	A_4	a ‖ a	a ‖ o	o ‖ a	o ‖ o		a ‖ a	a ‖ o	o ‖ a	o ‖ o		a ‖ a	a ‖ o	o ‖ a	o ‖ o		a ‖ a	a ‖ o	o ‖ a	o ‖ o	
k+1	C.8	o ‖ a	o ‖ a				o ‖ o	o ‖ a	a ‖ o	a ‖ a		o ‖ a	o ‖ o	a ‖ a	a ‖ o		a ‖ o	a ‖ a	o ‖ o	o ‖ a		a ‖ a	a ‖ o	o ‖ a	o ‖ o	

[1] Cross types at two marker loci, *k* and *k+1*, respectively. Notation according to Wu et al. (2002a) (Table 3.1).

[2] By analogy with Table 3.1, the alleles at loci *k* and *k+1* were symbolised by the same letters, *i.e.*, *a* and *o* for presence/absence of amplification, respectively.

[2] The alleles at loci *k* and *k+1* were symbolised by the same letters, *i.e.*, *a* and *o* for presence/absence of amplification, respectively.

[4] No. of assignment depending on the allelic arrangement in the parents. The nos. of assignment correspond to those in Fig. 3.4b.

[5] Genotypes of the offspring. The detectable recombinants are in grey.

Table 3.3 (continued)

Marker locus	Cross type [1]	Parents [2]				Offspring [2,5]			
		Genotypes		LP [3]		No recombination	Recombination in parent *Q*	Recombination in parent *P*	Rececombination in both parents
		P	*Q*	*P*	*Q*	Assign-ment [4]			
k	C.8	a ‖ o	a ‖ o	C	-	A_1 or A_2	a ‖ a a ‖ o o ‖ a o ‖ o	a ‖ a a ‖ o o ‖ a o ‖ o	a ‖ a a ‖ o o ‖ a o ‖ o	a ‖ a a ‖ o o ‖ a o ‖ o
k+1	D1.13	a ‖ o	o ‖ o				a ‖ o a ‖ o o ‖ o o ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o	o ‖ o o ‖ o a ‖ o a ‖ o	o ‖ o o ‖ o a ‖ o a ‖ o
k	C.8	a ‖ o	a ‖ o	-	C	A_3 or A_4	a ‖ a a ‖ o o ‖ a o ‖ o	a ‖ o a ‖ a o ‖ o o ‖ a	a ‖ a a ‖ o o ‖ a o ‖ o	a ‖ a a ‖ o o ‖ a o ‖ o
k+1	D1.13	o ‖ a	o ‖ o				o ‖ o o ‖ o a ‖ o a ‖ o	o ‖ o o ‖ o a ‖ o a ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o
k	D1.13	a ‖ o	o ‖ o	C	-	A_1 or A_2	a ‖ o a ‖ o o ‖ o o ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o
k+1	D1.13	a ‖ o	o ‖ o				a ‖ o a ‖ o o ‖ o o ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o	o ‖ o o ‖ o a ‖ o a ‖ o	o ‖ o o ‖ o a ‖ o a ‖ o
k	D1.13	a ‖ o	o ‖ o	-	C	A_3 or A_4	a ‖ o a ‖ o o ‖ o o ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o
k+1	D1.13	o ‖ a	o ‖ o				o ‖ o o ‖ o a ‖ o a ‖ o	o ‖ o o ‖ o a ‖ o a ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o
k	D1.13	a ‖ o	o ‖ o	-	-	-	a ‖ o a ‖ o o ‖ o o ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o	a ‖ o a ‖ o o ‖ o o ‖ o
k+1	D2.18	o ‖ o	a ‖ o				o ‖ a o ‖ o o ‖ a o ‖ o	o ‖ o o ‖ a o ‖ o o ‖ a	o ‖ a o ‖ o o ‖ a o ‖ o	o ‖ o o ‖ a o ‖ o o ‖ a

[1] Cross types at two marker loci, *k* and *k+1*, respectively. Notation according to Wu et al. (2002a) (Table 3.1). The C.8-D2.18 and D2.18-D2.18 combinations (not shown) are reciprocals of the C.8-D1.13 and D1.13-D1.13 combinations, respectively.

[2] The alleles at loci *k* and *k+1* were symbolised by the same letters, *i.e.*, *a* and *o* for presence/absence of amplification, respectively.

[3] Linkage phases (LP) between the alleles *a* at loci *k* and *k+1* in parents *P* and *Q*, respectively. C and R account for coupling and repulsion, respectively (the alleles *a* at loci *k* and *k+1* are located on the same chromosome or on distinct chromosomes in a given parent, respectively). '-' stands for undefined LP (locus *k* or *k+1* is homozygous).

[4] No. of assignment depending on the allelic arrangement in the parents. The nos. of assignment correspond to those in Fig. 3.4b.

[5] Genotypes of the offspring. The detectable recombinants are in grey.

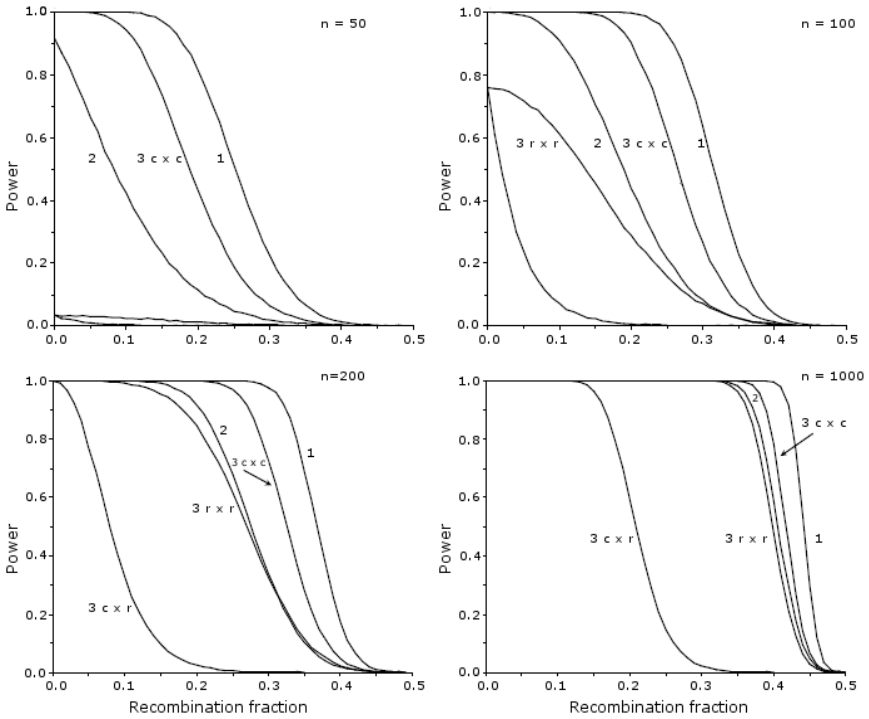

Fig. 3.5 Power of detecting linkage between two markers with D1.13 (or D2.18) and C.8 cross types (Table 3.1). The power accounts for the fraction of 20 000 simulation runs for which a recombination fraction associated with a LOD score > 3.0 was obtained. Four population sizes were tested (n = 50, 100, 200 and 1000). The configuration 1 refers to the cross-types combination D1.13-D1.13, the configuration 2, C.8-D1.13, and the configuration 3, C.8-C.8 in C x C, C x R (or R x C) or R x R parental linkage phases, where C and R account for coupling and repulsion phases, respectively (from Maliepaard et al. 1997). See Table 3.3 for the complete linkage analysis between two markers displaying the present three cross types.

From two-point to multi-point analysis

The two-point analysis can be extended to make simultaneous inferences concerning multi-locus recombination fractions. If the markers provide all genetic information, the two-point estimates of recombination fraction are identical to the multipoint estimates (van Ooijen 2011). However, this is rarely the case, because of the presence of missing data or the use of a dominant marker system. A general algorithm was then derived to calculate the multipoint recombination fractions in presence of missing data in populations derived from two inbred lines (Jiang and Zeng 1997). This method uses Markov chains to link multiple intervals. Wu et al. (2002a) developed a three-point analysis to combine all markers of a linkage group and perform a multi-locus linkage phase inference. The R package OneMap, which was used in the present study, merges the methodology of Wu et al. (2002a) and the computation of multipoint likelihoods through Hidden Markov models (Jiang and Zeng 1997) to construct linkage maps in outcrossing species (Margarido et al. 2007). OneMap was successfully applied for linkage mapping in soybean (Garcia et al. 2006; Pastina et al. 2012) and *Passiflora* (Oliveira et al. 2008).

3.2.5. Linkage maps in hemp

The first linkage map that has been developed in hemp was based on an F_1 progeny of 40 individuals derived from a cross between a monoecious plant and a female plant of a dioecious cultivar (Mandolino and Ranalli 2002). Two maps were created, one for each parent. The maternal map included 66 RAPD markers segregating 1:1 distributed on 11 co-segregation groups, and the paternal one, 43 markers on 9 co-segregation groups. The female and monoecious individuals segregated in a 1:1 ratio in the F_1, suggesting the existence of a monoecious trait inherited by monohybridism. However, no locus responsible for the monoecy was mapped. The second map was obtained from a cross between a male plant and a female one of a dioecious accession, thus including the Y chromosome. This map included 122 AFLP markers distributed along 10 linkage groups (Peil et al. 2001 cited by Mandolino and Carboni 2004).

3.3. IDENTIFICATION OF QUANTITATIVE TRAIT LOCI (QTLS)

3.3.1. Statistical methods

The simplest method of associating markers with quantitative variation is to test the segregation of a phenotype with respect to a marker genotype (Soller and Brody 1976). This **single-marker** (SM) analysis can be performed using a *t*-test, analysis of variance or regression analysis of multiple markers. The progenies are classified according to their marker genotype, and the phenotypic means are compared between classes. A significant difference indicates the presence of a QTL linked to the marker under investigation (Doerge 2002).

The single-marker (SM) analyses are useful to show the existence of a genetic control. In addition, they do not require a genetic map. However, these analyses fail to provide an accurate estimate of the QTL location since the QTL effect and location are unable to be estimated separately. A significant difference in phenotypic means between the marker genotypes can be due to a tight linkage between the marker and QTL with small effect or to a loose linkage and large QTL effect (Doerge 2002; Wu et al. 2007).

While single-marker analyses investigate individual markers independently, without reference to their position, the **interval-mapping** (IM) approach (Lander and Botstein 1989) uses an estimated genetic map as framework for the location of QTLs. The intervals that are defined by a pair of consecutive markers are searched in increments (*e.g.*, 1 or 2 cM). Statistical methods are then used to test whether a QTL is likely to be present at each incremental position across the ordered markers in the genome. The results of the tests are expressed as LOD (logarithm of odds) scores, which compare the evaluation of the likelihood function under the null hypothesis (there is no QTL at the position under study) with the alternative hypothesis (there is a QTL). The LOD score values obtained for each increment are then plotted against the linkage map positions, expressed in cM. The most probable position of the QTL is given by the position of the highest LOD score exceeding a pre-defined critical threshold (Zeng 1994; Doerge 2002).

Compared to the SM analysis, the IM method allows (i) the inference of the probable position of a QTL between marker loci and (ii) the improved estimates of QTL effects (Lander and Botstein 1989; Zeng 1994). However, IM is limited by both the model that defines it as a single QTL method, and by the one-dimensional search that does not allow interactions between multiple QTLs to be considered (Doerge 2002). Indeed, even if there is no QTL at a given position, the LOD score can still exceed the threshold due to the presence of a QTL at some nearby region on the chromosome. If there is only one QTL on a chromosome, this effect does not matter since the most likely QTL location is given at the position with maximum significant likelihood profile. However, the number of QTLs on a chromosome is unknown, and, if there is more than one QTL on a chromosome, the test statistic at the position under study will be affected by those other QTLs (Zeng 1994). Therefore, the IM method can lead to biased estimates when multiple QTLs occur on the same chromosome (Knott and Haleyt 1992; Haley and Knott 1992; Martínez and Curnow 1992; Jansen 1993; Zeng 1994; Jansen and Stam 1994).

Faced with this issue, the **composite-interval-mapping** (CIM) combines multiple linear regression methods with the conventional interval mapping: additional markers – beyond those flanking the interval of interest – are used as covariates in a multiple regression in order to absorb the effect of other QTLs (Jansen 1993; Zeng 1994). By conditioning the test on linked markers, the sensitivity of the test statistic to the position of individual QTLs is increased, resulting in an improved precision of QTL mapping. Besides, unlinked markers can be added to the analysis, reducing the residual genetic variation and further increasing the power of QTL mapping (Zeng 1994).

While the CIM confines the QTL analysis to one region at a time, the **multiple-interval-mapping** (MIM) uses multiple marker intervals simultaneously to fit multiple putative QTLs directly in the model for mapping QTL (Kao et al. 1999). The MIM allows thereby the full estimation of the genetic architecture of a quantitative trait in terms of the number of underlying QTLs, their genetic effects, pleiotropic effects, and epistatic network among different QTLs (Wu et al. 2007).

3.3.2. Interval mapping in classical populations *vs.* F_1 segregating populations

Suppose the presence of a QTL in two parental diploid lines denoted as *P* and *Q*. In populations derived from inbred lines, only two alleles segregate at the QTL locus, P_1 and Q_1 derived from parents *P* and *Q*, respectively. In a backcross population (Fig. 3.6a), the segregation of the QTL in the progenies results in two possible QTL genotypes, P_1P_1 and P_1Q_1.

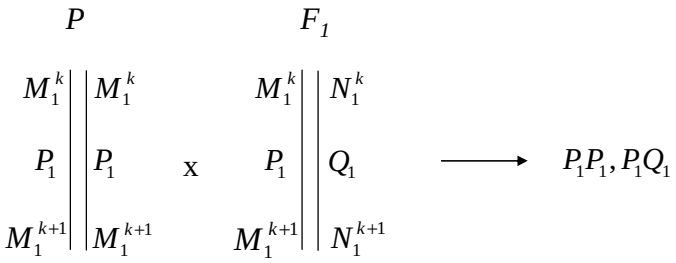

Fig. 3.6 Backcross of the F_1 resulting from a cross between two diploid inbred lines, *P* and *Q*, with the parental line *P* (Fig. 3.4a). The parental lines are characterized by two marker loci, *k* and *k+1*, and a QTL. Two QTL genotypes are present in the backcross population: P_1P_1 and P_1Q_1. Adapted from Lin et al. (2003) and Pastina et al. (2012).

The traditional approach for QTL mapping involves a linear regression of phenotype on genotype. In a backcross population, the phenotype and genotype of individual *i* at a given locus are assumed to be related by the following equation:

$$y_i = \mu_{P_1P_1} + \alpha g_i + \varepsilon_i \qquad (3.5)$$

where y_i is the phenotypic value of individual *i*, $\mu_{P_1P_1}$ is the mean phenotypic value of individuals with genotype P_1P_1, g_i is an indicator variable whose value depends on the genotype of individual *i*, α is the additive effect of a single allelic substitution (effect of having Q_1 instead of P_1), and ε_i is a random variable with

mean 0 and variance σ^2 (Lander and Botstein 1989). At the positions corresponding to a marker locus along the genetic map, g_i takes the value of 0 or 1 depending on the observed genotype, P_1P_1 and P_1Q_1, respectively. However, at any other position, the genotype of individual i and thus the value of g_i are unknown. Therefore, probabilities are used: the probability of having a given QTL genotype is inferred from the genotypes of the flanking markers and the relative genetic distances between these markers and the putative QTL (Doerge and Zeng 2011). Statistically, the linear model (3.5) becomes a mixture model since probability distributions replace the unknown independent variables (genotypes) (Lander and Botstein 1989; Zeng 1994). The linear regression solutions ($\hat{\mu}_{P_1P_1}$, $\hat{\alpha}$, $\hat{\sigma}^2$) are maximum likelihood estimates, *i.e.*, the values that maximize the probability, or likelihood function L, that the observed data would have occurred:

$$L(\mu_{P_1P_1}, \alpha, \sigma^2) = \prod_{i=1}^{N} [P_i(g_i = 0)L_i(0) + P_i(g_i = 1)L_i(1)] \qquad (3.6)$$

where $P_i(g_i = 0)$ is the probability that the QTL genotype of individual i, g_i, is P_1P_1 conditional on the genotypes and positions of the flanking markers, and $L_i(0) = z(y_i - \mu_{P_1P_1}, \sigma^2)$ is the likelihood function for individual i assuming that $g_i = 0$ and with z denoting the normal distribution with mean 0 and variance σ^2. Similarly, $P_i(g_i = 1)$ is the conditional probability that the QTL genotype of individual i is P_1Q_1, and $L_i(1) = z(y_i - \mu_{P_1P_1} - \alpha, \sigma^2)$, the likelihood function assuming that $g_i = 1$ (Lander and Botstein 1989). The test statistic is the LOD score, constructed using a likelihood ratio:

$$LOD = \log_{10} \frac{L(\hat{\mu}_{P_1P_1}, \hat{\alpha}, \hat{\sigma}^2)}{L(\hat{\hat{\mu}}_{P_1P_1}, \alpha = 0, \hat{\hat{\sigma}}^2)} \qquad (3.7)$$

where L is the likelihood function (3.6), $\hat{\mu}_{P_1P_1}$, $\hat{\alpha}$ and $\hat{\sigma}^2$ are the maximum likelihood estimates of $\mu_{P_1P_1}$, α and σ^2 under the alternative hypothesis H_1: $\alpha \neq 0$ (there is a QTL), and $\hat{\hat{\mu}}_{P_1P_1}$ and $\hat{\hat{\sigma}}^2$ are the respective estimates obtained under the null hypothesis H_0: $\alpha = 0$ (there is no QTL) (Lander and Botstein 1989; Doerge and Zeng 2011).

The application of the interval-mapping methodology in backcross, F_2 and F_1 segregating populations differs from each other by (i) the number of distinct genotypes that can be found at each locus and (ii) the number of distinct QTL effects that can be estimated.

(i) Number of distinct genotypes at each locus

In a backcross design, there are two possible genotypes at each locus. Thus, one can distinguish two possible QTL genotypes – P_1P_1 and P_1Q_1 – and four possible combinations of genotype for the two markers flanking the QTL, or *marker classes*. The likelihood function (3.6) is constructed by considering eight distinct conditional probabilities, each of them corresponding to the probability of having one of both possible QTL genotypes given one of the four marker classes (Wu et al. 2007).

In a F_2 design, three possible genotypes are possible at each locus – these are P_1P_1, P_1Q_1 and Q_1Q_1 at the QTL –, and thus the likelihood function includes 27 conditional probabilities, each of them corresponding to the probability of having one of the three possible QTL genotypes given one of the nine marker classes.

In an F_1 segregating population, the presence of four possible genotypes at each locus – these are P_1Q_1, P_1Q_2, P_2Q_1 and P_2Q_2 at the QTL – generates a total of 64 distinct conditional probabilities, each of them accounting for a distinct marker-QTL-marker genotypic combination. However, under a fixed allelic configuration at marker loci k and $k+1$, the allelic configurations between the QTL and flanking markers on a single chromosome are *a priori* unknown. Four assignments of the QTL alleles relative to the two flanking markers can be distinguished (Fig. 3.7), each of them generating a distinct set of 64 conditional probabilities (Lin et al. 2003). The foundation of the statistical framework developed by Lin et al. (2003) for QTL mapping in F_1 segregating populations is presented here.

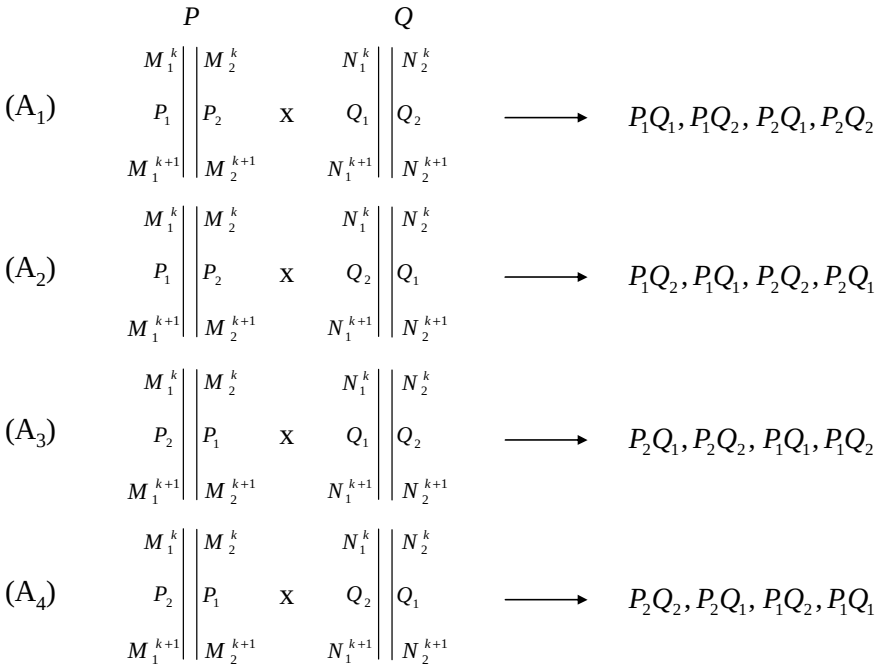

Fig. 3.7 Cross between two diploid outbred lines, *P* and *Q*, characterized by two marker loci, *k* and *k+1*, and a QTL. Under a fixed allelic configuration at marker loci *k* and *k+1*, four assignments of the QTL alleles relative to the flanking markers can be distinghuished. Each assignment generates the same four QTL genotypes; however, the conditional probability of a given QTL genotype upon the genotypes at marker loci vary depending on the assignment (Table 3.4). From Lin et al. (2003).

Let r, r_1 and r_2 denote the recombination fraction between loci k and $k+1$, between locus k and the QTL and between the QTL and locus $k+1$, respectively, with $r = r_1 + r_2 - 2r_1r_2$. Each parent generates eight M^k–QTL–M^{k+1} haploid gametes with the gametic probability depending on the assignment (Fig. 3.7). Under assignment A_1, the gamete probabilities are calculated as:

$$
\begin{aligned}
g_{111} &= P(M_1^k P_1\, M_1^{k+1}) = P(N_1^k Q_1\, N_1^{k+1}) = \frac{1}{2}(1-r_1)(1-r_2),\\
g_{112} &= P(M_1^k P_1\, M_2^{k+1}) = P(N_1^k Q_1\, N_2^{k+1}) = \frac{1}{2}(1-r_1)r_2,\\
g_{121} &= P(M_1^k P_2\, M_1^{k+1}) = P(N_1^k Q_2\, N_1^{k+1}) = \frac{1}{2}r_1r_2,\\
g_{122} &= P(M_1^k P_2\, M_2^{k+1}) = P(N_1^k Q_2\, N_2^{k+1}) = \frac{1}{2}r_1(1-r_2),\\
g_{211} &= P(M_2^k P_1\, M_1^{k+1}) = P(N_2^k Q_1\, N_1^{k+1}) = \frac{1}{2}r_1(1-r_2),\\
g_{212} &= P(M_2^k P_1\, M_2^{k+1}) = P(N_2^k Q_1\, N_2^{k+1}) = \frac{1}{2}r_1r_2,\\
g_{221} &= P(M_2^k P_2\, M_1^{k+1}) = P(N_2^k Q_2\, N_1^{k+1}) = \frac{1}{2}(1-r_1)r_2,\\
g_{222} &= P(M_2^k P_2\, M_2^{k+1}) = P(N_2^k Q_2\, N_2^{k+1}) = \frac{1}{2}(1-r_1)(1-r_2)
\end{aligned}
\qquad (3.8).
$$

The combination of the eight distinct gametes from each parent, *P* and *Q*, generates a total of 64 zygotic gentotypes. The joint probabilities of the 64 genotypes for the two markers and the QTL are calculated on the basis of the *g*'s probabilities (3.8) under each of the four assignments of QTL alleles and flanking markers (Table 3.4). Under a fixed allelic configuration at marker loci k and $k+1$, the conditional probability of each QTL genotype upon the marker interval genotypes is obtained from the M^k–QTL–M^{k+1} joint probabilities (Table 3.4) and M^k–M^{k+1} joint probabilities calculated from Table 3.2 by applying Bayes's theorem (Lin et al. 2003).

According to Lin et al. (2003), the correct characterization of the linkage phases between QTL and markers in an F_1 segregating population allows a more accurate estimation of the QTL effects and location. However, their statistical framework seems to have been little applied to date.

Table 3.4 Joint probabilities of 64 zygotic genotypes derived from a cross between two diploid outbred lines characterized by two marker loci, *k* and *k+1*, and a QTL. The marker alleles M_1 and M_2 derived from parent *P*, and N_1 and N_2, from parent *Q*. P_1, P_2, Q_1 and Q_2 are the alleles of the QTL. The 64 probabilities are generated by the combination of 16 marker classes and 4 QTL genotypes. These probabilities vary according to the assignment (Ass.) of the QTL alleles relative to the flanking markers (Fig. 3.7). From Lin et al. (2003)

		QTL				
		Ass.	Genotypes			
		A_1	P_1Q_1	P_1Q_2	P_2Q_1	P_2Q_2
		A_2	P_1Q_2	P_1Q_1	P_2Q_2	P_2Q_1
Marker		A_3	P_2Q_1	P_2Q_2	P_1Q_1	P_1Q_2
No	Genotypes	A_4	P_2Q_2	P_2Q_1	P_1Q_2	P_1Q_1
1	$M_1^k N_1^k M_1^{k+1} N_1^{k+1}$		g_{111}^2	$g_{111}g_{121}$	$g_{121}g_{111}$	g_{121}^2
2	$M_1^k N_1^k M_1^{k+1} N_2^{k+1}$		$g_{111}g_{112}$	$g_{111}g_{122}$	$g_{122}g_{112}$	$g_{121}g_{122}$
3	$M_1^k N_1^k M_2^{k+1} N_1^{k+1}$		$g_{111}g_{112}$	$g_{112}g_{121}$	$g_{121}g_{111}$	$g_{121}g_{122}$
4	$M_1^k N_1^k M_2^{k+1} N_2^{k+1}$		g_{112}^2	$g_{112}g_{122}$	$g_{122}g_{112}$	g_{122}^2
5	$M_1^k N_2^k M_1^{k+1} N_1^{k+1}$		$g_{111}g_{211}$	$g_{111}g_{221}$	$g_{121}g_{211}$	$g_{121}g_{221}$
6	$M_1^k N_2^k M_1^{k+1} N_2^{k+1}$		$g_{111}g_{212}$	$g_{111}g_{222}$	$g_{121}g_{212}$	$g_{121}g_{222}$
7	$M_1^k N_1^k M_2^{k+1} N_1^{k+1}$		$g_{112}g_{211}$	$g_{112}g_{221}$	$g_{122}g_{211}$	$g_{122}g_{221}$
8	$M_1^k N_2^k M_2^{k+1} N_2^{k+1}$		$g_{112}g_{212}$	$g_{112}g_{222}$	$g_{122}g_{212}$	$g_{122}g_{222}$
9	$M_2^k N_1^k M_1^{k+1} N_1^{k+1}$		$g_{211}g_{111}$	$g_{211}g_{121}$	$g_{221}g_{111}$	$g_{221}g_{121}$
10	$M_2^k N_1^k M_1^{k+1} N_2^{k+1}$		$g_{211}g_{112}$	$g_{211}g_{122}$	$g_{221}g_{112}$	$g_{221}g_{122}$
11	$M_2^k N_1^k M_2^{k+1} N_1^{k+1}$		$g_{212}g_{111}$	$g_{212}g_{121}$	$g_{222}g_{111}$	$g_{222}g_{121}$
12	$M_2^k N_1^k M_2^{k+1} N_2^{k+1}$		$g_{212}g_{112}$	$g_{212}g_{122}$	$g_{222}g_{112}$	$g_{222}g_{122}$
13	$M_2^k N_2^k M_1^{k+1} N_1^{k+1}$		g_{211}^2	$g_{211}g_{221}$	$g_{221}g_{211}$	g_{221}^2
14	$M_2^k N_2^k M_1^{k+1} N_2^{k+1}$		$g_{211}g_{212}$	$g_{211}g_{222}$	$g_{221}g_{212}$	$g_{221}g_{222}$
15	$M_2^k N_2^k M_2^{k+1} N_1^{k+1}$		$g_{211}g_{212}$	$g_{212}g_{221}$	$g_{222}g_{211}$	$g_{221}g_{222}$
16	$M_2^k N_2^k M_2^{k+1} N_2^{k+1}$		g_{212}^2	$g_{212}g_{222}$	$g_{222}g_{212}$	g_{222}^2

(ii) Number of QTL effects

A search for QTLs by interval mapping in a backcross design allows the estimation of only one additive QTL effect due to a single allelic substitution, as illustrated by the linear regression (3.1). In a F_2 population, two QTL effects can be estimated, one additive (effect of having the QTL genotype Q_1Q_1 instead of P_1P_1) and one dominance effect, which results from an intra-locus interaction between the additive effects in each parent (effect of having P_1Q_1 instead of P_1P_1 or Q_1Q_1). In an F_1 segregating population, two additive effects, one for each parent, and one dominance effect can be estimated. The additive effect due to parent *P* is the effect of having P_1Q_1 or P_1Q_2 instead of P_2Q_1 or P_2Q_2, while the dominance effect is the effect having P_1Q_1 or P_2Q_2 instead of P_1Q_2 or P_2Q_1 (Lin et al. 2003; Wu et al. 2007).

QTL mapping in F_1 segregating populations by using the mixture model has been implemented in the software MapQTL® (van Ooijen 2009).

3.3.3. QTL mapping by regression analysis

An approximation of the linear regression model (3.5) was proposed by Martinez and Curnow (1992) and Haley and Knott (1992). Instead of treating g_i as missing data and using a mixture model via maximum likelihood for the analysis, the Haley-Knott approximation uses the probability p_i of having a given QTL genotype conditional on the marker data directly as known variable in the model. For the backcross example, the model (3.5) becomes:

$$y_i = \mu_{P_1P_1} + \alpha\, p_i + \varepsilon_i \tag{3.8}$$

where $p_i = P(g_i = 1 | m^{k}, m^{k+1}, \theta)$, with θ accounting for the position of the QTL relatively to the flanking marker loci *k* and *k+1*.

The equation (3.8) is a simple regression model, and the statistical analysis is therefore straightforward (Doerge and Zeng 2011). It has been shown that this procedure gives a very good approximation of the likelihood profile (Haley and Knott 1992; Rebai et al. 1995). One additional advantage of this method is that the analysis can be performed using any general statistical

package (Haley and Knott 1992). Moreover, the formulation of QTL mapping as a regression problem allows a great flexibility for incorporating environmental co-variables into the analysis and adequately modelling the residual genetic variation (Malosetti et al. 2004; Boer et al. 2007; Malosetti et al. 2007). A variety of mixed models is currently used for QTL mapping (Xu and Yi 2000; Gilmour 2007; Cooper et al. 2009; van Eeuwijk et al. 2005, 2010). These models use *genetic predictors* as explanatory variables, which consist of a function of the conditional probabilities of the QTL genotype given flanking marker information and represent the number and origin (parent/ founder) of QTL alleles (Lynch and Walsh 1998; van Eeuwijk et al. 2010).

The regression model was used for QTL mapping in an F_1 segregating population in sugarcane (Pastina et al. 2012). Because of its flexibility for incorporating additional factors and the possibility of implementation in general statistical softwares, the QTL analysis performed in the present study was conducted using a regression model in a way similar to Pastina et al. (2012).

PART II – EXPERIMENTATIONS

AND RESULTS

CHAPTER IV

THE RELATIONSHIP OF STEM AND SEED YIELDS TO FLOWERING PHENOLOGY AND SEX EXPRESSION IN MONOECIOUS HEMP

The present chapter is a modified version of a research paper entitled:

A.-M. Faux [a], X. Draye [a], R. Lambert [a,b], R. d'Andrimont [a], P. Raulier [a], P. Bertin [a] (2013). The relationship of stem and seed yields to flowering phenology and sex expression in monoecious hemp (*Cannabis sativa* L.). Eur J Agron 47:11-22.

[a] Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve

[b] Centre de Michamps, Bastogne

ABSTRACT

Flowering phenology and sexual dimorphism are two major features that affect stem and seed production in hemp (*Cannabis sativa* L.), a short-day naturally dioecious plant. The sowing time is of primary importance because it affects flowering time and thereby influences stem yield. In spite of their unstable sexual phenotype, monoecious cultivars facilitate the harvest of both stems and seeds by reducing crop heterogeneity. The main objective of this paper was to determine the stem and seed yields for five monoecious hemp cultivars in relation to their flowering phenology and sex expression.

Sowing was carried out on five distinct dates during 2007 and 2008 at two sites in Belgium. The duration from sowing to flowering in days, both stem and seed yields and the seed harvest index decreased when sowing was postponed from mid-April to the end of June. The stem and seed yields from the mid-April sowing (approximately 12.5 and 1.9 t ha^{-1}, respectively) were within the ranges that were reported for fibre and both fibre and seeds production, respectively, in monoecious hemp. No interaction was observed between the sowing date and cultivar for both yields. Sex expression varied among cultivars, indicating that it might be selected, and was partly linked to earliness. Stem yields were lowest in the earliest cultivar (Uso 31) and highest in the latest one (Epsilon 68) while seed yields were lowest in the most masculinised and earliest cultivar (Uso 31) and highest in the most feminised and early (Fedora 17) or mid-early (Felina 32) ones. Both stem and seed yields correlated best with the duration from sowing to full female flowering or from sowing to the end of male flowering.

According to our results, harvesting the seeds in addition to the stems in monoecious hemp requires early sowing and the selection of feminised early or mid-early cultivars, earliness depending on the climatic conditions in the cultivation area. Therefore, it might be agriculturally valuable to take sex expression into account in addition to earliness during the selection of cultivars that are adapted to a dual purpose.

Keywords: Hemp (*Cannabis sativa* L.) · Flowering · Sex · Photoperiod · Stem yield · Seed yield.

4.1. INTRODUCTION

Increasing attention is being devoted to new crops because biobased products could play a key role in the search for energy independence and provide new opportunities for farmers (Janick and Whipkey 2007). A recently renewed interest in hemp, *Cannabis sativa* L., has been triggered by its agronomic characteristics and by the diversity of renewable resources that it produces. Hemp is a relatively high yielding crop with great agro-ecological plasticity, low or no pesticide requirement and modest demands for fertilizer (van der Werf et al. 1996; Struik et al. 2000). From lime-hemp concretes in the home insulation sector (Evrard 2006) to hempseed oil in food (Callaway 2004), a wide variety of products is derived from hemp fibre, woody core, cannabinoids or seeds (Struik et al. 2000). In this context, harvesting the seeds in addition to the stems can improve the profitability of the farming operation if the seed yield and market price are sufficiently high (Bocsa and Karus 1998). This paper focuses on the cultivation of hemp for a dual purpose, a practice that is found in countries like France (Bocsa and Karus 1998; Bouloc 2006; Beherec 2000) and Canada (Vera and Hanks 2004).

The control of flowering is essential for the genetic improvement and cultivation of hemp because its yields are affected by two major features of the flowering system: dioecy *vs.* monoecy and photoperiodism. Firstly, hemp is a naturally dioecious plant with sexual dimorphism. Dioecious hemp cultivars provide relatively high stem yields; however, the earlier flowering and senescence of male plants in comparison to female ones make the harvest of both the stems and the seeds difficult in these cultivars (Bocsa and Karus 1998). Monoecious hemp cultivars facilitate the simultaneous production of stems and seeds by reducing crop heterogeneity, but they have the disadvantage of an unstable sexual phenotype (Bocsa and Karus 1998). The monoecious state varies from predominantly male to predominantly female extreme phenotypes, and dioecious male plants occur at ever-increasing rates from one generation to another. The seed production of monoecious cultivars therefore requires the elimination of male dioecious plants prior to flowering and the control of the monoecious state at each generation (Beherec 2000). Additionally, hemp sex expression is known to be widely influenced by external factors such as hormonal treatments, photoperiod or nitrogen status (Freeman et al. 1980).

Secondly, hemp is a quantitative short-day plant, with a critical photoperiod of approximately 14 hours (Amaducci et al. 2008a; Borthwick and Scully 1954; Lisson et al. 2000b). By modulating the flowering time, photoperiodic conditions can have a key influence on crop yield because radiation-use efficiency (RUE) is lower after flowering than before flowering (van der Werf et al. 1994). The link between stem yield and flowering in fibre hemp has been reported in several studies. Stem yields can be improved by breeding late-flowering cultivars (van der Werf et al. 1994; Meijer et al. 1995; Lisson and Mendham 2000), and maximum stem and fibre yields are obtained when the harvest is carried out at the peak of flowering (Mediavilla et al. 2001; Amaducci et al. 2008c). On the other hand, studies related to the simultaneous production of stems and seeds are scarce. To our knowledge, the sex expression and flowering phenology in monoecious hemp in relation to both stem and seed production have never been investigated. This study could establish new criteria for the development of cultivars that are adapted for a dual purpose and support the crop management and thus the yields of dual purpose hemp.

Morphologically speaking, the inflorescences of monoecious hemp are similar to those of dioecious female plants (Bocsa and Karus 1998), with flowers arising in crowded racemes with leafy bracts (Mohan Ram and Nath 1964). A five-point scale established by Sengbusch (1952) classifies monoecious hemp forms according to their ratio of male/female flowers. As the number of flowering plants progressively increases after the appearance of the first flower, several floral developmental stages can be found simultaneously in the hemp crop (Bocsa and Karus 1998). The duration from emergence to 50% flowering has been modelled in response to temperature and photoperiod (Lisson et al. 2000b; Amaducci et al. 2008a). Flowering dynamics (% of flowering plants *vs.* time) were modelled with a bi-logistic curve, which indicated the presence of two processes corresponding to successive male and female flowerings in dioecious cultivars (Amaducci et al. 2008b). Two scales describing the phenology of monoecious hemp have been identified. The first one, which is used by the Fédération Nationale des Producteurs de Chanvre (FNPC, France), assigns seven stages without accounting for the sex of the plant (1-emergence, 2-canopy closure, 3-active growth, 4-start of flowering, 5-full flowering, 6-end of flowering, 7-seed maturity) (Picault 2006). The second one (Mediavilla et al. 1998) divides the life cycle of the plant into four main stages (1-germination and emergence, 2-vegetative stage, 3-flowering and seed formation, 4-senescence) and distinguishes the male and female flowers (stage

3) in monoecious plants. The second scale was considered to be more appropriate for this study to characterize both the phenology and sex expression in monoecious hemp.

The main objective of this research was to determine the stem and seed yields in monoecious hemp cultivars in relation to their flowering phenology and sex expression. Environmental variation was created by the selection of two contrasting trial sites with a range of sowing dates. Finally, the relationships between floral development and yields are discussed with an eye towards improving crop management for hemp that is cultivated for both stem and seed production.

4.2. MATERIALS AND METHODS

4.2.1. Genetic material

Seeds were imported from the Fédération Nationale des Producteurs de Chanvre (FNPC), which is located in Le Mans (France). A set of five monoecious hemp cultivars was selected to cover a wide range of earliness: Uso 31 (very early), Fedora 17 (early), Santhica 27 and Felina 32 (mid-early), and Epsilon 68 (late). Cultivar use varied according to the latitude because the time of anthesis is primarily daylength-dependent (de Meijer and Keizer 1994). At our latitude (approximately 50° N), the Uso 31, Fedora 17 and Santhica 27 cultivars were *a priori* considered for the production of both stems and seeds, and Felina 32 and Epsilon 68 were considered for stem production only.

4.2.2. Trials

The trials were carried out during 2007 and 2008 at two sites in Belgium with distinct edapho-climatic conditions: Corroy-le-Grand (50°40’N, 4°38’E, 110 m a.s.l.) in the loamy region and Michamps (50°02’N, 5°48’E, 500 m a.s.l.) in the Ardennes, which are characterized by a stony soil with slightly developed shale substrate and a silty loam texture (FAO 2006). Both soils were characterized by a volumetric water content (θ) of approximately 40% between

the field capacity (pF = 2; θ ~ 45%) and the permanent wilting point (pF = 4.2; θ ~ 5%). On average, the temperature is 3°C lower and the annual rainfall 200 mm higher in the Ardennes than in the loamy region (IRM 2010). The daylength pattern is identical for both locations as they are located roughly at the same latitude.

In each of the four combinations of year and site, the trial was arranged in a split-plot design that included three blocks. The main plots in each block were occupied by five sowing dates and the subplots by the five cultivars, generating a total of 300 plots (2 years x 2 sites x 3 blocks x 5 sowing dates x 5 cultivars). The sowing dates ranged from mid-April to mid-July. In 2007, the plot size was relatively high (7 x 6 m²) to prevent crop heterogeneity. In 2008, it was reduced to 6 x 3 and 6 x 4 m² in Corroy-le-Grand and Michamps, respectively. The effect of the plot size reduction on plot heterogeneity was assessed by testing the model (4.1) on the plant density by considering a decreasing number of randomly selected plant density measurements per plot in the analysis. We verified that the interpretation of the effect of experimental factors on plant density was not modified after a reduction of on average 45% of the number of measurements per plot that were integrated into the analysis, and concluded that such a plot size reduction should not increase the heterogeneity. This result was confirmed by the a posteriori determination of the mean variation coefficient of plant density within each plot, which was similar in both years (CV = 0.28 and 0.25 in Corroy-le-Grand and 0.34 and 0.32 in Michamps in 2007 and 2008, respectively). The row width was 12 cm. In 2007, the seeding rate was 55 kg ha^{-1} for *a priori* mixed cultivars Uso 31, Fedora 17 and Santhica 27 and 65 kg ha^{-1} for *a priori* stem cultivars Felina 32 and Epsilon 68. The seeding rate was fixed to 60 kg ha^{-1} for all cultivars during 2008 for the sake of convenience. Tillage consisted of ploughing in January followed by two passages with a rotary harrow in Corroy-le-Grand and one passage with an extirpator and one with a spring-tooth harrow in Michamps. The second passage was performed just before sowing at both sites. The previous crops were spelt and winter barley in Corroy-le-Grand and spelt and grass in Michamps during 2007 and 2008, respectively. N was supplied according to hemp cultivation references in France (Desanlis 2006), while P_2O_5 and K_2O were supplied according to the results of soil analysis. N was manually applied just after each sowing (as NH_4NO_3 27%) and there was a basal application before tillage for the P_2O_5 (as superphosphate) and K_2O (as KCl 40%) for ratios of 120-0-160 and 120-0-165 kg ha^{-1} in Corroy-le-Grand and 120-100-190 and 120-120-168 kg ha^{-1}

in Michamps in 2007 and 2008, respectively. No pesticide was applied and no weeding was performed.

4.2.3. Data collection

Climatic data

Hourly measurements of the temperature (°C) and rainfall (mm) were provided by the agro-meteorological sensors from a regional network with locations at both sites (PAMESEB 2008).

Phenology and sex expression

The time of full canopy closure and seven floral developmental stages (three male and four female, Table 4.1) was recorded. The five-point scale by Sengbusch (1952) was used to characterize sex expression, or degree of monoecy, in the hemp crops (Table 4.2). Approximately 25 to 30 plants were randomly taken at each observation time to determine the plot's developmental stage and monoecy degree. During 2007, observations at the first, third and fifth 'sowing date' treatments were restricted to three (Uso 31, Santhica 27 and Epsilon 68) and two (Santhica 27 and Epsilon 68) cultivars in Corroy-le-Grand and Michamps, respectively. Records of the flowering phenology and sex expression were made throughout the growing season, from mid-June (start of flowering in the mid-April sowings) to mid-September (before the first harvest) once every 10 days on average in 2007 and every 6 days in 2008. Full canopy closure was considered to be reached when sowing lines were no longer distinguishable, and floral developmental stages were said to be achieved when observed in about 50% of the plants. The median time was used when the same stage was observed at distinct consecutive times in about 50% of the plants. The duration of a phenophase was defined as the duration between two consecutive developmental stages.

Table 4.1 Developmental stages used to describe the flowering phenology in monoecious hemp

	Developmental stage	Code of Mediavilla et al. (1998) [1]	Description
Male phenology	Start of flowering (SMF)	2303	Presence of closed male flowers, easily visible.
	Full flowering (FMF)	2304	Male flowers are open and disperse pollen. Huge quantities of pollen are usually produced.
	End of flowering (EMF)	-	Most male flowers are withering. Petals turn brown and fall.
Female phenology	Start of flowering (SFF)	2301	Presence of white styles at leaf axils.
	Full flowering (FFF)	2302	Coexistence of pollinated flowers (with a swelling basis) and unpollinated flowers (white styles). The leaves of the top of the stem are flattening and form a plateau.
	End of flowering (EFF)	2305	Presence of seeds, still green. White styles are scarce.
	Fruit set (FS)	2306	Fruit ripening. Seeds are hardening. Bracts are tight.

[1] The phenological scale of Mediavilla et al. (1998) is presented in Table 2.1.

Table 4.2 Sex expression scale in monoecious hemp based on Sengbusch (1952)

Monoecy degree	Sex ratio
1	80 - 100% of ♂ flowers (strongly masculinised plot)
2	60 - 80% of ♂ flowers (masculinised plot)
3	40 - 60% of ♂ and ♀ flowers
4	60 - 80% of ♀ flowers (feminised plot)
5	80 - 100% of ♀ flowers (strongly feminised plot)

Plant density and yields

The plant density was evaluated at approximately 60 days after sowing by counting the number of plants in 12 1 m-long lines for each plot. The mean plant density across the four year*site combinations was 19.7 ± 6.5, *i.e.*, approximately 167 plants m^{-2}. However, we chose not to present the observations of plant density because no significant effect of the plant density on yields was observed ($\alpha = 0.05$). This was tested for each year*site combination independently by adding a continuous fixed 'plant density' effect and its interactions in the model (4.1).

The plots corresponding to the same sowing date were harvested at the same date once the seed maturity – or senescence if the seed production was null or not significant – had been reached. Three samples were harvested per plot, each having a surface of 1 m². For each sample, the fresh weight of the stems was recorded and a sub-sample of the stems as well as the entire quantity of seeds were dried at 70°C for 72 hours to compute dry matter yields.

4.2.4. Statistical analyses

Statistical analyses were performed to assess the effects of the year, site, cultivar and sowing date on the duration of phenophases, monoecy degree, stem and seed yields and seed harvest index. The phenophases duration was expressed in days and degree-days when the rate of development was correlated to the temperature. In this latter case, a base temperature (T_b) was determined by the variation coefficient method and the thermal duration of the phenophase was calculated as the time integral of $(T_{min} + T_{max})*0.5 - T_b$, with T_{min} and T_{max} being the minimum and maximum daily temperatures. As a time series, the monoecy data were split into sub-datasets according to the developmental stage of the plants at the time of observation, and each sub-dataset was independently analyzed. The flowering phenology and yields were analyzed by using the MIXED procedure in the SAS statistical package while the monoecy degree, which was a discrete variable with five levels of response, was analyzed by using the GLIMMIX procedure with multinomial distribution for the response variable (DIST = MULTI) and cumulative logit link function (LINK = CUMLOGIT) (SAS Institute Inc. 2010).

The ‘year’, ‘site’ and ‘cultivar’ (cvar) factors were considered to be discrete and fixed because the trial only covered two years and the sites and cultivars were chosen for their *a priori* contrasting characteristics. The ‘block’ was considered to be discrete and random and the ‘sowing date’ (SD) was continuous and fixed. Because the early limit of hemp sowing in the region of study was unknown, the sowing date was counted as the delay in days after the earliest sowing date of the trial (April 16 was the first sowing date in Corroy-le-Grand in 2007). On the basis of our data, the response of the observed variables to the sowing date was considered to be linear. An initial model that included all possible interactions was tested first. The non-significant interactions ($\alpha = 0.05$) were removed as far as the interpretation of the main effects and their interactions was not modified (‘year*cvar’, ‘year*site*cvar’, ‘year*cvar*SD’, ‘year*site*cvar*SD’). The following model was used:

$$\begin{aligned} y_{ijkm} = {} & \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\beta\gamma)_{jk} + \\ & \left[\delta + (\alpha\delta)_i + (\beta\delta)_j + (\gamma\delta)_k + (\alpha\beta\delta)_{ij} + (\beta\gamma\delta)_{jk}\right] * SD_m + \\ & random\ \ effects + e_{ijkm} \end{aligned} \qquad (4.1)$$

where y_{ijkm} was the response of the ith cultivar sown at the jth site during the kth year at the mth sowing date; μ was the general mean response; α_i, β_j and γ_k were the main effects of the ith cultivar, jth site and kth year; $(\alpha\beta_{ij})$ and $(\beta\gamma)_{jk}$ were the effects of the interactions between the ith cultivar and jth site and between the jth site and kth year, respectively; δ was the mean effect of delaying the sowing by one day and $(\alpha\delta)_i$, $(\beta\delta)_j$, $(\gamma\delta)_k$, $(\alpha\beta\delta)_{ij}$ and $(\beta\gamma\delta)_{jk}$, were its deviations for every fixed factor; SD_m was the delay in days of the m sowing date after April 16. According to the split-plot design, the random effects in the model (4.1) included a ‘block’ effect that was nested into the year and site and, with the MIXED procedure only, its interaction with the sowing date which was the main plot effect. The LSMEANS estimates of phenophases duration and yields were computed for significant fixed discrete effects ($\alpha = 0.05$) and compared with the PDIFF method while the CONTRAST statement was used to examine significant interactions including the ‘sowing date’ continuous effect. For the monoecy degree, the MEANS procedure and the CONTRAST statement from the GLIMMIX procedure were used to compute means and pairwise comparisons, respectively, because LSMEANS computations are not supported by the GLIMMIX procedure for multinomial models (SAS Institute Inc. 2010).

4.3. RESULTS

4.3.1. Daylength, climatic conditions and culture periods

The daylength increased from 13.9 hours on April 16, the earliest sowing date, to 16.5 hours around June 21 at summer solstice before decreasing to 16.1 hours on July 16, the latest sowing date of the trials (Fig. 4.1). The intermediate sowing dates (from mid-May to mid-June) had the longest days during the first 30 days after sowing (Fig. 4.2a). With the exception of lower temperatures in April of 2008, the air temperature conditions were similar for both years during the trial periods (Fig. 4.1). The same temperature pattern was observed for both sites with temperatures on average 2.4°C lower in Michamps. During both years, the temperature sum that was calculated from the sowing time showed increasing differences between sowing dates from approximately 20 days after sowing, with lower temperature sum for mid-April-sown treatments, and between sites from approximately 50 days after sowing, with lower temperature sum in Michamps (Fig. 4.2b-c). The maximum daily temperatures were 30.7 and 31.1°C in Corroy-le-Grand and 28.8 and 28.5°C in Michamps in July 2007 and 2008, respectively. Lower daily temperatures were observed during spring or autumn depending on the sowing date, with 2.7 and 3.6°C in mid-April of 2007 and 2008 in Corroy-le-Grand, respectively, 4.6 and 0.3°C around May 30, 2008 in Corroy-le-Grand and Michamps, respectively, and -0.6 and -2°C in late October in Corroy-le-Grand (2007) and Michamps (both years). Negative temperatures were registered after October 15 during the trial periods, once in Corroy-le-Grand in 2007 *vs.* five and three times in Michamps in 2007 and 2008, respectively. With the exception of unusually dry and warm weather in April of 2007 (IRM 2010), the precipitation was distributed throughout the whole growing season and was higher in Michamps (575.5 and 621.4 mm in Corroy-le-Grand and 690.7 and 659.9 mm in Michamps, from April to November in 2007 and 2008, respectively). The mean relative air humidity followed the same pattern (with an average of 81.8 and 83.8% in Corroy-le-Grand and Michamps, respectively).

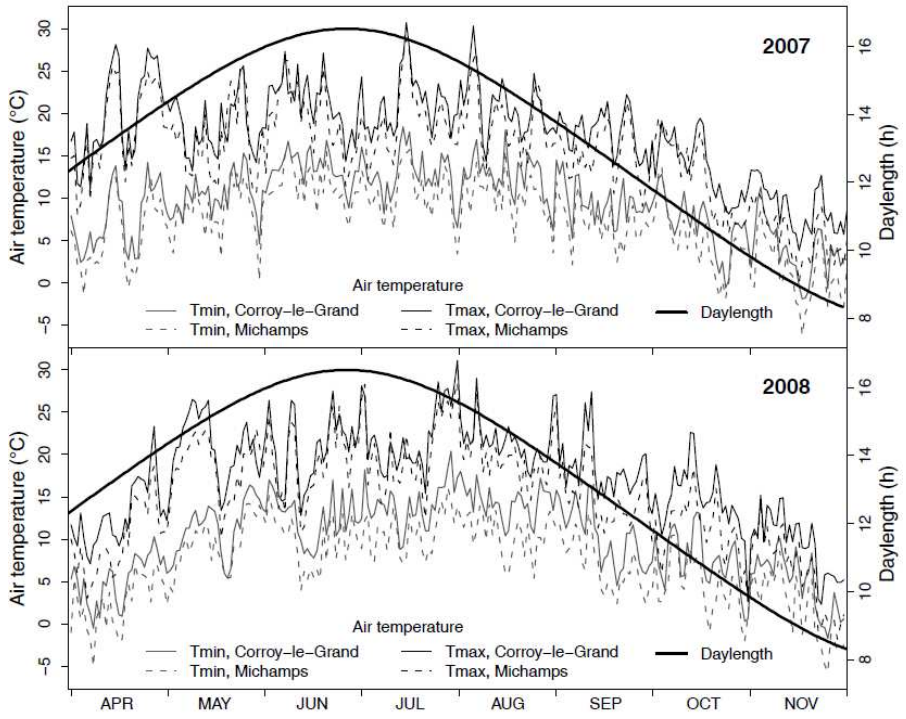

Fig. 4.1 Minimum and maximum daily air temperature and daylength from April to November at the two study sites in Belgium, Corroy-le-Grand and Michamps, in 2007 (above) and 2008 (below).

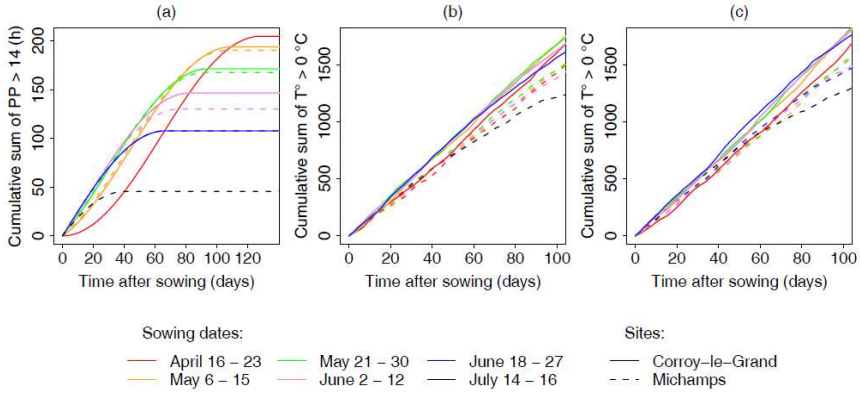

Fig. 4.2 Daylength and temperature conditions experienced by hemp plants sown at distinct dates ranging from Mid-April to Mid-July (colour) in two sites (line type) in Belgium. (a) Cumulative sum of photoperiods above 14 h, *i.e.*, the critical photoperiod for hemp (Amaducci et al. 2008a; Borthwick and Scully 1954; Lisson et al. 2000b). (b-c) Cumulative sum of temperatures above 0°C in 2007 (b) and 2008 (c).

Because of the climatic conditions, sowing was completed earlier in Corroy-le-Grand than in Michamps (Table 4.3). The drier weather in April of 2007 was responsible for the larger interval between the first two sowing dates. In Michamps, the low temperatures caused a delay in the first sowing during 2008 and precipitation caused a delay in the last sowing for both years. Harvest times varied from September 25 to November 12 depending on the sowing date. September and October of 2008 were relatively rainy and the harvest had to be postponed in Michamps. Both stems and seeds were harvested from all treatments in Corroy-le-Grand. In Michamps, the stems were harvested from the four (2007) and three (2008) first 'sowing date' treatments and seeds were harvested from the first three alone in 2007. Rainfall and high relative humidity (which was higher than 93% on average in Michamps in October and November of 2008) postponed the harvest and caused stem damage (lignification, lodging) and significant seed losses as a consequence of advanced maturity. This damage was too large to provide reliable results in the treatments from which stems or seeds were not harvested.

Table 4.3 Dates of sowing and harvest, delay of the sowing dates after April 16 and duration of the crop cycles in days for trials carried out at two sites in Belgium, Corroy-le-Grand and Michamps, in 2007 and 2008

	Corroy-le-Grand				Michamps			
	Sowing	Harvest	Delay of SD after April 16 [a]	Duration of crop cycle	Sowing	Harvest [b]	Delay of SD after April 16 [a]	Duration of crop cycle [b]
2007	April 16	Sep. 25	0	162	April 23	Oct. 3	7	163
	May 15	Oct. 9	29	147	May 12	Oct. 10	26	151
	May 29	Oct. 16	43	140	May 30	Oct. 31	44	154
	June 11	Oct. 30	56	141	June 12	Nov. 12	57	153
	June 27	Nov. 5	72	131	July 16	-	91	-
2008	April 17	Sep. 25	2	161	May 9	Oct. 28	24	172
	May 6	Sep. 26	21	143	May 23	Nov. 7	38	168
	May 21	Oct. 2	36	134	June 9	Nov. 7	55	151
	June 2	Oct. 3	48	123	June 18	-	64	-
	June 18	Oct. 9	64	113	July 14	-	90	-

[a] The delay of SD after April 16 (in days) accounts for the 'sowing date' effect in the statistical analysis. It was increased by one day in 2008, which was a leap year.

[b] Last sown treatments were not harvested due to high crop damage resulting from unfavourable climatic conditions.

4.3.2. Phenology

The duration from sowing to full canopy closure (CC) decreased from 32 to 23 days when the sowing date was postponed from mid-April to mid-July regardless of the site and cultivar (r^2 = 0.790, P <0.001, Tables 4.4 and 4.5). A base temperature of 4.7°C was computed by using the variation coefficient method (CV min = 0.056 ± 0.026). The mean thermal durations from sowing to CC were longer in Corroy-le-Grand (299°Cd) than in Michamps (247°Cd) (Table 4.4). Although the effect was lower, the influence of sowing date on the duration from sowing to CC remained significant at both sites despite the use of degree-days (P <0.01 in Corroy-le-Grand and P <0.001 in Michamps, Table 4.5).

Table 4.4 Date of canopy closure (CC) and duration from sowing to canopy closure in days and degree-days (base 4.7°C) according to the sowing date in two sites in Belgium, Corroy-le-Grand and Michamps, in 2007

Corroy-le-Grand				Michamps			
Sowing	Canopy closure	Duration sowing - CC		Sowing	Canopy closure	Duration sowing - CC	
		days	°Cd			days	°Cd
April 16	May 18	32	306.8	April 23	May 24	31	270.65
May 15	June 10	26	296.9	May 12	June 8	27	251.7
May 29	June 22	24	315.05	May 30	June 23	24	274.4
June 11	July 5	24	286.7	June 12	July 6	24	225.25
June 27	July 21	24	299.6	July 16	August 8	23	236.05

The male and female developmental stages occurred close together, with the start of male flowering happening slightly earlier than the female one. Thus, only the development of female flowers is illustrated here.

Table 4.5 P-values of fixed effects (SD = sowing date, cvar = cultivar) for the phenology (duration of phenophases), sex expression (monoecy degree at six developmental stages), yields in stem and seed and seed harvest index

			Year	Site	Cvar	Year x Site	Cvar x Site	SD	SD x Year	SD x Site	SD x Cvar	SD x Year x Site	SD x Cvar x Site
Phenology [a]	Sowing - CC [b]	days		0.92	0.633		0.822	<0.001		0.84	0.933		0.84
	Sowing - SMF		0.113	0.559	<0.001	0.612	0.129	<0.001	0.639	0.405	<0.001	0.228	0.591
	Sowing - SFF		0.618	0.599	<0.001	0.33	0.037	<0.001	0.671	0.45	0.004	0.195	0.842
	SMF - FMF		0.276	0.363	<0.001	0.955	0.17	0.009	0.049	0.827	0.007	0.393	0.139
	SFF - FFF			0.926	0.001		0.042	0.967		0.41	0.026		0.119
	FMF - EMF		0.863	0.593	0.178		0.024	0.59	0.527	0.775	0.307		0.085
	FFF - EFF		0.636	0.611	0.692		0.242	0.284	0.329	0.895	0.894		0.335
	Sowing - CC b	°Cd [c]		<0.001	0.999		1	<0.001		0.002	0.269		0.766
	Sowing - SMF		0.099	<0.001	<0.001	0.879	0.259	0.003	0.298	0.655	<0.001	0.884	0.507
	Sowing - SFF		0.273	<0.001	<0.001	0.718	0.208	0.018	0.531	0.588	0.001	0.751	0.972
Sex expression [b]	SMF			0.858	0.001		0.05	0.001		0.85	0.144		0.266
	SFF			0.321	0.002		0.153	0.011		0.714	0.289		0.399
	FMF			0.201	0.038		0.139	0.065		0.297	0.244		0.445
	FFF			0.939	0.196		0.927	0.119		0.968	0.588		0.872
	EMF			0.179	0.718		0.874	0.108		0.918	0.637		
	EFF			0.405	0.892		0.512	0.295		0.511	0.659		0.475
Yields [d]	Stem yield		0.023	0.992	<0.001		0.003	<0.001	<0.001	0.028	0.121		0.254
	Seed yield		<0.001		0.002			<0.001	0.056		0.158		
	Seed harvest index		0.006		0.037			<0.001	0.03		0.194		

[a] For abbreviations of floral developmental stages see Table 4.1. CC = full canopy closure.

[b] Statistical analysis of time to canopy closure was performed on the data of 2007 and statistical of sex expression, on the data of 2008.

[c] Durations in °Cd were computed for phenophases whose rate of development was linked to temperature (base 4.4 °C for CC and 10.2 °C for SMF and SFF).

[d] Statistical analysis of yields was performed on the data from both sites in 2007 and from Corroy-le-Grand only in 2008 for stem yields and on the data from orroy-le-Grand only in both years for seed yields and seed harvest index.

The calendar dates for all floral developmental stages from the start of flowering to fruit set were delayed during both years and at both sites for all cultivars when the sowing date was postponed from mid-April to mid-July (Table 4.6). The delay in calendar dates of developmental stages was however less than the delay in sowing dates, so the number of days from sowing to each developmental stage decreased when sowing was postponed. The last sowing date of each trial year (mid-July), which occurred only in Michamps, was not integrated in the following statistical analysis because it modified the interpretation of fixed effects. The year, site and their interactions with the sowing date had no significant effect on the duration of phenophases as expressed in days (Fig. 4.3, Table 4.5). The duration from sowing to the start of male flowering (SMF) and from sowing to the start of female flowering (SFF) in days decreased significantly with the sowing date in all cultivars (P <0.001), but the intensity of the response varied with the cultivar as follows: Uso 31 was significantly less sensitive to the sowing date than Fedora 17 (P <0.05), Santhica 27, Epsilon 68 and Felina 32 (P <0.01), the latter four of which showed no significant difference between them. All the cultivars flowered relatively early when sown in mid-July (not shown), *i.e.*, under the shortest photoperiods under evaluation (Fig. 4.2a).

Table 4.6 Dates of female developmental stages for two hemp cultivars, 'Uso 31' and Epsilon 68, *a priori* very early and late, respectively, in two sites in Belgium in 2008 (*note*: the present table was published as supplementary data, available on the web only)

		Start of flowering		Full flowering		End of flowering		Fruit set	
		Uso 31	Eps. 68	Uso 31	Eps.68	Uso 31	Eps.68	Uso 31	Eps.68
Corroy-le-Grand	April 17	26/06 [a]	30/07	18/07	06/08	26/07	16/08	11/08	28/08
	May 6	13/07	02/08	26/07	09/08	04/08	24/08	25/08	
	May 21	20/07	09/08	02/08	20/08	18/08	27/08	06/09	
	June 2	28/07	19/08	11/08	29/08	23/08	06/09		
	June 18	09/08	27/08	20/08	01/09	30/08			
Michamps	May 9	16/07	02/08	31/07	12/08	12/08	22/08	27/08	02/09
	May 23	27/07	05/08	09/08	19/08	22/08	28/08	02/09	06/09
	June 9	04/08	15/08	17/08	27/08	29/08	06/09	09/09	
	June 18	14/08	25/08	25/08	03/09	06/09	10/09		
	July 14	03/09	06/09						

[a] Mean date of the three observations (one per block) corresponding to the same 'year*site*SD*cvar' treatment.

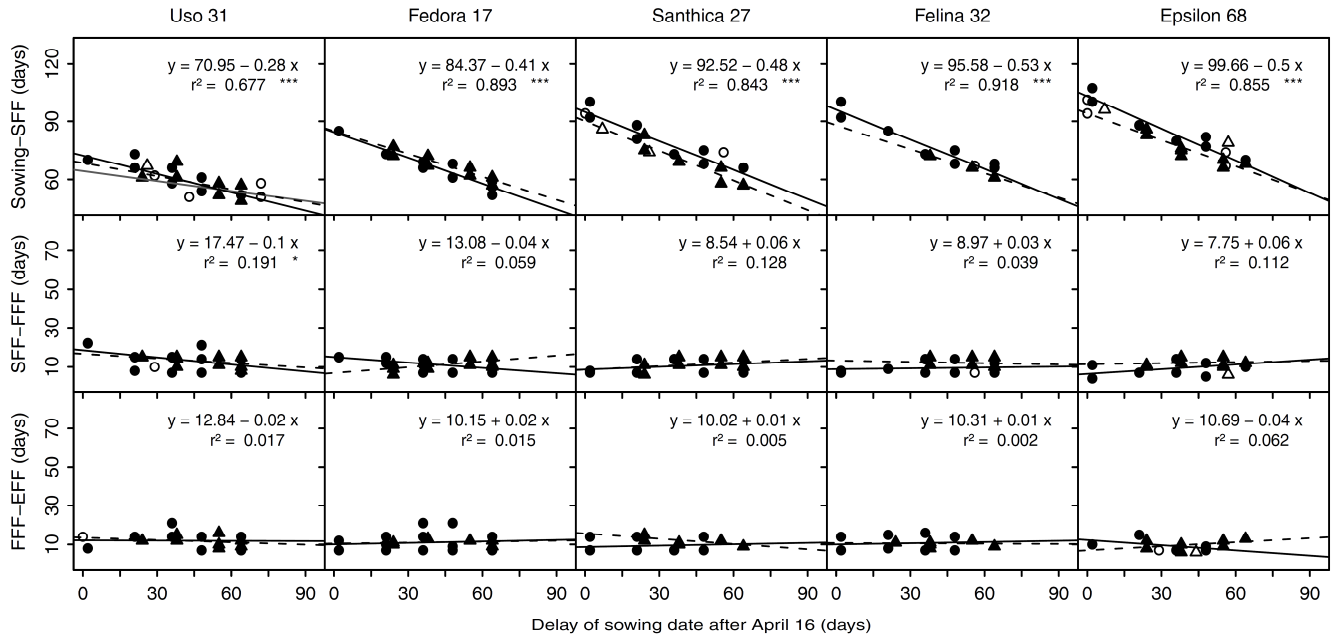

Fig. 4.3 Effect of delaying the sowing date on the duration in days of three female phenophases for five monoecious hemp cultivars at two sites in Belgium in 2007 and 2008 (SFF = start of female flowering; FFF = full female flowering; EFF = end of female flowering). The length of the *y* axis represents the same time interval (60 days) regardless of the phenophase. The equation and r^2 of the overall regression (not shown) are given. One point represents the observation of a single plot ('year*site*block*SD*cvar' treatment). ○ 2007 Corroy-le-Grand, Δ --- 2007 Michamps, ● — 2008 Corroy-le-Grand, ▲ --- 2008 Michamps.

A base temperature of 10.2°C was found for both durations from sowing to SMF and from sowing to SFF by the variation coefficient method (CV min = 0.069 ± 0.023 and 0.068 ± 0.021, respectively) (Fig. 4.4a). This base temperature was close to the values found by regressing the rate of development on temperature in each site independently and estimating the base temperature by extrapolating the relationship to its intercept with the temperature axis (Ong 1983; Villalobos and Ritchie 1992; van der Werf et al. 1995a) (Fig. 4.4b). The "site" effect on the relation between development rate and mean temperature resulted from similar durations sowing – start of flowering in days but distinct temperatures between sites.

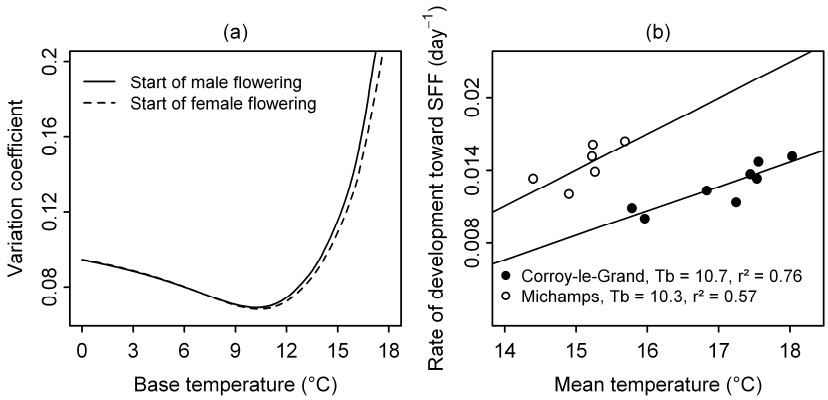

Fig. 4.4 (a) Mean variation coefficient of the durations from sowing to start of male/female flowering expressed in degree-days in function of the base temperature. Variation coefficients were computed in each 'year*site*cultivar*block' treatment so that the sowing date was the only controlled source of time variation. (b) Rate of development toward the start of female flowering (SFF) as a function of mean temperature for the mid-early cultivar Santhica 27. Base temperature and r^2 are given in each trial site

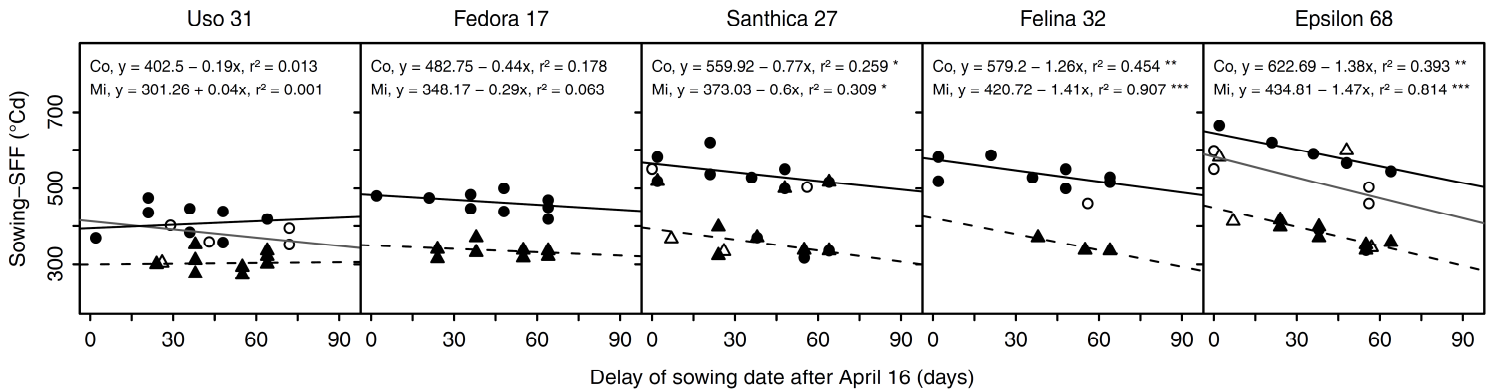

Fig. 4.5 Effect of delaying the sowing date on the duration in degree-days (base 10.2°C) from sowing to the start of female flowering for five monoecious hemp cultivars at two sites in Belgium in 2007 and 2008. The regression equation and r^2 are given at each site for both years together. One point represents the observation of a single plot ('year*site*block*SD*cvar' treatment). ○ — 2007 Corroy-le-Grand, Δ --- 2007 Michamps, ● ▬ 2008 Corroy-le-Grand, ▲ --- 2008 Michamps.

The effect of experimental factors on the thermal durations from sowing to the start of flowering was tested by using the linear model (4.1) after verifying that the quadratic effect of the sowing date and its interactions with the site, year and cultivar were not significant. Thermal durations from sowing to SMF and from sowing to SFF were higher in Corroy-le-Grand than in Michamps (P <0.001) as a result of higher temperatures in Corroy-le-Grand that were associated with similar durations from sowing to the start of flowering in days between sites (Fig. 4.5, Table 4.5). The effect of the sowing date was reduced with time as expressed in thermal units, with significant effects in Santhica 27 (P <0.05), Felina 32 (P <0.01) and Epsilon 68 (P <0.001) only.

Table 4.7 (a) Duration of phenophases for five monoecious hemp cultivars in 2008 (LSMEANS estimates ± SE). (b) Duration from emergence to flowering for five hemp cultivars as estimated from Amaducci et al. (2008a) in base 10.2°C and presented for earliness comparison

(a)		Uso 31	Fedora 17	Santhica 27	Felina 32	Epsilon 68
Sowing - SMF	days	59.84 ± 0.71 a	67.19 ± 0.71 b	73.07 ± 0.78 c	74.03 ± 1.04 c	78.91 ± 0.82 d
Sowing - SFF		61.78 ± 0.64 a	68.47 ± 0.64 b	73.72 ± 0.7 c	74.03 ± 0.94 c	78.86 ± 0.69 d
SMF - FMF		13.94 ± 0.68 a	11.44 ± 0.68 b	9.29 ± 0.72 c	10.12 ± 0.88 bc	10.2 ± 0.76 bc
SFF - FFF		13.65 ± 0.71 a	11.07 ± 0.7 b	10.67 ± 0.75 b	10.9 ± 0.94 b	10.74 ± 0.75 b
FMF - EMF		11.44 ± 0.64	10.26 ± 0.67	10.86 ± 0.69	9.75 ± 0.67	9.4 ± 0.68
FFF - EFF		12.06 ± 0.74	11.03 ± 0.78	10.99 ± 0.82	10.68 ± 0.81	9.29 ± 0.81
EFF - FS [a]		16.91 ± 1.43	17.55 ± 1.88	11.33 ± 2.53		
Sowing - SMF [b]	°Cd	341 ± 5 a	394 ± 5 b	439 ± 6 c	452 ± 8 c	486 ± 6 d
Sowing - SFF [b]		354 ± 5 a	401 ± 5 b	445 ± 6 c	452 ± 8 c	485 ± 6 d
(b)		Felina 34	Futura 77	Tiborszallasi	Fibranova	Carmagnola
Emergence – Flowering, mean (min – max) [c]	°Cd	604 (329 – 1136)	1158 (451 – 1471)	1484 (439 – 1913)	1485 (504 – 1787)	1535 (505 – 1743)

For abbreviations of floral developmental stages see Table 4.1.

Data on the same line and followed by a different letter are significantly different (P <0.05).

[a] The duration from EFF to FS was not included in the statistical analysis (Table 4.5) because the fruit set was recorded for the first three (Uso 31), first two (Fedora 17 and Santhica 27) or first (Felina 32 and Epsilon 68) sowing dates only.

[b] Durations in °Cd were computed for phenophases whose rate of development was linked to the temperature (base 10.2°C).

[c] The values reported here represent the sum of the durations in degree-days for the BVP, PIP and FDP (mean, minimum and maximum, respectively) given by Amaducci et al. (2008a) in their Table 5 after conversion from base temperatures of 1.9, 11.3 and 11.3°C, respectively, into a base temperature of 10.2°C by using the respective durations in days.

All cultivars had distinct floral responses except Santhica 27 and Felina 32, which were not significantly different from one other (Table 4.7a). Uso 31 was the earliest cultivar. It was followed by Fedora 17 (early), Santhica 27 and Felina 32 (mid-early) and Epsilon 68 (late). This order remained consistent across the different sowing dates despite the significant 'SD*cvar' interaction and agreed with the result from the FNPC. The start of male flowering preceded the female one only in the earliest cultivars (by 2 days in Uso 31 and 1 day in Fedora 17).

With the exception of Uso 31, for which the duration from the start of flowering to full flowering decreased with the sowing date (P <0.01), no sowing date effect was observed on the duration of phenophases following the start of flowering (Fig. 4.3). The duration from start to full flowering, male and female, was longer in Uso 31 (on average 13.8 days) than in later cultivars (on average 10.5 days) (P <0.01; Table 4.7a) and the duration from full flowering to the end of flowering was 10.5 days on average, regardless of the cultivar. Nevertheless, the order of cultivars according to earliness that was observed at the start of flowering was maintained at later developmental stages in spite of differences in phenophases durations. No base temperature was calculated for completion of phenophases following the start of flowering as the relation between the rate of development and temperature was mostly insignificant.

Table 4.8 Monoecy degree (Table 4.2) at seven developmental stages for five hemp cultivars in 2008 (mean ± standard deviation)

Stage	Uso 31	Fedora 17	Santhica 27	Felina 32	Epsilon 68
Start of male flowering	1.72 ± 0.79 a	2.46 ± 0.76 bc	2.24 ± 0.72 ab	2.64 ± 0.7 c	2.78 ± 0.67 c
Start of female flowering	1.85 ± 0.77 a	2.5 ± 0.69 b	2.27 ± 0.72 ab	2.68 ± 0.69 b	2.82 ± 0.72 b
Full male flowering	2.48 ± 0.64 a	3.04 ± 0.61 c	2.73 ± 0.53 ab	3 ± 0.69 abc	2.84 ± 0.62 bc
Full female flowering	2.56 ± 0.65	3.04 ± 0.62	2.74 ± 0.53	3 ± 0.71	2.96 ± 0.55
End of male flowering	3.52 ± 0.59	3.52 ± 0.6	3.39 ± 0.61	3.7 ± 0.66	3.84 ± 0.76
End of female flowering	3.6 ± 0.76	3.64 ± 0.73	3.76 ± 0.83	4.1 ± 0.7	4.05 ± 0.76
Fruit set	4.6 ± 0.51	4.67 ± 0.49	4.6 ± 0.52	4.78 ± 0.44	4.8 ± 0.45

Data on the same line followed by a different letter are significantly different (P <0.05).

4.3.3. Sex expression

The degree of monoecy of the plants (Table 4.2) evolved from the start of flowering to fruit set in both years and sites (Table 4.8 in 2008). At the start of flowering, the first flowers were generally male (with a degree of monoecy of on average 2.4, depending on the cultivar). Afterwards, the newly emerging flowers were mostly female, leading to a monoecy degree of on average 2.8 at full flowering and 4.7 at fruit set. The withering and eventual fall of male flowers at fruit set could influence the perceived evolution of sex expression at this stage, but it was not the main cause of feminisation.

The GLIMMIX procedure converged when the analysis was restricted to the 2008 data, in which the monoecy degree was observed in all treatments. The analysis results were supported by the 2007 data, in which the degree of monoecy was observed in only three cultivars. The monoecy degree did not vary with the site but rather with the cultivar at the start of male and female flowering (P <0.01) and at full male flowering (P <0.05) (Table 4.5). At these developmental stages, Uso 31, which was the earliest cultivar, was the most masculinised one with a monoecy degree ranging from 1.7 to 2.5. It was followed by Santhica 27 (from 2.2 to 2.7) and then Fedora 17, Felina 32 and Epsilon 68 (from 2.5 to 3), which were the most feminised cultivars (Table 4.8).

Plants tended to be more feminised in the last sowings at the start of male (P <0.01) and female (P <0.05) flowering. This trend was observed in the Uso 31 and Santhica 27 cultivars in particular, although the 'SD*cvar' interaction was not significant. In 2008, the monoecy degree at the start of flowering, male or female, increased from 1 (sowing of April 17 in Corroy-le-Grand and May 9 in Michamps) to 2.3 (sowing of June 18 in Corroy-le-Grand and July 16 in Michamps) for Uso 31 on average and from 1.5 (sowing of April 17) to 2.5 (sowing of June 18) in Corroy-le-Grand and from 2 (sowing of May 9) to 3 (sowing of July 16) in Michamps for Santhica 27. The same observation was made in 2007 with the Uso 31 and Santhica 27 cultivars.

Table 4.9 Stem and seed dry matter yields (t ha^{-1}) and seed harvest index (%) for five hemp cultivars from two sites in Belgium (Corroy-le-Grand and Michamps) in 2007 and 2008 (LSMEANS estimates ± SE)

	Year	Site	Uso 31	Fedora 17	Santhica 27	Felina 32	Epsilon 68
Stem yield	2007	Co.	8.06 ± 0.26 a	8.22 ± 0.26 ab	9.48 ± 0.26 d	8.77 ± 0.26 bc	9.05 ± 0.26 cd
		Mi.	7.69 ± 0.34 a	8.74 ± 0.34 b	8.33 ± 0.34 ab	8.54 ± 0.34 ab	10.49 ± 0.34 c
	2008	Co.	8.13 ± 0.33 a	9.04 ± 0.33 b	10.39 ± 0.33 c	9.21 ± 0.33 b	10.76 ± 0.33 c
		Mi. [b]	7.03 ± 1.83	7.21 ± 1.86	8.72 ± 1.83	7.77 ± 1.83	8.9 ± 1.87
Seed yield	2007	Co.	0.93 ± 0.05 a	1.03 ± 0.05 a	0.98 ± 0.05 a	1.21 ± 0.05 b	0.92 ± 0.05 a
		Mi. [b]	0.14 ± 0.12	0.29 ± 0.12	0.19 ± 0.12	0.26 ± 0.12	0.4 ± 0.12
	2008 [a]	Co.	1.49 ± 0.09 a	1.75 ± 0.09 b	1.56 ± 0.09 ab	1.75 ± 0.09 b	1.61 ± 0.09 ab
Seed harvest index	2007	Co.	9.81 ± 0.52 ab	10.71 ± 0.52 bc	8.61 ± 0.52 ad	11.31 ± 0.53 c	8.28 ± 0.52 d
		Mi. [b]	1.37 ± 0.77	2.47 ± 0.77	2.23 ± 0.77	2.4 ± 0.77	2.91 ± 0.77
	2008 [a]	Co.	15.03 ± 0.76 a	16.05 ± 0.75 a	12.72 ± 0.75 b	15.78 ± 0.75 a	12.49 ± 0.75 b

Data on the same line followed by a different letter are significantly different (P <0.05).

[a] Seed yields were not available for Michamps in 2008 due to high crop damage resulting from unfavourable climatic conditions.
[b] Contrasts were not performed on 2008 stem yields and 2007 seed yields from Michamps because of significant losses attributed to hail and birds, respectively.

4.3.4. Stem and seed yields

There was a decrease in the stem and seed dry matter yields as well as the seed harvest index when the sowing date was postponed from mid-April to the end of June for all year*site combinations. However, the 2008 stem yields and the 2007 seed yields from Michamps (Table 4.9) were not included in the statistical analysis because of significant stem damage attributed to hail in May and June of 2008 and seed losses due to birds, respectively.

The 'SD*cvar' interaction had no significant effect on stem yields, indicating that the cultivars had a similar response to the sowing date (Fig. 4.6, Table 4.5). Significant 'year' (P <0.05) and 'SD*year' (P <0.001) effects were observed in the stem yields as a result of the relatively high stem yields that were obtained with Santhica 27, Felina 32 and Epsilon 68 in late May and early June 2008 sowings in Corroy-le-Grand. Intercepts (on average 13.2 and 12.3 t ha^{-1}, respectively, for a mid-April sowing) as well as slopes (on average -0.111 and -0.082 t ha^{-1} with a one-day delay in the sowing date, respectively) of the regressions for stem yields on the sowing date in Corroy-le-Grand were higher in absolute value in 2007 than in 2008. No 'site' effect was observed on the regression intercepts, but it was present on their slopes (P <0.05) and was higher in absolute value in Michamps (on average -0.131 t ha^{-1} with a one-day delay in the sowing date) than in Corroy-le-Grand in 2007. Nevertheless, the variation in stem yields with the sowing date appeared to be congruent between both years and sites (Fig. 4.6). The 'cvar*site' effect (P <0.01) was found in Santhica 27 in 2007, with relatively low stem yields in Michamps (on average 8.3 t ha^{-1}) in comparison to Corroy-le-Grand (on average 9.5 t ha^{-1}) (Table 4.9). As a result, the stem yields over both years and sites increased from Uso 31 (on average 8.0 t ha^{-1}), which was the earliest cultivar, followed by Fedora 17, Felina 32 and Santhica 27, to Epsilon 68 (on average 10.1 t ha^{-1}), which was the latest cultivar. The yields in the three intermediate cultivars were not significantly different from one other over both years and sites.

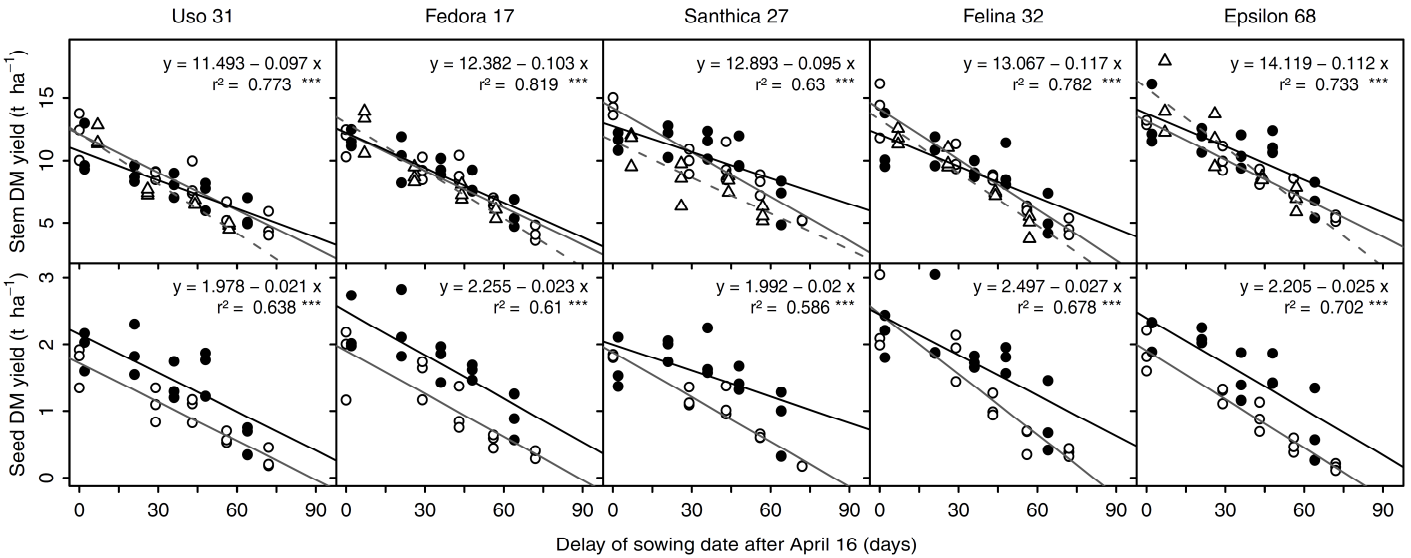

Fig. 4.6 Effect of delaying the sowing date on the stem (above) and seed (below) dry matter yields for five monoecious hemp cultivars at two sites in Belgium in 2007 and 2008. The equation and r^2 of the overall regression (not shown) are given. One point represents the observation of a single plot ('year*site*block*SD*cvar' treatment). ○ — 2007 Corroy-le-Grand, Δ --- 2007 Michamps, ● ▬ 2008 Corroy-le-Grand, ▲ --- 2008 Michamps.

The seed yields had a similar response to the sowing date regardless of the cultivar and year (Fig. 4.6, Table 4.5), decreasing by 0.023 t ha^{-1} with a one-day delay in the sowing date, on average. They were higher in 2008 than in 2007 (on average 2.3 and 2.0 t ha^{-1}, respectively, for a mid-April sowing in Corroy-le-Grand). The seed yields over both years were lowest in Uso 31 (on average 1.2 t ha^{-1}) and highest in Fedora 17 and Felina 32 (on average 1.4 and 1.5 t ha^{-1}, respectively), the latter two of which showed no significant difference between them. The seed yields in Epsilon 68 and Santhica 27 were not significantly higher than Uso 31 and only significantly lower than Felina 32 (Table 4.9).

The seed harvest index was lower in 2007 than 2008 (P <0.05) (on average 13.9 and 16.4% in 2007 and 2008, respectively, for a mid-April sowing in Corroy-le-Grand), and its decrease with the sowing date was higher in absolute value in 2007 than in 2008 (P <0.05) (on average -0.106 and -0.06% with a one-day delay in the sowing date in 2007 and 2008, respectively). The seed harvest index varied with the cultivar (P <0.05). It was lowest in Epsilon 68 and Santhica 27, followed by Uso 31, and highest in Fedora 17 and Felina 32 (Table 4.9).

Table 4.10 R-squared values (r^2) of the relations between dry matter yields in stems and seeds and the durations from sowing to six different floral developmental stages in days from two sites in Belgium during 2007 and 2008

Duration from sowing to	Corroy-le-Grand 2007			Michamps 2007		Corroy-le-Grand 2008		
	n [a]	Stem yield	Seed yield	n	Stem yield	n	Stem yield	Seed yield
Start of male flowering	31	0.347 ***	0.138 *	17	0.206	63	0.409 ***	0.237 ***
Start of female flowering	19	0.622 ***	0.362 **	10	0.416 *	68	0.482 ***	0.28 ***
Full male flowering	31	0.619 ***	0.521 ***	17	0.492 **	69	0.481 ***	0.308 ***
Full female flowering	18	0.685 ***	0.681 ***	14	0.676 ***	65	0.542 ***	0.334 ***
End of male flowering	20	0.523 ***	0.343 **	14	0.401*	61	0.646 ***	0.485 ***
End of female flowering	20	0.481 ***	0.274 *	18	0.572 ***	63	0.532 ***	0.389 ***

[a] Number of "block*SD*cvar" treatments included in the computation of r^2.

Significance levels: * P <0.05, ** P <0.01 and *** P <0.001.

4.3.5. Stem and seed yields in relation to flowering phenology

The stem and seed dry matter yields were related to the duration of each phenophase as expressed in days. Correlations were calculated by considering each 'block*SD*cvar' treatment individually for each year*site combination. No relation was observed between yields and durations of phenophases following the start of flowering (*i.e.*, durations from the start of flowering to full flowering and from full flowering to the end of flowering). On the other hand, both stem and seed yields were positively correlated with the duration from sowing to the start of flowering, from sowing to full flowering and from sowing to the end of flowering. Indeed, the earliest sowing dates (mid-April and early May) were associated with relatively long durations from sowing to each developmental stage and high stem and seed yields and, conversely, the latest ones (from mid-June) had relatively short durations and low yields. In all year*site combinations, the lowest r-squared values of the relationships between yields, in stems and seeds, and durations from sowing to a given developmental stage were observed in the duration from sowing to the start of male flowering, which was the earliest floral developmental stage (Table 4.10). In Corroy-le-Grand in 2008, the highest r^2 values were observed in the duration from sowing to the end of male flowering, followed by the duration from sowing to full female flowering or end of female flowering. The 2007 results were based on a lower number of observations but showed a similar tendency, with highest r^2 values observed with the duration from sowing to full female flowering.

4.4. DISCUSSION

4.4.1. Phenology

Although it was shorter, the mean duration from sowing to full canopy closure that was found in this study (273°Cd in base 4.7°C or 330°Cd in base 2.5°C) was in the same range as the duration reported by van der Werf et al. (1996), with 56°Cd (base 3°C) from sowing to emergence and 340°C (base 2.5°C) from emergence to 90% intercepted radiation. The difference was attributed to the lower density of plants in van der Werf et al. (1996) (64 plants m^{-2}).

A longer thermal duration from sowing to canopy closure was found in Corroy-le-Grand compared to Michamps (Table 4.4). For the first sowing date, it is possible that the canopy establishment was further limited by the dry weather of April 2007 in Corroy-le-Grand. Indeed, no rainfall occurred between April 5 and May 7 in Corroy-le-Grand, and between March 31 and April 29 in Michamps. This latter rainfall has likely positively affected the development of the hemp plants sown on April 23. However, differences in thermal duration from sowing to canopy closure between sites were also observed for the remaining sowing dates. Three of these sowings were performed at very close dates (between May 12 June 12), making the assumption of an effect of distinct radiation intensities unlikely. On the opposite, it is possible that the use of a more objective criterion for the assessment of the time of canopy closure, such as the time of 90% intercepted light (van der Werf et al. 1995a), would have allowed a more accurate comparison of the duration from sowing to canopy closure between sites.

The use of the median time between consecutive observations improved the observational accuracy of flowering phenology. The developmental stage time errors were estimated to be ± 5 and 3 days in 2007 and 2008, respectively. For this reason, we concluded that the observations provided reasonable developmental stage times for the purpose of the present study, *i.e.*, the characterization of the flowering phenology as a function of cultivar and environmental variation.

The sowing date and cultivar were the major factors in determining the flowering phenology as expressed in days (Fig. 4.3, Table 4.5). The congruence of the phenological development as expressed in days over years can be attributed to the primacy of the photoperiod in the determination of flowering time (de Meijer and Keizer 1994), in addition to the similar temperature pattern from both years (Fig. 4.1). The absence of a significant 'site' effect on the duration from sowing to the start of flowering in days may be explained by low temperature sum variations between sites during the period corresponding to the temperature-dependent basic vegetative phase (BVP) of development [max. 30 days after sowing in our conditions as estimated according to Lisson et al. (2000b) and Amaducci et al. (2008a)] in combination with the identical daylength pattern at both sites (Figs. 4.1 and 4.2). Indeed, once the BVP is complete, development is mainly determined by photoperiod sensitivity (Lisson et al. 2000b; Amaducci et al. 2008a) so it would barely have been affected by increasing differences in temperature sum between the sites (Fig. 4.2 b, c).

The use of degree-days revealed that the flowering response was determined by both temperature and photoperiod variations among sowing dates in combination with the cultivar (Fig. 4.5, Table 4.5). The long duration in days from sowing to the start of flowering that was observed in mid-April sowings was attributed to the relatively lower temperatures, which resulted in a longer BVP that stretched into the time of year when the photoperiod is long and not favourable for flowering. Later sowings experienced higher temperatures and, for treatments sown after early June, decreasing photoperiods, inducing shorter BVP and photoperiod-induced phase (PIP) of development, respectively.

The base temperature for the duration from sowing to the start of flowering (10.2°C) was close to the value of 11.3°C that was identified by Amaducci et al. (2008a) for both PIP and flower development phase (FDP) in hemp. Thermal requirements for the completion of flowering that was reported by Amaducci et al. (2008a) from field trials in Bologna were converted into a base of 10.2°C for comparison with those obtained in this study (Table 4.7b). The monoecious cultivar Felina 34, which was the earliest from those tested by Amaducci et al. (2008a), was the closest to the cultivars tested in this study in terms of earliness. According to Picault (2006), both Felina 32 and Felina 34 are mid-early cultivars. In this case, the shorter duration to flowering in days that was observed for Felina 34 by Amaducci et al. (2008a) (on average 49 days from emergence) would result from the shorter photoperiod in Bologna (from 0

to 1 h 20 minutes shorter between March 21 and June 21) and from the effect of very early (March) or very late (August) sowing dates (Amaducci et al. 2008a), while the longer thermal duration to flowering (Table 4.7b) would be a consequence of the higher temperatures in Bologna (on average from 5.5 to 8°C higher between April and August). The disparity in thermal durations to flowering between both studies still demonstrates the primacy of the photoperiod in determining hemp flowering time. Similarly, Amaducci et al. (2008a) noted a limited sensitivity to the base temperature for the PIP and FDP in their model predicting flowering time in response to temperature and photoperiod.

With the exception of very early cultivar Uso 31, the variations in both temperature and photoperiod during the phenophases following the start of flowering (*i.e.*, from the start of flowering to full flowering and from full flowering to the end of flowering) were likely too low to imply a differential response to the sowing date. Lisson et al. (2000b) also did not observe any clear response to a range of photoperiods from 10 to 15 hours for the duration from flowering to harvest in cultivars Futura 77 and Kompolti under controlled conditions.

4.4.2. Sex expression

Plants from the five cultivars were progressively feminised from the start of flowering to fruit set. The start of male flowering preceded the female one in the earliest cultivars (Table 4.7a). The relative earliness of male flowers suggests that protandry, which is well known in dioecious hemp (Bocsa and Karus 1998), may also be found in monoecious hemp. This finding confirms the hypothesis by Amaducci et al. (2008b) that early plants in monoecious hemp cultivars have male characteristics.

The sex expression varied significantly according to the cultivar at the start of flowering – male and female – and at full male flowering. This finding suggests that sex expression in monoecious hemp has a genetic basis and that the sexual phenotype of monoecious hemp could be selected. Borthwick and Scully (1954) noted a variable tendency to form male flowers among different dioecious female lines and concluded that the occurrence of monoecy in dioecious hemp is, to some extent, hereditary. Moreover, the order of cultivars

according to sex expression, ranging from the most masculinised one Uso 31, followed by Santhica 27, to the three most feminised ones Fedora 17, Felina 32 and Epsilon 68, was partly consistent with their order according to earliness, ranging from the earliest one Uso 31, followed by Fedora 17 and then Santhica 27 and Felina 32, to the latest one Epsilon 68. This observation suggests that the activity of genes that regulate flowering time may be involved in the determinism of the sex expression of monoecious hemp cultivars.

Plants were more feminised at the start of flowering in the last sowings, *i.e.*, under relatively shorter photoperiods. Borthwick and Scully (1954) attributed the excess of male over female plants that were observed in long photoperiods in a dioecious hemp cultivar to the fact that flowering is less delayed by long photoperiods in male plants than in female plants. According to Freeman et al. (1980), long day conditions can be considered a stress factor which is able to induce sex change by inclining the sex ratio toward males in hemp. They showed that this system of labile sexual expression that reduced the proportion of females in times of stress has a survival value for species in which the cost (in either increased mortality or decreased competitive ability) of reproducing as a female significantly exceeds that of reproducing as a male. In addition, the ability to change sex expression with sowing date, and thus with photoperiod, appeared to be linked to the expression of maleness, as the feminisation of plants at the start of flowering under shorter photoperiods was observed in the most masculinised cultivars, Uso 31 and Santhica 27 in particular. Further studies are clearly necessary to explore the genetic basis of sex expression in monoecious hemp.

4.4.3. Stem and seed yields and their relation to flowering phenology and sex expression

Stem yields in treatments that were sown before June 2 in Corroy-le-Grand and before May 12 in Michamps (Fig. 4.6) were within the ranges reported by Lisson and Mendham (2000) for monoecious cultivars and by Struik et al. (2000) with the monoecious Fedora 19 and Felina 34, all of which were cultivated for fibre only. The observation of the highest stem yields in early sowings may be explained by a larger amount of intercepted radiation, resulting

in a higher availability in carbon and accumulation of dry matter. This explanation is argued as follows.

In the earliest sown treatments, the canopy closure was reached around Mid-May, while the onset of senescence occurred from early to Mid-August depending on the cultivar (Table 4.4; time of end of flowering in Table 4.6). In the later sowings, both the canopy closure and onset of senescence occured later in terms of calendar dates. However, the duration from canopy closure to the onset of senescence decreased with sowing date (*e.g.*, it decreased from 80 to 60 days between the first and third sowing dates in Corroy-le-Grand). Besides, in the present area of cultivation, the intensity of the incident radiation is highest between May and July (data not shown). Therefore, as a result of both a longer duration from the canopy closure to the onset of senescence and the presence of a fully established canopy during the period of high intensity of incident radiation, the amount of radiation intercepted by the canopy would have been largest with early sowings. According to Monteith (1977), this larger amount of radiation intercepted by the canopy would have result in higher availability in carbon and dry matter accumulation in the earliest sown treatments.

The decrease in the stem yield with the delay of the sowing date (on average -0.102 t ha^{-1} with a one-day delay in the sowing date) was relatively close to the values of -2.3 and -1.4 t ha^{-1} that were reported by van der Werf et al. (1996) in the Netherlands when sowing was postponed from March 16 and 31 to April 15, respectively (*i.e.*, -0.076 and -0.093 t ha^{-1}, respectively with a one-day delay in the sowing date). Increasing stem yields therefore seems possible by sowing before mid-April, but the potential benefit must be weighed against the increased risk of frost damage (van der Werf et al. 1996).

Stem yields had a similar response to the sowing date regardless of the cultivar while the flowering response to the sowing date (duration from sowing to the start of flowering) was significantly lower in the earliest cultivar, Uso 31 (Table 4.5). This finding indicates that stem yield formation in Uso 31 still proceeded after the start of flowering and agreed with results from de Meijer and Keizer (1994), who stated that the earlier the start of flowering, the greater the stem length grew after anthesis. The variation of stem yields among cultivars was attributed to their earliness because no time variation in full canopy closure was observed among them. The lowest stem yields obtained with the earliest flowering cultivar (Uso 31) and, conversely, the highest ones obtained with the

latest flowering cultivar (Epsilon 68) were therefore explained by the decrease of stem growth after flowering (van der Werf et al. 1994).

Seed yields in treatments that were sown before June 2 (Fig. 4.6) in Corroy-le-Grand were greater than or close to the average values given by Bouloc (2006) in France (0.92 and 1.07 t ha^{-1} in Barrois and Champagne, respectively) for dual production systems. The reduction in the seed harvest index with a later sowing date indicates that the decrease in seed yields with the sowing date was higher than that in stem yields. The mean temperature that was experienced by plants between the end of flowering and harvest, *i.e.*, during seed maturation, was lower for treatments sown after mid-June than in mid-April (approximately 13.4 *vs.* 16.4°C). However, whether such temperature differences can affect seed yields in hemp is unknown. Moreover, the temperature may affect seed yield through opposite effects on the seed-fill duration and seed growth rate (Egli 2004). Plants from the late sowings were smaller and had shorter inflorescences than plants from the early sowings, which could negatively affect the number of flowers. According to Amaducci et al. (2008b), the flowering duration, as computed by the difference between the date of the last and the first flowering plant in a plot, was shorter in late sowings under our conditions, which could also result in a lower production of flowers. It must be noted that a long flowering duration inversely increases the probability of heterogeneous seed maturity, and thereby increases the number of green seeds at harvest so that excessively early sowing (before April in France) must be avoided (Picault 2006).

The order of cultivars according to seed yields was not affected by the sowing date, as already described for stem yields. Seed yields increased from Uso 31 (very early), followed by Epsilon 68 (late) and Santhica 27 (mid-early) and then Fedora 17 (early), to Felina 32 (mid-early). Schumann et al. (1999) also observed low seed yields in late flowering cultivars but did not always find higher yields for early cultivars. In the case of a competition between male and female flowers on a monoecious plant, the variation of seed yields among the cultivars of interest could be explained by their sex expression in addition to their earliness. Indeed, with the exception of Epsilon 68, the order of cultivars according to sex expression (from Uso 31, the most masculinised one, followed by Santhica 27, to Fedora 17, Felina 32 and Epsilon 68) corresponded to their order according to seed yield. Fournier and Beherec (2006) recognized that the masculinised types of monoecious hemp are less fruiting, and additionally, more

prone to developing fungal infections from *Botrytis cinerea*, which can compromise their potential seed yield. The consideration of sex expression could therefore be valuable when selecting a monoecious cultivar for seed production. However, this finding must be confirmed by observing the sex expression and seed yields at the plant level.

Both stem and seed yields were positively correlated with the duration from sowing to the start of flowering, full flowering and end of flowering. de Meijer and Keizer (1994) observed that the day of anthesis (*i.e.*, start of male flowering) is a useful indicator for stem yield potential. The higher correlation observed in this study between stem yield and the duration from sowing to full female flowering and end of male flowering (Table 4.10) suggests that the observation of developmental stages following anthesis could still improve the assessment of stem yield potential. According to Mediavilla et al. (2001), peaks of stem and fibre yields in dioecious hemp are reached at the time of full female flowering and the end of male flowering, with both stages occurring at the same time. On the other hand, the end of male and female flowering corresponds to a decrease in the formation of new seeds due to the withering of male flowers and scarcity of new pistils, respectively. These observations suggest that the time of end of male flowering could provide an indicator for the evaluation of both potential stem and seed yields in monoecious hemp. Additional field trials including a wider range of sowing dates and cultivars would be necessary to assess the predictive value of such an indicator.

From a practical point of view, producing both stems and seeds implies several antagonisms.

Firstly, van der Werf et al. (1994) observed that stem yields increased while the allocation of dry matter to floral parts was greatly reduced under a 24-hour daylength. For a given cultivar, producing seeds at low latitude and stems at high latitude is therefore recommended (de Meijer and Keizer 1994). The observation of stem and seed yields within the range of reported yields in monoecious hemp indicates that the simultaneous production of stems and seeds is feasible at the latitude of this study. Secondly, no cultivar among those tested provided the highest stem and seed yields simultaneously. Indeed, increasing stem yield in fibre hemp relies on late flowering cultivars (van der Werf et al. 1994; Meijer et al. 1995; Lisson and Mendham 2000) while the production of seed requires early to mid-maturing cultivars (Bocsa and Karus 1998). Thirdly,

harvesting the seed requires the grower to wait for seed maturity, which leads to decreasing stem and fibre yields as a result of senescence (Mediavilla et al. 2001). In addition, our results showed that the calendar date of the end of flowering, and thus the onset of senescence and beginning of seed maturity (Mediavilla et al. 2001), is postponed by a delay in the sowing date. Fourthly, the industrial harvest of seeds requires the loss of the upper part of the stems by mowing, unlike in this study in which the top of the stems was included in the computation of stem yields. However, the top third of the stem does not account for much biomass and could be used for other purposes (Mediavilla et al. 2001).

Considering these four points, harvesting the seeds in addition to the stems can be agriculturally significant, provided that the seed yield is sufficiently high and the stem quality is preserved until harvest. According to our results, the simultaneous production of stems and seeds relies on early sowing dates and on the selection of feminised early or mid-early cultivars, the earliness of which depends on the climatic conditions of the cultivation area. Early sowing resulted in the highest stem and seed yields and the resulting plants senesced earlier in the growing season, thereby allowing an earlier harvest and a lower risk of stem quality loss as a result of less favourable climatic conditions at the end of the crop cycle. Among the cultivars under study, the mid-early cultivar Felina 32 would be preferred to the mid-early Santhica 27 under the climatic conditions of Corroy-le-Grand, given its higher seed yield. The late Epsilon 68 would be discarded as it increases the probability of late harvest. In Michamps, earlier cultivars would be selected because of the higher probability of a delayed sowing date and of less favourable climatic conditions at the end of the crop cycle in terms of temperature and relative humidity. Among the cultivars that were tested, the early Fedora 17 provides higher stem and seed yields than the very early Uso 31, which does, however, offer the highest guarantee of harvest under favourable conditions thanks to its earliness.

4.5. CONCLUSIONS

This paper determines the stem and seed yields in five monoecious hemp cultivars that were sown on distinct dates in relation to their flowering phenology and sex expression with the intention of dual purpose production.

Early sowing dates are associated with a longer duration from sowing to flowering, with higher yields in both the stem and seed and a higher seed harvest index. Stem yields vary among cultivars according to their earliness while seed yields would be affected by sex expression in addition to earliness.

Sex expression in monoecious hemp could be selected because it varies significantly among cultivars. Additionally, sex expression was partly linked to earliness, the earliest cultivar being the most masculinised. More research is necessary to understand the genetic control of sex expression in monoecious hemp.

Our results suggest that the observation of developmental stages following anthesis could still improve the assessment of stem yield potential, while the time of end of male flowering might provide an indicator for the evaluation of both potential stem and seed yields in monoecious hemp. However, further trials are needed to assess the predictive value of such an indicator.

Harvesting seeds in addition to the stems in monoecious hemp may be profitable, provided that the seed yields are sufficiently high and the stem quality is preserved until harvest. According to our results, the simultaneous production of stems and seeds in monoecious hemp requires early sowing and the selection of feminised early or mid-early cultivars, earliness depending on the climatic conditions in the cultivation area. Thus, taking the sex expression into account during the selection of cultivars that are adapted to a dual purpose, in addition to earliness, might be valuable.

Acknowledgments

The authors gratefully thank Prof. Jean-François Ledent and Dr. Nathalie Wuyts for their valuable advice, the staff of the Centre de Michamps and of the Centre A. de Marbaix for their assistance with the trials that were reported in this study, the FNPC (Fédération Nationale des Producteurs de Chanvre, France) for genetic material and advice, and the SMCS (Support in Methodology and Statistical Computation) of the UCL. This research was supported by the FNRS (1.5.279.08).

CHAPTER V

QUANTITATIVE VARIATION OF THE SEX EXPRESSION IN MONOECIOUS HEMP

The present chapter is a modified version of a research paper entitled:

A.-M. Faux [a], A. Berhin [a], N. Dauguet [b], P. Bertin [a] (2014) Sex chromosomes and quantitative sex expression in monoecious hemp (*Cannabis sativa* L.). Euphytica 196:183–197.

[a] Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve

[b] de Duve Institute, Woluwe-Saint-Lambert

This paper was split into two distinct chapters: the assessment of the quantitative variation of the sex expression in monoecious hemp is reported in the present chapter, while the determination of its sex chromosomes is reported in Chapter VI.

ABSTRACT

Hemp (*Cannabis sativa*) has a highly variable sexual phenotype. In dioecious hemp, the sex is controlled by heteromorphic sex chromosomes according to an X-to-autosomes equilibrium. However, in monoecious hemp, the genetic determinism of the sex expression remains widely unknown and has never been related to a quantitative approach of sex expression. The present study aims to contribute to the comprehension of the determinism of the sex expression in monoecious hemp by assessing the genotypic variability of its sex expression.

Five monoecious were grown in controlled conditions under several photoperiods. The monoecy degree of 194 monoecious plants was recorded at each node by a figure ranging from 0 (male flowers only) to 6 (female flowers only). The monoecy degree varied significantly among monoecious cultivars from 3.36 ± 2.28 in 'Uso 31' to 5.70 ± 0.81 in the most feminised 'Epsilon 68'. The variation of monoecy degree among cultivars remained consistent across trials despite a significant 'cultivar x trial' interaction and partly agreed with their earliness.

These results strongly support that the sex expression of monoecious hemp cultivars has a genetic basis.

Keywords: *Cannabis sativa* · Monoecy · Photoperiod · Sex expression.

5.1. INTRODUCTION

In a context of increased attention for alternative crops, the potential of the non-food species hemp (*Cannabis sativa* L.) has been raised in the production of fibre, bio-composites or paper pulp (van der Werf et al. 1996; Struik et al. 2000; Ranalli and Venturi 2004).

Agriculturally, growing hemp is significantly affected by its photoperiodism and reproductive features. Firstly, hemp is a short-day plant with a critical photoperiod of approximately 14 hours (Amaducci et al. 2008a; Borthwick and Scully 1954; Lisson et al. 2000b). By modulating the flowering time, the photoperiodic conditions can have a key influence on the crop yield (van der Werf et al. 1994). Secondly, the species is naturally dioecious and characterized by sexual dimorphism, in plant size and precocity in particular. Monoecious cultivars of hemp have been developed from plants bearing hermaphrodite flowers or bisexual inflorescences (Moliterni et al. 2004). These cultivars display several agronomical advantages in comparison to the dioecious ones, such as higher seed yields, higher crop homogeneity and easier mechanical harvest due to synchronised maturity (Mandolino and Carboni 2004). However, their sexual phenotype is unstable. The monoecious state presents a continuous distribution between the male and female extreme phenotypes (Bocsa and Karus 1998), and the multiplication of monoecious hemp seeds requires a strict elimination of the sporadically occurring dioecious male plants in order to prevent a gradual return to the dioecy. In addition, the sexual phenotype of hemp is affected by external factors, such as hormonal treatments, the photoperiod or nitrogen status (Freeman et al. 1980a). In spite of the diversity and plasticity of the intersexual forms of hemp, significant variations of the sex expression were observed among monoecious hemp cultivars in field trials, and higher seed yields were obtained with the early and mid-early feminised cultivars, suggesting that the sex expression of monoecious plants could affect the seed yields (Faux et al. 2013). Investigating the potential genetic basis of the sex expression in hemp appears therefore valuable for the improvement and cultivation of monoecious hemp.

The present study aims to contribute to the comprehension of the genetic determinism of the sex expression in monoecious hemp by assessing the

genotypic variability of the sexual phenotype expressed as a quantitative variable. In contrast to previous field trials (Faux et al. 2013), the characterization of the sexual phenotype was performed under several photoperiodic controlled conditions, while the sex expression of the monoecious hemp plants was recorded at the plant-node level.

5.2. MATERIALS AND METHODS

5.2.1. Genetic material

Six hemp cultivars were used. Five of them were monoecious and covered a wide range of earliness: 'Uso 31' (very early), 'Fedora 17' (early), 'Santhica 27' and 'Felina 32' (mid-early), and 'Epsilon 68' (late). The sixth one, 'Carmagnola', was dioecious and cultivated for the purpose of cytometric and molecular analyses (Chapter VI). All five monoecious cultivars were obtained from the Fédération nationale des Producteurs de Chanvre (FNPC), Le Mans (France), and the dioecious one was obtained from Assocanapa, Carmagnola (Italy).

5.2.2. Growth conditions

Three trials were conducted successively, the first two in a greenhouse and the third one in a phytotron to apply short photoperiods (Table 5.1). Sowing was performed on 13 September 2011, 2 February 2012 and 18 may 2012 in each trial, respectively. The trial 1 included 20 plants from each of the five monoecious hemp cultivars with the exception of 'Santhica 27' with 15 plants. Both trials 2 and 3 included 10 plants per monoecious cultivar with the exception of 'Epsilon 68' in trial 3 in which a male plant was found. In total, 95, 50 and 49 plants of monoecious hemp were grown in trials 1, 2 and 3, respectively. In addition, trial 3 included 10 plants of the dioecious cultivar 'Carmagnola'.

The set-point temperatures were 25/20°C day/night in all of the trials. The photoperiod was firstly 16/8 h during 22, 60 and 20 days and then shortened

to 14/10, the natural daylength and 12/12 h in order to promote flowering in trials 1, 2 and 3, respectively. In trial 2, the natural daylength increased from 13 h at the time of photoperiod shortening (2 April) to 14.5 h at the end of the trial (27 April). In the greenhouse, the plants received the natural daylight, and the daylength was extended by Philips HPLR lamps (400 W). The daylight intensity decreased from on average 580 to 170 $\mu mol.m^{-2}.s^{-1}$ between September and November (trial 1) and, conversely, increased from on average 250 to 520 $\mu mol.m^{-2}.s^{-1}$ between February and April (trial 2). The light intensity used to extend the natural daylength was approximately 40 $\mu mol.m^{-2}.s^{-1}$ in both trials 1 and 2. In the phytotron, lighting was performed by Philips HPI-T plus lamps (400 W, ~ 135 $\mu mol.m^{-2}.s^{-1}$). There were 4 or 5 plants per lamp in both greenhouse and phytotron.

For each trial, the sowing was performed in dimpled germination plates. After 10 days, the seedlings were transplanted into 30 cm (height) by 18 cm (diameter) pots. The substrate consisted of a 3:1:1 mixture of soil, sand and loam with 2.8 g per pot of Osmocote® (14:13:13 NPK with slow release of nutrients for up to six months) added just before transplanting. The plants were watered by capillarity twice per week.

Table 5.1 The experimental protocol applied for the three trials: genetic material and environmental conditions

	Trial 1	Trial 2	Trial 3
Genetic material (number of plants)			
Monoecious cultivars			
'Uso 31'	20	10	10
'Fedora 17'	20	10	10
'Santhica 27'	15	10	10
'Felina 32'	20	10	10
'Epsilon 68'	20	10	9
Total	95	50	49
Dioecious cultivar			
'Carmagnola'	-	-	10
Type of environment	greenhouse	greenhouse	phytotron
Sowing date	Sept. 13, 2011	Feb. 2, 2012	May 18, 2012
Trial duration (days)	63	85	55
Temperature	25/20°C day/night	25/20°C day/night	25/20°C day/night
Photoperiod (PP)			
PP1 (day/night hours)	16/8	16/8	16/8
PP2 (day/night hours)	14/10	increasing from 13/11 to 14.5/9.5	12/12
Duration of PP1 treatment (days)	22	60	20
Light			
Type	natural daylight + Philips HPLR lamps 400 W (daylength extension)		Philips HPI-T plus (400 W)
Intensity (μmol m^{-2} s^{-1})	170 - 580 [1] (natural daylight), 40 (daylength extension)	250 - 520 [1] (natural daylight), 40 (daylength extension)	~ 135
Quality (emission spectrum)	continuous emission between ~ 400 and 700 nm (natural daylight), peaks of emission at 360, 420, 540 and 590 nm (daylength extension)		peaks of emission at 460, 540 and 600 nm

[1] Range of mean monthly light intensity during the trial period.

5.2.3. Plant phenotyping

The sex of the dioecious hemp plants was noted. In the monoecious hemp, the time of flowering, flowering duration and sex expression were recorded. The flowering time was defined as the time at which closed male flowers or white styles were easily visible at leaf axils. The sex expression was characterized by the degree of monoecy modified from the scale of Sengbusch (1952) by varying from 0 to 6 according to the ratio between the female and male flowers (Table 5.2). The monoecy degree was recorded at each flowering node independently once a week during the flowering duration, *i.e.*, from the first to last flower appearance. However, the plants were checked until the end of flowering, *i.e.*, when male flowers were withering and green seeds were present at most nodes. The experiments stopped when this developmental stage was reached, *i.e.*, 63, 85 and 55 days after sowing in trials 1, 2 and 3, respectively. All of the monoecious plants were phenotyped, *i.e.*, 194 plants.

Table 5.2 Sex-expression scale in monoecious hemp modified from Sengbusch (1952)

Monoecy degree	Sex ratio
0	100% of male flowers
1	80 – 99% of male flowers (strongly masculinised node)
2	60 – 80% of male flowers (masculinised node)
3	40 – 60% of female and male flowers
4	60 – 80% of female flowers (feminised node)
5	80 – 99% of female flowers (strongly feminised node)
6	100% of female flowers

5.2.4. Statistical analyses

The statistical analyses were carried out to assess the effects of the cultivar and trial on the flowering time, flowering duration and sex expression in monoecious hemp.

The effect of the experimental factors on the flowering time and flowering duration was assessed by using the following model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk} \quad (5.1)$$

where y_{ijk} was the flowering time or flowering duration of the *k*th hemp plant of cultivar *i* in trial *j*; μ was the general mean response; α_i, β_j and $(\alpha\beta)_{ij}$ were the fixed effects of cultivar *i*, trial *j* and their interaction, respectively; e_{ijk} was the error associated to the *k*th plant of cultivar *i* in trial *j*. The model (5.1) was tested with the GLM procedure from the SAS statistical package (SAS Institute Inc. 2012).

The sex expression, quantified by the degree of monoecy (Table 5.2), is a discrete variable with seven response levels and was therefore statistically analysed using the GLIMMIX procedure with a multinomial response distribution (DIST=MULTI) and cumulative logit link function (LINK=CUMLOGIT) (SAS Institute Inc. 2012). The analysis was performed with the model (5.2) below, which integrates random effects of the 'plant' and 'node' nested to the plant in addition to the 'cultivar' and 'trial' fixed effects:

$$y_{ijknp} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + d_{ijk} + f_{ijkn} + e_{ijknp} \quad (5.2)$$

where y_{ijknp} was the monoecy degree of the *n*th node of the *k*th hemp plant of cultivar *i* in trial *j* observed at time *p*; μ, α_i, β_j and $(\alpha\beta)_{ij}$ had the same meaning as in model (5.1); d_{ijk} was the random effect associated to the *k*th plant of cultivar *i* in trial *j*; f_{ijkn} was the random effect associated to the *n*th node of the *k*th plant of cultivar *i* in trial *j*; e_{ijknp} was the error associated to the *n*th node of the *k*th plant of cultivar *i* in trial *j* observed at time *p*. The option MAXOPT=100 was used to obtain the convergence of the model.

For each observed variable – flowering time, flowering duration or monoecy –, the MEANS procedure and the CONTRAST statement of the GLM or GLIMMIX procedures were used to compute means and pairwise comparisons, respectively.

5.3. RESULTS

5.3.1. Flowering phenology and records of the sex expression

The plants started to flower 34, 57 and 35 days after sowing in trials 1, 2 and 3, respectively (as a reminder, the photoperiod was shortened 22, 60 and 20 days after sowing in trials 1, 2 and 3, respectively). The appearance of new flowers was noted until 55, 77 and 49 days after sowing in each trial, respectively. The mean flowering time was 36.1, 63.0 and 37.3 days, and the mean flowering duration was 18.6, 10.6 and 7.7 days in trials 1, 2 and 3, respectively (Fig. 5.1a, b).

The differences in flowering time among trials were significant only between trial 2 and both trials 1 and 3 (P <0.001; Table 5.3). The cultivar significantly affected the flowering time in trial 2 only (P <0.001). However, the ranking of cultivars according to their flowering time was similar across trials, resulting in no significant 'cultivar x trial' interaction and an overall significant 'cultivar' effect (P <0.001). The relative flowering times of the five monoecious cultivars were in agreement with their earliness as stated by the FNPC: 'Uso 31' was the earliest one, followed by 'Fedora 17', 'Felina 32', 'Santhica 27' and 'Epsilon 68', the latest one.

The differences in flowering duration were significant among all three trials (P <0.001; Table 5.3). Similarly to the flowering time, the flowering duration was affected by the cultivar in trial 2 only (P <0.05), and no significant 'cultivar x trial' interaction was found. The flowering duration was on average longer in 'Fedora 17' and 'Uso 31', the earliest cultivars, than in 'Santhica 27' and 'Epsilon 68', the latest ones, and intermediate in 'Felina 32' over the three trials.

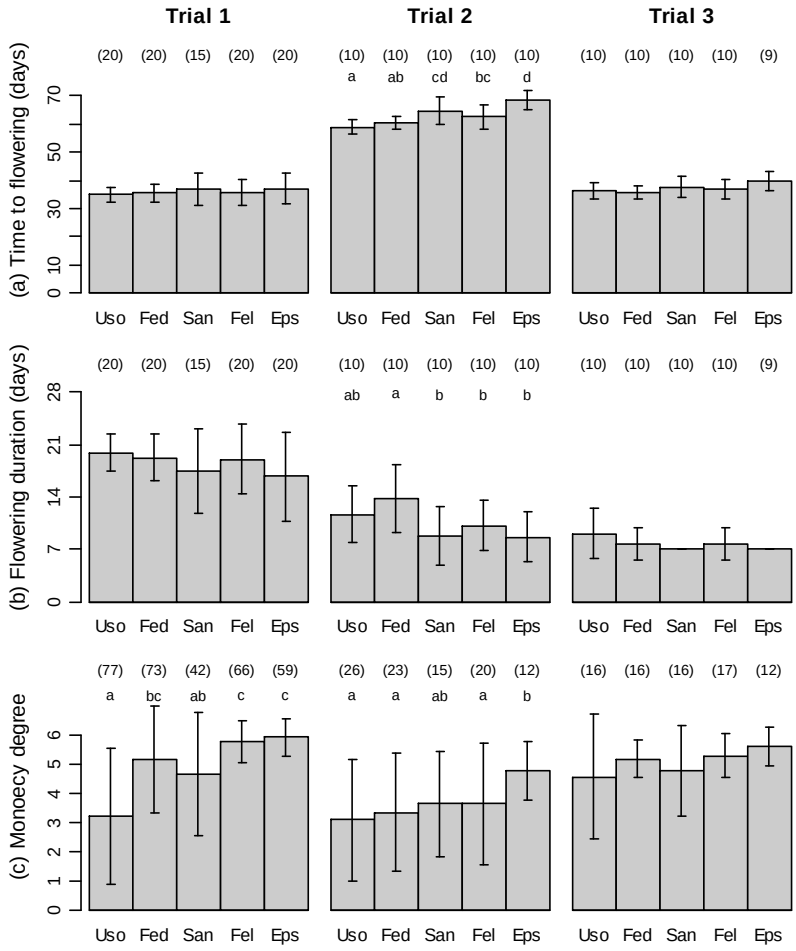

Fig. 5.1 Flowering phenology and sex expression in five monoecious hemp cultivars. (a) Time to flowering, (b) flowering duration and (c) degree of monoecy (Table 5.2) (mean ± sd). The photoperiod was 16/8 h during 22, 60 and 20 days and then 14/10, the natural daylength and 12/12 h in trials 1, 2 and 3, respectively. In trial 2, the natural daylength increased from 13 h at the time of photoperiod shortening (2 April) to 14.5 h at the end of the trial (27 April). 'Uso' to 'Eps' refer to the cultivars 'Uso 31', 'Fedora 17', 'Santhica 27', 'Felina 32' and 'Epsilon 68', respectively. Significant differences among cultivars (P <0.05) were indicated by different letters above the bars when the 'cultivar' effect was significant (P <0.05). The number of data points is given between brackets for each 'cultivar x trial' treatment. For the monoecy degree, each data point corresponds to the mean value computed on all flowering nodes of a given plant at a given observation time.

Table 5.3 Flowering phenology: ANOVA table for the time of flowering and flowering duration

Dependent	Source	DF	SS	MS	FValue	ProbF
Time of flowering	Cultivar	4	544.1	136	9.1	<0.001
	Trial	2	25517.4	12758.7	853.6	<0.001
	Cultivar x trial	8	232.7	29.1	1.9	0.056
Flowering duration	Cultivar	4	239.8	60	4.2	0.003
	Trial	2	4184.7	2092.3	145.2	<0.001
	Cultivar x trial	8	77.4	9.7	0.7	0.716

As a result of the flowering phenology, the sex expression was recorded four times in trial 1 at 7 days intervals (34, 41, 48 and 55 days after sowing), three times in trial 2 (57, 62 and 70 days after sowing) and only twice in trial 3 (38 and 45 days after sowing). From the first to the last observation time, the number of phenotyped nodes per plant varied from 1 to 14 in trial 1 (mean ± sd = 5.8 ± 2.9), from 1 to 16 in trial 2 (8.4 ± 3.6) and from 1 to 12 in trial 3 (5.7 ± 2.7).

5.3.2. Sex expression

The dioecious cultivar ‘Carmagnola’, grown in trial 3, included six female plants, characterized by racemes with leafy bracts and only female flowers, and four male plants, characterized by hanging panicles with few or no leaves and only male flowers (Fig. 5.2a, b). All of the plants from the five monoecious hemp cultivars showed inflorescences similar to those of female plants from dioecious hemp with both female and male flowers arising in variable amounts (Fig. 5.2c).

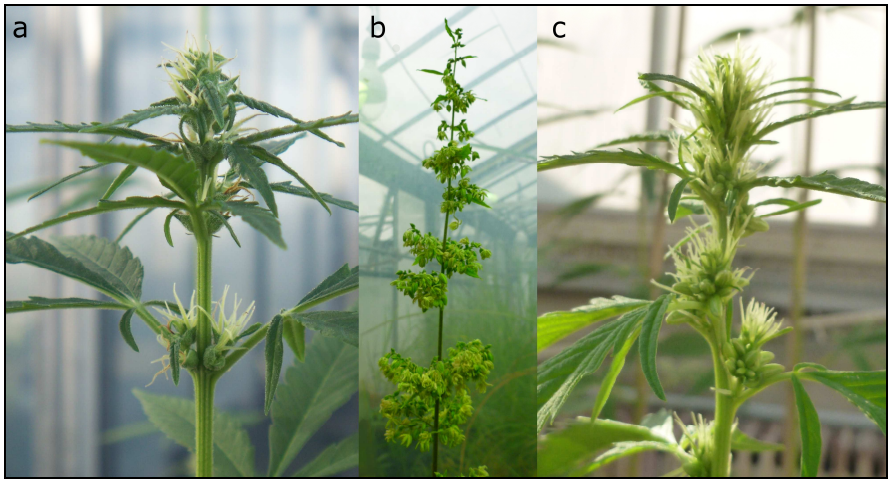

Fig. 5.2 Inflorescences found in dioecious and monoecious hemp: (a) female and (b) male plants from the dioecious cultivar 'Carmagnola' and (c) a plant from the monoecious cultivar 'Uso 31'.

The sex expression of the monoecious hemp plants was affected by the cultivar and trial (P <0.001) as well as by their interaction (P <0.01) (Table 5.4). The plants were more masculinised in trial 2 (mean monoecy degree ± sd = 3.55 ± 1.95), *i.e.*, in the trial in which the duration of the long-photoperiod treatment was longer, than in both trials 1 (4.88 ± 1.98) and 3 (5.05 ± 1.34) (P <0.001; Fig. 5.1c). However, the monoecy degree varied significantly among cultivars only in trials 1 (P <0.001) and 2 (P <0.05) and among trials only in three cultivars, 'Fedora 17' (P <0.01) and both 'Felina 32' and 'Epsilon 68' (P <0.001), resulting in a significant 'cultivar x trial' interaction. Nevertheless, the ranking of the cultivars according to their monoecy degree remained consistent across the trials (Fig. 5.1c). The cultivar 'Uso 31' was the most masculinised one (mean monoecy degree ± sd = 3.36 ± 2.28). It was followed by 'Santhica 27' (4.47 ± 1.97), 'Fedora 17' (4.78 ± 1.91), 'Felina 32' (5.28 ± 1.39) and 'Epsilon 68' (5.70 ± 0.81), the most feminised cultivar. Thus, the rankings of cultivars according to earliness and sex expression agreed with each other, with the exception of 'Fedora 17' and 'Santhica 27', the former being earlier and more feminised than the latter. The differences in monoecy degree between the cultivars were highly significant (P <0.001) between 'Uso 31' and both 'Felina 32' and 'Epsilon 68' and between 'Epsilon 68' and both 'Fedora 17' and 'Santhica 27'. The differences were also significant (P <0.05) between 'Uso 31'

and both ‘Santhica 27’ and ‘Fedora 17’ as well as between ‘Santhica 27’ and ‘Felina 32’.

Table 5.4 Sex expression in monoecious hemp: test of fixed effects on the monoecy degree (Table 5.2)

Effect	NumDF	DenDF	FValue	ProbF
Cultivar	4	159.99	8.9	<0.001
Trial	2	170.51	15.05	<0.001
Cultivar x trial	8	157.79	2.66	0.009

5.4. DISCUSSION

5.4.1. Flowering phenology

The variation of flowering time among trials was in agreement with the short-day photoperiodic reaction of hemp: the plants started to flower earlier in the trials in which the photoperiod was shortened earlier (trials 1 and 3). The late flowering observed in trial 2 should be due to a longer photoperiod-induced phase (PIP) of development, the duration of which is determined by the number of hours exceeding the critical photoperiod [approximately 14 h in hemp (Amaducci et al. 2008a; Borthwick and Scully 1954; Lisson et al. 2000b)] and the genotypic sensitivity to the photoperiod (Lisson et al. 2000b). Thus, the genes conferring the photoperiod sensitivity were likely further stimulated by the long 16-h day treatment of trial 2, so that the flowering was delayed and the variation of flowering time among cultivars was more significant in this trial.

The differences in flowering duration observed among trials (on average 18.6, 10.6 and 7.7 days, respectively) could be due to the photoperiod experienced by the plants after the flowering time, which was 14 and 12 h in trials 1 and 3, respectively, and increased from 13 to 14.5 h during the flowering period in trial 2 (between 60 to 85 days after sowing). In field trials, Amaducci et al. (2008b) observed that the flowering duration of hemp plants was shorter in

late sown treatments, which flowered under decreasing daylengths. However, a relatively high difference in flowering duration was observed between trials 1 and 2, which experienced relatively close photoperiods. It is possible that the flowering duration was affected by the time of photoperiod shortening – which occurred 14 days before and close to the flowering time in trials 1 and 2, respectively –, or that the response of the flowering duration to the photoperiod is not linear. In soybean, a short-day species, the flowering duration is affected by photoperiods above a critical threshold, which varies according to the genotype (Summerfield et al. 1998). In addition, the flowering duration was longer in the earliest cultivars ('Uso 31', 'Fedora 17') than in the latest ones ('Santhica 27', 'Epsilon 68'). Amaducci et al. (2008b) observed a large genotypic variation in the flowering duration among hemp cultivars; however, no link between earliness and flowering duration was drawn. Additional trials in controlled conditions are needed to explain the response of the flowering duration to the photoperiod and genotype.

5.4.2. The sex expression in response to the genotype

Both previous field trials (Faux et al. 2013) and the present study revealed (i) a significant effect of the cultivar on the sex expression of monoecious hemp plants and (ii), despite a significant 'cultivar x trial' interaction in the present study, the same ranking of cultivars according to their sex expression. The consistency of the genotypic effect on the sex expression across environments supports that the sex expression in monoecious hemp has a genetic basis.

The rankings of cultivars according to their sex expression and earliness partly agreed with each other. In dioecious hemp, the association of earliness and maleness is known: the male plants flower generally earlier than the female ones (Bocsa and Karus 1998; Struik et al. 2000). Borthwick and Scully (1954) attributed the excess of male over female plants observed in long photoperiods to the lower sensitivity to the photoperiod of the male plants. In monoecious hemp, Amaducci et al. (2008b) suggested that early plants have male characteristics. These observations support the hypothesis that both traits share a common basis in their genetic determinism. Effects of exogenous hormones on sex have been reported in hemp: gibberellin induces predominantly male plants,

while auxin, cytokinin and abscisic acid induce the feminisation of the plants (Heslop-Harrison 1956; Mohan Ram and Jaiswal 1972; Chailakyan and Khryanin 1978, 1979; Freeman et al. 1980). In addition, Chailakyan and Khryanin (1978, 1979) observed that plants treated with gibberellin produced floral buds earlier than the controls, whereas the auxin treatment delayed the flowering. These observations suggest that the gibberellin would be associated with more masculinised phenotypes and earlier flowering, while, conversely, the auxin would be associated with more feminised phenotypes and later flowering. Therefore, investigating the role of gibberellin and auxin in hemp flowering might be an interesting pathway to dissect the relation between earliness and sex expression. Besides, in the long-day model plant *Arabidopsis thaliana*, gibberellin is required for flowering under non-inductive short days (Wilson et al. 1992). Although the comparison is highly speculative, the requirement of gibberellin in hemp for flowering under non-inductive conditions, *i.e.*, long-day conditions, could provide an explanation to the masculinising effect associated with long photoperiods. Therefore, it could be worthwhile to integrate the photoperiod in the study of the hormonal regulation of flowering and sex expression in hemp.

5.4.3. The sex expression in response to the environment

The pattern of variation of the sex expression among trials suggested that the 'trial' effect on the sex expression was primarily due to the photoperiod (Fig. 5.1c). Indeed, a masculinising effect of long photoperiods has been observed in hemp (Borthwick and Scully 1954; Arnoux 1966b; Arnoux and Mathieu 1969; Freeman et al. 1980). In the present study, the duration of the 16-h day treatment was the longest in trial 2 (60 days *vs.* 22 and 20 days in trials 1 and 3, respectively), resulting in a higher production of male flowers in this trial.

However, the significant 'cultivar x trial' interaction pointed out, firstly, that the 'cultivar' effect on the monoecy degree decreased from trial 1 to trial 3 (Fig. 5.1c; Table 5.4). This observation could be explained by the flowering duration or by both the photoperiodic and light conditions. Indeed, the 'cultivar' effect on the monoecy degree varied among trials similarly to the flowering duration, suggesting that a long flowering duration would allow a higher differentiation of the sex expression among the monoecious hemp cultivars. In

addition, according to Borthwick and Scully (1954), a high light intensity induces a greater production of male flowers on female plants. The light conditions of trials 1 and 2 – performed in the greenhouse under natural light artificially extended – would therefore be more favourable to the production of male flowers than those of trial 3 – performed in phytotron under artificial light only. However, the photoperiodic conditions were less masculinising in trials 1 and 3 than in trial 2, as a result of their shorter duration of 16-h day treatment (Freeman et al. 1980). Therefore, the light and photoperiodic conditions would have opposite effects on the sex expression in trial 1. It would be possible that the cultivars were affected differentially by these conditions, resulting in a relatively high genotypic variability of the sex expression in this trial. On the opposite, the light and photoperiodic conditions would be both inductive for the production of male flowers in trial 2, and, conversely, both relatively little inductive in trial 3, inducing thereby less variable genotypic responses. Further studies are clearly necessary to elucidate how the genotypic response of the sex expression is affected by the environment.

Secondly, the sex expression varied significantly among trials only in the most feminised cultivars (‘Fedora 17’, ‘Felina 32’ and ‘Epsilon 68’). On the opposite, in field trials performed with the same five present cultivars (Faux et al. 2013), the variation of sex expression with sowing date was the highest in the most masculinised cultivars (‘Uso 31’ and ‘Santhica 27’). In view of the effect of the light intensity on the sex expression in hemp (Borthwick and Scully 1954), the inconsistency observed between field and controlled conditions could be due to the light quality and intensity. In the field, the plants grew under natural daylight only, while they received both natural and artificial light in the greenhouse and artificial light only in the phytotron. In addition, the daylight intensity is higher during the field growth season of hemp (from April to September) than during the periods of the present greenhouse trials (from September to April). According to Borthwick and Scully (1954), the cultural conditions could also affect the sex expression of the plants, since differences in the production of male flowers by female hemp plants were observed between field and greenhouse experiments conducted simultaneously under natural daylight and photoperiod.

The observation of a higher genotypic variability for the sex expression in trial 1 suggests that the photoperiodic conditions of this trial – 16-h day treatment during 22 days followed by a 14-h day treatment – would be the most

favourable ones among those tested here to express the genetic potential of monoecious hemp cultivars with respect to their sexual phenotype and investigate the genetic determinism of the trait. However, the dissection of the genetic basis of the observed 'genotype x environment' interactions would necessarily require the test of distinct environmental conditions.

5.5. CONCLUSIONS

The present study provided new insights for further studies on the genetic determinism of the sex expression in monoecious hemp. Firstly, despite the observation of 'genotype x environment' interactions, the ranking of cultivars according to their sex expression was maintained across environments. This result supports that, although highly sensitive to the environment, the sex expression in monoecious hemp has a genetic basis and therefore should have a non-null broad-sense heritability.

Secondly, consistent variations of sex expression and earliness among cultivars were found, suggesting that genes responsible for the earliness are involved in the determinism of the sex expression in monoecious hemp.

Given that the sex expression in monoecious hemp varies quantitatively and appears to rely on a genetic basis, investigating its genetic determinism through the identification of quantitative trait loci (QTL) seems relevant. To this purpose, the characterization of distinct monoecious hemp cultivars made in the present study can be useful to select contrasting parents for the creation of a segregating population.

Acknowledgements

We thank Prof. Bernadette Govaerts and the platform Statistical Methodology and Computing Support (UCL, Belgium) for their valuable advice on statistics and the Fédération Nationale des Producteurs de Chanvre (France) for the genetic material. This work was supported by the Fonds de la Recherche Scientifique [grant number 1.5.279.08].

CHAPTER VI

SEX CHROMOSOMES OF MONOECIOUS HEMP

The present chapter is a modified version of a research paper entitled:

A.-M. Faux [a], A. Berhin [a], N. Dauguet [b], P. Bertin [a] (2014) Sex chromosomes and quantitative sex expression in monoecious hemp (*Cannabis sativa* L.). Euphytica 196:183–197.

[a] Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve

[b] de Duve Institute, Woluwe-Saint-Lambert

This paper was split into two distinct chapters: the assessment of the quantitative variation of the sex expression in monoecious hemp was reported in Chapter V, while the determination of its sex chromosomes is reported in the present chapter.

ABSTRACT

Hemp (*Cannabis sativa*) has a highly variable sexual phenotype. In dioecious hemp, the sex is controlled by heteromorphic sex chromosomes according to an X-to-autosomes equilibrium. However, in monoecious hemp, the genetic determinism of the sex expression remains widely unknown. The present study aims to contribute to the comprehension of the determinism of the sex expression in monoecious hemp by establishing its sex chromosomes.

Five monoecious and one dioecious cultivars were grown in controlled conditions under several photoperiods. The genome size of 55 plants was determined by flow cytometry, and the DNA of 115 monoecious plants was screened with the male-associated marker MADC2. The genome size of monoecious plants (1.791 ± 0.017 pg) was not different from that of females (1.789 ± 0.019 pg) but significantly lower than that of males (1.835 ± 0.019 pg). MADC2 was absent from all monoecious plants.

These results demonstrate that cultivars of monoecious hemp have the XX constitution for the sex chromosomes.

Keywords: *Cannabis sativa* · Genome size · Monoecy · Sex chromosomes.

6.1. INTRODUCTION

Hemp (*Cannabis sativa* L.) is naturally dioecious and characterized by sexual dimorphism, in plant size and precocity in particular. The species is diploid (2n = 20) and includes sex chromosomes (Hirata 1924). The chromosomes XX are found in female plants, and XY are present in male plants, with the Y chromosome larger than the X one and larger than the autosomes (Yamada 1943 cited by Sakamoto et al. 1995). The genome sizes of diploid female and male plants have been estimated using flow cytometry as 1636 ± 7.2 and 1683 ± 13.9 Mbp, respectively, and the size difference between them (47 Mbp, *i.e.*, 2.8% of the genome size of male plants) has been attributed to the large long arm of the Y chromosome (Sakamoto et al. 1998). The sex determinism in dioecious hemp would be based on an X-to-autosomes equilibrium and not on a Y-active mechanism (Westergaard 1958; Ainsworth 2000).

In addition to dioecious hemp cultivars, monoecious cultivars have been developed. However, the determinism of monoecious forms of hemp remains widely unknown. Although Hoffman (1961 cited by Truta et al. 2007) assumed the existence of XX, XY and YY forms, Menzel (1964) observed presumably XX chromosomes in monoecious hemp plants, and recent studies did not reveal the presence of male-associated DNA markers (MADC, for male-associated DNA in *Cannabis*). However, the karyotype was limited to the cultivar 'Kentucky' (Menzel 1964) while very few monoecious plants have been screened with MADC markers to date, *i.e.*, 20 plants of cultivars 'Bialobrzeskie' and 'Beniko' (Mandolino et al. 1999) and 6 plants of cultivars 'Fibrimon' and 'Fleischmann' (Torjek et al. 2002). Mandolino et al. (2002) stated that there is no specific report on the chromosome set in monoecious hemp.

Flow cytometry constitutes an indirect but advantageous technique to show dissimilarities in chromosomal composition. Indeed, flow cytometry is able to analyse quickly large populations of cells (Dolezel and Bartos 2005) and highlight very small genome-size differences (Costich et al. 1991; Vagera et al. 1994; Dolezel and Göhde 1995). Given the larger size of the Y chromosome of hemp (Sakamoto et al. 1998), the presence, in monoecious hemp, of a Y chromosome similar to that found in male plants could be inferred from the

comparison of the genome sizes of monoecious, female and male plants. In addition, several DNA markers closely linked to the male phenotype have been developed in hemp (Sakamoto et al. 1995; Mandolino et al. 1999; Torjek et al. 2002; Peil et al. 2003; Sakamoto et al. 2005; Rode et al. 2005). Among them, the MADC2 marker proved to be completely associated to the male phenotype and should therefore be located on the Y chromosome in a region excluded from recombination during meiosis (Mandolino et al. 2002). The amplification of this marker by monoecious plants would indicate the existence of a male-associated fragment in the sex chromosomes of monoecious hemp.

The present study aims to establish the constitution in sex chromosomes of monoecious hemp through the evaluation of a large number of plants from distinct cultivars. Flow cytometric analyses and the MADC2 DNA marker (Mandolino et al. 1999) were used to address the present objective.

6.2. MATERIALS AND METHODS

6.2.1. Genetic material

Six hemp cultivars were used, one dioecious, 'Carmagnola', and five monoecious: 'Uso 31', 'Fedora 17', 'Santhica 27', 'Felina 32' and 'Epsilon 68'. The dioecious cultivar was obtained from Assocanapa (Italy), and the monoecious ones were obtained from the Fédération nationale des Producteurs de Chanvre (FNPC, France).

6.2.2. Growth conditions

Three trials were conducted successively, the first two in a greenhouse and the third one in a phytotron to apply short photoperiods. The trial 1 included 20 plants from each of the five monoecious hemp cultivars, and both trials 2 and 3 included 10 plants per monoecious cultivar, with the exception of 'Santhica 27' in trial 1 with 15 plants and 'Epsilon 68' in trial 3 with 9 plants. In total, 95, 50 and 49 plants of monoecious hemp were grown in trials 1, 2 and 3, respectively. In addition, trial 3 included 10 plants of the dioecious cultivar

'Carmagnola'. The growth conditions are further detailed in Chapter V (section 5.2.2, Table 5.1).

6.2.3. Estimation of the genome sizes using flow cytometry

The nuclear genome size of monoecious and dioecious hemp plants was measured using flow cytometry, simultaneously with *Glycine max* (soybean) and *Zea mays* (maize) as internal reference standards to compare the peaks from distinct samples of hemp nuclei. Both reference standards were chosen from those recommended by Praça-Fontes et al. (2011) for their genome size, which is *a priori* relatively close and distant from that of hemp for soybean and maize, respectively (Sakamoto et al. 1998). Otto's buffers (Otto 1992; Dolezel and Gohde 1995) were used for nuclei isolation as recommended by Loureiro et al. (2006) for plant species with low nuclear DNA content (between 1.30 and 1.96 pg per 2C).

Fresh leaf samples of hemp (20 mg), soybean (20 mg) and maize (50 mg) were collected from mature leaves, mixed and chopped for 120 s with a razor blade in a plastic Petri dish on ice containing 1 ml of the Otto I buffer used for nuclei isolation. A larger proportion of maize leaf tissue was needed to obtain sufficient nuclei. The resulting suspension rested for 15 min with shaking every 5 min to facilitate the nuclei extraction. The mixture was thereafter filtered twice through 30 µm (Sakamoto et al. 1998) and 10 µm nylon meshes successively (Partec GmbH, Münster, Germany). A volume of 200 and 500 µl of Otto I buffer was poured on the first and second filter, respectively, to recover as many nuclei as possible. The solution was then centrifuged for nuclei precipitation (2,500 rpm, 5 min). The resulting pellet was diluted in 100 µl of Otto I buffer before being filtered through a 10 µm mesh. A volume of 800 µl of modified Otto II buffer (20 ml of 0.4 M $Na_2HPO_4.12H_2O$, 1 ml of PI, 1 ml of RNAse and 40 µl of β-mercaptoethanol), used as staining buffer, was added upon the filter. After 10 min, the solution was analysed with the LSRFortessa cell analyser (Becton Dickinson Benelux N.V., Erembodegem) from the de Duve Institute (Brussels, Belgium). For each sample, a minimum of 10,000 events were analysed at a maximum rate of 300 events s^{-1}. The fluorescence was measured using the 575/26 BP filter, specific to the wavelengths emitted by propidium iodide. The median, standard deviation and coefficient of variation

(CV) of the peaks were extracted using the BD FACsDiva Software. For each sample, the genome size of the hemp nuclei was estimated by dividing the median of fluorescence of the hemp peak by that of each standard peak multiplied by the genome size of the standard. Genome sizes of 2.41 ± 0.170 and 5.57 ± 0.146 pg per 2C DNA for soybean and maize, respectively, were used to calibrate the hemp peaks (Praça-Fontes et al. 2011).

The protocol (nuclei isolation and flow cytometry) was developed on plants from trials 1 and 2, while the presented results were obtained from 55 plants from trial 3: 10 'Carmagnola' (6 female and 4 male), 9 'Uso 31', 9 'Fedora 17', 8 'Santhica 27', 10 'Felina 32' and 9 'Epsilon 68'. The cytometric analyses were spread out among six dates as a result of logistical constraints linked to the use of the flow cytometer. For each plant on each date of analysis, two distinct leaf samples were probed independently at the flow cytometer.

6.2.4. PCR with male-associated DNA marker

Leaf DNA was extracted according to Murray and Thompson (1980) with modifications as detailed below. Frozen leaves were ground to a fine powder in liquid nitrogen. The resultant powder was mixed with 900 µl of extraction buffer (10% Tris-HCl, 10% EDTA, 2% CTAB, 50% NaCl, 2% PVPP and 1% β-mercaptoethanol) and heated at 55°C for 1 h with upside-down shaking every 10 min. A volume of 500 µl of a 24:1 mixture of chloroform-isoamyl alcohol (v/v) was added, shaken for 10 s and centrifuged (13,000 rpm for 10 min). The upper phase was collected, and the extraction with chloroform-isoamyl alcohol was repeated once. The final upper phase was collected, mixed with 350 µl of isopropanol at 4°C, shaken for 10 sec, rested for 30 min and centrifuged (13,000 rpm for 10 min). The supernatant was drained, and the resultant DNA pellet was mixed with 500 µl of 70% ethanol at -20°C and centrifuged (13,000 rpm for 5 min). This last step was repeated once before dissolving the pellet in 50 µl TE buffer (10 mM Tris-HCl pH8, 1 mM EDTA pH8). The extracted DNA was kept at -20°C.

The presence of the male-associated DNA marker MADC2 (Mandolino et al. 1999) was tested. The reaction volume was 12.5 µl and contained 6.25 ng of genomic DNA, 0.4 µM of the MADC2 primer (Eurogentec SA, Seraing, Belgium), 80 µM of each dNTP, 1.6 mM of $MgCl_2$, 1.25 U of GoTAQ and 1X

of GoTAQ buffer (Promega Corp, Madison, USA). The PCR amplifications were performed with a PTC 100 thermal cycler (MJ Research, Waltham, Mass): 10 min at 95°C, 40 cycles [30 s at 95°C, 45 s at 58°C and 2 min at 72°C] and 4 min at 72°C. The amplification products and a DNA ladder (Smartladder, Eurogentec SA, Seraing, Belgium) were separated on 1% agarose gel run at a constant voltage of 90 V. The gels were stained with ethidium bromide and displayed under UV light.

The MADC2 primer was preliminarily tested on ten plants of the dioecious cultivar 'Carmagnola' (five of each sex) and then applied on 115 plants of monoecious hemp (from 21 to 26 plants per cultivar).

6.2.5. Statistical analyses

The statistical analyses were carried out to assess the effects of the sex (monoecious, female or male), cultivar and date of analysis on the genome size of hemp nuclei. The statistical analysis was firstly simplified by incorporating the 'sex' effect of the dioecious cultivar into the 'cultivar' effect, which consequently included seven levels (one for each of the five monoecious cultivars and two for each sex of the dioecious cultivar), and secondly by using the mean genome size computed for each plant on each analysis date. The effect of the experimental factors on the genome size was tested in the GLM procedure (SAS Institute Inc. 2012) by using the following model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk} \quad (6.1)$$

where y_{ijk} was the genome size of the kth hemp plant of cultivar or sex i analysed at the jth date, α_i, β_j and $(\alpha\beta)_{ij}$ were the fixed effects of cultivar or sex i, date j and their interaction, respectively, and e_{ijk} was the error associated to the kth plant of cultivar or sex i analysed at the jth date. An entire mixed model, which included the fixed effects of 'sex', 'cultivar', 'date' and their interactions as well as the random 'plant' effect and its interaction with 'date', was firstly tested, and we verified that the interpretation of the fixed effects was the same as in model (6.1). The MEANS procedure and the CONTRAST statement of the GLM procedure were used to compute means and pairwise comparisons, respectively.

6.3. RESULTS

6.3.1. Genome size

The analysis of each sample by flow cytometry allowed the detection of three peaks that corresponded to hemp, soybean and maize nuclei successively. The CVs of the peaks ranged between 1.5 and 3.4% for maize, 2.6 and 4.7% for soybean and 2.6 and 5.0% for hemp.

Despite absolute values that were very close to each other, the genome size of the hemp nuclei was significantly affected by the sex of the plant, defined as monoecious, female or male, regardless of the reference standard (Fig. 6.1; Table 6.1). The genome size of the male 'Carmagnola' plants was significantly larger ($P < 0.001$) than that of all of the other plants, *i.e.*, both female 'Carmagnola' plants and the plants from all of the monoecious cultivars. The female plants were not significantly different from the monoecious plants, nor were the monoecious cultivars significantly different from each other. The mean genome sizes estimated with soybean as the reference standard were 1.791 ± 0.017, 1.789 ± 0.019 and 1.835 ± 0.019 (pg) in monoecious, female and male plants, respectively, and the corresponding sizes estimated with maize as the reference standard were 1.838 ± 0.020 (pg), 1.835 ± 0.017 and 1.883 ± 0.018, respectively. The date of analysis significantly affected the genome size of the nuclei (Table 6.1). However, no interaction was observed between the fixed effects ('cultivar or sex' and 'date').

Table 6.1 Flow cytometry: ANOVA table for the genome size (GS) of hemp nuclei estimated using soybean or maize as the reference standard

Dependent variable	Source [a]	DF	SS	MS	FValue	ProbF
GS - soybean	Cultivar or sex	6	0.017	0.003	8.78	<0.0001
	Date	5	0.009	0.002	5.89	0.0003
	(Cultivar or sex) x date	16	0.004	0	0.86	0.6157
GS - maize	Cultivar or sex	6	0.019	0.003	7.98	<0.0001
	Date	5	0.01	0.002	5.33	0.0006
	(Cultivar or sex) x date	16	0.006	0	0.98	0.4949

[a] The analysis was performed considering seven levels for the "cultivar or sex" factor: one for each of the five monoecious cultivars and two for each sex of the dioecious cultivar.

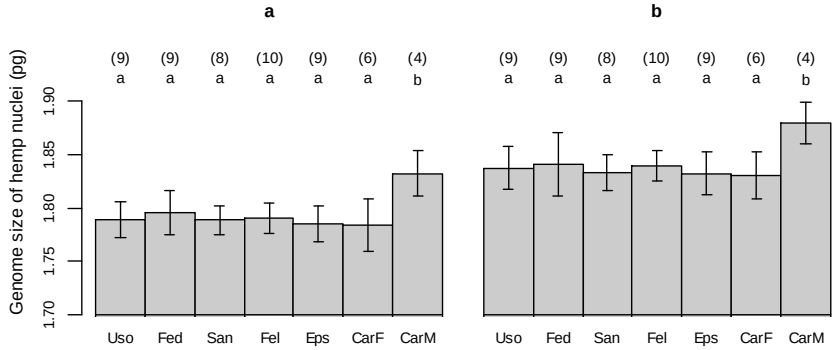

Fig. 6.1 Genome size of monoecious and dioecious hemp nuclei (mean ± sd) estimated using (a) soybean and (b) maize as the reference standard. 'Uso' to 'Eps' refer to the five monoecious cultivars 'Uso 31, 'Fedora 17', 'Santhica 27', 'Felina 32' and 'Epsilon 68', respectively, and 'CarF' and 'CarM' refer to female and male plants of the dioecious cultivar 'Carmagnola'. The values are the overall mean genome size to which the residual values obtained from an ANOVA including 'date' as cofactor were added (Costich et al. 1991). Different letters above the bars indicate significant differences ($P < 0.001$). The number of data points is given between brackets for each cultivar.

6.3.2. Male-associated DNA marker

The MADC2 primer amplified one male-associated band of 390 bp – the MADC2 marker – in the male plants of the dioecious cultivar 'Carmagnola', as expected (Mandolino et al. 1999). This male-associated band was absent from all of the female 'Carmagnola' plants (Fig. 6.2a). Moreover, the MADC2 marker was absent from all of the plants belonging to monoecious hemp cultivars (Fig. 6.2b for cultivar 'Fedora 17').

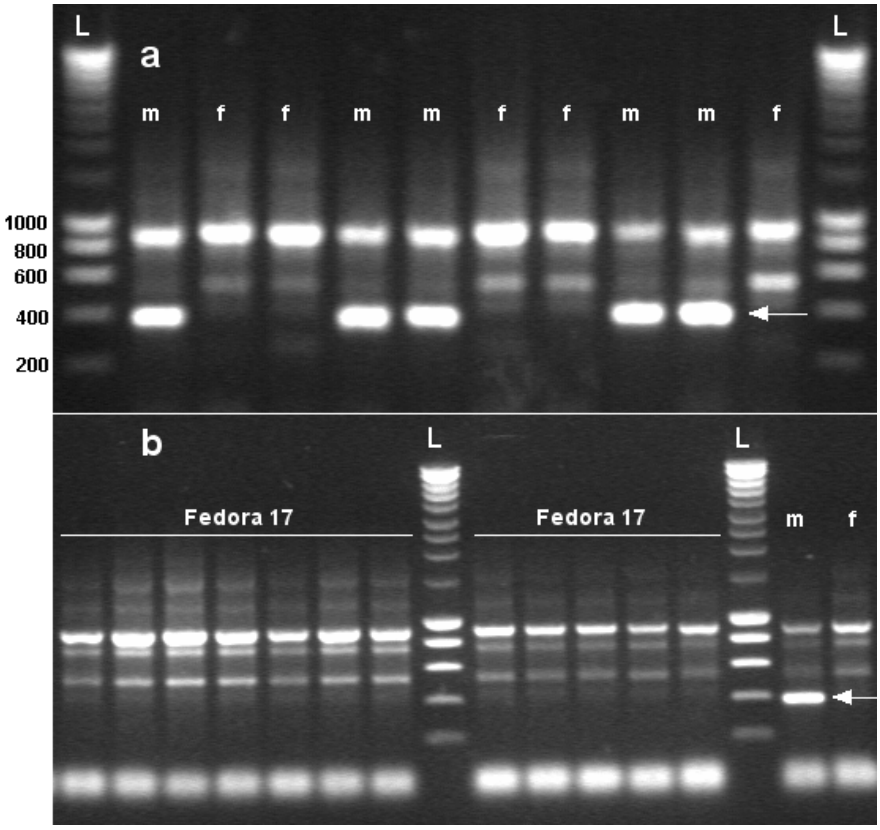

Fig. 6.2 PCR products obtained with the male-associated DNA marker MADC2 applied on (a) ten plants of the dioecious hemp cultivar 'Carmagnola', male (m) and female (f), and (b) twelve plants of the monoecious hemp cultivar 'Fedora 17'. The arrows indicate the male-specific band of 390 bp long. L: Smartladder (bp).

6.4. DISCUSSION

6.4.1. Flow cytometry: practical considerations

The quality of the nuclei suspension prepared for the purpose of flow cytometry is properly judged by analysing the histogram of the relative fluorescence intensity, which accounts for the relative DNA content (Dolozel and Bartos 2005; Loureiro et al. 2006). The CVs of the peaks obtained in the present study were acceptable according to the 5% threshold given by Dolezel and Bartos (2005).

Flow cytometry has been widely used to determine the absolute nuclear DNA content of various plant species (Dolezel and Bartos 2005; Praça-Fontes et al. 2011). In this context, the choice of an internal reference standard is considered to be one of the main issues (Dolezel and Bartos 2005). Here, the genome size of hemp nuclei was function of the reference standard – soybean or maize. This observation was due to a difference between the peak ratio between both reference standards (average ratio = 0.44) and their genome size ratio given by Praça-Fontes et al. (2011) (2.41/5.57 = 0.43) and used here to calibrate the hemp peaks. Such discrepancies among laboratories are common issues in flow cytometry (Dolezel et al. 1998; Praça-Fontes et al. 2011). According to Dolezel and Bartos (2005), the genome sizes estimated using soybean as a reference standard were more accurate that those obtained using maize since soybean and hemp have closer genome sizes.

The effect of the date of analysis on the genome size observed in this study (Table 6.1) was similar to the 'run' effect reported by Costich et al. (1991) and Taliaferro et al. (1997). According to Taliaferro et al. (1997), the differences in genome size among dates of analysis could result from nuances associated with sample preparation or machine calibration. Moreover, cytosolic compounds might interact with PI fluorescence, as noted with anthocyanin in *Euphorbia pulcherrima* by Bennett et al. (2008). Hemp produces a variety of secondary metabolites, such as cannabinoids, unique to the species (de Meijer et al. 2003), flavonoids, stilbenoids, terpenoids, alkaloids and lignans (Flores-Sanchez and Verpoorte 2008), which could cause significant variations in the genome-size estimation. In the present study, the significant 'date' effect was removed

according to Costich et al. (1991) for the computation of the mean genome size of hemp nuclei (Fig. 6.1).

6.4.2. Genome size of the monoecious hemp

The use of flow cytometry revealed significant differences in genome size between male hemp plants *vs.* both monoecious and female ones. To our knowledge, the present study constitutes the first report on the genome size of monoecious hemp. Considering a conversion factor of 1 bp = $0.978*10^9$ DNA content (pg) (Dolezel and Greilhuber 2010), the genome sizes calculated using soybean as the reference standard were 1751 ± 16, 1750 ± 18 and 1795 ± 18 Mbp for monoecious, female and male hemp plants, respectively. Thus, the genome size of monoecious hemp was similar to that of female plants.

Though higher, the values of genome size obtained in the present study are in the range of those found by Sakamoto et al. (1998) for male and female hemp plants (1683 ± 13.9 and 1636 ± 7.2 Mbp, respectively). The difference between both studies could firstly be explained by the use of distinct reference standards (Dolezel and Bartos 2005), *i.e.*, *Arabidopsis thaliana* by Sakamoto et al. (1998) *vs.* soybean and maize in the present study. Secondly, Sakamoto et al. (1998) used the DAPI fluorochrome, which preferentially binds to AT bases and therefore leads to DNA-content estimations that should be interpreted with caution (Dolezel et al. 1992). Thirdly, the genome size of 130 Mbp that was considered for *Arabidopsis thaliana* by Sakamoto et al. (1998) has been reassessed at 157 Mbp by the Arabidopsis Genome Initiative (2000). Nevertheless, the difference in percentage of the genome size between male and female plants found in our study (2.7%) was very close to the 2.8% reported by Sakamoto et al. (1998) and to the 2.7% found by J. Suda (Charles University, Czech Republic, ‘pers. comm.’).

6.4.3. Male-associated DNA marker

The male-associated band amplified by the MADC2 primer corresponded to the marker described by Mandolino et al. (1999). The adequacy of this marker to discriminate male from female plants in the dioecious cultivar 'Carmagnola' as observed here confirmed the results of Mandolino et al. (1999; 2002).

The present study showed the absence of the MADC2 marker in 115 plants obtained from monoecious hemp cultivars. This result amplified the absence of this marker observed in the monoecious cultivars 'Bialobrzeskie' and 'Beniko' by Mandolino et al. (1999).

6.5. CONCLUSIONS

The species *Cannabis sativa* is characterized by heteromorphic sex chromosomes. Female plants of dioecious hemp have XX chromosomes, and male ones have XY chromosomes. The present study showed a similitude of genome size between monoecious plants and female plants of dioecious hemp and the absence of the male-associated DNA marker MADC2 in monoecious hemp, which is likely located on the Y chromosome in a region excluded from recombination (Mandolino et al. 2002). These findings suggest that monoecious hemp has basically the same sex chromosomes than female plants of dioecious hemp, *i.e.*, XX chromosomes, and confirm the cytological observations made in the 'Kentucky' cultivar by Menzel (1964).

Finally, the present study determined the constitution in sex chromosomes of monoecious hemp and thereby established fundamental information for further studies on the genetic determinism of is sex expression.

Acknowledgements

We thank Prof. Bernadette Govaerts and the platform Statistical Methodology and Computing Support (UCL, Belgium) for their valuable advice on statistics, Dr. Olivier Hardy (ULB, Belgium) and Prof. Jan Suda (Charles University, Czech Republic) for their advice on flow cytometry, and the Fédération Nationale des Producteurs de Chanvre (France) for the genetic material. This work was supported by the Fonds de la Recherche Scientifique [grant number 1.5.279.08].

CHAPTER VII

LINKAGE MAPS OF SEX EXPRESSION IN DIOECIOUS AND MONOECIOUS HEMP

The present chapter has been submitted as a research paper under the title:

A.-M. Faux [a], X. Draye [a], M.-C. Flamand [b],·A. Occre [a], P. Bertin [a]. Linkage maps of sex expression in dioecious and monoecious hemp (*Cannabis sativa* L.). Submitted to Euphytica.

[a] Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve

[b] Institut des sciences de la vie, Université catholique de Louvain, Louvain-la-Neuve

Abstract

Hemp (*Cannabis sativa* L.) is diploid (2n = 20) and includes both dioecious and monoecious cultivars with hetero- and homomorphic sex chromosomes, respectively. The sex expression varies not only with environmental factors but also with genotypes, suggesting that it might be selected. The present paper aimed to develop a linkage map for investigating the genetic determinism of sex expression in hemp.

Two dioecious ('Carmagnola' ♀ x 'Carmagnola' ♂) and one monoecious ('Uso 31' x 'Fedora 17') F_1 segregating populations were used. After selection of independently segregating markers, the linkage maps contained 93, 92 and 86 AFLP markers – 225 distinct markers in total – distributed along 11, 16 and 10 co-segregation groups (CGs), respectively. The markers' integration resulted in five homology groups derived from at least two distinct maps and seven CGs derived from single maps. A 'sex' phenotypic marker was mapped in each dioecious population. The sex CG of each dioecious map was divided into four regions: (1) a central region including markers mapped on the Y and/or X chromosomes, (2 and 3) two flanking regions including markers mapped on X chromosomes, and (4) a terminal pseudoautosomal region recombining with the sex locus at a rate ≥ 0.41. Five sex-linked markers were mapped in the monoecious population. Two of them revealed homologies between the male-plant Y and monoecious-plant X chromosomes.

Given the detection of sex-linked markers, we believe that the present unsaturated maps may be of interest for identifying QTL involved in the determinism of the sex expression in hemp.

Keywords: AFLP · *Cannabis* · Dioecious · Linkage map · Monoecious · Outcrossing · Sex.

7.1. INTRODUCTION

Hemp, *Cannabis sativa* L., is a non-food crop currently grown for a wide range of end products derived from its fibre, wooden core or seeds, such as textiles, paper pulp, lime-hemp concretes or hempseed oil (Struik et al. 2000). After a decline in its cultivation during the last two centuries, the species is now perceived as a potentially profitable crop with the right profile to fit into sustainable farming systems (van der Werf et al. 1996).

Agriculturally, the cultivation of hemp is significantly affected by the plant's reproductive features. Although the species is naturally dioecious, both dioecious and monoecious cultivars exist. Dioecious cultivars provide relatively high stem yields. However, male plants flower and senesce earlier than female ones, which makes the mechanical harvest of the seed difficult (Bocsa and Karus 1998). Monoecious cultivars display higher seed yields and allow the mechanical harvest of both stem and seed due to the synchronised maturity of the plants (Mandolino and Carboni 2004). However, the sex expression in monoecious hemp varies continuously between the extreme male and female phenotypes. The masculinised types set few seeds and are more prone to develop fungal infections (Fournier and Beherec 2006). Therefore, the seed production of monoecious cultivars requires the elimination of the strongly masculinised plants (Beherec 2000). A previous study showed that the sexual phenotype of monoecious hemp varies according to the cultivar, suggesting that it has a genetic basis (Faux et al. 2014). Higher seed yields were obtained with the early and mid-early feminised cultivars (Faux et al. 2013). Given the practical implications of the monoecy in both cultivation and breeding, investigating sex expression in hemp is considered to be one of the most important tasks for its genetic improvement (Mandolino and Carboni 2004; Ranalli 2004).

Hemp is diploid (2n = 20), with heteromorphic sex chromosomes and sexual dimorphism. The male plants of dioecious hemp are characterized by hanging panicles with few or no leaves and bearing staminate flowers. The female plants have a distinct appearance, characterized by racemes with leafy bracts and pistillate flowers. The female and male plants of dioecious hemp have XX and XY chromosomes, respectively (Yamada, 1943 cited by Sakamoto et al.

1995), and the sex would be controlled by an X-to-autosomes equilibrium rather than by a Y-active system (Westergaard 1958; Ainsworth 2000). The Y chromosome is larger than the X chromosome and the autosomes (Yamada 1943 cited by Sakamoto et al. 1995), and the size difference between the X and Y chromosomes has been attributed to the large long arm of the Y chromosome (Sakamoto et al. 1998). The plants of monoecious hemp bear racemose inflorescences similar to those of female plants, with both male and female flowers arising in variable amounts. Their sex chromosomes are XX (Faux et al. 2014). However, the determinism of their sexual phenotype remains widely unknown. Moreover, the sex expression of both dioecious and monoecious hemp is sensitive to external factors such as photoperiod, nitrogen status and hormonal treatments (Freeman et al. 1980; Arnoux 1966). The diversity of the intersexual forms, the bipotency of sexually predetermined plants and the occurrence of fertile male flowers on female plants of dioecious hemp (Schaffner 1921; Borthwick and Scully 1954) suggested that the sex expression in hemp would be determined by the activity of genes that are located not only on the sex chromosomes but also on the autosomes (Grisko, 1937 cited by Truta et al. 2007; Sengbusch, 1952 cited by Westergaard, 1958; Rath, 1968 cited by Truta et al. 2007; Migail 1986 cited by Mandolino and Ranalli 2002). Assuming that the sex expression in hemp has a genetic basis (Faux et al. 2014), we propose to approach its determinism through the identification quantitative trait loci (QTLs) linked to the sexual phenotypes.

A very high degree of polymorphism and heterozygosity was found in hemp, suggesting that molecular mapping in F_1 progenies would be possible provided that the parental strains were chosen from the most heterozygous populations (Forapani et al. 2001). However, the use of an F_1 population derived from outbred parents makes the construction of a linkage map much more challenging than the use of segregating populations derived from inbred lines, such as F_2 or backcross populations. Indeed, in F_1 populations, the number of segregating alleles can vary from one locus to another, resulting in a mixture of markers with distinct segregating patterns, and the linkage phases between markers in the parents are unknown (Wu et al. 2002a; Lin et al. 2003). The traditional construction of genetic maps with F_1 populations is performed with 1:1 segregating markers and presents the disadvantage of resulting in two distinct maps, one for each parent (Grattapaglia and Sederoff 1994), as performed by the software MapMaker (Lander et al. 1987). Wu et al. (2002a) developed a method based on a maximum-likelihood approach to construct

genetic maps that combine a set of markers with various segregating patterns in outcrossing species. For a given pair of markers, their methodology simultaneously estimates the maximum-likelihood recombination fraction and the parental linkage phase by using an iterative expectation-maximisation algorithm (Dempster 1977). The advantages of this methodology have been established in the determination of the most likely assignment of markers within a co-segregation group (Wu et al. 2002a) and in terms of numbers of mapped markers and obtained co-segregation groups (Garcia et al. 2006).

Sex-specific AFLP markers were successfully detected in hemp by Flachowsky et al. (2001). The AFLP technique proved its utility in the detection of sex-specific markers in *Dioscorea tokoro* (Terauchi et al. 1998) and *Asparagus officinalis* (Reamon-Büttner and Jung 2000) and is known for its reliability and the potentially high number of amplified fragments, which makes it an useful tool for the construction of genetic maps (Vos et al. 1995; Mueller and Wolfenbarger 1999; Meudt and Clarke 2007). AFLP amplifications reveal a polymorphism of presence/absence, with 1:1 or 3:1 segregating patterns depending on the presence of the allele as a single copy from only one parent or from both parents.

Two non-saturated molecular maps have been published so far in hemp. The first (Mandolino and Ranalli 2002) was based on an F_1 progeny of 40 individuals derived from a cross between a monoecious plant CAN 19/86 and a female plant of the dioecious cultivar ‘Carmagnola’, thus not including the Y chromosome. Sixty-six RAPD markers segregating 1:1 were distributed among 11 co-segregation groups in the ‘Carmagnola’ parent map, and 43 markers were distributed among 9 co-segregation groups in the monoecious parent map. The second map was obtained from a cross between a male plant and a female plant of a dioecious accession. A total of 122 AFLP markers was distributed among 10 linkage groups (Peil et al. 2001 cited by Mandolino and Carboni 2004). Peil et al. (2003) studied the segregation of 66 sex-linked markers in male and female progenies and concluded that there was a pseudoautosomal region allowing recombination between the X and Y chromosomes of hemp. In addition, polymorphism between the X chromosomes has been reported in hemp (Peil et al. 2003; Rode et al. 2005).

The present paper is dedicated to the development of a linkage map for investigating the genetic determinism of sex expression in hemp. To this

purpose, distinct segregating populations derived from dioecious and monoecious hemp cultivars were used. In particular, the use of a dioecious segregating population allows the mapping of a 'sex' phenotypic marker and thus the identification of sex-linked markers as performed by Peil et al. (2003). In this case, joining monoecious and dioecious maps can offer an overview of the mapping work. However, the integration of linkage maps assumes that the markers are homologous, *i.e.*, derived from one unique locus that is identical among all individuals showing the DNA fragment in the distinct maps. Rouppe van der Voort et al. (1997) concluded from the comparison of the sequence of co-migrating bands that the homology of the AFLP products is nearly always valid. However, according to Qi et al. (1998), the locus specificity of the AFLP markers is limited to populations belonging to the same species or to very closely related species. Thus, the locus specificity should hold true here.

The present study constitutes the first report of the construction of a linkage map derived from a cross between two cultivars of monoecious hemp. In addition, two maps derived from a dioecious hemp cultivar were constructed. Therefore, the specific objectives of the present study were as follows: (i) the construction of genetic maps in dioecious and monoecious hemp and (ii) the identification of sex-linked markers and co-segregation groups putatively located on the sex chromosomes.

7.2. MATERIALS AND METHODS

7.2.1. Genetic material

Three cultivars were used: one dioecious, 'Carmagnola', and two monoecious, 'Uso 31' and 'Fedora 17'. Seeds of the dioecious cultivar were obtained from Assocanapa (Carmagnola, Italy), and seeds of both monoecious cultivars were obtained from the Fédération nationale des Producteurs de Chanvre (Le Mans, France). Three populations were created, all consisting of the full-sib F_1 progeny of a cross between two outbred parents. The first two populations (C1 and C2) were derived from a cross between male and female 'Carmagnola' plants. The third (UF) was obtained from a cross between 'Uso 31' and 'Fedora 17' used as male and female parents, respectively. 'Uso 31' has been described as more masculinised than 'Fedora 17'. However, both cultivars exhibit a relatively close earliness (Faux et al. 2013, 2014), which allowed their flowering periods to synchronise. The C1, C2 and UF populations included 77 (43 females and 34 males), 76 (48 females and 28 males) and 167 individuals, respectively. The relatively poor seed set in the controlled crosses in 'Carmagnola' made the use of two populations necessary.

7.2.2. Growth conditions

Sowing was performed on 6 October 2009 for the UF population and on 10 September 2010 for both C1 and C2 populations. Seeds were sown in dimpled germination plates. After 10 days, the seedlings were transplanted to pots of 30 cm height by 18 cm diameter in a greenhouse. The substrate consisted of a 3:1:1 mixture of soil, sand and loam to which an amount of 2.8 g per pot of Osmocote® (14:13:13 NPK with slow release of nutrients for up to six months) was added just before transplanting. The plants were watered by capillarity twice per week. The set point temperatures were 25/20°C day/night. The photoperiod was 16 h during the whole growing period for the UF population and 16 h during 69 days and 8 h thereafter to promote flowering for both C1 and C2 populations.

7.2.3. Molecular data

DNA extraction

Leaf DNA was extracted according to Murray and Thompson (1980) with modifications as detailed below. Frozen leaves were ground to a fine powder in liquid nitrogen. The resultant powder was mixed with 900 µl of extraction buffer (10% Tris-HCl, 10% EDTA, 2% CTAB, 50% NaCl, 2% PVPP and 1% β-mercaptoethanol) and heated at 55°C for 1 h with upside-down shaking every 10 min. A volume of 500 µl of chloroform-isoamyl alcohol (24:1, v/v) was added, and the mixture was shaken for 10 s and centrifuged (13,000 rpm for 10 min). The upper phase was collected, and the extraction with chloroform-isoamyl alcohol was repeated once. The final upper phase was collected, mixed with 350 µl of isopropanol at 4°C, shaken for 10 sec, rested for 30 min and centrifuged (13,000 rpm for 10 min). The supernatant was drained, and the resultant DNA pellet was mixed with 500 µl of 70% ethanol at -20°C and centrifuged (13,000 rpm for 5 min). This last step was repeated once before dissolving the pellet in 50 µl TE buffer (10 mM Tris-HCl pH8, 1 mM EDTA pH8). The extracted DNA was kept at -20°C.

AFLP amplifications

The AFLP amplifications were performed according to Vos et al. (1995) with slight modifications according to Flachowsky et al. (2001) and A. Peil (pers. comm. 2009). All primer and adapter sequences were designed by Eurogentec SA (Seraing, Belgium). Genomic DNA (0.25 µg) was incubated at 37°C for 3 h with 2.5 U of HindIII, 2.5 U of Tru91 (same restriction site as MseI), 2.5 µg of acetylated BSA and 1X Multi-CoreTM buffer (Promega Corp, Madison, USA) in a final volume of 25 µl. Adaptor ligation was achieved by adding 2.5 pmol of the HindIII-adaptator, 25 pmol of the Tru91-adaptator, 1 U of T4-DNA ligase and 1X T4-DNA ligase reaction buffer (Biolabs Inc, Ipswich, USA) to the digested DNA (30 µl final volume) and incubated at 37°C for 3 h. The reaction mixture was diluted to a final volume of 50 µl with sterile H_2O. Preamplifications were performed in a 25-µl volume containing 2.5 µl of digested and ligated DNA, 37.5 ng of HindIII-primer, 37.5 ng of Tru91-primer, both primers having one selective nucleotide (A), 0.2 mM of each dNTP, 1.5 mM of MgCl2, 1.25 U of GoTAQ and 1X of GoTAQ buffer (Promega Corp,

Madison, USA). The PCR was run on a PTC 100 thermal cycler (MJ Research, Waltham, Mass) with the following cycle profile: 25 cycles of denaturation at 94°C for 60 s, annealing at 60°C for 60 s and extension at 72°C for 2 min. The preamplification mixture was diluted 20 times with sterile H_2O. The selective amplifications (25 µl volume) differed from the preamplifications as follows: (1) 2.5 µl of diluted preamplified DNA, 5 ng of labelled HindIII-primer (*i.e.*, the rare cutter) and 30 ng of Tru91-primer - both primers having three selective nucleotides – were used (Table 7.1), and (2) the cycle profile was 1 cycle of denaturation at 94°C for 60 s, annealing at 65°C for 60 s and extension at 72°C for 90 s, followed by 9 cycles with the same conditions as the first except that the annealing temperature dropped by 1°C at each cycle, and finally 40 cycles of denaturation at 94°C for 60 s, annealing at 56°C for 30 s and extension at 72°C for 60 s. The selective amplifications were performed using eight distinct primer combinations (Flachowsky et al. 2001) (Table 7.1). The amplification products were fractionated with an automated ABI Prism 3100 Genetic analyser (Applied Biosystems, Warrington, UK) according to the manufacturer's instructions. Their sizes were scaled with the molecular standard GeneScan-500 Rox. The size and intensity of the amplified fragments were visualised with the Peak ScannerTM v1.0 free software (Applied Biosystems 2006), and the fragments were scored between 40 and 400 bp.

Table 7.1 Selective nucleotides of each primer combination used to generate AFLP markers and number of markers selected from each primer combination

No. [1]	Primer combination		Number of selected markers [2]
	*Hind*III	*Mse*I	
1	AGA	AAC	58
2	AGA	AGG	36
3	AGA	AAG	43
4	ACC	AAG	45
5	ACC	AGA	46
6	ACC	ATC	62
7	ACT	ACA	48
8	ACT	ACC	47

[1] Number of primer combination used to label the markers.
[2] The total number of selected markers is 385, as shown in Table 7.3.

Segregation of AFLP markers

Each amplification was scored as a dominant marker – *i.e.*, following its presence (allele *a*) or absence (allele *o*) – in each individual. The markers were labelled by the number of the primer combination (Table 7.1) followed by the molecular weight of the DNA fragment (in bp). A segregation type was attributed to each marker in each population by testing its segregating pattern against both possible ratios (1:1 and 3:1) using chi-square tests (χ^2). The *single-dose* markers were polymorphic between parents and segregated 1:1 in the offspring. Following Wu et al. (2002a), they were denoted D1.13 or D2.18 depending on the presence of the allele *a* in the female (configuration '*ao* x *oo*') and male (configuration '*oo* x *ao*') parent, respectively, with the location of the alleles *a* and *o* being interchangeable between both homologous chromosomes in each parent. The *double-single-dose* markers were monomorphic between parents (configuration '*ao* x *ao*'), segregated 3:1 and were denoted C.8 (Wu et al. 2002a).

7.2.4. Identification of sex-linked markers

In the dioecious populations, the sex of each plant was coded as 0 (female) or 1 (male) and used as a phenotypic marker. Sex-linked AFLP markers were detected by using the five classes of markers segregating with the 'sex' phenotypic marker as defined by Peil et al. (2003) (Table 7.2). The classes A, B and C contain markers that are polymorphic on the X chromosome with no fragment on the Y. The class A markers result from '*ao* x *ao*' or '*oa* x *ao*' crosses, segregate 1:1 in male progenies and are always present in female progenies. The class B markers derive from '*ao* x *oo*' or '*oa* x *oo*' crosses and segregate 1:1 in both male and female progenies. The class C markers are characterized by their location on the X chromosome of the male parent ('*oo* x *ao*' cross configuration) and are always present in female progenies and always absent from male progenies. Class D includes markers that are polymorphic on X and present on Y ('*ao* x *oa*' or '*oa* x *oa*' cross configurations), indicating a pseudoautosomal region. They segregate 1:1 in the female progenies and are always present in male ones. Class E includes male-associated markers, present on Y and absent from X ('*oo* x *oa*' cross configuration). Recombinants between class B and sex cannot be detected because of their cross type, D1.13 and D2.18,

respectively (Maliepaard et al. 1997; Wu et al. 2002a). Therefore, the markers of class B are identified by linkage to class A in the male progenies and to class D in the female ones (Peil et al. 2003). In the present study, the markers that responded to the cross type and segregation criteria of class B (Table 7.2) but were not linked to a class A or D were referred to as 'putative class B' markers. However, these markers can be located on both sex chromosomes and autosomes.

The markers were assigned to a class of sex-linked markers by testing the equality of their segregation ratios in male and female progenies against the expected segregation ratios using a chi-square test. Deviations from the expected segregation ratios at alpha = 0.01 were discarded.

Table 7.2 Five classes of sex-linked markers, as defined by Peil et al. (2003)

	Parents			Expected segregation ratios in the F_1	
Class [1]	X | | X [2]	X | | Y	Cross-type	♀ (a:o)	♂ (a:o)
A	a | | o	a | | o	C.8	1:00	1:01
B	a | | o	o | | o	D1.13	1:01	1:01
C	o | | o	a | | o	D2.18	1:00	0:01
D	a | | o	o | | a	C.8	1:01	1:00
E	o | | o	o | | a	D2.18	0:01	1:00

[1] The class B markers are detected by linkage to class A in the male and to class D in the female progenies, in addition to falling into the expected cross-type and segregation ratios.
[2] The location of the alleles *a* and *o* is interchangeable between the X chromosomes in the female parent.

7.2.5. Map construction

The linkage analysis and map construction were performed independently for each segregating population using the OneMap package (Margarido et al. 2007), which implements the methodology developed by Wu et al. (2002a) and consists of a set of functions for the freely available software R (R Development Core Team 2010).

For each pair of markers, the *rf.2pts* function of OneMap computed the maximum likelihood estimate of the recombination fraction and its LOD score by considering four distinct assignments of alleles, each of them corresponding to one of the two possible linkage phases – coupling and repulsion – in each of both parents. The most probable recombination fraction and parental linkage phases were detected based on the maximum posterior probability of a given assignment conditional on the observed phenotypes for the markers (Wu et al. 2002a). Markers were then assigned to linkage groups by using the *group* function. The linkage analysis was first performed under stringent conditions to identify the sets of markers that formed tight co-segregation groups. Distinct threshold values were tested independently in each population; however, the same statistical stringency was applied for the construction of the three maps for more meaningful comparisons among the populations. The LOD score and recombination fraction threshold used for mapping were 3.5 and 0.5, respectively. These thresholds allowed gathering distinct sex-linked markers as detected according to Peil et al. (2003) in the same co-segregation group, whereas lower LOD score thresholds dramatically increased the size of the largest group in each population and were therefore avoided.

The ordering and positions of the markers within each co-segregation group were determined by using the *order.seq* function of OneMap with the same LOD score value as for linkage. This function first compares all possible orders among a user-defined number of markers that are selected according to their degree of informativeness, with the C.8 markers selected prior to D1.13 and D2.18 markers (Margarido et al. 2012). Here, we set the number of markers to 6. The maximum likelihood order among these six markers was chosen, and the remaining markers were positioned sequentially at the set of locations that maximises the map LOD score while keeping the original map unaltered

(Margarido et al. 2012). Map distances were based on the Kosambi function. Marker orderings and parental linkage phases were verified with the *ripple* and *map* functions, which check for alternative orders in windows of four markers and estimate the multipoint log-likelihood, linkage phases and recombination frequencies for a given order of markers, respectively (Margarido et al. 2012). When distinct marker orders maximised the LOD score, the selected order was the one that provided the best agreement among the distinct maps. The three maps, C1, C2 and UF, were integrated by using the markers that were positioned in at least two distinct maps and by gathering the co-segregation groups into putative homology groups.

Table 7.3 The number of markers scored and selected for mapping in each population

	C1	C2	UF	All populations [1]
All scored markers	410	415	357	480
Segregating markers [2]	274 (67)	182 (44)	184 (52)	385
Polymorphic markers between parents	121	83	98	
Monomorphic markers between parents	153	99	86	
All independently segregating markers (1:1 or 3:1) [3]	193 (70)	147 (81)	115 (63)	319
Single-dose markers (segregation 1:1)	71	53	64	
D1.13 [4]	46 (24)	33 (22)	23 (20)	
D2.18 [4]	25 (13)	20 (14)	41 (36)	
Double-single-dose markers (segregation 3:1) – C.8 [4]	122 (63)	94 (64)	51 (44)	
All markers with distorted segregation	81	35	69	

[1] Total number of distinct markers regardless of the population.
[2] The segregating markers had a segregating pattern significantly different from 1:0 (*i.e.*, marker always present) as tested by a Chi² test with Bonferroni correction at alpha=0.05. Between brackets, % of segregating markers in relation to the total number of scored markers.
[3] In parentheses, % of independently segregating markers in relation to the total number of segregating markers.
[4] In parentheses, % of markers of each cross type (D1.13, D2.18 and C.8 following the notation of Wu et al. 2002a) in relation to the total number of independently segregating markers.

7.3. RESULTS

7.3.1. AFLP markers

The eight primer combinations (Table 7.1) allowed the detection of 410, 415 and 357 markers in the C1, C2 and UF segregating populations, respectively, generating a total of 480 distinct AFLP markers (Table 7.3). Out of these, 287 (60%) were scored in all three populations. The percentage of markers shared by C1 and C2 (81%) was higher than that shared by C1 and UF (67%) or C2 and UF (68%).

Among these 480 detected markers, 274 (67% of the total number of detected markers), 182 (44%) and 184 (52%) segregated in C1, C2 and UF, respectively. They accounted for a total of 385 distinct segregating AFLP markers, regardless of the population. Of these, only 54 markers were present in all three populations. The sets of selected markers comprised 121, 83 and 98 polymorphic markers between parents and 153, 99 and 86 monomorphic markers between parents, respectively (Table 7.3).

Among the 385 segregating markers, 193 (70% of 274 markers), 147 (81% of 182 markers) and 115 (63% of 184 markers) segregated independently (1:1 or 3:1) in the C1, C2 and UF populations, respectively. The remaining markers showed a distorted segregation and were discarded. The independently segregating markers consisted of 71, 53 and 64 single-dose markers and 122, 94 and 51 double-single-dose markers in C1, C2, and UF, respectively (Table 7.3), accounting for a total of 319 distinct markers available for mapping. Of these, 98 were found in two distinct populations, and only 19 were found in all three populations.

A cross type was assigned to each independently segregating marker in each mapping population. The double-single-dose markers, which received the C.8 cross type (‘*ao* x *ao*’ cross configuration, with the location of the alleles *a* and *o* being interchangeable), accounted for 63, 64 and 44% of the total number of independently segregating markers in C1, C2 and UF, respectively (Table 7.3).

7.3.2. Linkage maps

The linkage analysis resulted in the construction of three maps that included 93, 92 and 86 markers detected in the C1, C2 and UF segregating populations, respectively (Table 7.4). The markers were assigned to 11, 16 and 10 co-segregation groups, respectively (Fig. 7.1).

The linked markers accounted for 48, 63 and 75% of the 193, 147 and 115 independently segregating markers in C1, C2 and UF, respectively (Table 7.4). The lower percentage of linked markers in C1 was partly because fifteen C.8 markers were discarded, the most weakly linked in the largest CG of the map (C1-3). Nevertheless, the C.8 cross type remained the most represented in both dioecious maps, while the single-dose markers (D1.13 or D2.18 cross types) constituted the majority (60%) of the linked markers in the UF map.

Table 7.4 Distribution of the linked markers according to their cross type (notation following Wu et al. 2002a) in each mapping population

Cross type [1]	Number of linked markers		
	C1	C2	UF
D1.13 ('*ao* x *oo*') [2]	24 (26)	18 (20)	17 (20)
D2.18 ('*oo* x *ao*') [2]	17 (18)	12 (13)	34 (40)
C.8 ('*ao* x *ao*') [2]	52 (56)	62 (67)	35 (41)
Total [3]	93 (48)	92 (63)	86 (75)

[1] Cross type following Wu et al. (2002a).
[2] In parentheses, % of markers of each cross type in relation to the total number of linked markers.
[3] In parentheses, % of linked markers in relation to the total number of independently segregating markers (*i.e.*, 193, 147 and 115, respectively).

Each of the three maps contained several groups with only 2 or 3 markers – there were 6, 8 and 2 such groups in C1, C2 and UF, respectively. The largest groups comprised 29, 25 and 15 markers in C1, C2 and UF, respectively. The percentages of C.8 markers in the largest CG of each map were 96% (C1-3), 95% (C2-4), and 40 and 20% (UF-3 and UF-9, respectively). Some CGs included a high proportion of markers derived from the same primer

combination. This was observed in CGs containing only two markers (C1-2, 8, 9 and 10 and C2-3) as well as in larger CGs (C1-4 and 5, C2-13 and 14 and UF-7, in particular) (Fig. 7.1).

The C1, C2 and UF maps covered a total distance of 1196.7, 878.2 and 1571.0 cM, respectively, with an average distance between markers of 12.7, 9.4 and 18.3 cM and a maximum of 43.5 (UF-5), 41.3 (C2-11) and 65.0 cM (UF-9), respectively. The largest gaps observed (54.93 cM in UF-4 and 65.04 cM in UF-9) were due to the ignorance of the recombination fraction between D1.13 and D2.18 markers. With the exception of one CG with null length in each C1 and C2, the size of the CGs varied from 6.1 to 327.6, 5.0 to 247.3 and 15.4 to 323.0 cM in the C1, C2 and UF maps, respectively. The size of the CGs varied the least in the UF map.

The three maps included a total of 225 distinct markers, with only five shared by all three maps and 36 by two of them: 22 in both C1 and C2, 12 in both C1 and UF, and 17 in both C2 and UF. These markers were used as anchors to integrate the maps.

7.3.3. Integrated map

The integrated map assembled five HGs constructed by joining CGs that shared from 1 to 6 common markers and seven groups including 2 or 3 markers present in only one map (Fig. 7.1). The HGs included from 17 (HGIV) to 73 (HGV) markers, accounting for a total of 209 markers out of the 225 distinct mapped markers. The largest HG gathered 32% of the mapped markers, including the sex marker, all class A, D and E markers and 17 putative class B markers.

HGI included 23 distinct markers distributed along 3 distinct CGs (Fig. 7.1). HGII contained 61 distinct markers distributed along 7 CGs. The largest CGs gathered in HGII (C1-3, C2-4 and UF-3) included two markers present in the three maps, in addition to four markers common to C1-3 and C2-4 and two markers common to C1-3 and UF-3. The four additional groups joined in HGII shared one or two common markers. However, inconsistencies were found in the order of markers across the three maps for HGII (for the markers 8_190 in C1-3, 4_123 in C2-4, 7_254 in C2-3 and 8_87 in C2-4). HGs III and IV included 35

and 17 markers distributed along 4 and 2 CGs, respectively, which shared only one common marker. Details on HGV are provided in the next two sections.

Seven putative class B markers were found in HGs II, III and IV (Fig. 7.1). Three of these were mapped in the same CG (C2-6 in HGIV). In addition, four putative class B markers were detected in CGs that were not assigned to any homology group (C1-8, C1-11 and C2-15). As stated before, these markers can be located on both sex chromosomes and autosomes.

The figure 7.1 is presented on the next three pages.

Fig. 7.1 Integration of three linkage maps of the sex expression in hemp. The C1 and C2 maps were derived from a dioecious population ('Carmagnola' x 'Carmagnola') and included a 'sex' phenotypic marker in C1-7 and C2-14, respectively. The UF map was derived from a monoecious population ('Uso 31' x 'Fedora 17'). The integrated map includes five homology groups (HGI to HGV) and seven groups which were not assigned to a HG. The name of the segregating population and number of co-segregation group are given above each co-segregation group. The markers are labelled according to the number of primer combination (Table 7.1) and molecular weight of the corresponding DNA fragment. The font of the marker name indicates the cross type: normal for D1.13, italic for D2.18 and bold for C.8. Their colour refers to the classes of sex-linked markers (Table 7.2). Map distances are given in cM. The Kosambi map function was used. LOD and maximum recombination fraction were 3.5 and 0.5, respectively.

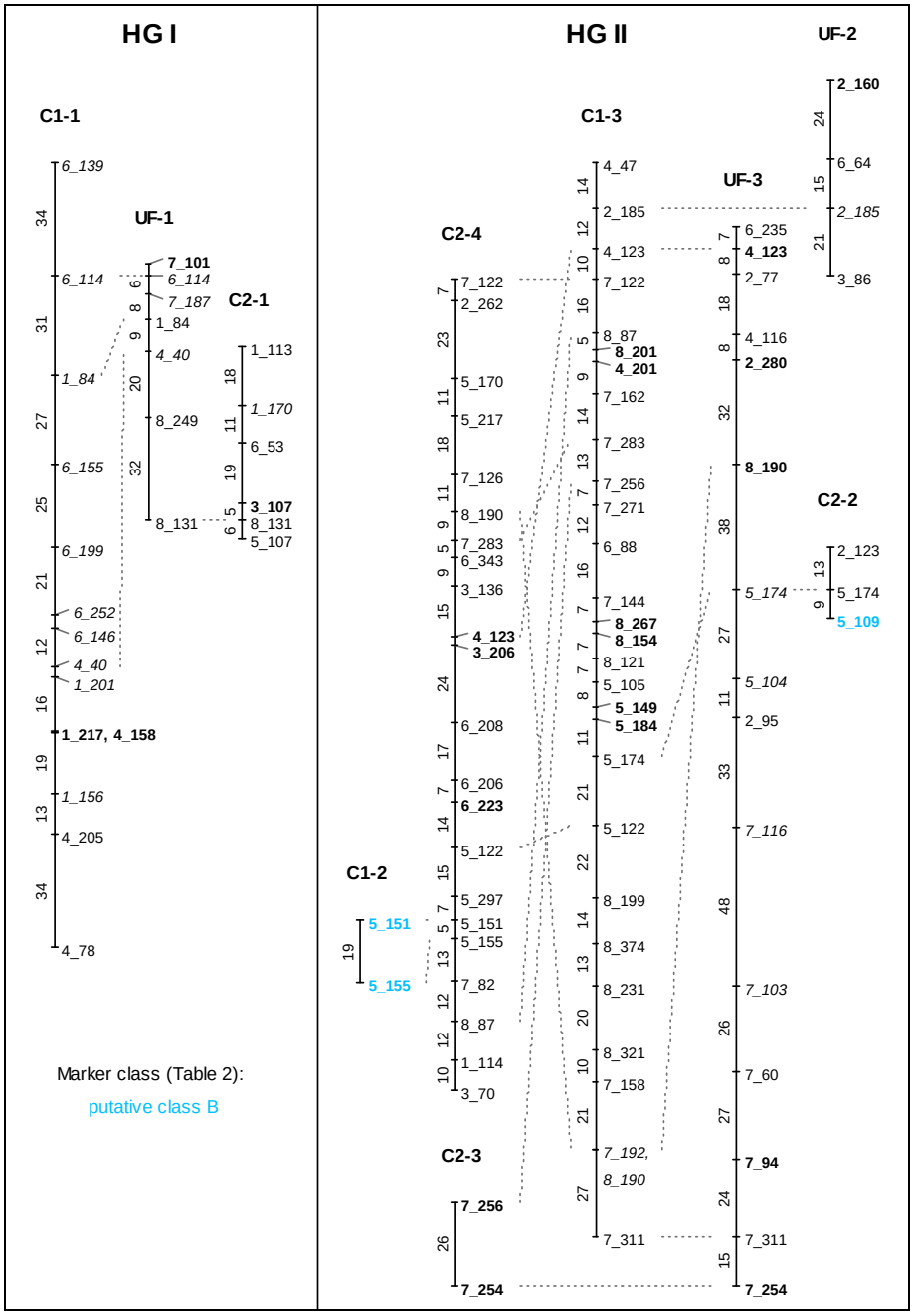
HG I
HG II
UF-2
C1-1
C1-3
UF-1
UF-3
C2-1
C2-4
C2-2
C1-2
C2-3
6_139
6_114
1_84
6_155
6_199
6_252
6_146
4_40
1_201
1_217, 4_158
1_156
4_205
4_78
7_101
6_114
7_187
1_84
4_40
8_249
8_131
1_113
1_170
6_53
3_107
8_131
5_107
4_47
2_185
4_123
7_122
8_87
8_201
4_201
7_162
7_283
7_256
7_271
6_88
7_144
8_267
8_154
8_121
5_105
5_149
5_184
5_174
5_122
8_199
8_374
8_231
8_321
7_158
7_192,
8_190
7_311
7_122
2_262
5_170
5_217
7_126
8_190
7_283
6_343
3_136
4_123
3_206
6_208
6_206
6_223
5_122
5_297
5_151
5_155
7_82
8_87
1_114
3_70
7_256
7_254
6_235
4_123
2_77
4_116
2_280
8_190
5_174
5_104
2_95
7_116
7_103
7_60
7_94
7_311
7_254
2_160
6_64
2_185
3_86
2_123
5_174
5_109
5_151
5_155
Marker class (Table 2):
putative class B

Fig. 7.1

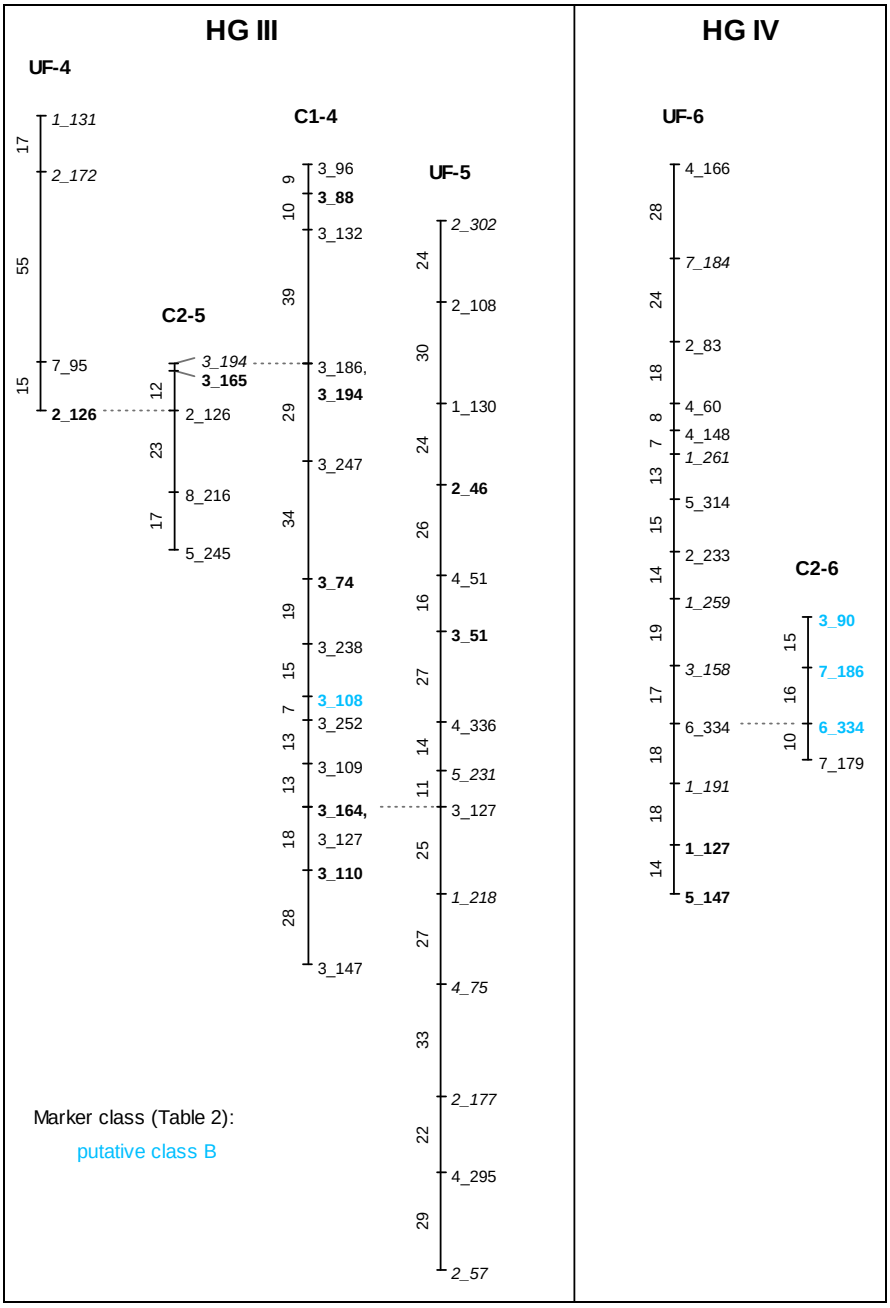

Fig. 7.1 (continued)

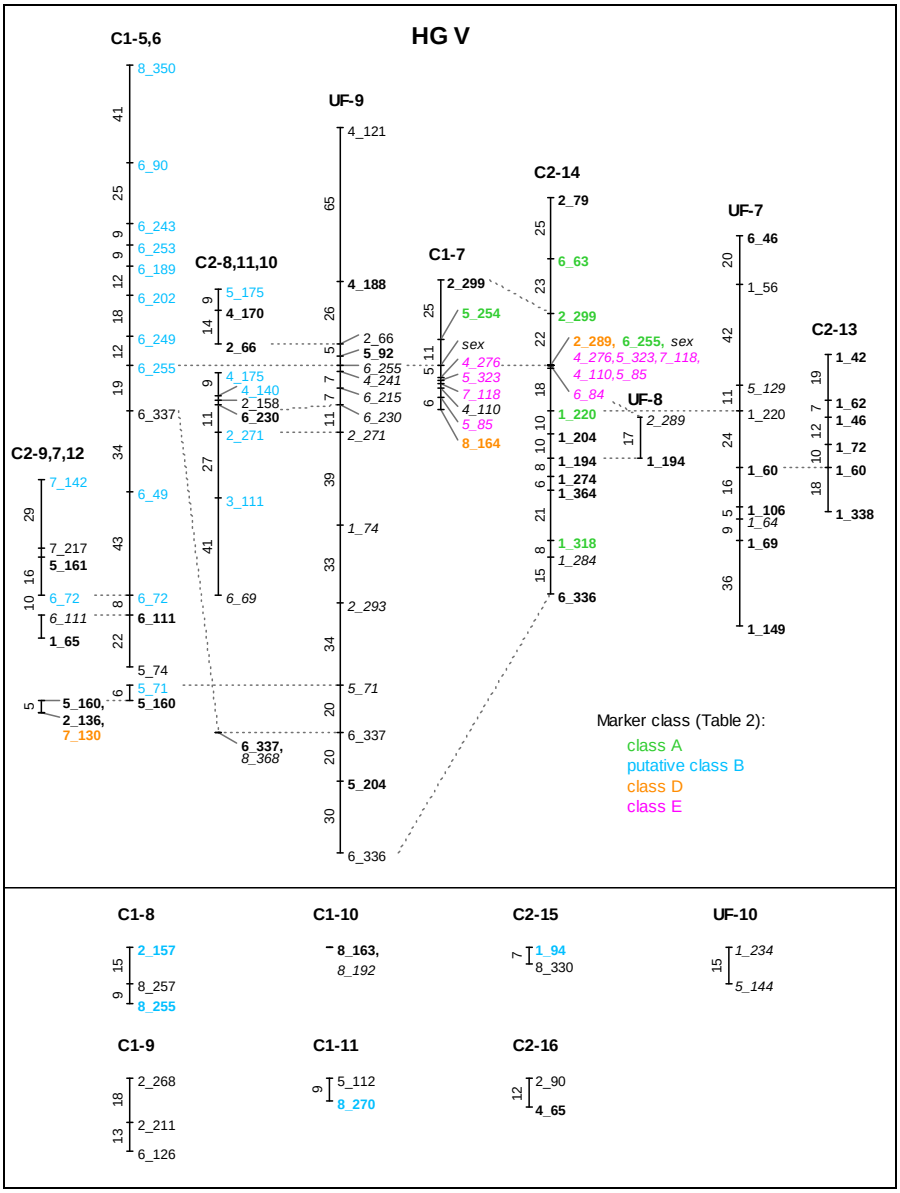

Fig. 7.1 (continued)

7.3.4. Structure of the sex CG in the dioecious maps

The 'sex' phenotypic marker was mapped in the dioecious maps, in C1-7 and C2-14 (Table 7.5). These CGs, referred to as 'sex CGs', included respective totals of 8 and 19 AFLP markers – 21 distinct markers. Among them, 6 and 12 markers –14 distinct markers – were assigned to classes A, D or E as defined by Peil et al. (2003) according to their cross type and segregation ratios in the male and female progenies (Table 7.2) in each map. All of the remaining markers linked in the sex CGs had segregation ratios in the female and male progenies that significantly differed from those expected for sex-linked markers at alpha = 0.01. They were referred to as 'non-classed' markers.

Four successive regions were determined in the sex CGs on the basis of the recombination rate of the markers with sex (Table 7.5): (1) a region including class A markers in addition to non-classed markers, (2) a central region surrounding the sex locus and including all of class E and some class A and D markers, (3) a second region including class A and non-classed markers, and (4) a terminal region that recombined completely with the sex locus.

The markers of regions 1 and 3 were mapped on the X parental chromosomes (Tables 7.2 and 7.6). The recombination rates observed between the markers mapped in these regions and sex ranged from 0.11 to 0.3. In C1, the regions 1 and 3 included one marker of class A (5_254) and one of class D (8_164), respectively. This latter indicated the presence of a fragment common to an X chromosome of the female parent and the Y chromosome. In C2, seven markers were mapped in regions 1 and 3. Two of these markers belonged to class A (6_63 and 1_318). All of the class A markers mapped in regions 1 and 3 of one of both dioecious maps were absent from some female progenies, indicating the presence of recombination with sex in the male parent. The segregating patterns of the five remaining markers mapped in regions 1 and 3 in C2 were close to class A though significantly different from those expected for this class. This was partly due to distorted segregations (2_79 was present in 58% of the progenies, and 1_204, 1_194, 1_274 and 1_364 in ~ 90% of progenies, instead of the 75% expected for C.8 markers). All of these five markers in addition to the class A 1_318 were linked to sex through intermediate markers only.

Region 2 included the sex locus in addition to markers derived from the parental X and/or Y chromosomes. The markers mapped in this region had recombination rates with sex ranging from 0 to 0.07. The recombination rate observed between the class A marker 2_299 and sex was lower in C2 (0.04) than in C1 (0.17), such that this marker was mapped in distinct regions in C1 (region 1) and C2 (region 2). In C2, all class E markers but one completely segregated with sex (Table 7.5). In contrast, 3 to 5 recombinants between class E and sex were found in C1. Moreover, one class E marker (5_85) was found in a female plant in C1. In C2, the class A 6_255 accounted for a fragment mapped on X parental chromosomes at a locus corresponding to the sex locus, while the class D 2_289 indicated the presence of a fragment common to an X of the female parent and the Y chromosome of the male parent at the sex locus (Table 7.5).

Region 4 was determined by two markers mapped on the X parental chromosomes at an extremity of the sex CG in the C2 map (1_284 and 6_336; Table 7.5). Both markers completely recombined with sex, displaying recombination rates of 0.41 and 0.50 with sex, respectively, and thus indicating the presence of a pseudoautosomal region. The marker 1_284 was mapped in the sex CG by linkage in coupling phase with the class A 1_318 in the male parent. Thus, it was mapped onto the X chromosome of the male parent. The second marker (6_336) was linked in coupling phase with 1_284 in the male parent and thus also mapped onto its X chromosome.

Table 7.5 Structure of the co-segregation groups (CGs) that included the 'sex' locus in the dioecious maps, C1-7 and C2-14. Region in the CG [1], marker name, position along the CG, recombination rate with sex, class (Table 7.2), significance level at which the observed segregation ratios in the female and male progenies were not different from the expected ones for a given class of sex-linked markers (alpha), number of male and female progenies according to the presence (M) / absence (m) of the marker and parental genotypes *(this table is split into two parts; the second one is on the next page)*

Map - CG	Region [1]	Marker [2]	Position (cM)	Rec. rate with sex [3]	Class	Alpha	Female progenies		Male progenies		Parents	
							M	m	M	m	X ‖ X	X ‖ Y
C1-7	1	**2_299**	0	0.17	-	-	36	7	9	25	a ‖ o	a ‖ o
		5_254	25.02	0.11	A	0.01	39	4	10	24	a ‖ o	a ‖ o
	2	**sex**	35.57		-	-	0	0	34	0	o ‖ o	o ‖ a
		4_276	40.94	0.05	E	0.05	0	43	30	4	o ‖ o	o ‖ a
		5_323	42.28	0.04	E	0.05	0	43	31	3	o ‖ o	o ‖ a
		7_118	43.61	0.05	E	0.05	0	43	30	4	o ‖ o	o ‖ a
		4_110	44.95	0.06	-	-	0	43	29	5	o ‖ o	o ‖ a
		5_85	48.97	0.05	E	0.05	1	42	31	3	o ‖ o	o ‖ a
	3	8_164	54.47	0.19	D	0.05	23	20	30	4	a ‖ o	o ‖ a

[1] The regions were determined on the basis of the recombination rate of the markers with sex.

[2] The markers in bold face were mapped in both C1 and C2, and the markers in italic were mapped in UF and in C1 or C2.

[3] Recombination rate with sex as estimated by the function *rf.2pts* of OneMap (Margarido et al. 2007).

Table 7.5 (continued)

Map - CG	Region [1]	Marker [2]	Position (cM)	Rec. rate with sex [3]	Class	Alpha	Female progenies		Male progenies		Parents	
							M	m	M	m	X ‖ X	X ‖ Y
C2-14	1	2_79	0	0.25	-	-	37	11	7	21	a ‖ o	a ‖ o
		6_63	25.38	0.16	A	0.01	42	6	9	19	a ‖ o	a ‖ o
	2	**2_299**	48.28	0.04	A	0.05	47	1	15	13	a ‖ o	a ‖ o
		2_289	70.13	0.00	D	0.05	29	19	28	0	a ‖ o	o ‖ a
		6_255	70.13	0.00	A	0.05	48	0	18	10	o ‖ a	a ‖ o
		sex	70.13		-	-	0	48	28	0	o ‖ o	o ‖ a
		4_276	70.13	0.00	E	0.05	0	48	28	0	o ‖ o	o ‖ a
		5_323	70.13	0.00	E	0.05	0	48	28	0	o ‖ o	o ‖ a
		7_118	70.13	0.00	E	0.05	0	48	28	0	o ‖ o	o ‖ a
		4_110	70.13	0.00	E	0.05	0	48	28	0	o ‖ o	o ‖ a
		5_85	70.13	0.00	E	0.05	0	48	28	0	o ‖ o	o ‖ a
		6_84	71.67	0.01	E	0.05	0	48	27	1	o ‖ o	o ‖ a
		1_220	89.46	0.07	A	0.01	47	1	20	8	o ‖ a	a ‖ o
	3	1_204	99.04	0.23	-	-	45	3	23	5	o ‖ a	a ‖ o
		1_194	109.29	0.22	-	-	46	2	25	3	o ‖ a	a ‖ o
		1_274	117.25	0.25	-	-	45	3	24	4	o ‖ a	a ‖ o
		1_364	122.87	0.3	-	-	44	4	24	4	o ‖ a	a ‖ o
		1_318	143.63	0.18	A	0.01	45	3	20	8	o ‖ a	a ‖ o
	4	1_284	151.16	0.41	-	-	32	16	15	13	o ‖ o	a ‖ o
		6_336	166.28	0.50	-	-	34	14	19	9	a ‖ o	a ‖ o

[1] The regions were determined on the basis of the recombination rate of the markers with sex.

[2] The markers in bold face were mapped in both C1 and C2, and the markers in italic were mapped in UF and in C1 or C2.

[3] Recombination rate with sex as estimated by the function *rf.2pts* of OneMap (Margarido et al. 2007).

7.3.5. Identification of CGs putatively located on the sex chromosomes

Distinct CGs putatively located on the sex chromosomes were identified in each map through the presence of markers mapped in the sex CG of the dioecious maps. These CGs were referred to as 'putative sex CGs'. They were gathered in HGV and are described in this section.

As expected, all of the male-specific markers (class E) were absent from the UF population. However, five markers mapped in the sex CG of the C2 map were also found in UF: two class A, one class D and two non-classed markers (Table 7.5; Fig. 7.1). Two of them (6_255 and 2_289) were mapped at the sex locus in C2, and one (6_336) in the pseudoautosomal region. However, these 5 markers did not segregate together in the UF population; rather, they were mapped in 3 distinct putative sex CGs (UF-7, 8 and 9; Fig. 7.1).

Two putative sex CGs were found in the C1 map (Table 7.6). C1-5 included a marker that was linked to the sex and assigned to the class A in C2 (6_255). Among the 14 markers linked in C1-5, 12 consisted of an allele derived from the female parent only and 10 were putative class B markers. One of them (6_72) was also putative class B in C2. The second CG putatively located on the sex chromosomes in C1 was identified through the presence of one marker common to a putative sex CG in UF (5_71, present in both C1-6 and UF-9). This marker was putative class B in C1 and linked to a class D marker (7_130) in C2-12.

Seven CGs putatively located on sex chromosomes were identified in the C2 map from the presence of markers common to putative sex CGs in C1 or UF (Table 7.6). In C2-8, one C.8 marker (2_66) was completely linked to the sex in the male parent, but its segregation in the female progenies did not correspond to the expected 1:1 ratio of class D (Table 7.2). C2-8, 9 and 11 included 1, 2 and 4 putative class B markers, respectively. C2-12 included a marker for which the cross type and segregation ratios in the female and male progenies corresponded to class D at alpha = 0.01 (7_130).

Table 7.6 Structure of the co-segregation groups (CGs) putatively located on the sex chromosomes in the dioecious maps, C1 and C2. Marker name, position along the CG, marker class (Table 7.2), significance level at which the observed segregation ratios in the female and male progenies were not different from those expected for a given class of sex-linked markers (alpha), number of male and female progenies according to the presence (M) / absence (m) of the marker and parental genotypes *(this table is split into two parts; the second one is on the next page)*

Map - CG	Marker [1]	Position (cM)	Class [2]	Alpha	Female progenies		Male progenies		Parents	
					M	m	M	m	Female	Male
C1-5	5_74	0	-	-	32	11	20	14	a || o	o || o
	6_111	21.95	-	-	38	5	32	2	a || o	o || a
	6_72	29.78	B*	0.05	27	16	21	13	a || o	o || o
	6_49	73.27	B*	0.05	16	27	14	20	o || a	o || o
	6_337	107.73	-	-	12	31	11	23	o || a	o || o
	6_255	126.78	B*	0.01	15	28	14	20	o || a	o || o
	6_249	138.69	B*	0.05	19	24	17	17	o || a	o || o
	6_253	156.26	B*	0.05	27	16	22	12	o || a	o || o
	6_202	168.17	B*	0.05	22	21	20	14	o || a	o || o
	6_189	177.36	B*	0.05	22	21	19	15	o || a	o || o
	6_243	186.56	B*	0.05	21	22	15	19	o || a	o || o
	6_90	211.9	B*	0.05	20	23	16	18	o || a	o || o
	8_350	252.93	B*	0.05	17	26	13	21	o || a	o || o
C1-6	*5_71*	0	B*	0.01	28	15	21	13	a || o	o || o
	5_160	6.12	-	-	42	1	25	9	a || o	a || o

[1] The markers in bold face were mapped in both C1 and C2, and the markers in italic were mapped in C2 and UF.

[2] B* markers are putative class B markers: they responded to the cross type and segregation criteria of class B but were not linked to class A or D markers.

Table 7.6 (continued)

Map - CG	Marker [1]	Position (cM)	Class [2]	Alpha	Female progenies		Male progenies		Parents	
					M	m	M	m	Female	Male
C2-7	1_65	0	-	-	31	17	14	14	a || o	a || o
	6_111	10.03	-	-	27	21	18	10	o || o	o || a
C2-8	*2_66*	0	-	-	37	11	28	0	a || o	o || a
	4_170	13.79	-	-	35	13	21	7	a || o	o || a
	5_175	22.83	B*	0.05	21	27	10	18	a || o	o || o
C2-9	**6_72**	0	B*	0.05	28	20	10	18	a || o	o || o
	5_161	15.95	-	-	34	14	26	2	o || a	o || a
	7_217	19.77	-	-	25	23	21	7	o || a	o || o
	7_142	49.02	B*	0.01	31	17	19	9	o || a	o || o
C2-10	***6_337***	0	-	-	32	16	22	6	a || o	o || a
	8_368	0	-	-	15	33	11	17	o || o	o || a
C2-11	4_175	0	B*	0.05	23	25	18	10	a || o	o || o
	4_140	9.32	B*	0.05	21	27	11	17	o || a	o || o
	2_158	11.43	-	-	32	16	18	10	o || a	o || a
	6_230	13.53	-	-	31	17	19	9	o || a	o || a
	2_271	21.99	B*	0.01	16	32	11	17	o || a	o || o
	3_111	49.46	B*	0.05	19	29	13	15	a || o	o || o
	6_69	90.75	-	-	33	15	18	10	o || o	a || o
C2-12	7_130	0	D	0.01	31	17	24	4	a || o	o || a
	2_136	0	-	-	35	13	18	10	a || o	a || o
	5_160	5.02	-	-	37	11	18	10	a || o	a || o
C2-13	1_42	0	-	-	39	9	24	4	a || o	a || o
	1_62	19.19	-	-	45	3	26	2	a || o	a || o
	1_46	26.44	-	-	46	2	25	3	a || o	a || o
	1_72	37.99	-	-	44	4	25	3	a || o	a || o
	1_60	47.68	-	-	46	2	26	2	a || o	a || o
	1_338	65.95	-	-	40	8	25	3	a || o	a || o

[1] The markers in bold face were mapped in both C1 and C2, and the markers in italic were mapped in C2 and UF.

[2] B* markers are putative class B markers: they responded to the cross type and segregation criteria of class B but were not linked to class A or D markers.

7.4. DISCUSSION

We report the construction of three genetic maps of the sex expression in hemp: two from a cross between female and male plants of the dioecious cultivar 'Carmagnola' and one from a cross between the monoecious cultivars 'Fedora 17' and 'Uso 31'. To our knowledge, the present paper constitutes the first study to integrate linkage maps from both dioecious and monoecious hemp. This strategy allowed the identification of markers putatively located on the sex chromosomes of the monoecious population.

7.4.1. Segregation analysis

The dioecious maps were based on segregating populations including 77 and 76 individuals, respectively, and the monoecious map was based on 167 individuals. Two dioecious populations instead of one were used to compensate for the relatively low number of seeds obtained from the 'Carmagnola' crosses. This low number of seeds was due to poor synchronisation of the flowering times of the male and female plants; synchronisation was performed by transferring the female plants in phytotrons under 12-h photoperiod and resulted in relatively small inflorescences. Nevertheless, the present 'Carmagnola' population sizes were of the same magnitude as those previously used in hemp mapping from crosses between male and female plants of dioecious accessions, *i.e.*, 66 (Mandolino and Ranalli 2002) and 80 individuals (Peil et al. 2003).

A relatively high proportion of markers was scored simultaneously in both dioecious populations (81%), as expected from their common 'Carmagnola' origin. However, the proportions of markers scored in one dioecious population and in the monoecious one were also large (67% by C1 and UF, and 68% by C2 and UF), consistent with the structure of the genetic diversity reported in hemp. Indeed, from the analysis of six hemp cultivars – four dioecious, one monoecious and a female inbred line – Forapani et al. (2001) concluded the existence of a widely shared gene pool with limited genetic separations among groups.

The percentage of segregating markers was lower in UF (52%) than in C1 (67%). This may result from the narrower genetic base of monoecious hemp,

as a result of its ability to self-pollinate (Bocsa and Karus, 1998) and from the selection pressure needed to maintain the monoecious trait (Forapani et al. 2001; Mandolino and Carboni 2004). The relatively low percentage of segregating markers (44%) found in C2 compared to C1 (67%) indicated that the proportion of loci that were homozygous dominant ('*aa*') in one or both parents was higher in C2.

The percentages of segregating markers (from 44 to 67% according to the population) were relatively high compared to the percentage of RAPD-segregating loci found in a hemp progeny obtained from a cross between a 'Carmagnola' female plant and a monoecious plant (39.1%) by Carboni et al. (2000), who judged the revealed polymorphism fairly high already. Therefore, we concluded that the degree of polymorphism detected in the three segregating populations was satisfactory. However, the percentages of independently segregating markers (from 63 to 81% according to the population) were lower than that (84%) obtained by Carboni et al. (2000) but in the same range as in sugarcane with AFLP markers (65%) (Garcia et al. 2006).

7.4.2. Linkage analysis

The percentages of linked markers (48, 63 and 75% of the independently segregating markers for the C1, C2 and UF maps) were higher than that reported by Garcia et al. (2006) in sugarcane with AFLP markers (33%) using the methodology developed by Wu et al. (2002a). Compared to UF, both the C1 and C2 maps were characterized by a lower percentage of linked markers and a higher number of small CGs (there were 6, 8 and 2 CGs including only 2 or 3 markers in C1, C2 and UF, respectively). These observations were explained by a lower probability of detecting linkage in C1 and C2, as reflected by the maximum recombination fractions detected at LOD $\geq$ 3.5 in each population (0.25 in both C1 and C2 and 0.34 in UF). Indeed, according to Maliepaard et al. (1997), the probability of detecting linkage was negatively affected in both C1 and C2 by their higher percentage of C.8 markers (63, 64 and 44% in C1, C2 and UF, respectively; Table 7.3) and smaller progeny size (77, 76 and 167 individuals, respectively) compared to UF.

The consecutive mapping of markers derived from the same primer combination as observed in few CGs may be due to the presence of repetitive

DNA sequences. In hemp, Peil et al. (2000) also observed clusters of AFLP markers along linkage groups. Similarly, a non-random distribution of AFLP loci amplified from distinct selective primer combinations across linkage groups was found by Rogers et al. (2007) in whitefish (*Coregonus clupeaformis*). They raised the possibility that the occurrence of repetitive AFLP sequences would be influenced by the presence of repetitive DNA in the genome, which would itself be partially due to transposable elements. In *Asparagus officinalis*, Reamon-Büttner et al. (1999) observed hybridisation signals that were present as clusters on chromosomes, suggesting that the AFLP fragments are part of repetitive sequence families. According to Sakamoto et al. (2005), multiple sequences encoding retrotransposable elements may exist ubiquitously in the hemp genome.

The C1, C2 and UF maps presented in this study included 92, 95 and 87 markers and covered a total distance of 1196.7, 878.2 and 1571.0 cM, respectively. The longer distance covered by the UF map could be attributed to its higher proportion in D1.13 and D2.18 markers. Indeed, the maximum recombination fraction that can be detected among the present marker types is highest between two D1.13 markers or two D2.18 markers (Maliepaard et al. 1997). All three maps were unsaturated, as indicated by the presence of unlinked markers or co-segregation groups that included only two or three markers. As a result, with the exception of the sex CGs (C1-7 and C2-14), the location of the CGs on autosomes or sex chromosomes cannot be ascertained.

7.4.3. Integrated map

The integrated map included 5 HGs, which consisted of 3 to 14 CGs derived from distinct maps and 7 CGs from only one map. However, although the number of identified groups was higher than the haploid number of chromosomes in hemp, it is possible that one or more chromosome(s) were not screened by the AFLP markers, given the unsaturated nature of the maps.

The number of markers that were found in two distinct maps and could be used as anchors to integrate the maps was low (from 12 to 23, depending on the populations). This was explained by the successive steps of marker selection. Indeed, the selection of segregating markers reduced the number of common markers among the three populations from 287 to 54, among which 19

segregated independently and only 5 were mapped in all three populations. Similarly, Waugh et al. (1997) identified only eight AFLP markers that segregated in three populations of barley among totals of 234, 194 and 376 mapped AFLP markers, while Hoarau et al. (2001) used a relatively low number of anchoring markers (45 AFLPs) to join two maps that included 887 AFLP and 408 RFLP markers in sugarcane.

The integration of the three maps revealed distinct conflicts among the maps, with respect to the markers ordering (as observed in HGII) or linkage (as raised between markers mapped together in the sex CG in C2 but in distinct putative CGs in UF). These inconsistencies were attributed to the low proportion of common markers between the maps and to the accuracy of the recombination fraction estimates. Indeed, the linkage between two markers and their alignment along the map can be affected by the absence of a given co-segregating marker, or, conversely, the presence of additional co-segregating markers. In addition, the accuracy of the recombination fraction estimates depends on the progeny size, cross types of the markers and the parental linkage phase between them, in addition to the recombination fraction itself in the progeny (Maliepaard et al. 1997). In particular, as discussed above, a high uncertainty encompasses the linkages among C.8 markers, and this can lead to difficulties in establishing the correct ordering of markers (Maliepaard et al. 1997). Waugh et al. (1997) attributed the ordering conflicts observed when merging three AFLP maps in barley to computational problems in estimating the most likely order.

7.4.4. Identification of sex-linked markers and CGs putatively located on the sex chromosomes

Four male-specific markers detected in the present study had sizes similar to male-specific AFLP markers reported in hemp: the markers 4_251, 4_276, 5_323 and 6_292 (not mapped) would correspond to ACC*AAG250, ACC*AAG275, ACC*AGA323 and ACC*ATC292, respectively, in Flachowsky et al. (2001). However, two markers identified as male-specific by Flachowsky et al. (2001) were not specific to the male plants in the present study (5_163 or ACC*AGA163, and 2_101 or AGA*AGG101).

A large proportion of markers (33%) was gathered in HGV, *i.e.*, the homology group that included the sex marker and CGs putatively located on the sex chromosomes. This observation could be partly explained by the use of AFLP primer combinations that proved to provide sex-specific markers, as detected by Flachowsky et al. (2001). In *Asparagus officinalis*, Reamon-Büttner and Jung (2000) noted that the AFLP technique is an efficient method to detect markers tighly linked to the sex compared to the RAPD and RFLP techniques. In *Carica papaya* (2n = 18), Liu et al. (2004) identified 1501 AFLP markers among which 23% were mapped in the co-segregation group that included the sex locus, *i.e.*, a higher proportion than expected if the markers were equally distributed along the 9 chromosomes (~11%). Based on these references, we assumed that the presence of discrepancies in the distribution of AFLP markers among chromosomes in favour of the sex chromosomes would be possible and could explain the observed high proportion of markers in the HG including the sex locus.

The detection of sex-linked markers allowed (i) the characterization of the structure of the sex CG in the dioecious maps and (ii) the identification of markers and CGs putatively located on the sex chromosomes in the monoecious map. As in Peil et al. (2003), we identified a region including markers of classes A, D and E closely linked to the sex locus (region 2), as well as a region including markers that completely recombined with the sex locus at an extremity of the sex CG (region 4). These observations confirmed the presence of common fragments between X and Y – represented by class D – and the existence of a pseudoautosomal region in the sex chromosomes of hemp as reported by Peil et al. (2003). However, our results differed from those of Peil et al. (2003) in the observation of higher recombination rates between the markers mapped in each of the four regions and sex. First, up to 5 recombinants were observed between class E and sex in C1 (Table 7.5). In contrast, Peil et al. (2003) detected only one such recombinant and therefore assumed that their class E markers could be located in the non-pairing portion of the Y chromosome. The recombinations observed between our class E markers and sex suggest that they are located in a region where pairing between X and Y chromosomes would be allowed. Secondly, the recombination rates found here between the pseudoautosomal markers and sex (≥ 0.41) were relatively high compared to the values of 0.25 and 0.27 reported in hemp by Peil et al. (2003) and Rode et al. (2005), respectively. Third, two additional regions were distinguished in the sex CGs (regions 1 and 3; Table 7.5). These regions included markers mapped on the X

chromosomes. Without considering the markers showing evident distorted segregations, the recombination rates with sex observed in regions 1 and 3 ranged from 0.11 to 0.19. This result suggests that the X and Y chromosomes of hemp recombine with each other between the sex locus and the pseudoautosomal region. The recombination of the X and Y chromosomes in the male parent as observed by Peil et al. (2003) and in the present study was supported by the pairing of both chromosomes at the short arm of the Y chromosome reported by Sakamoto et al. (2000).

The recombination rates observed between the sex chromosomes in hemp contrasted with the situation in *Silene dioica* and *Silene latifolia*, two well-studied closely related dioecious species with heteromorphic sex chromosomes (Filatov et al. 2001). Indeed, a recombination rate of 0.15 was reported between a pseudoautosomal marker and the differentiating portion of the Y chromosome in *S. dioica* (Di Stilio et al. 1998). In *S. latifolia*, pairing between the X and Y chromosomes during the male meiosis is confined to the tips, and recombination is absent from most of the Y chromosome, such that 90% of the Y chromosome is considered to be degenerated (Charlesworth 2002). In contrast, in *Papaya carica*, the size of the male-specific region of the Y chromosome was estimated to be only 10% of the chromosome size, suggesting the presence of an incipient sex chromosome evolution (Liu et al. 2004). Nevertheless, from the absence of hermaphrodite relatives in papaya, Charlesworth et al. (2005) raised the possibility of an ancient origin for the sex chromosomes; in this case, the small size of the non-recombining region would indicate that suppressed recombination can remain limited in extent and thus that the degeneration of the entire Y chromosome would not be inevitable. In hemp, Peil et al. (2003) concluded that a typical sex behaviour was present, *i.e.*, almost no recombination, from the observation of very few recombinants with the sex locus. The higher recombination rates observed here in the sex CGs (Table 7.5) suggest that the divergence between the X and Y chromosomes in hemp is not as advanced as would be expected from the results of Peil et al. (2003). According to Charlesworth et al. (2005), the suppression of recombination between the sex chromosomes in dioecious species would result from the presence of sex-determining genes and the evolution of Y-linked genes that benefit male but not female functions, both effects causing a selection against recombinants. In contrast to the dioecious species *S. latifolia*, in which the genetic basis of sex determination is strong and there is little evidence for lability or environmental effects (Ainsworth 2000), the diversity of intersexual forms observed in hemp

could indicate that hemp individuals bearing recombined sex chromosomes still have a relatively high adaptive value.

The present study reported the presence of sex-linked markers in monoecious hemp. Indeed, five markers mapped in the sex CG of a dioecious population were detected in the monoecious population (Table 7.5). Four of these were mapped on X chromosomes, including one in the pseudoautosomal region, and one was mapped on both X and Y chromosomes (class D). However, the sex chromosomes of monoecious hemp have the XX constitution (Menzel 1964; Faux et al. 2014). Therefore, the mapping of pseudoautosomal and class D markers in monoecious hemp suggests the presence of homologous fragments between the Y chromosome of dioecious hemp and the X chromosomes of monoecious hemp. Thus, it would be worthwhile to study the sexual phenotypes associated with these markers and with those mapped in CGs putatively located on the sex chromosomes.

7.5. CONCLUSIONS

We present the construction of three AFLP linkage maps of the sex expression in hemp, two derived from dioecious populations and one from a monoecious population. A previous study demonstrated that the sexual phenotype of monoecious hemp plants varies quantitatively among cultivars, suggesting that sex expression in monoecious hemp has a genetic basis (Faux et al. 2014). The present maps are still in their infancy and a larger number of primer combinations should be tested to obtain their saturation. However, although they are unsaturated, we believe that these maps could be of interest for identifying quantitative loci involved in the genetic determinism of sex expression in hemp because they include sex-linked markers, sex CGs and several CGs putatively located on the sex chromosomes. No such study has previously been performed. However, sych analysis faces two major difficulties: (i) the use of F_1 mapping populations derived from outbred parents and (ii) the complexity of the studied trait. Although various statistical solutions to the first obstacle have been developed (Knott and Haley 1992; Schafer-Pregl et al. 1996; Lin et al. 2003; Wu et al. 2007; van Ooijen 2009), the second issue requires dissecting the variability of the sexual phenotype observed in monoecious hemp.

This work was developed in two papers dedicated to the identification of QTLs linked to the sex expression in hemp.

Acknowledgments

We gratefully thank Dr H. Flachowsky and Dr A. Peil (Julius Kühn-Institut, Germany) for sharing their AFLP protocol and V. Van Hese and P. Van Thorre (UCL, Belgium) for their valuable help in the laboratory work and cultivation of hemp plants, respectively.

CHAPTER VIII

MODELLING APPROACH FOR THE QUANTITATIVE VARIATION OF SEX EXPRESSION IN MONOECIOUS HEMP

The present chapter has been submitted as a research paper under the title:

A.-M. Faux, P. Bertin [a]. Modelling approach for the quantitative variation of sex expression in monoecious hemp (*Cannabis sativa* L.). Submitted to Plant Breeding.

[a] Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve

ABSTRACT

Sex expression is of primary importance for the genetic improvement and production of monoecious hemp because masculinised and feminised phenotypes are associated with higher fungal sensitivity and seed yields, respectively. However, sex expression varies quantitatively among the plants and nodes and with time. Here, we developed eight variables that characterize the sex expression in monoecious hemp in order to dissect its genetic determinism.

The monoecy degree (MD), ranging from 1 (mostly male flowers) to 5 (mostly female flowers), was recorded for each node of 167 plants, at 6 times at 1-week intervals. Two types of longitudinal variables were constructed: 'synthesis' (mean MD and percentages of nodes of each MD) and 'structure'. The latter consisted of the parameters of a logistic curve describing MD as a function of the node position. An *r*-square of 0.97 was obtained between the estimated and observed MD values, and the logistic parameters were weakly correlated with each other and with the synthesis variables. Therefore, we conclude that the present modelling approach is relevant to the characterization of the sex expression in monoecious hemp.

Keywords: *Cannabis sativa* · Hemp · Logistic · Modelling · Monoecy · Quantitative trait · Sex expression.

8.1. INTRODUCTION

Hemp (*Cannabis sativa* L.) is used for a multitude of end products that are derived from its fibre, wooden core, seeds and cannabinoids (Struik et al. 2000). In addition, it is a low-input crop with low environmental impact (van der Werf 2004). According to van der Werf et al. (1996), hemp has the right profile to fit into a sustainable farming system. However, the production of hemp is greatly affected by its reproductive features. Both dioecious and monoecious cultivars exist; only the monoecious cultivars allow the mechanical harvest of both stems and seeds. However, sex expression in monoecious hemp varies quantitatively between the predominantly masculinised and the predominantly feminised extreme phenotypes. In addition, this is sensitive to environmental factors such as the photoperiod and nitrogen status (Freeman et al. 1980; Fournier and Beherec 2006; Truta et al. 2007). This quantitative variation of the sex expression has significant implications for both breeding and cultivation. The strongly masculinised plants are eliminated during the seed production process because they are less fruiting and more sensitive to fungal infections (Fournier and Beherec 2006). Despite its high environmental sensitivity, sex expression varies among monoecious cultivars, and higher seed yields are obtained from the early and mid-early feminised cultivars (Faux et al. 2013, 2014). The observation of genotypic variability for sex expression led us to assume that this trait has a genetic basis and, hence, could be investigated by searching for quantitative trait loci (QTL) associated with variation in sex-expression related variables (Faux et al. 2014).

QTL analyses of traits related to the sex expression have been conducted in distinct species including *Citrullus lanatus* (Prothro et al. 2013), *Cucumis melo* (Diaz et al. 2011), *Cucumis sativus* (Serquen et al. 1997; Fazio et al. 2003; Yuan et al. 2008; Miao et al. 2011) and *Vitis vinifera* (Marguerit et al. 2009). In these studies, the quantitative variation of the sex expression was characterized by the percentage of male, female or hermaphrodite flowers (Yuan et al. 2008; Prothro et al. 2013), or by the number or percentage of pistillate nodes (Serquen et al. 1997; Fazio et al. 2003). In monoecious hemp, the plants bear racemose inflorescences with male and female flowers arising in ratios that vary among nodes along the stem and with time, in addition to being affected by environmental factors. A feminisation of the monoecious hemp plants between

the start of flowering and fruit set was observed in previous field trials (Faux et al. 2013). Unlike in the family of *Cucurbitaceae*, counting the number of flowers of each sex (male or female) in hemp is made difficult by the small size of the flowers and the density of the inflorescence. Alternatively, the sex expression can be assessed at each plant node by using an ordinal scale based on the ratio between the female and male flowers (Sengbusch 1952). However, with the aim of investigating the genetic determinism of the sex expression in monoecious hemp, there is a need for phenotypic variables that efficiently summarise the variation of the sex expression among the nodes and/or with time. Such variables could be used to test the relationship between the phenotypic and genotypic variations in the framework of a QTL analysis. According to Ma et al. (2002) and Malosetti et al. (2006), integrating the effects of node and/or time on the sex expression in the QTL analysis would allow to consider the development of the sexual phenotype as a continuous process, thereby leading to more biologically-sound conclusions on the genetic basis of its variation.

In the present study, we constructed distinct variables that characterize the sex expression of monoecious hemp plants in order to dissect its genetic determinism. The originality of our approach is in the modelling of the sex expression along the stem of the monoecious plants and in the use of the parameters for characterization of the sex expression.

8.2. MATERIALS AND METHODS

8.2.1. Plant materials

Phenotypic data were collected in an F_1 segregating population of 167 individuals that were derived from a cross between two monoecious cultivars, 'Uso 31' and 'Fedora 17', used as male and female parents, respectively. The seeds were obtained from the Fédération nationale des Producteurs de Chanvre (France). 'Uso 31' has been described as being earlier and more masculinised than 'Fedora 17' (Faux et al. 2013, 2014).

8.2.2. Growth conditions

The plants were cultivated in a greenhouse. The seeds were sown in germination dimpled plates. After 10 days, the seedlings were transplanted into 30 cm height by 18 cm diameter pots in a greenhouse. The substrate consisted of a 3:1:1 mixture of soil, sand and loam to which 2.8 g per pot of Osmocote® (14:13:13 NPK with slow release of nutrients for up to six months) was added just before the plants were transplanted. The plants were watered by capillarity twice per week. The set-point temperatures were 25/20°C day/night. The photoperiod was set at 16 h during the entire trial period.

8.2.3. Phenotyping

The quantitative variation of the sex expression was characterized by the monoecy degree (MD), ranging from 1 to 5 according to the ratio between the female and male flowers (Sengbusch 1952; Table 8.1). MD was scored for each flowering node at six times during the flowering period, *i.e.*, 43, 50, 57, 64, 71 and 78 days after sowing. An average of 73 data points was recorded per plant, with each data point corresponding to the MD of a given node at a given time (Fig. 8.1a). Two types of variables were defined in order to summarise the sex expression of each monoecious plant (Table 8.2): 'synthesis' variables, including the mean monoecy degree and percentages of nodes having a given

MD, and 'structure' variables, describing the distribution of male and female flowers along the stem. Thus, the structure variables integrate the node position at which the MD was recorded, while the synthesis variables do not. All of the variables were longitudinal, *i.e.*, they included a series of six phenotypic values – one per observation time – for each plant.

Table 8.1 Sex expression scale for monoecious hemp based on Sengbusch (1952)

Monoecy degree	Sex ratio
1	80 - 100% of ♂ flowers (strongly masculinised node)
2	60 - 80% of ♂ flowers (masculinised node)
3	40 - 60% of ♂ and ♀ flowers
4	60 - 80% of ♀ flowers (feminised node)
5	80 - 100% of ♀ flowers (strongly feminised node)

Table 8.2 Eight variables that were constructed to characterize the sex expression of the monoecious hemp at a given observation time

Type [1]	Name	Description [2]
Synthesis	mMD	Mean MD
	%MD1	% of nodes with MD = 1
	%MD5	% of nodes with MD = 5
	%MDinter	% of nodes with 1 < MD < 5
Structure	log_k	Logistic-function parameter determining the curvature of the function.
	log_NDm	Logistic-function parameter accounting for the node at which the variation of MD along the stem is maximum.
	log_Mmin	Logistic-function parameter accounting for the minimum MD along the stem.
	log_Mmax	Logistic-function parameter accounting for the maximum MD along the stem.

[1] The structure variables consisted of the parameter estimates of a logistic function describing the variation of the monoecy degree (Table 8.2) along the stem (Equation 1).

[2] MD = monoecy degree (Table 8.2). For each plant, the monoecy degree was scored for each flowering node at six distinct observation times.

Synthesis variables

For each monoecious plant at a given time, *mMD* was the mean MD computed over all flowering nodes, and *%MD1*, *%MD5* and *%MDinter* were the percentages of the strongly masculinised nodes (MD = 1), strongly feminised nodes (MD = 5) and nodes with an intermediate MD (1 < MD < 5), respectively.

Structure variables

Similar to previous field observations (Faux et al. 2013), a process of feminisation of the plants with time was noted in the present study (Fig. 8.1). This process could be explained by two underlying time-dependent effects: (i) the proportion of female flowers increases with the order number of the nodes, and (ii) the proportion of female flowers at a given node increases with time. Because the first effect was the most significant one (Fig. 8.1b, c), we chose to model MD as a function of the plant node rather than as a direct function of the time. Based on our data, the variation of MD along the stem was assumed to fit the following logistic function:

$$y_{ijp} = M_{\min_{ij}} + \frac{M_{\max_{ij}} - M_{\min_{ij}}}{1 + e^{-k_{ij}(ND_p - ND_{m_{ij}})}} \quad (8.1)$$

where y_{ijp} is the MD at the pth node, ND_p, for the individual i at time j, $M_{\min_{ij}}$ and $M_{\max_{ij}}$ are the minimum and maximum MD, k_{ij} is a constant that determines the curvature of the pattern of the MD variation along the stem, and $ND_{m_{ij}}$ is the plant node at which the MD variation along the stem is the maximum. At node ND_m, MD is $(M_{\min} + M_{\max})/2$, *i.e.*, at the midpoint between $M_{\min}$ and $M_{\max}$. Therefore, the parameters k_{ij} and $ND_{m_{ij}}$ can be viewed as the rate of the feminisation process along the stem and the node at which half of the feminisation process had been completed, respectively, for an individual i at time j.

The parameters of the logistic function (1) were estimated by using the NLIN procedure in the SAS statistical package (SAS Institute Inc. 2012). For each plant, at each observation time, a start value was assigned to each parameter as follows. $M_{\min}$ and $M_{\max}$ were set to their respective observed

values. For k and ND_m, we first computed the MD difference (*dMD*) between all pairs of consecutive nodes. The start value of ND_m was the mean node at which *dMD* was the maximum. Based on comparison with the relative growth rate (RGR) of a growth curve (Yin et al. 2003), the relative rate of the MD variation (RMDR) is $k/2$ at ND_m. Therefore, the start value of k was twice the RMDR at ND_m, *i.e.*, 2max(*dMD*) divided by the mean MD of the nodes at which *dMD* was the maximum. In addition to the start values, bounds were provided for each parameter: $M_{\min} \geq M_{\min_{observed}}$, $M_{\max} \leq M_{\max_{observed}}$, $M_{\min} \leq M_{\max}$, $-8 \leq k \leq 8$, *i.e.*, the extreme values that can be theoretically reached by k (when $dMD = -4$ or 4 and mean MD = 1), and $7 \leq ND_m \leq 33$, *i.e.*, the lowest and highest observed flowering nodes. The logistic function (1) was fitted to a total of 930 profiles, with each of them corresponding to one individual at one observation time. However, ND_m was listed as missing for 182 profiles that displayed a constant MD along the stem because it has no fixed value in this case.

The correlations among the eight phenotypic variables that were constructed (Table 8.2) were assessed by Pearson correlation coefficients. Only significant correlations were reported in the text ($P < 0.001$).

8.3. RESULTS

The variation in time of the mean MD (*mMD*), percentage of strongly masculinised nodes (*%MD1*) and percentage of strongly feminised nodes (*%MD5*) reflected the feminisation of the plants that was observed during the flowering period (Fig. 8.1 and 8.2). *mMD* increased on average from 1.9 to 4.0 and *%MD5* from 11 to 64%, while *%MD1* decreased from 62 to 12% between 43 and 78 days after sowing. The percentage of nodes with an intermediate monoecy degree (*%MDinter*) remained constant over time, with a value of 25% on average (Fig. 8.2).

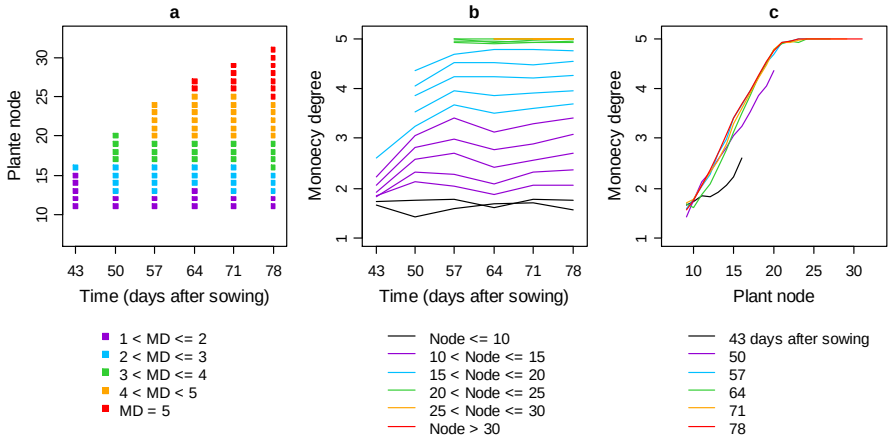

Fig. 8.1 Mean monoecy degree as a function of the observation time and the node in a monoecious hemp plant. (a) For each time value, each square corresponds to a distinct plant node of a mean monoecy degree that is indicated by the colour. (b) The evolution of the mean monoecy degree as a function of the time (*x*-axis) and plant node (colour). (c) The evolution of the mean monoecy degree as a function of the plant node (*x*-axis) and time (colour).

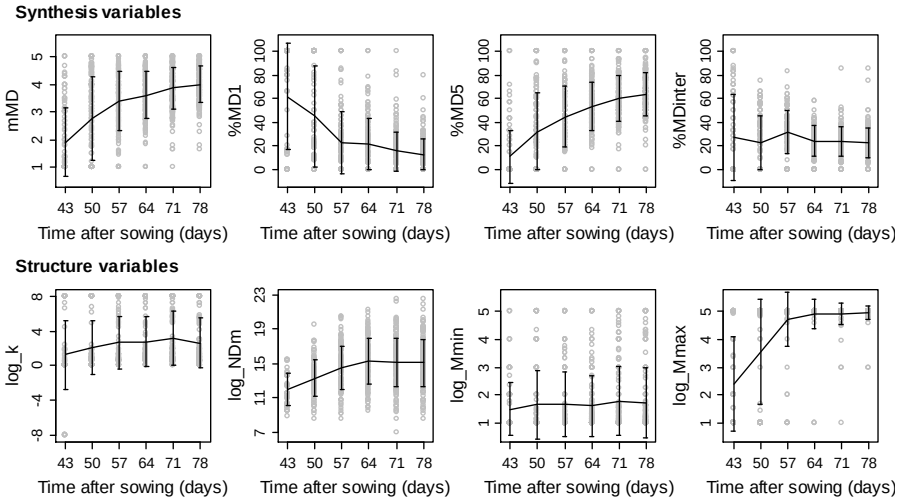

Fig. 8.2 The variation in time of eight phenotypic variables characterizing the quantitative variation of sex expression in the monoecious hemp (Table 8.2). Each point corresponds to one individual that was phenotyped at a given time. There were 118, 152, 163, 165, 166 and 166 individuals that were phenotyped at each of the six observation times, respectively, accounting for a total of 930 points. The line shows the mean ± sd.

The goodness of fit of the logistic function was established for modelling the variation of MD along the stem (Fig. 8.3). An *r*-square value of 0.97 was obtained between the estimated and observed monoecy degrees, regardless of the plant, node and time of observation (Fig. 8.4). The logistic parameters *log_k*, *log_ND*$_m$ and *log_M*$_{\max}$ varied with time (Fig. 8.2). The parameter *log_k*, which determines the curvature of the logistic function, increased on average from 1.3 to 3.2 between 43 and 71 days after sowing and then decreased to 2.6 at 78 days after sowing. *Log_ND*$_m$, the node at which the variation of MD along the stem is the maximum, increased on average from 12 to 15 between 43 and 64 days after sowing, and then remained constant. *Log_M*$_{\max}$, the maximum MD along the stem, increased on average from 2.4 to 5 between 43 and 64 days after sowing. In contrast, *log_M*$_{\min}$, the minimum MD along the stem, did not vary with time but rather maintained an average value of 1.7.

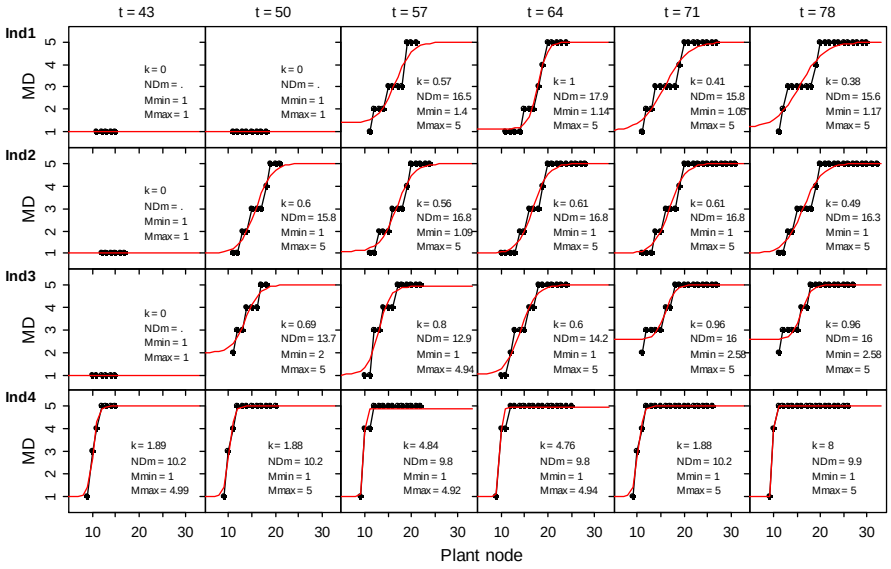

Fig. 8.3 The fit of a logistic curve (equation 8.1, red line) describing the monoecy degree as a function of the node position in four distinct monoecious hemp plants (Ind1 to Ind4) at six distinct observation times (t = 43 to 78 days after sowing). The four logistic parameters (k, ND_m, $M_{\min}$ and $M_{\max}$) are defined in Table 8.2.

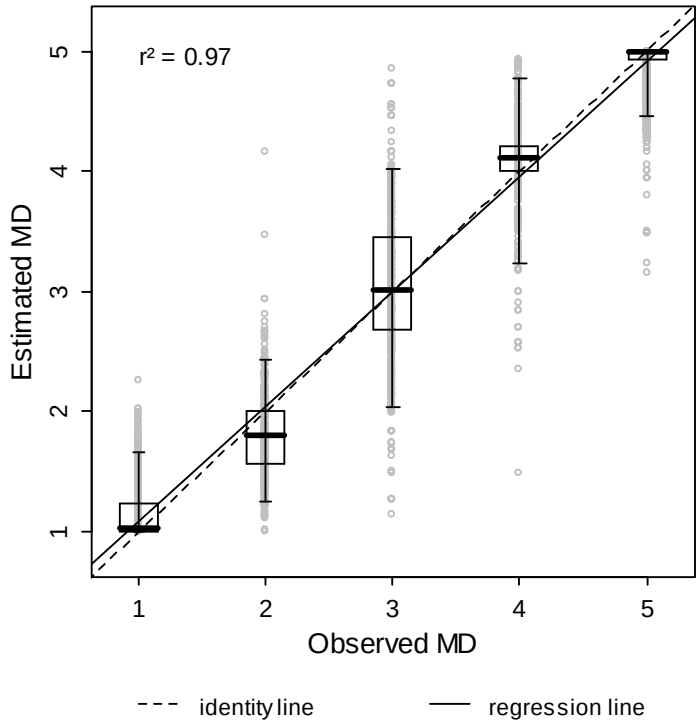

Fig. 8.4 Estimated *vs.* observed values of the monoecy degree (MD). The estimated values of the MD were obtained by describing the MD as a logistic function of the node position at each observation time (equation 8.1). The figure includes a total of 12,152 points, with each point corresponding to one node of a given plant at a given observation time. The boxplots indicate the 0.5, 0.25, 0.5, 0.75 and 0.95 quantiles of the estimated values of the MD.

As expected, the synthesis variables *mMD*, *%MD1* and *%MD5* were highly correlated with each other. *mMD* and *%MD5* were positively correlated with each other (r = 0.94), and both were negatively correlated with *%MD1* (r = -0.94 and -0.8, respectively). *%MDinter* was weakly correlated with the other variables (r = 0.18, -0.5, 0.12 and 0.36 with *mMD*, *%MD1* and *%MD5* and log_M_{max}, respectively). In contrast, all of the logistic parameters were weakly correlated with each other ($r \leq 0.32$ in absolute value) as well as with the synthesis variables. Log_M_{min} and log_M_{max} were relatively higher in more feminised plants, being positively correlated with *mMD* (r = 0.54 and 0.8,

respectively), while log_ND_m tended to be higher in less feminised plants, being negatively correlated with *mMD* (r = -0.5). The correlations involving *log_k* were low ($r \leq 0.26$ in absolute value).

8.4. DISCUSSION

In the present study, we developed eight variables that dissect the quantitative variation of the sex expression in monoecious hemp. Four of the variables consisted of the parameters of a logistic curve describing the distribution of the male and female flowers as a function of the node position along the stem. The logistic function is commonly used to describe the growth and development processes of living organisms over time (Ma et al. 2002). This function has been applied to identify QTLs for stem growth in poplar (Ma et al. 2002; Wu et al. 2004) and the senescence stage of leaves in potato (Malosetti et al. 2006). In the present situation, the use of a logistic function assumed the presence of asymptotic minimum and maximum values of monoecy degree (MD =1 and MD = 5, respectively). We think that this assumption makes sense for modelling the sex expression in monoecious hemp given the ability of the monoecious plants to produce both types of flowers and to modify their sex expression in response to environmental factors, so that the extreme male and female phenotypes would never be reached. In addition, the sex expression was described as a function of the node position along the stem rather than directly as a function of time. From a modelling point of view, this approach enables a link between the development of the sexual phenotype and plant growth through the process of node formation.

The logistic modelling of the sex expression resulted in satisfactory estimations of the variation of monoecy degree along the stem (Fig. 8.3 and 8.4). Therefore, we conclude that the present modelling approach is relevant for characterizing the quantitative variation of the sex expression in monoecious hemp plants. Moreover, low correlations were observed between the synthesis variables and the logistic parameters, suggesting that both the synthesis variables and the logistic parameters should be considered for dissecting the genetic basis of sex expression.

With the aim of investigating the genetic determinism of the sex expression, the step following the construction of the phenotypic variables consists of the expression of these variables in terms of QTLs. A such two-step procedure has been applied for QTL mapping for leaf growth in rice (Wu et al. 2002b), leaf elongation in maize (Reymond et al. 2003) or flowering time in barley (Yin et al. 2005). Here, due to the availability of repeated observations, the QTL analysis of each variable will be performed by allowing non-null correlations between the values that are estimated at distinct times for a given plant. To this end, the use of mixed models will be beneficial because of their great flexibility for incorporating co-variables into the analysis and for adequately modelling the residual genetic variation (Malosetti et al. 2006; Boer et al. 2007; Pastina et al. 2012).

Acknowledgements

We thank the platform SMCS (Support in Methodology and Statistical Computation) of the UCL for their valuable advice. This work was supported by the Fonds de la Recherche Scientifique [grant number 1.5.279.08].

CHAPTER IX

QTL ANALYSIS FOR THE SEX EXPRESSION, EARLINESS AND YIELDS IN STEM AND SEED IN HEMP

The present chapter is a research paper in preparation.

A.-M. Faux, P. Bertin [a].

[a] Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve

ABSTRACT

Hemp (*Cannabis sativa* L.) includes both dioecious and monoecious cultivars with hetero- and homomorphic sex chromosomes, respectively, and displays a highly variable sex expression. The sex expression is of primary importance for hemp cultivation: the mechanical harvest of both stem and seed is feasible only with monoecious cultivars, the seed yield varies with both sex expression and earliness, while the stem yield varies with earliness. In this paper, we dissected the genetic basis of the sex expression, earliness and stem and seed yields in hemp by the identification of QTLs.

Three unsaturated AFLP linkage maps derived from two dioecious (C1 and C2) and one monoecious (UF) F_1 populations were used. The sex expression in C1 and C2 was characterized by the plant sex – male or female – and percentage of nodes bearing flowers of the sex opposite to the dominant sex of the plant. In UF, it was characterized by 'synthesis' (mean sex expression, percentages of nodes with a given sex expression) and 'structure' variables. The latter consisted of the parameters of a logistic function modelling the sex expression along the stem. A total of 6, 8 and 13 QTLs associated with variation in sex expression was identified in C1, C2 and UF, respectively. In UF, 9 of them were detected for the model parameters. Twelve QTLs for the sex expression in C1 or C2 and 3 in UF were mapped in co-segregation groups putatively located on the sex chromosomes. Seven markers segregating with sex in C1 or C2 were associated with variation of the sex expression in UF. Four QTLs were detected for the time delay between the starts of male and female flowering in UF. Two of them were closely linked to QTLs for the sex expression. Five QTLs were identified for the stem yield, and 4 for the seed yield regardless of the population. One major QTL for the seed yield (r^2 = 17.8%) was mapped close to a marker segregating with sex.

This study supports the relevance of a quantitative approach for investigating the genetic determinism of the sex expression in hemp. In this context, dissecting the genetic basis of sex expression, earliness and yields in stem and seed might be valuable for both fundamental and agronomical purposes.

Keywords: *Cannabis sativa* · Dioecy · Earliness · Monoecy · QTL · Sex chromosomes · Sex expression · Seed yield · Stem yield.

9.1. INTRODUCTION

Hemp (*Cannabis sativa* L.) is a multi-use crop cultivated for a wide range of end products derived from its fibre, seeds or wooden core, such as textile, paper pulp, cooking oil, livestock seeds and bedding, cosmetics, biocomposites for the automobile industry and lime-hemp concretes for the building sector (Struik et al. 2000; Ranalli and Venturi 2004). However, the production of hemp is greatly affected by its sensitivity to the photoperiod and its reproductive features. Firstly, the species is a short-day plant, with a critical photoperiod of approximately 14 hours (Borthwick and Scully 1954; Lisson et al. 2000b; Amaducci et al. 2008a). By modulating the flowering time, the photoperiodic conditions significantly affect the crop yield (van der Werf et al. 1994). Higher stem yields are obtained with late-flowering cultivars (van der Werf et al. 1994; Meijer et al. 1995; Lisson and Mendham 2000). Secondly, hemp is a naturally dioecious plant with high plasticity of sexual phenotype. The monoecious hemp cultivars facilitate the simultaneous production of stems and seeds, thereby offering a dual income opportunity to the farmer. However, the monoecious state varies from predominantly male to predominantly female extreme phenotypes. Strongly masculinised plants are less fruiting and more prone to developing fungal infections from *Botrytis cinerea*, and must therefore be eliminated during the seed production process (Fournier and Beherec 2006). Despite its high plasticity, the sex expression, defined as the ratio between the female and male flowers, varies significantly among cultivars, suggesting that it has a genetic basis and might be bred (Faux et al. 2013, 2014). Besides, higher seed yields were obtained with the early and mid-early feminised cultivars (Faux et al. 2013). Therefore, investigating the genetic determinism of the sex expression in monoecious hemp appears valuable for both hemp breeding and cultivation.

Hemp is a diploid species (2n=20) including heteromorphic sex chromosomes and characterized by sexual dimorphism. The male plants of dioecious hemp have XY sex chromosomes, and the female plants, XX (Yamada 1943 cited by Sakamoto et al. 1995). According to Westergaard (1958) and Ainsworth (2000), the sex in dioecious hemp would be controlled by an X-to-autosomes equilibrium rather than by a Y-active system. The inflorescence of the male plants consists of panicles with generally few or no leaves and bears

flowers with five stamens with large and longitudinally dehiscing anthers (Mohan Ram and Nath 1964). The male plants flower and senesce earlier than the female ones (Bocsa and Karus 1998; Struik et al. 2000). The female inflorescence is a raceme that develops at the apex of the plant or at the axils of leaves or lateral branches (Moliterni et al. 2004) and bears flowers consisting of an ovary enclosed in a leafy glandular bract with a reduced style and a bifid stigma (Mohan Ram and Nath 1964). Despite the presence of sex chromosomes, the production of flowers of the opposite sex is sporadically observed on dioecious hemp plants and is influenced by external factors, such as the photoperiod or hormonal treatments (Freeman et al. 1980).

Monoecious hemp has homomorphic, XX, sex chromosomes (Menzel 1964; Faux et al. 2014). However, the sex expression is highly variable, and the genetic determinism of its variation is unknown. The plants of monoecious hemp bear inflorescences similar to those of female plants. However, male and female flowers arise in ratios that vary among nodes along the stem, plants, cultivars (Faux et al. 2014), external factors (Arnoux et al. 1966a, b; Arnoux and Mathieu 1969; Truta et al. 2007) and time. Indeed, the plants of monoecious hemp tend to feminize from the start of flowering to fruit set. The quantitative variation of the sex expression along the stem of monoecious hemp plants has been successfully described as a logistic function of the node position (Chapter VIII). In addition, the rankings of cultivars according to their sex expression and earliness were consistent with each other for most cultivars (Faux et al. 2013, 2014), suggesting that the activity of genes regulating the flowering time may be involved in the determinism of the sex expression of monoecious hemp cultivars. The observation of genotypic variability for the sex expression in monoecious hemp led us to assume that the trait can be inherited and therefore could be investigated through the search for QTLs (Faux et al. 2014).

Molecular mapping in F_1 progenies derived from outbred lines has been suggested in hemp given the very high degree of polymorphism and heterozygosity of the species (Forapani et al. 2001). However, using an F_1 derived from outbred lines makes the linkage analysis and QTL mapping much more challenging than with segregating populations derived from inbred lines, such as F_2 or backcross populations. Indeed, the use of outbred lines involves the presence of a variable number of segregating alleles at each locus and, therefore, a mixture of markers with distinct segregating patterns, and the ignorance of the linkage phases between markers in the parents (Wu et al.

2002a; Lin et al. 2003). Distinct strategies have been developed for QTL mapping in outcrossing species (Schäfer-Pregl et al. 1996; Song et al. 1999; Johnson et al. 1999; Lin et al. 2003; Pastina et al. 2012). In the present study, the sporadic production of flowers of the opposite sex by dioecious plants (Flachowsky et al. 2001; Mandolino et al. 2002) and the variation of the sex expression of monoecious plants with time (Faux et al. 2013, 2014) suggest to integrate of a 'time' effect in the QTL analysis of the sex expression. In this context, the use of mixed models for QTL mapping is advantageous because it allows a great flexibility for incorporating co-variables into the analysis and adequately modelling the residual genetic variation (Malosetti et al. 2004, 2006, 2007; Boer et al. 2007; Pastina et al. 2012).

The main objective of the present paper was to identify QTLs associated with the sex expression in hemp. In addition to the sex expression, the earliness and stem and seed yields were recorded because of the existing relationships between sex expression, earliness and seed yields, on the one hand, and between stem yields and earliness, on the other hand (de Meijer and Keizer 1994; van der Werf et al. 1994; Faux et al. 2013, 2014). The present study was carried out on one monoecious and two dioecious F_1 segregating populations for each of which a linkage AFLP map was constructed (Chapter VII). Two specific objectives were pursued: (i) the search for QTLs linked to the sex expression, earliness and yields by interval mapping using a mixed model approach in each segregating population independently and (ii) the comparison of the results across the three populations.

9.2. MATERIALS AND METHODS

9.2.1. Genetic material and growth conditions

Phenotypic and molecular data were collected in three F_1 segregating populations. The first two populations (C1 and C2) derived from a cross between a female and a male plant from the dioecious hemp cultivar 'Carmagnola', and the third one (UF) was obtained from a cross between two monoecious cultivars, 'Uso 31' and 'Fedora 17', used as male and female parents, respectively. Seeds were obtained from Assocanapa (Italy) for the cultivar 'Carmagnola' and from the Fédération nationale des Producteurs de Chanvre (France) for both monoecious cultivars. 'Uso 31' has been described as earlier and more masculinised than 'Fedora 17' (Faux et al. 2013, 2014). The C1, C2 and UF populations included 77 (43 females and 34 males), 76 (48 females and 28 males) and 167 individuals, respectively.

The plants were cultivated in a greenhouse. Sowing was performed on 6 October 2009 for the UF population and on 10 September 2010 for both C1 and C2 populations. The set-point temperatures were 25/20°C day/night. The photoperiod was set at 16 h during the entire trial period for the UF population, and 16 h during 69 days and 8 h thereafter in order to promote flowering for both C1 and C2 populations. The growth conditions were further detailed in Chapter VII.

9.2.2. Phenotyping

The sex expression, earliness and yields in stems and seeds were recorded in each population and summarised by distinct variables (Table 9.1).

(i) Sex expression

The sex expression was characterized distinguishingly according to the reproductive morphology of the plants, dioecious or monoecious. In the dioecious populations, the plants were phenotyped at three times, *i.e.*, 77, 84 and

98 days after sowing. The sex of each plant *(sex)* was recorded as a binary variable, *i.e.*, male or female. However, flowers of the sex opposite to the dominant sex of the plant were observed on some plant nodes. For each male plant at each observation time, *%F-*♂ was the percentage of nodes bearing female flowers. Similarly, for each female plant at each observation time, *%M-*♀ was the percentage of nodes bearing male flowers.

In the monoecious population, the continuous variation of the sex expression was characterized by the monoecy degree (MD), a five-point scale ranging from 1 (mostly male flowers) to 5 (mostly female flowers) (Sengbusch 1952). MD was scored for each flowering node at six times, *i.e.*, 43, 50, 57, 64, 71 and 78 days after sowing. Eight phenotypic variables were used to summarise the sex expression of each monoecious plant at each observation time (Table 9.1; see Chapter VIII for details). These variables are of two types: 'synthesis' and 'structure'. The synthesis variables include the mean MD (*mMD*) and percentages of strongly masculinised nodes (*%MD1*), strongly feminised nodes (*%MD5*) and nodes with intermediate MD (*%MDinter*). The structure variables differ from the synthesis ones by the integration of the node position at which MD is recorded. The structure variables consist of the parameters of a logistic curve describing MD as a function of the node position along the stem (Chapter VIII). These parameters are *log_k*, which determines the curvature of the logistic function, log_ND_m, which accounts for the node at which the variation of MD along the stem is maximum, and log_M_{min} and log_M_{max}, which are the minimum and maximum MD along the stem, respectively. Because of the feminisation of the plants observed along the stem (Chapter VIII), *log_k* and log_ND_m can be viewed as the rate of the feminisation process along the stem and node at which half of the feminisation process has been completed, respectively.

With the exception of *sex*, all variables relative to the sex expression were longitudinal, *i.e.*, they included a series of phenotypic values – one per observation time – for each plant.

Table 9.1 Phenotyped traits and phenotypic variables used for the QTL analysis

Trait	Pop. [1]	Variable			
		Name	Longitu-dinal [2]	Type [3]	Description [4]
(1) Sex expression	Dioecious	sex	No	-	Sex of the plant, male or female
	Dioecious, ♂ plants	%F-♂	Yes	-	% of nodes bearing female flowers on male plants
	Dioecious, ♀ plants	%M-♀	Yes	-	% of nodes bearing male flowers on female plants
	Monoecious	mMD	Yes	Synthesis	Mean MD
		%MD1	Yes		% of nodes with MD = 1
		%MD5	Yes		% of nodes with MD = 5
		%MDinter	Yes		% of nodes with 1 < MD < 5
		log_k	Yes	Structure	Logistic-function parameter determining the curvature of the function
		log_NDm	Yes		Logistic-function parameter accounting for the node at which the variation of MD along the stem is maximum
		log_Mmin	Yes		Logistic-function parameter accounting for the minimum MD along the stem
		log_Mmax	Yes		Logistic-function parameter accounting for the maximum MD along the stem
(2) Earliness	Monoecious	tf	No	-	Time at which closed male flowers or white styles at leaf axils are easily visible
		ΔMF	No	-	Time difference between the starts of female and male flowering (time of start of female flowering - time of start of male flowering)
(3) Yields	Dio- and monoecious	DMstem	No	-	Dry matter yield in stem (g)
		DMseed	No	-	Dry matter yield in seeds (g)

[1] Population(s) in which the trait was scored.
[2] The longitudinal variables were determined at several times and included a series of distinct values – one per observation time – for each plant .
[3] The structure variables integrated the node position at which MD was recorded, while the synthesis variables did not. They consisted of the four parameters of a logistic curve describing MD as a function of the node position along the stem (Chapter VIII).
[4] MD = monoecy degree.

(ii) Earliness

The earliness was characterized by the time of first flower appearance *(tf), i.e.*, the time at which closed male flowers or white styles at leaf axils were easily visible. However, *tf* was analysed for the monoecious population only because all dioecious plants flowered simultaneously as a result of the photoperiod shortening. The starts of male and female flowering were observed independently, and the time difference between both of them was computed (*ΔMF*, in days).

(iii) Yields

The harvest was performed 141 days after sowing for both dioecious populations and 139 days after sowing for the monoecious population. The stems and seeds were dried at 70°C for 72 hours to compute the dry matter yields.

9.2.3. Molecular data and linkage maps

The molecular data were provided by AFLP markers (Chapter VII). A linkage map was constructed for each of the three populations. DNA restriction was performed with the *Hind*III and *Mse*I enzymes. The pre-selective amplifications were performed with one selective nucleotide (A), and the selective amplifications with pairs of primers with three selective nucleotides. Eight primers pairs were tested, and the detected markers were labelled according to the number of the primer combination followed by the molecular weight of the corresponding DNA fragment (Chapter VII). The AFLP markers reveal a polymorphism of presence/absence with 1:1 or 3:1 segregating patterns depending on the presence of the allele in single copy in only one parent or in both of them. According to the notation of Wu et al. (2002a), the markers segregating 1:1 received the ‘D1.13’ or ‘D2.18’ cross types following their exclusive presence in the female and male parent, respectively, and the markers detected in both parents and segregating 3:1 received the ‘C.8’ cross type.

The C1, C2 and UF linkage maps included 93, 92 and 86 linked markers, respectively, accounting for a total of 225 distinct markers. The markers were distributed along 11, 16 and 10 co-segregation groups (CGs) in C1, C2 and UF, respectively. All three maps were unsaturated. Their integration through anchoring markers resulted in 5 homology groups (HGs) including from 17 to 73 markers and 7 single CGs including 2 or 3 markers only. In each dioecious map, a phenotypic marker accounting for the sex of the plants – male or female – was mapped. The CG including the 'sex' marker was referred to as 'sex CG' and assumed to be located on the sex chromosomes. The HG that included the sex CGs (HGV) was the largest one, with 33% of the mapped markers. The CGs that were gathered in this HG in addition to the sex CGs were considered as 'putative sex CGs'.

9.2.4. QTL analysis

The search for QTLs linked to the earliness, sex expression and yields was performed in each segregating population independently. Putative QTLs were firstly identified by interval mapping (Lander and Botstein 1989). Then, they were tested in a multiple regression, similarly to the composite interval mapping (Zeng 1994) and multi-QTL mapping (Jansen and Stam 1994) approaches and as applied in an F_1 segregating population by Pastina et al. (2012).

(i) Computation of genetic predictors

Genetic predictors were computed according to Lynch and Walsh (1998) and Pastina et al. (2012). Consider two outbred parental lines denoted as *P* and *Q*, each containing two homologous chromosomes. Suppose a QTL, with alleles P_1 and P_2 in parent *P* and Q_1 and Q_2 in parent *Q*. The segregation of the QTL in the progenies will result in four possible genotypes, P_1Q_1, P_1Q_2, P_2Q_1 and P_2Q_2 following a 1:1:1:1 segregation ratio. Three distinct QTL effects involving these four genotypic classes can be estimated: two additive effects, α_P in parent *P* and α_Q in parent *Q*, and one dominance effect, δ_{PQ}, resulting from an intra-locus interaction between the additive effects in each parent (Lin et al. 2003a; Pastina et al. 2012). Genetic predictors for the additive and dominance QTL effects were computed for a grid of evaluation points along the genome

and then introduced as explanatory variables in the QTL models (see below). For individual *i* and evaluation point *w* of the genome, the genetic predictors for the α_P, α_Q and δ_{PQ} QTL effects are, respectively:

$$x_{P_{iw}} = p(P_1Q_1|\mathbf{M}_i) + p(P_1Q_2|\mathbf{M}_i) - p(P_2Q_1|\mathbf{M}_i) - p(P_2Q_2|\mathbf{M}_i) \quad (9.1a)$$

$$x_{Q_{iw}} = p(P_1Q_1|\mathbf{M}_i) - p(P_1Q_2|\mathbf{M}_i) + p(P_2Q_1|\mathbf{M}_i) - p(P_2Q_2|\mathbf{M}_i) \quad (9.1b)$$

$$x_{PQ_{iw}} = p(P_1Q_1|\mathbf{M}_i) - p(P_1Q_2|\mathbf{M}_i) - p(P_2Q_1|\mathbf{M}_i) + p(P_2Q_2|\mathbf{M}_i) \quad (9.1c)$$

where $p(P_1Q_1|\mathbf{M}_i)$, $p(P_1Q_2|\mathbf{M}_i)$, $p(P_2Q_1|\mathbf{M}_i)$ and $p(P_2Q_2|\mathbf{M}_i)$ are the multipoint probabilities of having the QTL genotype P_1Q_1, P_1Q_2, P_2Q_1 and P_2Q_2 respectively, conditional upon the marker information $\mathbf{M}_i$ in a particular co-segregation group (Haley and Knott 1992; Martínez and Curnow 1992; Pastina et al. 2012). The conditional multipoint probabilities were calculated via hidden Markov chain model (OneMap package, Margarido et al. 2007) for each marker position and evaluation point with a step size of 1cM along the genome. Totals of 1286, 964 and 1654 points were evaluated in the C1, C2 and UF maps, respectively.

(ii) Identification of putative QTLs by interval mapping (IM)

For notation purpose, the random variables were underlined according to Boer et al. (2007). The presence of a putative QTL was assessed by testing the effect of the genetic predictors on each phenotypic variable at each position of the genome. The analysis of the non-longitudinal variables (Table 9.1) was performed as follows:

$$\underline{y}_i = \mu + x_{P_{iw}}\alpha_{P_w} + x_{Q_{iw}}\alpha_{Q_w} + x_{PQ_{iw}}\delta_{PQ_w} + \underline{e}_i \quad (9.2)$$

where $\underline{y}_j$ is the phenotypic value of individual *i*, μ is the overall mean phenotypic value, α_{P_w}, α_{Q_w} and δ_{PQ_w} are the effects of the additive genetic predictors for parent *P* and *Q* and dominance genetic predictor, respectively, at evaluation point *w* (Pastina et al. 2012) and $\underline{e}_i$ is the residual error associated with individual *i*, normally distributed with mean 0 and variance σ^2. The QTL analysis of *sex*, a binary variable, was performed by testing model (9.2) with the

GLIMMIX procedure with binomial distribution for the response variable and logit link function in SAS (SAS Institute Inc. 2012). The QTL analysis of all other non-longitudinal variables (Table 9.1) was conducted by the GLM procedure (SAS Institute Inc. 2012).

For the longitudinal variables (Table 9.1), the QTL analysis was performed by allowing non-null correlations between the observations made at distinct times on a same plant. To this purpose, the variance-covariance (VCOV) matrix of the residual errors was modelled according to the observation time by applying a first-order autoregressive structure [AR(1)]. This structure assumes that, for a given plant, the variances of observations made at distinct times are all equal to each other, while the covariance between them decreases exponentially with their separation in time (Brown and Prescott 2006). The used model was as follows:

$$\underline{y}_{ij} = \mu + x_{P_{iw}} \alpha_{P_w} + x_{Q_{iw}} \alpha_{Q_w} + x_{PQ_{iw}} \delta_{PQ_w} + \underline{e}_{ij} \qquad (9.3)$$

where $\underline{y}_{ij}$ and $\underline{e}_{ij}$ are the phenotypic value and residual error of individual i at time j, respectively, $\underline{e}_{ij} \sim N(0, \sigma^2)$, and $\text{cov}(y_{ij_a}, y_{ij_b}) = \text{cov}(e_{ij_a}, e_{ij_b}) = \sigma^2 \rho^{|j_a - j_b|}$, with j_a and j_b being two observation times and ρ the correlation between the phenotypic values at times j_a and j_b. The model (9.3) was tested by using the MIXED procedure with the REPEATED statement including type = AR(1) and subject = individual in SAS (SAS Institute Inc. 2012).

Similarly to Pastina et al. (2012), three successive searches for QTLs were performed for each phenotypic variable: (i) a search for additive effects from parent *P* by using the IM model – model (9.2) or (9.3) for the non-longitudinal and longitudinal variables, respectively – including only the α_P effect, (ii) a search for additive effects from parent *Q* by using the IM model including only the α_Q effect, and (iii) a search for dominance effect by using the IM model including all three effects, but only the δ_{PQ} effect was tested. We considered that a putative QTL was found at position *w* if α_P, α_Q or δ_{PQ} was significantly different from 0 at $P < 0.01$. The putative QTLs that corresponded to a local minimum of *p*-value (or maximum likelihood ratio) were retained for the multi-QTL analysis. For the non-longitudinal variables, the percentage of variation (r^2) explained by each additive effect was obtained from the fit of

model (9.2). In contrast, for the longitudinal variables, this was not possible due to the distinct method used to fit model (9.3) – *i.e.*, restricted maximum likelihood rather than least-squares (SAS Institute Inc. 2012).

(iii) Identification of QTLs by multi-QTL analysis

For each phenotypic variable, the significant genetic effects of each putative QTL were tested together in an additive multiple regression. For each non-longitudinal variable, the following model was used:

$$\underline{y}_i = \mu + \sum_{w\in(W_P,W_{PQ})} x_{P_{iw}}\alpha_{P_w} + \sum_{w\in(W_Q,W_{PQ})} x_{Q_{iw}}\alpha_{Q_w} + \sum_{w\in W_{PQ}} x_{PQ_{iw}}\delta_{PQ_w} + \underline{e}_i \quad (9.4),$$

where W_P, W_Q and W_{PQ} are the sets of positions along the genome accounting for a putative QTL with α_P, α_Q and δ_{PQ} effect, respectively. Both additive genetic predictors at the positions accounting for a putative QTL with dominance effect (W_{PQ}) were integrated into model (9.4) for a more powerful QTL detection (Pastina et al. 2012). For the longitudinal variables (Table 9.1), the VCOV matrix of residual errors was modelled by applying the AR(1) structure in model (9.4), similarly to model (9.3).

A multi-QTL model was constructed for each phenotypic variable through a procedure of backward selection of genetic effects starting from model (9.4). At each step, the model (9.4) was tested, and the effect associated with the highest non-significant (P >0.05) p-value was removed until no effect was associated with a non-significant p-value. During all steps, both additive effects of each putative QTL with dominance effect were maintained in the model regardless of their significance as far as δ_{PQ} was significant. The resulting model was referred to as multi-QTL model. The genomic positions selected in the multi-QTL model were considered to account for QTLs. The effect and standard error of each QTL were estimated from the multi-QTL model (Malosetti et al. 2006).

9.2.5. Cross-populations analysis

A cross-population (CP) analysis was conducted in order to assess the consistency of the effect of putative QTLs across the populations. This analysis was restricted to the putative QTLs that were detected at a locus corresponding to a marker. Indeed, the marker loci were the only positions that could be assumed as homologous between the distinct linkage maps. To this purpose, the markers accounting for a putative QTL for at least one phenotypic variable in at least one of the three populations were retained, and the effect of each of them was tested across the populations by single-marker analysis.

For *DMstem* and *DMseed*, which were observed in all three populations, a 'population' cofactor was incorporated into the analysis. The CP analysis of yields was performed as follows:

$$y_{ik} = \mu + \gamma_k + (\beta + \gamma\beta_k) \times M + e_{ik} \qquad (9.5)$$

where y_{ik} is the yield of individual i from population k, μ is the mean value over the three populations, γ_k is the effect of population k, β is the mean effect of the marker and $(\beta\gamma)_k$ its effect in population k, M is -1 or 1 according to the presence/absence of the marker in the ith individual of population k, and e_{ik} is the residual error associated with individual i in population k. The markers with a significant effect on *DMstem* or *DMseed* across the three populations were retained.

The remaining phenotypic variables – sex expression and earliness – varied depending on the reproductive morphology of the plants. The CP analysis of these variables was performed in both dioecious populations together by considering a 'population' cofactor, on the one hand, and in the monoecious population, on the other hand. The effect of the markers on *sex* was tested by using model (9.5). For the longitudinal variables (*%F-*♂ and *%M-*♀), the VCOV matrix of residual errors was modelled by applying the AR(1) structure in model (9.5) similarly to model (9.3). The CP analysis of the earliness and sex-expression related variables in the monoecious population was conducted similarly but without the 'population' cofactor. The GLM, GLIMMIX and MIXED procedures in SAS were used to carry out the SM analyses of each

phenotypic variable as previously described for the QTL analysis by interval mapping (SAS Institute Inc. 2012).

9.3. RESULTS

9.3.1. Phenotypic description

(i) Dioecious populations (C1 and C2)

In C1, 10 plants (7 male and 3 female) of a total of 77 showed flowers of the opposite sex. In C2, there were 8 such plants (4 male and 4 female) out of 76. The mean percentages of nodes bearing flowers of the opposite sex computed over both plants and observation times were low, being 1.1 and 3.1% for *%M-*♀ in C1 and C2, respectively, and 2.9% for *%F-*♂ in both populations. However, they were highly variable, ranging from 0 to 100% according to the plant and observation time, although no trend of variation with time was observed.

The dry matter yields in stem and seed were congruent between both dioecious populations, with on average 7.5 ± 2 and 8.5 ± 2.5 g plant^{-1} in stem and 3.1 ± 0.8 and 3.4 ± 1.2 g plant^{-1} in seeds in C1 and C2, respectively (mean ± sd). The stem yields were on average higher in the female plants than in the male ones (on average 7.6 and 7.2 g plant^{-1} in C1, 8.9 and 7.7, g plant^{-1} in C2), although the correlations between stem yield and sex were low (NS in C1; r = -0.234, P <0.05 in C2). The stem and seed yields were positively correlated with each other in the female plants of both populations (r = 0.415 and 0.513 in C1 and C2, respectively).

(ii) Monoecious population (UF)

The sex expression of the monoecious plants was described in Chapter VIII. The vast majority of the monoecious plants (91%) started to flower between 43 and 50 days after sowing, with average 45.9 ± 5.4 days (mean ± sd). The male flowers appeared before or simultaneously with the female ones, resulting in a positive mean value of *ΔMF* (4.9 ± 5.1 days). *ΔMF* was negatively

correlated with *mMD* (r = -0.59) and positively with log_ND_m (r = 0.57). Longer time differences between the starts of male and female flowering were associated with more masculinised plants and a higher node at which half of the feminisation process was performed.

The dry matter yields in stem obtained in UF (10 ± 6.6 g plant^{-1}) were on average higher than those obtained in both C1 and C2, while, conversely, the dry matter yield in seeds (1.6 ± 0.7 g plant^{-1}) were on average lower. Unlike in C1 and C2, *DMstem* and *DMseed* were not correlated with each other in UF. *DMseed* was positively correlated with *mMD* (r = 0.47), *i.e.*, with feminised phenotypes. The correlations involving *DMstem* were very low (r < 0.15). All correlations reported in the present section were significant at α = 0.001.

9.3.2. QTL analysis

The search for QTLs by interval mapping and multi-QTL analysis led to the identification of 7, 12 and 19 distinct QTLs in C1, C2 and UF, respectively – in the present study, the QTLs were labelled by the name of the closest marker, and the total number of QTLs per population was determined by counting once those that were detected for distinct phenotypic variables (Fig. 9.1; Tables 9.2 and 9.3). The number of QTLs detected per phenotypic variable varied from 0 (*DMstem* in C1; *mMD*, *%MD5*, log_M_{max} and *tf* in UF) to 5 (*log_k* in UF). On the basis of our results, we considered that the QTLs separated from each other by at most 25 cM defined a multi-QTL region.

(i) Dioecious populations (C1 and C2)

QTLs for the sex expression

Because of colinearity, the multi-QTL analysis for *sex* was performed after discarding the sex locus from the multi-QTL model. Six distinct QTLs for variables related to the sex expression (*sex*, *%F-*♂ or *%M-*♀) were identified in C1, and 8 in C2. In C1, all these QTLs were mapped in HGV (Table 9.2; Fig. 9.1a). In C2, 6 QTLs were mapped in HGV, 1 in HGIV and 1 in a CG not assigned to a HG (C2-15) (Table 9.2; Fig. 9.1b). As a reminder, HGV included

the sex CG of each dioecious map (C1-7 and C2-14) and CGs putatively located on the sex chromosomes (Chapter VII).

In C1, 2 QTLs were found for *sex* after discarding the sex locus from the multi-QTL model. Both were mapped in the sex CG (C1-7). One QTL with α_Q additive effect explained a large part of the sex variation (85.2%) and was detected at the locus of a marker mapped on the Y chromosome only (5_323). The second QTL for *sex* had a dominance effect and both α_P and α_Q additive effects explaining 23.6 and 50.5% of the sex variation, respectively. It was mapped at the locus of a marker present on the X chromosome in both the female and male parent (2_299). Two QTLs were found for *%F-*♂ and 4 for *%M-*♀. One of them (2_299) was a QTL for both *%F-*♂ and *%M-*♀ in addition to *sex*. This QTL had a α_Q additive effect on *%F-*♂ and both α_P additive and dominance effects on *%M-*♀. It must be noted that the alleles P_1 and Q_1 at this QTL were closely linked to the same DNA fragment (2_299, Table 9.2). Given its positive effect on *%F-*♂ and negative effect on *%M-*♀, this fragment could include feminising genetic factors mapped on X chromosomes. The 3 remaining QTLs for *%F-*♂ or *%M-*♀ were all mapped in a putative sex CG (C1-5). Two of them were found at the locus of markers that segregated 1:1 in both male and female progenies, which supported their presence on the sex chromosomes (6_243 and 6_253).

In C2, 3 QTLs were detected after discarding the sex locus from the multi-QTL model. All of them were mapped in the sex CG (C2-14) close to markers mapped on X parental chromosomes and had α_Q additive and dominance effects. High parts of phenotypic variation (>73%) were explained by the two closest QTLs from the sex locus (2_299 and 1_220). One of them was common to C1 (2_299) but, in contrast to C1, was not associated with variation in *%F-*♂ or *%M-*♀ in C2. Four QTLs were identified for *%F-*♂, while a single QTL with asymptotic effects was detected for *%M-*♀ in C2 (8_330).

Table 9.2 QTL effects estimated by multi-QTL analysis in the dioecious populations (C1 and C2)

Population C1									
Trait	Variable [1]	HG	CG [2]	Marker [3]		QTL			
				Name	Pos. (cM)	Pos. (cM)	Effect	Estimate ± SE [4]	r^2 [5]
Sex ex-pression	sex	V	7	2_299	0	0	αP	0.15 ± 0.06 **	23.6 ***
							αQ	-0.34 ± 0.06 ***	50.5 ***
							δPQ	-0.17 ± 0.06 **	-
				5_323	42.3	42.3	αQ	-0.38 ± 0.03 ***	85.2 ***
	%F-♂	V	5	6_337	108	110	αP	-4.27 ± 1.36 **	
			7	2_299	0	0	αQ	8.6 ± 2.72 **	
	%M-♀	V	5	6_253	156	156	αP	3.17 ± 1.03 **	
				6_243	187	187	αP	-2.57 ± 1.03 *	
			7	2_299	0	5	αP	-22.97 ± 5.82 ***	
							δPQ	-12.12 ± 4.1 **	
				sex	35.6	35	αP	622.34 ± 34.47 ***	
							αQ	342.48 ± 02.11 **	
							δPQ	644.98 ± 38.32 ***	
Yields	DMseed	III	4	3_147	234	222	αQ	-0.97 ± 0.42 *	4.5
							δPQ	-1.37 ± 0.48 **	-

[1] See Table 9.1 for details.

[2] C1-7 and C2-14 included the 'sex' locus. Therefore, the αP and αQ effects detected in these CGs were due to polymorphism between the X chromosomes in the female parent and polymorphism between the X and Y chromosomes in the male parent, respectively.

[3] Name and position of the closest marker from the QTL.

[4] Effect, standard error and significance of the QTLs estimated from the multi-QTL model.

[5] Part of the phenotypic variation (%) explained by the additive QTL effects, as computed by interval mapping for the non-longitudinal variables (model 9.2).

* P <0.05; ** P <0.01; *** P <0.001.

Table 9.2 (continued)

Population C2

Trait	Variable [1]	HG	CG [2]	Marker [3]		QTL			
				Name	Pos. (cM)	Pos. (cM)	Effect	Estimate ± SE [4]	r^2 [5]
Sex expression	sex [6]	V	14	2_299	48.3	48.3	αQ	-0.41 ± 0.05 ***	75.6 ***
							δPQ	-0.13 ± 0.04 **	-
				1_220	89.5	89.5	αQ	-0.29 ± 0.05 ***	73.2 ***
							δPQ	0.12 ± 0.05 *	-
				1_318	144	144	αQ	0.05 ± 0.02 *	6.2 *
							δPQ	-0.09 ± 0.04 *	-
	%F-♂	IV	6	7_186	14.9	20	αP	1.56 ± 0.76 *	
							αQ	53.5 ± 4.32 ***	
							δPQ	-54.19 ± 4.49 ***	
		V	13	1_62	19.2	19.2	αP	47.91 ± 12.71 ***	
							δPQ	17.19 ± 7.07 *	
				1_72	38	32	αP	-30.02 ± 5.94 ***	
				1_338	65.9	65.9	αP	-41.83 ± 14.02 **	
							δPQ	-43.59 ± 11.88 ***	
	%M-♀	-	15	8_330	6.6	6	αQ	390.21 ± 45.37 ***	
							δPQ	389.8 ± 45.58 ***	
Yields	DMstem	II	2	5_174	13	8	αP	1.08 ± 0.31 ***	8.6 **
		IV	6	7_186	14.9	14.9	δPQ	3.86 ± 1.42 **	-
		V	14	2_79	0	13	αQ	1.33 ± 0.38 ***	13.1 **
	DMseed	II	4	7_82	33.4	33.4	δPQ	1.28 ± 0.4 **	-
		V	9	5_161	15.9	15.9	αP	-0.47 ± 0.16 **	17.8 **

[1] See Table 9.1 for details.

[2] C1-7 and C2-14 included the 'sex' locus. Therefore, the *αP* and *αQ* effects detected in these CGs were due to polymorphism between the X chromosomes in the female parent and polymorphism between the X and Y chromosomes in the male parent, respectively.

[3] Name and position of the closest marker from the QTL.

[4] Effect, standard error and significance of the QTLs estimated from the multi-QTL model.

[5] Part of the phenotypic variation (%) explained by the additive QTL effects, as computed by interval mapping for the non-longitudinal variables (model 9.2).

[6] The markers mapped at the sex locus were not included in the multi-QTL model. There was no such marker in C1.

* P <0.05; ** P <0.01; *** P <0.001.

QTLs for the yields

The QTLs for the stem and seed yields identified in C1 and C2 were distributed among HGII, III, IV and V. In C2, three QTLs were detected for *DMstem*. Two of them (5_174 and 2_79) had additive effects, explaining 8.6 and 13.1% of the variation in stem yield. The QTL with highest r^2 (2_79) was located in the sex CG and due to an additive effect in the male parent, the allele associated with an increase in *DMstem* being mapped on the X chromosome (Table 9.2). In C1, one QTL with dominance effect was identified for *DMseed* (3_147). In C2, 2 QTLs were found for *DMseed* (7_82 and 5_161). One of them was mapped in HGV and explained a relatively large proportion of the variation in seed yield (17.8%).

Multi-QTL regions

In C1, a small multi-QTL region was identified between the sex locus and a QTL with α_Q effect on *sex* (Fig. 9.1a). In C2, two multi-QTL regions were defined in HGV (Fig. 9.1b). The first one included two QTLs associated with variation in *%F-♂* (C2-13), and the second one included the sex locus and QTLs with α_Q effects on *sex* (C2-14).

(ii) Monoecious population (UF)

QTLs for the sex expression

Thirteen QTLs were found for variables related to the sex expression. Five of them were mapped in HGIV, 4 in HGII, 3 in HGV and 1 in HGIII. Only one QTL was detected for *%MD1*, and three for *%MDinter*. Five QTLs were found for *log_k*, and 4 for *log_*ND_m, while one QTL was identified for *log_Mmin* very close to a QTL for *log_k* (7_116). Thus, nine distinct QTLs were detected for the parameters of the logistic function describing the monoecy degree along the stem.

Table 9.3 QTL effects estimated by multi-QTL analysis in the monoecious population (UF)

Trait	Variable [1]	HG	CG	Marker [2]		QTL			
				Name	Pos. (cM)	Pos. (cM)	Ef-fect	Estimate ± SE [3]	r^2 [4]
Sex	*Synthesis variables*								
ex-	%MD1	IV	6	1_259	126.8	126.8	αP	8.72 ± 2.7 **	
pres-	%MDinter	II	3	2_95	149.5	156	αP	6.5 ± 2.66 *	
sion							δPQ	7.75 ± 3.11 *	
				7_254	323	317	αQ	-5.78 ± 2.93 *	
							δPQ	-6.3 ± 2.94 *	
		IV	6	2_233	113	108	αP	-4.43 ± 1.17 ***	
	Structure variables								
	log_k	II	3	7_116	182.9	195	δPQ	-1.67 ± 0.5 ***	
		III	4	1_131	0	5	δPQ	-2.45 ± 1.22 *	
		IV	6	6_334	163.2	160	αP	-0.43 ± 0.17 *	
		V	7	1_149	164	151	αP	-0.93 ± 0.39 *	
							δPQ	1.75 ± 0.48 ***	
			9	4_121	0	22	δPQ	-2.79 ± 0.8 ***	
	log_NDm	II	3	6_235	0	2	αP	-0.54 ± 0.18 **	
		IV	6	7_184	27.7	16	αP	0.92 ± 0.41 *	
				4_60	69.8	69.8	αQ	-0.43 ± 0.2 *	
							δPQ	-0.62 ± 0.21 **	
		V	9	4_241	102.7	103	αQ	0.54 ± 0.16 **	
	log_Mmin	II	3	7_116	182.9	199	αQ	0.24 ± 0.09 **	
Earli-	ΔMF	II	3	7_103	231.3	243	δPQ	2.38 ± 0.83 **	-
ness		III	5	4_295	277.7	272	αP	9.51 ± 3.68 *	0.2
							δPQ	10.19 ± 3.87 **	-
		IV	6	3_158	146.3	139	αP	2.47 ± 0.62 ***	8 ***
		V	9	2_271	131.1	125	αQ	1.03 ± 0.42 *	4.3 **
Yields	DMstem	IV	5	4_336	146.2	143	αP	1.81 ± 0.77 *	5.6 **
		V	9	2_271	131.1	141	αQ	2.36 ± 0.72 **	9.1 ***
	DMseed	IV	6	5_147	212.6	212	αP	-0.17 ± 0.06 **	5.5 **

[1] See Table 9.1 for details.

[2] Name and position of the closest marker from the QTL.

[3] Effect, standard error and significance of the QTLs estimated from the multi-QTL model.

[4] Part of the phenotypic variation (%) explained by the additive QTL effects, as computed by interval mapping for the non-longitudinal variables (model 2).

* P <0.05; ** P <0.01; *** P <0.001.

QTLs for the earliness

Four QTLs were found for *ΔMF* (Table 9.3; Fig. 9.1c). Two of them had additive effects, explaining 8 and 4.3% of the variation in *ΔMF* (3_158 and 2_271). These QTLs were mapped in HGIV and HGV, respectively, and were closely mapped to QTLs associated with variation of the sex expression (2_233, 1_259 and 6_334 in HGIV, 4_241 in HGV).

QTLs for the yields

Two QTLs were identified for *DMstem*. They were located in HGIV and HGV and explained 5.6 and 9.1% of the variation in stem yield, respectively. Only one QTL was found for *DMseed*. It was mapped in HGIV and explained 5.5% of the variation in seed yield.

Multi-QTL regions

Two multi-QTL regions were defined in the UF map (Fig. 9.1c). The first region was mapped in HGIV (UF-6) and included 4 QTLs. All these QTLs had an additive effect due to the female parent: their respective alleles P_1 rather than P_2 increased both *%MD1* (1_259) and *ΔMF* (3_158) and, conversely, decreased both *%MDinter* (2_233) and *log_k* (6_334). The second multi-QTL region was defined in HGV (UF-9) and included 2 QTLs associated with variation in 3 distinct phenotypic variables. All of them had an additive effect due to the male parent: their respective alleles Q_1 rather than Q_2 increased *log_NDm* (4_241), *ΔMF* (2_271) and *DMstem* (2_271).

9.3.3. Cross-population analysis

Among the 225 distinct mapped markers, 86 accounted for a putative QTL as detected by interval mapping in at least one of the three maps, and 26 of these latter were associated with variation in at least one phenotypic variable in at least one population other than the one in which the putative QTL was detected (Tables 9.4, 9.5 and 9.6).

Table 9.4 Cross-population analysis of the sex expression: markers that were associated with variation across both dioecious populations only [†]

HG	CG no.			Marker	Variable [1]	Test of the effects [2]			Effect of the marker [3]		Variables affected by the marker locus [4]
	C1	C2	UF			M	Pop	M x Pop	C1	C2	
V	-	9	-	5_161	sex	*	NS	NS	53 - 68	71 - 93 *	DMseed
					%F-♂	**	NS	NS	- 2.65 ± 1.5	- 6.63 ± 2.94 *	
	-	11	-	4_140	%F-♂	*	NS	NS	1.48 ± 1.78	3.34 ± 1.62 *	%F-♂
	6	12	-	5_160	sex	**	*	NS	98 - 74 **	77 - 64	sex (C1)
	7	14	-	2_299	sex	***	NS	NS	84 - 26 ***	98 - 54 ***	%M-♀ (C1), %F-♂ (C1), sex (C1,C2)
	-	14	8	2_289	sex	***	NS	NS	74 - 85	60 - 100 ***	sex (C2)
	7	14	-	4_276	sex	***	NS	NS	0 - 88 ***	0 - 100 ***	sex (C1,C2)
	7	14	-	5_323	sex	***	NS	NS	0 - 91 ***	0 - 100 ***	sex (C1,C2)
	7	14	-	7_118	sex	***	NS	NS	0 - 88 ***	0 - 100 ***	sex (C1,C2)
	7	14	-	4_110	sex	***	NS	NS	0 - 85 ***	0 - 100 ***	sex (C1,C2)
	7	14	-	5_85	sex	***	NS	NS	2 - 91 ***	0 - 100 ***	sex (C1,C2)
	-	14	7	1_220	sex	***	NS	NS	35 - 18	98 - 71 **	sex (C2)
	-	14	-	1_274	%F-♂	*	NS	NS	4.32 ± 1.45 **	1.38 ± 2.4	sex
	7	-	-	8_164	sex	***	NS	NS	53 - 88 **	73 - 100 **	sex
-	-	15	-	1_94	%F-♂	*	NS	NS	- 2.13 ± 1.46	- 3.18 ± 1.61	%M-♀

[†] The markers that were associated with variation of the sex expression across both the dioecious and monoecious populations are shown in Table 9.5.

[1] Phenotypic variables showing variation associated with the presence/absence of the marker (model 9.5). See Table 9.1 for details.

[2] M = Marker; Pop = Population.

[3] Frequency of the marker in the female and male progenies for the variable "sex"; estimate ± SE for the other variables.

[4] Phenotypic variables showing variation associated with the genetic predictor at the marker locus (models 9.2 and 9.3). The population was mentioned only when the marker was mapped in more than one population.

NS $P > 0.05$; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

(i) Sex expression across both dioecious populations (C1 and C2)

Sixteen markers segregated with sex in C1 or C2, and 13 presented a congruent segregating pattern with sex across both dioecious populations (significant 'M' effect and not significant 'M x Pop' interaction in Tables 9.4 and 9.5; $P < 0.05$). Seven of them were mapped in the sex CGs (Table 9.4), five in putative sex CGs in HGV (Table 9.4; 2_66 in Table 9.5) and one in HGIV in UF (1_191 in Table 9.5). In addition, the significant 'M x Pop' interaction observed for one of the 3 remaining markers could be explained by its mapping on one female-parent X chromosome and absence from the male parent in C1 and on both one female-parent and the male-parent X chromosomes in C2 (6_255 in Table 9.5).

No marker showed congruent effects on *%M*-♀ across C1 and C2. On the opposite, four such markers were found for *%F*-♂ (Table 9.4). Two of them segregated 1:1 in both male and female progenies in C2 (4_140 and 1_94; Chapter VII). Another one was more frequent in the male plants and decreased the percentage of nodes bearing female flowers in male plants (*%F*-♂, 5_161). As a reminder, a QTL for *DMseed* was detected at this marker locus in C2 (Table 9.2).

Table 9.5 Cross-population analysis of the sex expression and earliness across both dioecious (C1 and C2) and monoecious (UF) populations: markers that were associated with variation in UF and segregated with sex and/or accounted for a putative QTL in C1 or C2

HG	CG no.			Marker	Dioecious populations					Monoecious population		Variables affected by the marker locus [4]
	C1	C2	UF		Test of the effects [2]			Frequency in ♀ - ♂ [3]		Variable [1]	Estimate ± SE	
					M	Pop	M x Pop	C1	C2			
IV	-	-	6	2_233	-	-	-	60 - 29 *	-	%MD1	-5.27 ± 2.16 *	%MD1, %MDinter, ΔMF
										%MDinter	3.18 ± 0.89 ***	
										log_NDm	-0.43 ± 0.21 *	
										log_Mmin	0.13 ± 0.06 *	
										ΔMF	-1.09 ± 0.43 *	
	-	-	6	1_259	*	NS	*	26 - 29	77 - 39 **	ΔMF	-0.85 ± 0.41 *	%MD1, %MDinter, log_k, ΔMF
	-	-	6	1_191	**	NS	NS	23 - 38	10 - 39 **	log_NDm	0.54 ± 0.18 **	log_NDm, ΔMF
										log_Mmin	-0.15 ± 0.07 *	
										ΔMF	0.92 ± 0.39 *	
V	-	-	9	4_121	*	NS	NS	84 - 94	88 - 100	-	-	log_k
	-	8	9	2_66	***	NS	NS	77 - 100**	77 - 100**	log_NDm	-0.52 ± 0.18 **	sex (C2), log_NDm (UF)
	-	11	9	2_271	NS	NS	NS	23 - 32	33 - 39	mMD	-0.21 ± 0.09 *	%F-♂ (C2), ΔMF (UF),
										%MD1	5.06 ± 2.05 *	DMstem (UF)
										log_NDm	0.41 ± 0.2 *	
										log_Mmax	-0.16 ± 0.07 *	
										ΔMF	1.08 ± 0.41 **	
	5	14	9	6_255	***	NS	***	35 - 41	100 - 64 ***	mMD	0.16 ± 0.08 *	sex (C2), log_NDm (UF)
										%MD1	-4.27 ± 1.96 *	
										log_k	0.28 ± 0.13 *	
										log_NDm	-0.53 ± 0.18 **	
										ΔMF	-0.8 ± 0.39 *	

[1,2,4] See legend of Table 9.4. [3] Frequency of the marker in the female and male progenies. NS P >0.05; * P <0.05; ** P <0.01; *** P <0.001.

(ii) Sex expression across the dioecious (C1 and C2) and monoecious (UF) populations

Seven markers were associated with variation or accounted for a putative QTL for variables related to the sex expression or earliness in both C1 or C2, on the one hand, and in UF, on the other hand (Table 9.5). Among them, 4 were mapped in a putative sex CG (HGV) and 3 in HGIV.

Six of these markers segregated with sex across both dioecious populations (significant 'M' effect). One of them was mapped on X chromosomes at the sex locus in C2 and segregated 1:1 in both male and female progenies in C1 (6_255; Chapter VII). In UF, this marker was associated with feminised phenotypes (positive effect on *mMD* and negative effect on *%MD1*) and a decreased log_ND_m. Two markers were more frequent in the female plants in C1 or C2 and were associated with feminised phenotypes in UF (negative effect on *%MD1*, 2_233) or negative effect on *ΔMF* (1_259). Both were mapped in the HGIV multi-QTL region of the UF map. One marker was present in all male plants and in 77% of the female ones in both C1 and C2 and was associated with variation in log_ND_m in UF (2_66). A marker more frequent in the male plants of both C1 (NS) and C2 (P <0.01) was associated with an increase in *ΔMF* and *log_NDm* in UF (1_191). Another marker more frequent in the male plants across both dioecious populations accounted for a QTL for *log_k* in UF (4_121). The remaining marker (2_271) segregated 1:1 in both male and female progenies in C2 and therefore was putatively located on the sex chromosomes (Chapter VII). This marker was associated with significant effects on sex-expression related variables and accounted for a QTL for log_ND_m in UF (Tables 9.3 and 9.5).

Table 9.6 Cross-population analysis of the dry matter yields in stem and seed: markers that were associated with variation across both dioecious (C1 and C2) and monoecious (UF) populations

HG	CG no.			Marker	Variable [1]	Test of the effects [2]			Effect of the marker (estimate ± SE) [3]			Variables affected by the marker locus [4]
	C1	C2	UF			M	Pop	M x Pop	C1	C2	UF	
IV	-	6	6	6_334	DMseed	**	***	NS	-	-0.37 ± 0.17 *	-0.1 ± 0.07	log_k (UF), ΔMF
					%MD1						4.34 ± 2.13 *	
					log_k						-0.38 ± 0.14 **	
					log_NDm						0.4 ± 0.2 *	
					ΔMF						1.28 ± 0.41 **	
V	-	13	-	1_62	DMseed	**	NS	*	0.13 ± 0.12	0.85 ± 0.34 *	-	%F-♂, DMseed
	-	14	-	6_84	DMstem	*	*	NS	-0.28 ± 0.28	-0.56 ± 0.29	-	Sex
					sex	***	NS	**	2 - 47 ***	0 - 96 ***	-	
	-	-	7	1_56	DMseed	*	***	NS	0.16 ± 0.12	0.14 ± 0.17	0.1 ± 0.06	DMseed
					log_NDm						0.41 ± 0.19 *	
	-	-	9	6_215	DMstem	*	***	NS	-0.39 ± 0.26	-	-1.55 ± 0.61 *	log_k, log_NDm

[1,2,3,4] See legend of Table 9.4.

[1] The sex-expression and earliness related variables that showed variation associated with the polymorphism of the markers retained from the CP analysis of yields (model 9.5) are also presented.

NS $P > 0.05$; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

(iii) Dry matter yields

Five markers had a significant main effect on the yields (Table 9.6). Two of them showed congruent effects on *DMstem* across two distinct populations. One of both was more frequent in the male plants in both C1 and C2 and was associated with a decrease in *DMstem* (6_84 in C2-14). The second one (6_215) was mapped at a locus with additive effect on *log_k* and log_ND_m and close to QTLs for log_ND_m, ΔMF and *DMstem* in UF-9 (4_241 and 2_271, Table 9.3 and Fig. 9.1c). Three markers had a significant main effect on *DMseed* across two (6_334 and 1_62) or all three populations (1_56). All of these markers were mapped in HGIV or HGV.

The figure 9.1 is presented on the next three pages.

Fig. 9.1 Likelihood ratio ($-\log_{10}$ *p*-value) *vs*. map distance (cM) in the co-segregation groups (CGs) in which putative QTLs were identified in the C1, C2 and UF populations (Fig. 9.1a, b and c, respectively). Marker names were indicated only for those retained from the multi-QTL (bold face; Tables 9.2 and 9.3) or cross-population (italic face; Tables 9.4, 9.5 and 9.6) analyses. The loci with significant effect as revealed by interval mapping were considered to be putative QTLs, and the putative QTLs with significant effect as revealed by the multi-QTL analysis were considered to be QTLs. The red lines on the *x*-axis indicate regions including several QTLs separated from each other by at most 25 cM. (*) indicates that the scale of the *y*-axis was modified due to very high likelihood ratios: it was in 0 – 140 for the α_Q effect on *sex* in C1-7, 0 – 210 for α_Q and 0 – 21 for δ_{PQ} on *sex* in C2-14, and 0 – 13 for δ_{PQ} on *%M-*♀ in C2-15. The CGs included in HGV were putatively located on sex chromosomes (Tables 7.5 and 7.6). Marker types were as in Wu et al. (2002a) (see Table 3.1 for details).

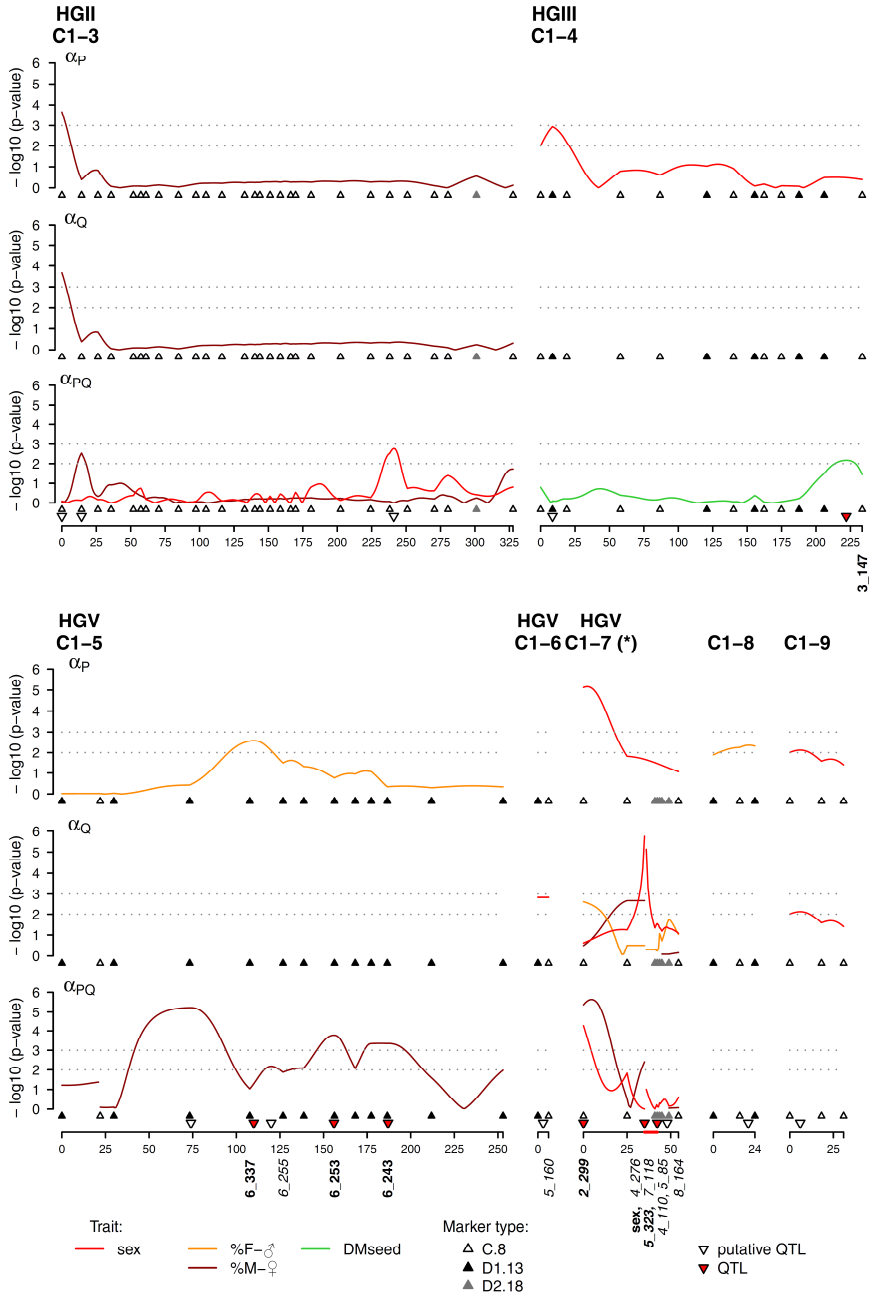

Fig. 9.1a QTL analysis in the dioecious population C1.

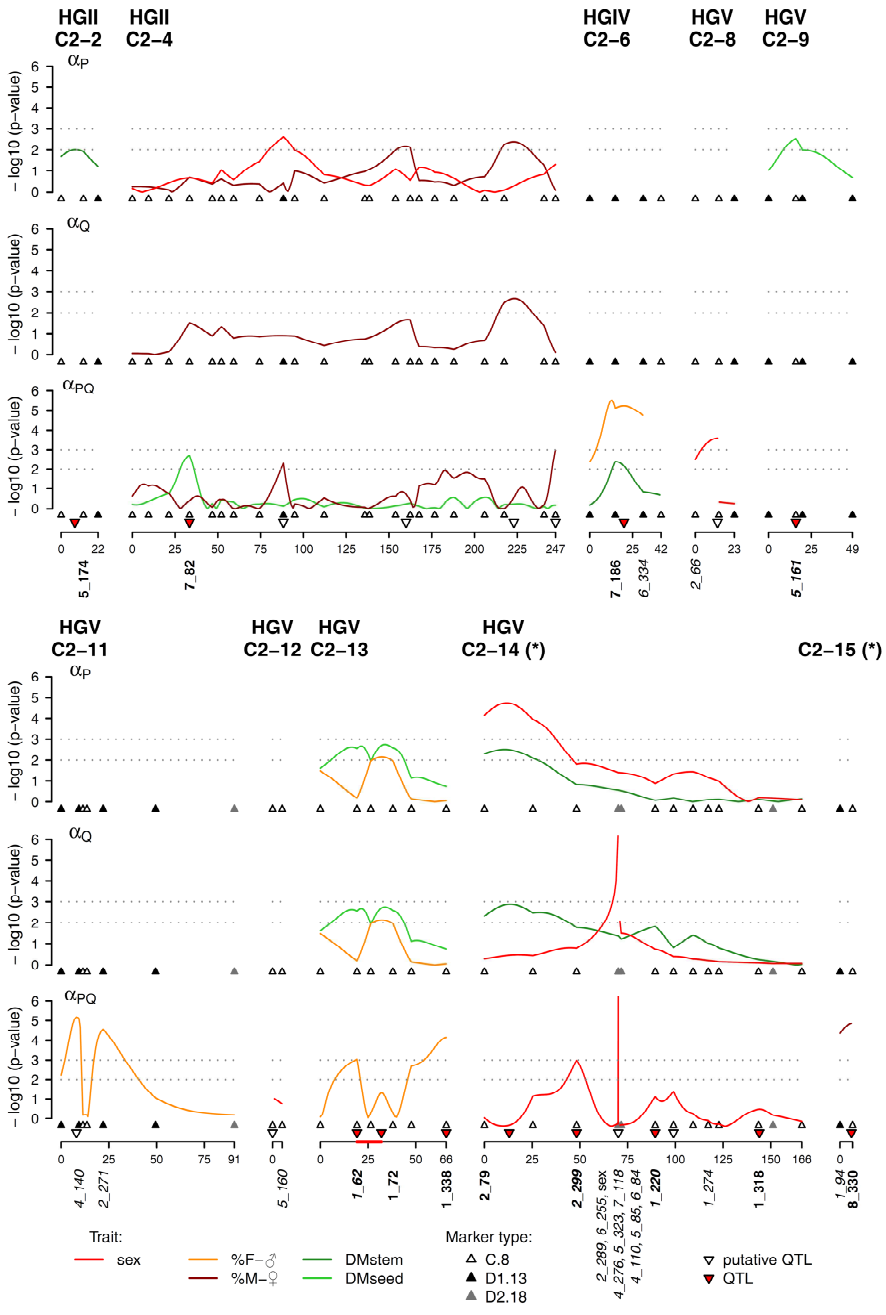

Fig. 9.1b QTL analysis in the dioecious population C2.

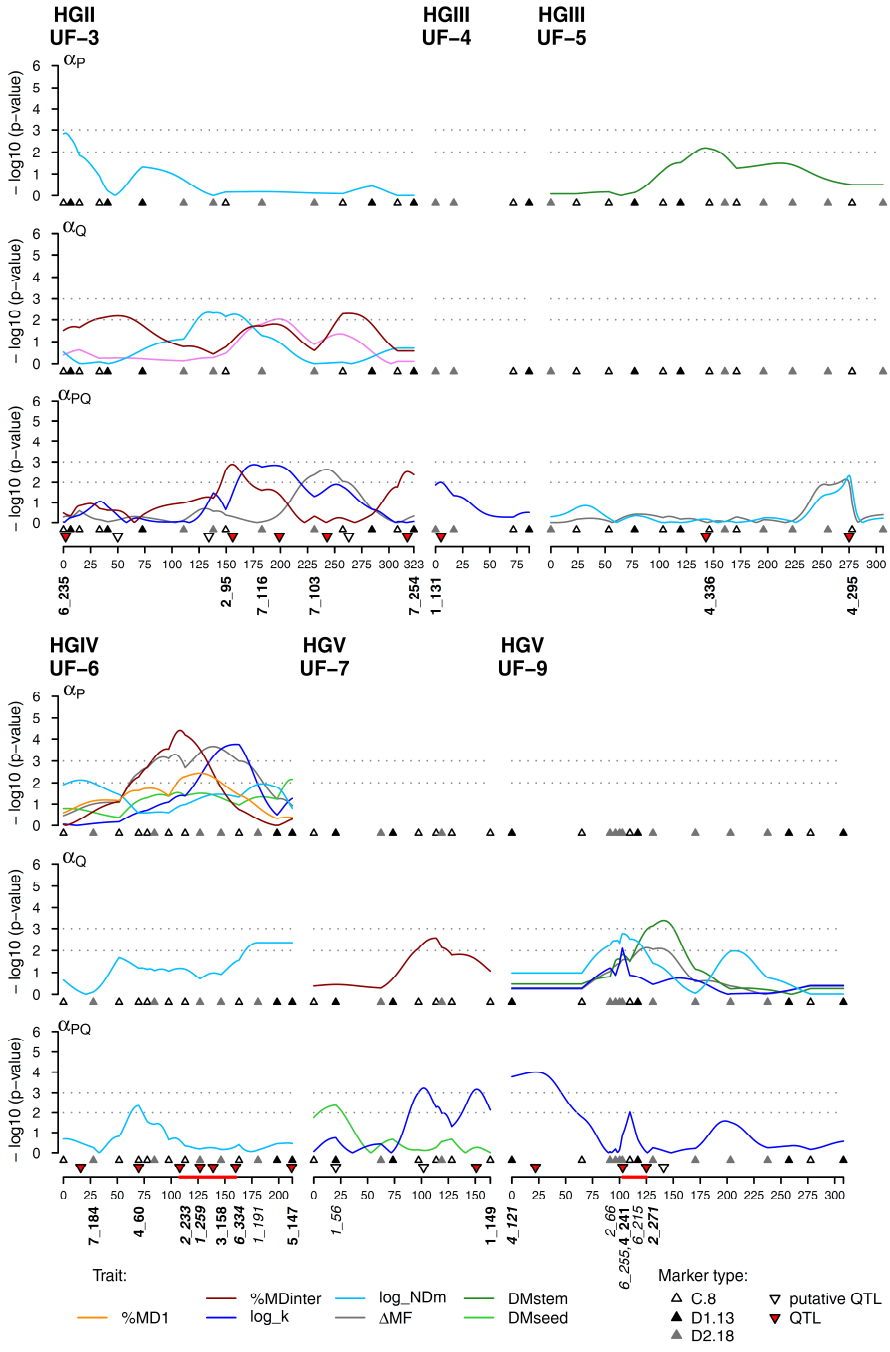

Fig. 9.1c QTL analysis in the dioecious population UF.

9.4. DISCUSSION

To our knowledge, the present study reported the first QTL analysis in hemp. It focused on the genetic determinism of the sex expression in both dioecious and monoecious forms of hemp. QTL analyses of traits related to the sex expression in monoecious plants have been reported in few species of the *Cucurbitaceae* family (Serquen et al. 1997; Fazio et al. 2003; Yuan et al. 2008; Diaz et al. 2011; Miao et al. 2011; Prothro et al. 2013) and in *Vitis vinifera* (Marguerit et al. 2009). In contrast to these species, the sexual phenotype in hemp is highly variable: plants of dioecious hemp are able to produce sporadically flowers of the opposite sex, while the distribution of male and female flowers varies widely among plants and nodes in monoecious hemp. This high variation of the sexual phenotype led us to develop distinct variables to characterize the sex expression (Table 9.1). In particular, the quantitative variation of the sex expression along the stem of monoecious plants was modelled by a logistic function of the node position (Chapter VIII). In addition, unlike in the aforementioned species, sex chromosomes are present in hemp (Yamada 1943 cited by Sakamoto et al. 1995; Westergaard 1958; Parker and Clark 1991). Therefore, the presence of both sex chromosomes and high variability of sexual phenotype in hemp and the search for QTLs for the parameters of a model describing the sexual phenotype of monoecious plants made the interest and originality of the present quantitative approach of sex determinism.

9.4.1. Determinism of the sex expression in dioecious hemp

Six and 8 QTLs associated with variables related to the sex expression were identified in two distinct dioecious hemp populations (Table 9.2). The high part of phenotypic variation explained by the QTLs for *sex* (from 73.2 to 85.2% for the closest QTLs from sex) could directly be attributed to the presence of heteromorphic sex chromosomes. These high values of r^2 were comparable to those obtained for major QTLs affecting the sex expression in *Cucumis sativus*, ranging between 60 (Yuan et al. 2008) and 74% (Serquen et al. 1997). The identification of QTLs for *%F-*♂ and *%M-*♀ supports previous assumptions on the existence of a genetic basis for the production of flowers of the opposite sex

in dioecious hemp [Borthwick and Scully 1954; Grassi and de Meijer (pers. comm.) cited by Moliterni et al. 2004].

Most of the QTLs detected for the sex expression in the dioecious maps (all of them in C1, 6 in C2) were mapped in the HG that included the sex locus (HGV). Among them, only one was detected in both maps (2_299, Table 9.2). This could be explained by the low number of markers shared by both dioecious maps (13.5% of the mapped markers). In C2, one QTL for the sex expression was identified in HGIV (7_186, Table 9.2). This HG included 6 markers that segregated with sex among a total of 17 distinct markers [these 6 markers were 3 markers segregating 1:1 in both male and female progenies in C2 ('putative class B' markers, Chapter VII) and 3 markers significantly segregating with sex in C1 or C2 but mapped in UF only (Table 9.5)]. It is therefore likely that HGIV derived from sex chromosomes similarly to HGV. However, additional markers are necessary to test this assumption.

Assuming that the CGs mapped in HGV derive from the sex chromosomes, our results suggest that the genetic determinism of the sex expression in dioecious hemp is primarily due to genetic factors located on the sex chromosomes. The implication of the sex chromosomes in the determinism of the sex expression in dioecious hemp could be as follows. The region surrounding the sex locus (region 2 in Table 7.5; multi-QTL region in C2-14 in Fig. 9.1b) would include genetic factors involved in the differentiation of the male plants and in the production of flowers of the opposite sex. This region included markers that completely segregated with the male phenotype (Table 7.5) and a QTL with additive effect due to the male parent in C1 (5_323, Fig. 9.1a). In addition, markers homologous between the X and Y chromosomes were identified in this region (Table 7.5), while a QTL for *%M-*♀ was mapped close to the sex locus in C1 (Table 9.2). These observations support that not only male-determining genes should be present in the region surrounding the sex locus. The regions 1 and 3 of the sex CGs (Table 7.5) would not be involved in the sex differentiation given the higher recombination rate observed with sex (Chapter VII) and the presence of only one QTL for *sex*, with both additive and dominance effects and low r^2 (1_318, Table 9.2). No QTL was detected in the terminal pseudoautosomal region of the sex CG in C2 (Chapter VII). However, beyond the four regions defined in the sex CGs (Table 7.5), our result suggest that additional regions of the sex chromosomes could carry genetic factors involved in the development of flowers of the opposite sex. Indeed, QTLs for

%F-♂ and *%M-*♀ (Table 9.2) and two markers associated with variation in *%F-*♂ across both dioecious populations (5_161 and 4_140, Table 9.4) were mapped in putative sex CGs in HGV. The QTLs with only an additive effect due to the female parent (6_337, 6_253, 6_243 in C1-5 and 1_72 in C2-13) would be located on X chromosomes only, while those with both additive and dominance effects (1_62 and 1_338 in C2-13) would be located on X and/or Y chromosomes.

9.4.2. Determinism of the sex expression in monoecious hemp

The present study identified 13 distinct QTLs associated with the sex expression in the monoecious population (Table 9.3). These QTLs were distributed along HGII (4 QTLs), HGIII (only 1 QTL), HGIV (6 QTLs) and HGV (4 QTLs). The identification of QTLs for the sex expression in HGV suggests the existence of genetic factors involved in the determinism of the sex expression in monoecious hemp on the sex chromosomes, and thus on X chromosomes (Faux et al. 2014). In addition, seven markers were identified that both segregated with sex in a dioecious population and were associated with variation of the sex expression of monoecious plants (Table 9.5). This result further supports the presence of genetic factors affecting the sex expression of monoecious hemp plants on the sex chromosomes.

The number of QTLs detected for each phenotypic variable related to the sex expression in the monecious population ranged from 0 to 5 (Table 9.3). The absence of QTL identified for log_M_{max} is not surprising given the relatively low variation that was observed for this parameter (Fig. 8.2, Chapter VIII). Also, no QTL was found for *mMD* and *%MD5*, suggesting that these variables do not efficiently characterize the sex expression of monoecious hemp plants. In total, the definition of synthesis variables allowed the detection of 4 QTLs, while the structure variables, which consisted of the parameters of a logistic curve describing the sex expression as a function of the node position (Table 9.1), resulted in the identification of 9 distinct QTLs. The higher number of QTLs found by the modelling approach tends to endorse its relevance for characterizing the variability of the sex expression among monoecious hemp plants.

According to Dellaporta and Calderon-Urrea (1993), the QTLs identified for the sex expression in the present study could include genetic factors that regulate programs of sexuality through a signal transduction mechanism that modifies endogenous hormonal levels. Indeed, the formation of male and female generative organs in hemp may be associated with an increased demand for gibberellin and auxins, respectively (Galoch 1980). More recently, cDNA-AFLP fragments differentially expressed in female and male apices in hemp were identified, and a similarity was found between these fragments and a Rac-GPT binding protein which plays a signalling role in auxin-regulated gene expression in *Arabidopsis* (Moliterni et al. 2004). These studies suggest that the sex expression in monoecious hemp could be related to the presence of hormonal gradients along the stem, which could provide a physiological interpretation for the present model parameters used to characterize the sex expression. Further studies combining physiological and QTL approaches are needed to test this assumption.

9.4.3. Sex expression and earliness in monoecious hemp

No QTL was found for the flowering time (*tf*). This could be due to the relatively close earliness of the parental moneocious cultivars, 'Uso 31' and 'Fedora 17' (Faux et al. 2013, 2014), resulting in a low variation of flowering time in the segregating population. On the opposite, four QTLs were detected for *ΔMF*. Among them, two were mapped in multi-QTL regions including QTLs for sex-expression related variables (3_158 in HGIV and 2_271 in HGV; Table 9.3, Fig. 9.1c). This finding could be expected from the correlation observed between *ΔMF* and sex expression, an increased delay of the start of female flowering compared to the male one being linked to a more masculinised sex expression. Relations between earliness and sex expression have been reported in several species. In *Cucumis sativus*, a QTL for days to anthesis was found close to a QTL for sex expression, indicating a possible association of earliness and sex expression (Fazio et al. 2003). In *Silene latifolia*, a well-studied species with heteromorphic sex chromosomes, the Y chromosome includes a locus which is responsible for early stamen development in addition to sex-determining loci (Donnison et al. 1996). In hemp, female plants exhibit higher auxin endogenous levels than male plants, which conversely have a higher gibberellin level (Galoch 1980). In addition, according to Chailakhyan and

Khryanin (1978, 1979), plants treated with exogenous gibberellin produce floral buds earlier than the controls, whereas the auxin treatment delay the flowering. These observations support the assumption that earliness and sex expression in monoecious hemp share a common basis of their genetic determinism and, therefore, the presence of pleiotropic QTLs for *ΔMF* and sex expression (Table 9.3, Fig. 9.1c). Further studies aiming to dissect the genetic basis of these traits could benefit from the statistical approaches that take into account the correlations among the phenotypic variables in the QTL analysis (Jiang and Zeng 1995; Malosetti et al. 2007).

9.4.4. QTLs for the dry matter yields in stem and seed

Five QTLs were identified for *DMstem*, and 4 for *DMseed* regardless of the population (Tables 9.2 and 9.3). The parts of phenotypic variation explained by the QTLs for *DMseed* (from 4.5 to 17.8%) were within the range reported for seed yield in *Triticum aestivum* (from 3.7 to 15.7%) by Groos et al. (2003) and in *Trifolium pratense* (from 7 to 15.3%) by Herrmann et al. (2006). In contrast, the parts of variation explained by the QTLs for *DMstem* (from 5.6 to 13.1%) were relatively low compared to the values reported for QTLs for the stem yield in *Miscanthus sinensis* (from 11.1 to 23.3%) by Atienza et al. (2003).

The most significant QTL found for *DMseed* was mapped in HGV, at the locus of a marker segregating with sex (5_161; Tables 9.2 and 9.4). In addition, two markers with congruent effect on *DMseed* across populations were mapped in HGV (1_62 and 1_56; Table 9.6). These findings suggest the existence of genetic factors affecting the seed yield on the sex chromosomes. However, the major QTL found for *DMseed* was associated with a marker that segregated with sex and significantly affected *%F-♂* across both dioecious populations (5_161, Table 9.4). This observation suggests the possibility of pleiotropic effects at this QTL, although no correlation could be *a priori* expected between *DMseed* and *%F-♂* since these variables characterize distinct plants in the dioecious populations, female and male, respectively.

The mapping of two QTLs with relatively high r^2 for *DMstem* in HGV (2_79 in C2 and 2_271 in UF; Tables 9.2 and 9.3) suggests that the sex chromosomes of hemp could include genetic factors involved in the determinism of the stem yield. In the dioecious population C2, an increase in *DMstem* was

associated with a QTL allele mapped on the X chromosome of the male parent (2_79, P <0.01, Table 9.2), although the correlation between *DMstem* and *sex* was relatively low (P <0.05). To our knowledge, no stem weight difference between male and female plants of dioecious hemp has been reported in the literature. However, the stem weight is more variable among female than male plants (van der Werf and van der Berg 1995), and the stem morphology varies according to the gender: the male plants are generally taller and have thinner stems than the female ones (Bocsa and Karus 1998; Schumann et al. 1999; Struik et al. 2000). Therefore, the stem diameter and stem length should be considered in further studies aiming to dissect the genetic basis of the stem yield and its relationship with sex in dioecious hemp.

In the monoecious population, QTLs with positive additive α_Q effect on *DMstem* and log_ND_m were mapped close to each other in the multi-QTL region defined in HGV (2_271 and 4_241; Fig. 9.1c). *DMstem* and log_ND_m were barely correlated with each other (r = 0.15), and log_ND_m only was correlated with *DMseed*, the correlation being negative (r = -0.42). Therefore, the close linkage found between QTLs for *DMstem* and log_ND_m suggests that breeding for *DMstem* can result in decreased seed yields by increasing the plant node at which half of the feminisation process is performed. The existence of such QTLs could provide an explanation to the observation that no cultivar among those tested in previous field trials maximized both stem and seed yields simultaneously (Faux et al. 2013).

9.5. CONCLUSIONS

The present study identified QTLs associated with the sex expression, earliness and yields in stem and seed in two dioecious and one monoecious segregating populations in hemp. Its main advances were as follows.

Firstly, QTLs associated with the presence of flowers of the sex opposite to the dominant one were detected in dioecious hemp. Secondly, QTLs associated with the variation of the sex expression in monoecious hemp were identified despite the high sensitivity of the trait to the environment. The relatively high number of QTLs detected for the parameters of a logistic curve modelling the sex expression as a function of the node position along the stem

tended to support the relevance of this approach for characterizing the variability of the sex expression among monoecious hemp plants. Thirdly, closely linked QTLs for the sex expression, earliness and yields in stem and seed were detected. Our results suggested that genetic factors located on the sex chromosomes of hemp would play a role in the determinism not only of the sex expression but also of earliness and yields.

Therefore, despite the high environmental sensitivity of the trait, there is scope for investigating the genetic determinism of the sex expression in hemp through a quantitative approach. In this context, dissecting the genetic basis of the earliness and yields in stem and seed in addition to the sex expression might be valuable for both fundamental and agronomical purposes.

Acknowledgments

We gratefully thank Prof. A. Garcia and Dr. M. Mollinari (Universidade de São Paulo, Brazil) for their support in the computation of genetic predictors with OneMap, V. Van Hese and P. Van Thorre for their help in the laboratory work and phenotyping of hemp plants, respectively, and the platform Statistical Methodology and Computing Support (UCL) for their advice on statistics. This work was supported by the Fonds de la Recherche Scientifique [grant number 1.5.279.08].

PART III – DISCUSSION AND PERSPECTIVES

CHAPTER X

GENERAL DISCUSSION

The present thesis addressed two main objectives:

(i) investigating the genetic determinism of the sex expression in monoecious hemp, and

(ii) establishing the relationships between sex expression, flowering phenology and yields in stem and seed in monoecious hemp.

Its originality is in the quantitative approach of the genetic determinism of the sex expression, which therefore could be investigated through the identification of quantitative trait loci (QTLs). Five specific objectives were defined in order to achieve the main objectives:

1. Assessing the genotypic variability of the sex expression in monoecious hemp (Chapters IV & V);

2. Establishing the relationships between sex expression, flowering phenology and stem and seed yields in monoecious hemp (Chapter IV & V);

3. Establishing the composition in sex chromosomes of monoecious hemp (Chapter VI);

4. Constructing linkage maps of the sex expression in hemp (Chapter VII);

5. Identifying QTLs involved in the determinism of the sex expression, earliness and yields in stem and seed in hemp (Chapters VIII and IX).

In the present part, the main advances of this thesis are presented with respect to the two main objectives and discussed in relation to the current knowledge on the determinism of the sex expression in hemp. Then, the shortcomings are discussed, and, finally, perspectives are suggested.

10.1. MAIN ADVANCES

10.1.1. Genetic determinism of the sex expression in monoecious hemp

In **Chapters IV & V**, the genotypic variability of the sex expression expressed as a quantitative trait was assessed among five monoecious hemp cultivars in the field (Chapter IV) and in controlled conditions (Chapter V). Environmental variation was created, on the one hand, by two contrasting trial sites and five sowing dates (Chapter IV) and, on the other hand, by three distinct photoperiodic controlled conditions (Chapter V). The sex expression was assessed by recording the degree of monoecy at the plot and plant-node levels in the field and controlled conditions, respectively.

In **Chapter VI**, the genome size of 55 monoecious hemp plants was determined by flow cytometry, and the DNA of 115 monoecious plants was screened with the male-associated marker MADC2.

In **Chapter VII**, three AFLP linkage maps of the sex expression in hemp were constructed from three F_1 populations: two derived from a cross between female and male plants of the dioecious cultivar 'Carmagnola' and one from a cross between the monoecious cultivars 'Fedora 17' and 'Uso 31'.

In **Chapter VIII**, eight variables were constructed to characterize the sex expression of monoecious hemp plants. The originality of our approach was in the modelling of the distribution of the female and male flowers along the stem and use of the function parameters for characterization of the sex expression.

In **Chapter IX**, a QTL analysis of the sex expression, earliness and stem and seed yields was performed in two dioecious and one monoecious F_1 populations in hemp. The three constructed AFLP linkage maps, one per population (Chapter VII), provided the framework for the location of QTLs, while the phenotypic variables constructed in Chapter VIII were used to identify QTLs associated with variation in the sex expression of the monoecious hemp plants.

The main advances with respect to our first main objective were as follows.

(1) The sex expression in monoecious hemp is affected by G x E interactions.

In accordance with the high sensitivity of the sex expression to the environment, our results documented the effect of G x E interactions on the sex expression in monoecious hemp.

The variation of the sex expression with the sowing date in the field (Chapter IV) and with the photoperiodic regime in controlled conditions (Chapter V) agreed with the masculinising effect of long photoperiods reported in hemp (Borthwick and Scully 1954; Arnoux 1966b; Arnoux and Mathieu 1969; Freeman et al. 1980). Furthermore, G x E interactions were observed. Firstly, in controlled conditions, the magnitude of the genotypic effect on the sex expression varied among the three trials. Secondly, the most sensitive cultivars to the environment were the most feminised ones in controlled conditions and, conversely, the most masculinised ones in the field. This apparent contradiction between controlled and field conditions suggested that the photoperiod was not the only environmental factor that explained the sex expression. Two hypotheses were proposed in regard to these G x E interactions. Firstly, a long flowering duration could allow a higher differentiation of the sex expression among the monoecious hemp cultivars. Secondly, the light conditions, which varied among the three trials in controlled conditions and among controlled and field conditions, could have affected the sex expression of the plants in addition to the photoperiod. This latter hypothesis relied on the masculinising effect of light intensity reported in hemp (Borthwick and Scully 1954).

(2) Though highly sensitive to the environment, the sex expression in monoecious hemp has a genetic basis.

Despite the observation of G x E interactions, both field and controlled-conditions trials revealed a significant effect of the cultivar on the sex expression of monoecious hemp plants and the same ranking of cultivars according to their sex expression. Given the consistency of the genotypic effect on the sex expression across environments, we concluded that the sex expression

in monoecious hemp has a genetic basis and might be selected (Chapters IV & V).

(3) Monoecious hemp has the XX constitution for the sex chromosomes.

The genome size of monoecious hemp plants (1.791 ± 0.017 pg) was not significantly different from that of female plants (1.789 ± 0.019 pg) but lower than that of male plants (1.835 ± 0.019 pg), while the male-associated DNA marker MADC2 (Mandolino et al. 1999, 2002) was absent from the monoecious hemp plants. Therefore, monoecious hemp has XX sex chromosomes. This result confirmed the karyological observations made in the 'Kentucky' cultivar by Menzel (1964) (Chapter VI).

(4) The X and Y chromosomes recombine with each other between the sex locus and the pseudoautosomal region.

Four regions were distinguished in the sex co-segregation groups of the dioecious maps: (I) a central region including markers mapped on the Y chromosome and/or X chromosomes, (II and III) two flanking regions including markers mapped on X chromosomes and (IV) a terminal pseudoautosomal region. Our results confirmed the presence of common fragments between the X and Y chromosomes and the existence of a pseudoautosomal region in the sex chromosomes of hemp reported by Peil et al. (2003). However, they differed from those of Peil et al. (2003) by the observation of higher recombination rates between the markers mapped in each of the four regions and the sex locus (Chapter VII).

(5) The X chromosomes of monoecious hemp present homologies with both the X and Y chromosomes of dioecious hemp.

Five markers that were mapped in the sex co-segregation group of one of both dioecious maps were also detected in the monoecious population. Among them, one recombined completely with sex and thus indicated the presence of a pseudoautosomal region, while another one was mapped on both X and Y chromosomes. Since the sex chromosomes of monoecious hemp have the XX constitution (Menzel 1964; Faux et al. 2014), the mapping of these markers in monoecious hemp suggested the presence of homologous fragments between the Y chromosome of dioecious hemp and the X chromosomes of monoecious hemp (Chapter VII).

(6) Modelling the distribution of male and female flowers along the stem might be valuable for characterizing the variability of the sex expression in monoecious hemp.

Eight phenotypic variables were constructed in order to summarise the sex expression of the monoecious plants. These variables were of two types: synthesis and structure. The synthesis variables included the mean monoecy degree and percentages of nodes having a given monoecy degree. The structure variables consisted of the parameters of a logistic curve describing the monoecy degree as a function of the node position along the stem. Thus the structure variables differed from the synthesis ones by the integration of the node position at which the monoecy degree was recorded. The logistic modelling of the sex expression resulted in satisfactory estimations of the variation of monoecy degree along the stem. In addition, a higher number of QTLs was detected for the structure variables (9 QTLs) compared to the synthesis ones (4 QTLs). This result tended to endorse the relevance of the present modelling approach for characterizing the variability of the sex expression among monoecious hemp plants (Chapter VIII).

(7) The sex chromosomes would include genetic factors involved in the determinism of the sex expression in both dioecious and monoecious hemp.

In dioecious hemp, most QTLs associated with variation in the percentage of flowers of the opposite sex were identified on the sex chromosomes in addition to QTLs associated with sex. In monoecious hemp, QTLs associated with variation of the sex expression were identified in the homology group that included the sex locus. In addition, we identified seven AFLP markers that segregated with sex in one of both dioecious populations and were associated with variation of the sex expression in the monoecious population. These results suggested the presence of genetic factors involved in the production of flowers of the opposite sex in dioecious hemp plants and in the determinism of the sex expression in monoecious hemp on the sex chromosomes (Chapter IX).

The observation of significant genotypic variability for the sex expression in monoecious hemp (Chapters IV & V) led us to investigate the genealogy of the five cultivars that were tested in the present study. The ancestors of four of them have been reported (de Meijer 1995), and two groups could be distinguished according to their breeding history. The first one included the single cultivar ‘Uso 31’, which derives from a cross including two Ukrainian cultivars. The second group included the cultivars ‘Fedora 17’, ‘Felina 32’ and ‘Epsilon 68’. All three derive from a cross between female plants of a dioecious cultivar and monoecious individuals of a cultivar directly selected from a same ‘truly’ monoecious strain (‘Fibrimon’) for diverging dates of maturity, followed by a back-cross of the F_1 with ‘Fibrimon’ intersex plants (de Meijer 1995). No common ancestor was found between ‘Uso 31’ and both ‘Felina 32’ and ‘Epsilon 68’. On the opposite, ‘Uso 31’ and ‘Fedora 17’ shared common ancestors of the third and second degrees, respectively (de Meijer 1995). Interestingly, the cultivars ‘Fedora 17’, ‘Felina 32’ and ‘Epsilon 68’, which all derived from a ‘Fibrimon’ parent, were more feminised than ‘Uso 31’. Therefore, we believe that the breeding history of the presently tested cultivars may support the genotypic variability of their sex expression as observed in the present work.

If the existence of a genetic basis of the sex expression in monoecious hemp is assumed, the main question is now how to explain the quantitative variations of the sex expression observed among the monoecious hemp cultivars. Although highly speculative, we propose here a hypothesis for the genetic determinism of the sex expression in monoecious hemp. Monoecious cultivars have been developed by inbreeding of plants that bore mutations resulting in the production of hermaphrodite flowers or bisexual inflorescences (Moliterni et al. 2004). Therefore, we could assume that the QTLs identified for the sex expression in the present study are related to the occurrence of these mutations. According to Forapani et al. (2001), these QTLs would be relatively more frequent at the homozygous state in monoecious hemp compared to dioecious hemp as a result of selfing. However, it is possible that the selection of QTLs present in the homozygous state differs among monoecious cultivars, resulting in the small but significant variations of the sex expression observed among them. In addition, according to Scheiner (1993), the presence of these QTLs in the homozygous state could confer a higher phenotypic plasticity to the monoecious plants compared to the dioecious ones. This latter assumption relies on the explanation of the genetic basis of the phenotypic plasticity by overdominance: the plasticity would evolve as an inverse function of heterozygosity such that the more heterozygous a genotype, the less plastic it will be, because heterozygosity helps to 'buffer' environmental influences (Scheiner 1993; Pigliucci 2005). With respect to the functions of the genes that could underlie the QTLs identified for the sex expression, the findings of Galoch (1980) and Moliterni et al. (2004) suggest that they might be related to the metabolism of auxins. Indeed, high levels of auxins have been observed in female plants just before their transition to flowering (Galoch 1980), while a DNA fragment differentially expressed in male and female floral apices has been found to show a similarity with a protein that plays a signalling role in auxin-regulated gene expression in *Arabidopsis* (Moliterni et al. 2004).

10.1.2. The relationships between sex expression, flowering phenology and yields in stem and seed in monoecious hemp

In **Chapters IV & V**, the relationship between sex expression and flowering phenology was assessed among five monoecious cultivars in the field (Chapter IV) and in controlled conditions (Chapter V). Both sex expression and flowering phenology were related to the yields in stem and seed in the field (Chapter IV).

In **Chapter IX**, a QTL analysis of the sex expression, earliness and yields in stem and seed was performed in two dioecious and one monoecious F_1 populations in hemp.

The main advances with respect to our second main objective were as follows.

(1) Monoecious hemp plants become more feminised with time under natural daylength and constant photoperiod.

A progressive feminisation of the plants was observed from start of flowering to fruit set in the field (Chapter IV). The same observation was made in the monoecious segregating population under controlled environment (Chapter VIII). These results suggested that the protandry, well known in dioecious hemp (Bocsa and Karus 1998), is also found in monoecious hemp. Given the sensitivity of the sex expression to the photoperiod (Freeman et al. 1980), it must be noted that these observations hold true under conditions of both natural daylength and constant photoperiod.

(2) The variations of sex expression and earliness among monoecious cultivars are partly consistent.

The rankings of cultivars according to sex expression and earliness were consistent with each other for most cultivars (Chapters IV & V). Indeed, the very early cultivar ‘Uso 31’ was the most masculinised one, the early cultivar ‘Fedora 17’ was feminised, and the mid-early ‘Felina 32’ and late ‘Epsilon 68’

were the most feminised ones. The only discrepancy was that 'Santhica 27' was later but more masculinised than 'Fedora 17'.

Four QTLs were identified for the time delay between the start of female and male flowering (*ΔMF*, Chapter IX). Two of them were closely linked to QTLs for the sex expression, reflecting the correlations that were observed between *ΔMF* and sex-expression related variables.

The general consistency observed between sex expression and earliness suggested that genes regulating the flowering time might be involved in the determinism of the sex expression. In hemp, female plants exhibit higher auxin endogenous levels than male plants, which conversely have a higher gibberellin level (Galoch 1980). In addition, according to Chailakhyan and Khryanin (1978, 1979), plants treated with exogenous gibberellin produce floral buds earlier than the controls, whereas the auxin treatment delay the flowering. These studies supported the hypothesis that sex expression and earliness share a common basis in their genetic determinism (Chapter V) and suggested that the close linkage observed between QTLs for *ΔMF* and sex-expression related variables might be due pleiotropic genes (Chapter IX).

(3) Higher seed yields are associated with (mid-) early feminised cultivars.

The seed yields were lowest in the most masculinised and earliest cultivar ('Uso 31') and highest in the feminised and early ('Fedora 17') or mid-early ('Felina 32') ones (Chapter IV). This positive correlation observed between feminised phenotype and seed yield was supported by the correlation found between mean monoecy degree and seed yield in the monoecious segregating population (r was 0.47, P <0.001, Chapter IX). These findings suggest that it could be valuable to take the sex expression into account when selecting a monoecious cultivar for seed production.

(4) No cultivar among those tested provided the highest stem and seed yields simultaneously.

The variation of stem yields among cultivars was attributed to their earliness, the stem yields being lowest in the earliest cultivar and highest in the latest one. This was in agreement with the conclusions of de Meijer and Keizer (1994) and van der Werf et al. (1994). However, the highest seed yields were observed in early or mid-early cultivars. The occurrence of highest stem and seed yields in distinct cultivars constitutes one of the four antagonisms to the simultaneous production of both stem and seed in monoecious hemp as discussed in Chapter IV.

In the monoecious segregating population, a close linkage was observed between two QTLs for the stem yield and the plant node at which half of the feminisation process is performed (ND_m), respectively (Chapter IX). These QTLs had both positive additive effects due to the male parent. However, a high position of ND_m was correlated with low seed yields. Therefore, the close linkage observed between QTLs for stem yield and ND_m could result in opposite genetic effects on the stem and seed yields in monoecious hemp and, thereby, provide an explanation to the observation that no cultivar among those tested in field trials maximized both the stem and seed yields simultaneously.

(5) The sex chromosomes would include genetic factors involved in the determinism of the stem and seed yields.

Three QTLs explaining relatively high part of the variation in stem or seed yields were identified in the homology group including the sex locus. Two of them were detected in a dioecious population. The first one was a major QTL for the seed yield, explaining 17.8% of its variation. The second one explained 13.1% of the variation in stem yield and was due to an additive effect in the male parent, the allele being located on the X rather than Y chromosome being linked to an increase in stem yield. The third QTL explained 9.1% of the variation in stem yield in the monoecious population.

The present thesis suggests two main practical advices for the cultivation of monoecious hemp for the production of both stem and seed.

Firstly, an **early sowing** should be performed, *i.e.*, as soon as the meteorological conditions are favourable for soil tillage in the spring. Indeed, the earliest sowing dates provide the highest stem and seed yields and, as a result of earlier plant senescence, allow an earlier harvest and a lower risk of stem quality loss due to less favourable climatic conditions at the end of the crop cycle. On the basis of the tested range of sowing dates, sowing from mid- to end of April would be adequate in the studied areas of cultivation. However, relatively high stem yields ($\geq$ 9 t ha^{-1}) could still be expected from sowings up to early June in the loamy region and to mid-May in the Ardennes. Indeed, these sowing dates provided stem yields that were within the range reported with monoecious cultivars cultivated for fibre production only (Struik et al. 2000). Similarly, according to our results, high seed yields can still be expected from sowings carried out up to early June in the loamy region (from 1.0 to 1.6 t ha^{-1}, as observed in 2007 and 2008, respectively).

Secondly, **early or mid-early** – depending on the climatic conditions of the cultivation area – **feminised cultivars** should be selected. The use of such cultivars is recommended for both their higher seed yields and relative earliness. This latter feature aims to allow harvesting under still favourable meteorological conditions at the end of the crop cycle. Among the cultivars under study, the feminised early and mi-early cultivars ‘Fedora 17’ and ‘Felina 32’ provided the highest seed yields. The use of ‘Felina 32’ would be recommended under the climatic conditions of the loamy region. In the Ardennes, the earlier feminised cultivar ‘Fedora 17’ would be selected because of the higher probability of a delayed sowing date and of less favourable meteorological conditions at the end of the crop cycle. The very early cultivar ‘Uso 31’ offers the highest guarantee of harvest under favourable conditions thanks to its earliness; however, it provides the lowest seed yield.

10.2. SHORTCOMINGS

The development of the present work was firstly limited by our **poor knowledge of hemp**. The species had never been studied in our laboratory, and the cultivation of hemp was inexistent in our country. However, the present thesis started during a period of renewed interest for hemp in our region (Chapelle and Bigaré 2001; Baudoin 2004; PSPc sa 2008). Meetings and field visits were organised with distinct scientific actors who intended to install field trials of hemp in early 2007. These exchanges of information were of great help for the establishment of a protocol for the field trials. In addition, preliminary trials were carried out in controlled conditions in 2007 and 2008 in order to apprehend the growth, development and sensitivity of the sex expression to the environment in monoecious hemp.

The linkage analysis carried out in the present study was limited by three shortcomings. The first limitation concerned the use of **F_1 segregating populations**. Indeed, the construction of linkage maps with such populations is complicated by the ignorance of both the number of alleles segregating at each locus and linkage phases between markers in the parents, and by the fact that the number of segregating alleles can vary from one locus to another (Maliepaard et al. 1997; Wu et al. 2002a). In dioecious hemp, the use of F_1 segregating populations was constrained by the reproductive features of the plants. In contrast, in monoecious hemp, the possibility of selfing should allow the creation of classical segregating populations – such as backcross populations, F_2 or recombined-inbred lines – for which the presence of only two segregating alleles and the knowledge of the parental linkage phases make the linkage analysis easier. The existence of inbred lines in drug *Cannabis* (Forapani et al. 2001; Mandolino and Carboni 2004) suggests that the creation of inbred lines in hemp should not be prevented by inbreeding depression. However, despite this possibility of selfing in monoecious hemp, we chose to work with F_1 segregating populations in both dioecious and monoecious hemp. Indeed, on the one hand, the time required to obtain inbred lines and create segregating populations could be relatively long due to the high percentage of polymorphic loci [6 cycles of inbreeding would be necessary to reduce the percentage of polymorphic loci

from 61.3% (Forapani et al. 2001) to 1% for the 'truly' monoecious cultivar 'Fibrimon']. On the other hand, Forapani et al. (2001) concluded to a very high degree of polymorphism and heterozygosity in hemp and suggested the use of F_1 progenies for genetic mapping provided that the parental strains are chosen among heterozygous populations. The relatively high percentage of polymorphic loci obtained in the present study (from 44 to 67% according to the population, Chapter VII) supported the use of F_1 segregating populations for mapping in both dioecious and monoecious hemp as suggested by Forapani et al. (2001).

The second limitation with respect to linkage mapping was the use of AFLP markers, *i.e.*, **dominant markers**. This choice was motivated by the low number of codominant markers published in hemp at the beginning of the thesis – 27 microsatellites were reviewed in *C. sativa* by Mandolino and Carboni (2004) –, the availability of AFLP primer pairs providing sex-linked markers in hemp (Flachowsky et al. 2001) and the advantages of the AFLP technique, *i.e.*, its ability to amplify many loci and ease of implementation (Vos et al. 1995; Mueller and Wolfenbarger 1999; Meudt and Clarke 2007). However, the use of a dominant marker system for mapping in F_1 segregating populations reduces inexorably the power of detecting linkage (Maliepaard et al. 1997).

The third constrain with respect to linkage mapping concerned the relatively low **number of individuals** obtained from each 'Carmagnola' cross. Indeed, according to Maliepaard et al. (1997), such population sizes (77 and 76 individuals) greatly limit the power of detecting linkage in the case of both dominant markers and unknown parental linkage phases (Fig. 3.5). However, as reported in Chapter VII, the present 'Carmagnola' population sizes were of the same magnitude than those previously used in hemp mapping.

According to Maliepaard et al. (1997), the presence of F_1 segregating populations and the use of a dominant marker system resulted in a relatively low power of detecting linkage between markers (Fig. 3.5). In particular, the power of detecting linkage is lowest between pairs of C.8 markers linked in CR, RC or still RR phases in the parents. This is illustrated by the linkage analysis performed between two C.8 markers for which no linkage was detected although both were linked to sex (Table 7.5) and their recombination rate under CR, RC or still RR linkage phases was very low (Table 10.1).

Table 10.1 Linkage analysis between two C.8 markers, 2_289 and 6_255, mapped in the sex CG of the C2 population (Table 7.5)

Linkage phase [1]	Theta	Posterior probatility	LODs
CC	0.950	0.249	1.213
CR	0.000	0.251	1.218
RC	0.000	0.251	1.218
RR	0.050	0.249	1.213

[1] C = coupling, R = repulsion.

[2] Theta is the estimated recombination fraction.

The frequency of such situation was likely higher in both dioecious populations (C1 and C2) due to their higher percentage of C.8 markers (Table 7.3) and smaller size compared to the monoecious population (UF). This situation could explain the structure of these maps, *i.e.*, the presence of six and eight co-segregation groups including only two or three markers and a lower percentage of linked markers compared to UF (48, 63 and 75% in C1, C2 and UF, respectively; Table 7.4).

Finally, the integration of three distinct maps highlighted that a high number of markers (33% of them) likely derived from the sex chromosomes and allowed the identification of co-segregation groups likely located on the sex chromosomes of the monoecious map. However, the **unsaturated nature of the maps** implied that, with the exception of the co-segregation groups including the sex locus, the location of co-segregation groups on autosomes or sex chromosomes could not be ascertained.

The main limitation of the QTL analysis carried out in the present study was the **absence of repetition**. Indeed, each plant was a one-shot test since no cloning was performed. Obviously, the absence of confirmation limits the scope of the presently identified QTLs.

10.3. PERSPECTIVES

Distinct perspectives may be proposed from the results of the present work and the literature review. These perspectives are organised around two directions: (i) improving our understanding of the genetic determinism of the sex expression in monoecious hemp, and (ii) supporting the cultivation of monoecious hemp for the production of stem and seed.

10.3.1. Genetic determinism of the sex expression in monoecious hemp

The present thesis started from the assumption that the sex expression in monoecious hemp is a quantitative trait that could be investigated by the identification of QTLs. Our results were encouraging with respect to this initial assumption. Indeed, monoecious cultivars could be differentiated according to their sex expression (Chapters IV & V) and QTLs linked to a quantitative expression of the sexual phenotype have been identified (Chapter IX). In addition, the present results suggested that genetic factors located on the sex chromosomes of monoecious hemp would be involved in the determinism of the sex expression. Therefore, we believe that further dissections of the genetic architecture of the sex expression through the identification of QTLs could be of great interest for understanding the genetic determinism of the trait. Saturated maps of the sex expression in hemp should be constructed with a double objective: providing saturated maps for further QTL analyses of traits related to the sex expression, and characterizing the structure of the sex chromosomes in both dioecious and monoecious hemp, as initiated in dioecious hemp by Peil et al. (2003).

In this section, several indications are proposed for further experimental protocols aiming at (i) the construction of genetic maps of the sex expression in hemp and (ii) the identification of QTLs associated with the sex expression in monoecious hemp.

(i) Construction of genetic maps of the sex expression in hemp

Genetic material

Two distinct crosses were investigated in the present thesis. The first one involved two plants of a dioecious cultivar, and the second one involved two distinct monoecious cultivars. The interest of each type of cross was demonstrated. Firstly, sex-linked markers were identified in both dioecious maps, and co-segregation groups putatively located on the sex chromosomes could be detected in the monoecious map through markers shared by both the dioecious and monoecious maps. Secondly, QTLs associated with the sex expression were found in each population, and markers segregating with the sex in the dioecious populations and associated with variation of the sex expression in the monoecious one were identified. However, performing a cross involving both dioecious and monoecious cultivars could be worthwhile. According to Mandolino and Ranalli (2002), such cross results in both female and monoecious progenies. In such case, it would be possible to study the segregation of QTLs associated with variation of the sex expression in monoecious hemp among both female and monoecious progenies.

In addition, when linkage and QTL mapping are carried out in distinct segregating populations – as performed in the present study –, it would be valuable to use related rather than independent populations. Two distinct approaches could be considered. The first one would include the cloning of parental individuals and their use in distinct crosses for generating different but related segregating populations. The second approach would include distinct generations, F_1 individuals obtained from an initial cross being used as parents in another cross. In both cases, the pedigree of the individuals should be integrated in the linkage and QTL analyses. The advantages of performing an integrated analysis would be (i) the estimation of the QTL location and effect directly across populations instead of in each population independently and (ii) a higher number of individuals included in the analysis (Wu et al. 2007). Methodologies for linkage and QTL analyses integrating the pedigree of the individuals have been developed for QTL mapping in human populations, in particular (Lander and Green 1987; Xu and Atchley 1995; Almasy and Blangero 1998).

Type of segregating population

In monoecious hemp, the possibility of selfing suggests the creation of classical mapping populations derived from inbred lines, for which the presence of only two segregating alleles and the knowledge of the linkage phases between markers make the linkage analysis and interpretation of QTL effects easier. Alternatively, homozygous lines could be obtained in both dioecious and monoecious hemp through the creation of doubled haploid populations from F_1 segregating progenies (Collard et al. 2005).

However, the potential of linkage mapping in F_1 segregating populations should not be underestimated. As previously mentioned, the creation of such population requires a relatively short time and, compared to the doubled haploids, no technical expertise, while the accuracy of the recombination fraction estimated in F_1 populations can be significantly improved by integrating both dominant and codominant markers (Wu et al. 2002a).

Molecular genetic markers

The AFLP technique proved to be efficient for detecting sex-linked polymorphism. However, additional primer combinations should be tested in order to obtain saturated maps and distinguish the ten expected linkage groups. In addition to AFLP markers, codominant markers, such as microsatellites or SNPs (single nucleotide polymorphism), should be used. High-throughout SNP genotyping is now an efficient tool for the construction of high-density linkage maps and precise characterization of the genetic variation associated with a trait of interest (Varshney et al. 2009). According to Maliepaard et al. (1997), codominant markers are of great help for linkage mapping in F_1 segregating populations, in particular, since they allow a more accurate characterization of the linkage phases and thus more accurate estimation of the recombination frequencies. The power of codominant markers for integrating homologous co-segregation groups derived from the respective parents has been emphasised in outbred progenies (Grattapaglia and Sederoff 1994; Ritter and Salamini 1996).

(ii) Identification of QTLs associated with the sex expression in monoecious hemp

Repetition of the QTL analysis and test of distinct environmental conditions

The possibility of repeating the phenotyping of a same genotype under a same environment is necessary to confirm the presence of the identified QTLs. For example, *n* replicates of each genotype should be cultivated and phenotyped simultaneously, and the experiment should be repeated in time, *e.g.*, as performed by Dufey et al. (2009) in their search for QTLs linked to resistance mechanisms to ferrous iron toxicity in rice. However, when F_1 segregating populations are used, repeating the QTL analysis implies the cloning of plants. Hemp cloning could be achieved by the creation and maintenance of cuttings for each plant (Peil et al. 2000) or *in vitro* culture of tissue (Feeney and Punja 2003), though this latter method includes a risk of somaclonal variation (Skirvin 1978; Larkin and Scowcroft 1981).

In addition, multi-environments should be tested. The evaluation of the trait of interest under distinct environments will allow assessing the stability of the QTLs across environments, or, in other words, the presence of QTL x environment interactions. Methodologies developed for QTL mapping in multi-environments make full use of the framework of mixed models (Malosetti et al. 2004; Boer et al. 2007; van Eeuwijk et al. 2010). The general procedure involves the fit of a model to the residual genetic variation, *i.e.*, the variation which is caused by not modelled QTLs elsewhere on the genome (Malosetti et al. 2004). This model will allow different genetic variances across environments and correlations between environments. Then, a search for QTLs is performed in each environment by interval mapping. Finally, the interactions between each QTL and the environment are tested. When the QTL x environment interaction is significant, a regression analysis can relate the sizes of QTL effects to environment variables (van Eeuwijk et al. 2010). In the case of hemp, testing for the presence of QTL x environment interactions under a range of photoperiodic conditions could be of interest given the significant effect of the photoperiod on both the flowering time and sex expression (Freeman et al. 1980; de Meijer and Keizer 1994; Lisson et al. 2000b). Such study could shed some light on how the genotypic response of the sex expression is affected by the environment.

Characterization of the sex expression of monoecious hemp plants

In the present work, we described the sexual phenotype of monoecious hemp plants by modelling the distribution of male and female flowers along the stem and expressing the parameters of the function in terms of QTLs. The relatively high number of QTLs identified following this approach supported its interest for characterizing the variability of the sex expression among monoecious hemp plants.

Actually, the idea underlying the modelling approach is close to the so-called *functional mapping* (Ma et al. 2002; Wu and Min 2006). Rather than mapping complex traits *per se*, we map parameters of response curves describing the behavior of the trait in relation to environmental and developmental factors (van Eeuwijk et al. 2010). According to Ma et al. (2002), results from functional mapping are closer to the biological reality because the underlying biological mechanisms are considered. However, in the present study, the use of a logistic function to describe the sex expression of monoecious hemp plants did not rely on any previously described biological process. Nevertheless, in view of the sensitivity of the sex expression to hormonal treatments, a physiological interpretation of the present modelling approach seems possible through the presence of hormonal gradients along the stem (Chapter VIII). Considering both the relatively high number of QTLs identified using the modelling approach and its possible biological meaning, we believe that similar strategies might be fruitful for further QTL analyses of the genetic determinism of the sex expression in monoecious hemp.

Finally, the consideration of the sex expression as a quantitative trait led us to the use of quantitative and modelling approaches. In this scope, the recent sequencing of the hemp genome (van Bakel et al. 2011) provides new insights for the study of the genetic determinism of the sex expression in hemp. As stated by van Bakel et al. (2011), a more complete assembly of the sequences on the X and Y chromosomes could provide valuable information on the mechanism of sex determination in hemp. To this purpose, the next-generation sequencing of monoecious cultivars would be worthwhile.

In addition, our understanding of the genetic determinism of the sex expression in monoecious hemp would certainly greatly benefit from the integration of distinct complementary approaches. The quantitative approach of the genetic determinism of sex expression should be linked to a transcriptomic approach of the sexual differentiation, as initiated by the work of Moliterni et al. (2004). The genetic differentiation of the male and female flowers should be investigated in both dioecious and monoecious plants, which would be simultaneously integrated in the quantitative study of the determinism of the sex expression. Last but not least, endogenous hormonal contents should be measured before initiation of flowering and at distinct times once flowering initiation occurred. This would notably allow to test the hypothesis of an effect of hormonal gradients along the stem on the sex expression in monoecious hemp.

10.3.2. The cultivation of monoecious hemp for the production of stem and seed

The literature review and Chapter IV showed that managing a dual-purpose hemp crop requires compromises for most agronomic factors. In particular, late flowering cultivars increase the stem yield by delaying the flowering, which is related to a lower allocation of dry matter to the inflorescence (van der Werf et al. 1994), while no cultivar among those tested in the present work provided the highest stem and seed yields simultaneously. However, the economic context can lead farmers to the simultaneous production of stem and seed in hemp (Fournier and Beherec 2006; Chanvre Wallon 2013). Although the use of monoecious cultivars is indicated to this purpose, the literature review (Chapter II) and results reported in the present work (Chapters IV & VIII) indicate that the success of a dual-hemp production could greatly benefit from both crop physiological and genetic studies of the relationship between stem and seed yield. These studies would investigate (i) the understanding of the mechanisms that govern the allocation of assimilates to the stem and to the seed and their responses to the environment and (ii) the dissection of the genetic basis of the yields in stem and seed, respectively.

A thorough crop physiological study of the fibre production in hemp has been carried out by van der Werf et al. (1994, 1995a, 1995b, 1995c, 1996).

On the opposite, the production of seed in hemp has received little attention. Therefore, the perspectives proposed here relatively to the crop physiological approach focus on the study of the seed production in hemp.

(i) Crop physiological appraisal of the production of seed in hemp

First and foremost, a crop physiological appraisal of the seed yield should include the characterization of the yield components. The seed yield of hemp plants could be divided into three components: the number of inflorescences per m², the number of seeds per inflorescence and the 1000-kernel weight, similarly to grain crops such as wheat or barley (Slafer 2007; Yin et al. 2002). Then, according to Egli (2004), an accurate definition of the seed-fill duration should be determined. In hemp, the seed-fill duration could be defined by the duration from end of female flowering to fruit set (Tables 2.1 and 4.1; Mediavilla et al. 1998). In addition, according to Picault (2006), the occurrence of green seeds at harvest should be assessed.

On the basis of the present results and literature review, the effect of four main factors should be investigated on the seed yield and its components, seed-fill duration and the occurrence of green seeds in field trials. Firstly, monoecious **cultivars** covering a wide range of both earliness and sex expression should be tested. Secondly, the effect of distinct **sowing dates** including very early sowings should be assessed. In the present work, the earliest sowing dates applied were performed in mid-April and provided the highest seed yields. The assessment of earlier sowing dates would allow to test, on the one hand, whether further gains of seed yield can be achieved by advancing the sowing date and, on the other hand, the effect of the sowing date on the seed-fill duration. Under the meteorological conditions of Corroy-le-Grand, sowing dates as early as mid-March might be tested. On the opposite, the latest sowings should be carried out around end of June in Corroy-le-Grand or mid-June in Michamps since later sowing dates resulted in too late harvests. Thirdly, distinct **seeding rates** should be investigated. According to Picault (2006), a thorough understanding of the effect of the plant density on the seed yield formation should include very diverse seeding rates, *e.g.*, ranging from 10 to 70 kg ha^{-1}. In the present field trials, seeding rates of 55, 60 and 65 kg ha^{-1} were used on the basis of the values suggested for dual production in hemp (Picault 2006). Fourthly, the effect of different levels of **nitrogen fertilization**

sould be considered. According to Picault (2006), an increased nitrogen fertilization would positively affect the seed yield. However, the management of the nitrogen fertilization for the production of seed in hemp has been poorly studied. In addition, although not consistent with each other, effects of the nitrogen status on the sex expression in hemp have been reported (Arnoux 1966a, b; Freeman et al. 1980; van der Werf and van der Berg 1995).

Finally, in addition to the seed yield and its components, the sex expression, flowering phenology and stem yield should be determined. Also, measurements of the plant density, LAI and intercepted PAR would be necessary to assess the relationships between plant senescence, seed-fill duration and seed yield. Not only the amount but also the quality of the production would be assessed. The determination of the fibre content of the stem and oil and protein contents of the seed would be valuable to support the crop management and ensure a quality of production responding to given industrial criteria.

(ii) Dissection of the genetic basis of the yields in stem and seed

The present work identified QTLs associated with variation in yields in stem or seed (Chapter IX). One major QTL was detected for the seed yield close to a marker that segregated with sex in one of both dioecious populations. While these results are promising for further studies aiming the identification of QTLs for yields in hemp, the study of the genetic basis of a complex such as yield could benefit from the dissection of the trait into 'component traits' (van Eeuwijk et al. 2010; Tardieu and Tuberosa 2010). This strategy includes two steps. The first step consists in identifying 'component' traits with properties exhibiting a higher heritability than the complex trait itself. Then, the 'functional mapping' approach is applied: the characterizations of the curves describing the behaviour of the component traits in response to an environmental variable are expressed in terms of QTLs (van Eeuwijk et al. 2010).

In hemp, the dissection of the genetic basis of the stem yield could include component traits such as the stem length, diameter and slenderness (g cm^{-2}), as suggested by the work of van der Werf et al. (1995b). In particular, the consideration of these traits could be of interest for investigating the relationship between stem yield and sex in dioecious hemp since the sexual dimorphism contributes to the variability in plant height and weight (van der Werf and van

der Berg 1995). For the seed yield, the number of seeds per inflorescence and the 1000-kernel weight could be used. In addition, variables related to the flowering phenology and sex expression should be integrated in the dissection of the genetic basis of yields, as suggested by the relationships between sex expression, earliness and yields reported here. However, when F_1 segregating populations are used, the dissection of the genetic architecture of the yields in stem and seed by the search of QTLs will firstly face with the possibility of repeating the analysis and testing distinct environments, as previously mentioned. In this case, the QTL analysis could be performed by using cloned segregating populations maintained under controlled conditions, while the selection of appropriate component traits and the study of their relationships with the yields would be carried out in field trials including a range of genotypes and environments.

CONCLUSIONS

The present thesis provided new insights to the understanding of the genetic determinism of the sex expression in monoecious hemp and its relationships with flowering phenology and stem and seed yields.

The present quantitative approach highlighted the existence of a genetic basis for the sex expression in monoecious hemp, stated its constitution in sex chromosomes and suggested the presence of genetic factors involved in the genetic determinism of the sex expression on the sex chromosomes. From a methodological point of view, the identification of QTLs linked to the sex expression encourages the use of QTL-based approaches for further investigations aiming to dissect the genetic basis of the sex expression in monoecious hemp. To this purpose, the present results suggest that modelling the distribution of male and female flowers along the stem might be a fruitful strategy for characterizing the variability of the sex expression among monoecious plants. Furthermore, sex expression and earliness appeared partly linked to each other in monoecious hemp, while higher seed yields were associated with (mid-) early feminised cultivars. Therefore, taking the sex expression into account could be valuable when selecting a monoecious cultivar for seed production.

Due to their intimate relations, integrating the dissections of the genetic basis of the sex expression, earliness and yields in stem and seed in hemp appears relevant from both fundamental and agronomical points of view. The success of such integrated approach will rely on the adequate combination of crop physiological and genetic studies with the constraints due to the reproductive morphology of hemp.

LITERATURE CITED

Ainsworth C (2000) Boys and girls come out to play: the molecular biology of dioecious plants. Ann Bot 86:211-221

Alghanim HJ, Almirall JR (2003) Development of microsatellite markers in *Cannabis sativa* for DNA typing and genetic relatedness analyses. Anal Bioanal Chem 376:1225-1233

Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62:1198-211

Amaducci (2003) HEMP-SYS. Design, development and up-Scaling of a sustainable production system for hemp textiles. An integrated quality systems approach. J Int Hemp Ass 8: 79-83

Amaducci S, Amaducci MT, Benati R, Venturi G (2000) Crop yield and quality parameters of four annual fibre crops (hemp, kenaf, maize and sorghum) in the North of Italy. Ind Crop Prod 11:179-186

Amaducci S, Colauzzi M, Bellocchi G, Venturi G (2008a) Modelling post-emergent hemp phenology (*Cannabis sativa* L.): theory and evaluation. Eur J Agron 28:90-102

Amaducci S, Colauzzi M, Zatta A, Venturi G (2008b) Flowering dynamics in monoecious and dioecious hemp genotypes. J Ind Hemp 13:5-19

Amaducci S, Zatta A, Pelatti F, Venturi G (2008c) Influence of agronomic factors on yield and quality of hemp (*Cannabis sativa* L.) fibre and implication for an innovative production system. Field Crops Res. 107:161-169

APG III (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III Bot J Linn Soc 161:105-121

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 108:796-815

Arnoux M (1966a) Influence des facteurs du milieu sur l'expression de la sexualité du chanvre monoïque (*Cannabis sativa* L.). II - Action de la nutrition azotée. Ann Amélior Plantes 16:123-134

Arnoux M (1966b) Influence des facteurs du milieu sur l'expression de la sexualité du chanvre monoïque (*Cannabis sativa* L.). III – Note sur l'interaction entre le cycle photopériodique et la nutrition azotée. Ann Amélior Plantes 16:259-262

Arnoux M, Mathieu G (1969) Influence du milieu sur le phénotype sexuel de descendances F_1 issues du croisement entre types dioïques et monoïque du chanvre (*Cannabis sativa* L.). Ann Amélior Plantes 19:53-58

Asins MJ (2002) Review present and future of quantitative trait locus analysis in plant breeding. Plant breeding 291:281-291

Asumadu H, Summerfield RJ, Ellis RH, Qi A (1998) Variation in the durations of the photoperiod-sensitive and photoperiod-insensitive phases of post-first flowering development in maturity isolines of soyabean [*Glycine max* (L .) Merrill] 'Clark'. Ann Bot 82:773-778

Atienza SG, Satovic Z, Petersen KK, *et al.* (2003) Identification of QTLs associated with yield and its components in *Miscanthus sinensis* Anderss. Euphytica 132:353-361

Barney's farm (2014) Website of Barney's farm Seedbank. Retrieved February 7, 2014 from http://www.barneysfarm.com/

Baudoin JG (2004) Les fibres végétales en Région wallonne. Les potentialités du chanvre et ses utilisations. Document réalisé dans le cadre du projet Farr-wal mené par la Faculté universitaire des Sciences agronomiques de Gembloux pour le compte de ValBiom, avec le soutien de la Région Wallonne - Direction générale de l'agriculture. 42 p

Beherec O (2000) FNPC hemp breeding and CCPSC seeds production (France). Proceedings of "Bioresource Hemp 2000", 13-16 September 2000, Wolfsburg, Germany

Belanger G, Gastal F, Warembourg FR (1994) Carbon balance of tall fescue (*Festuca arundinacea* Schreb.): effects of nitrogen fertilization and the growing season. Ann Bot 74:653–659

Bennett MD, Price HJ, Johnston S (2008) Anthocyanin inhibits propidium iodide DNA fluorescence in *Euphorbia pulcherrima*: implications for genome size variation and flow cytometry. Ann Bot 101:777-790

Bertoli A, Tozzi S, Pistelli L, Angelini LG (2010) Fibre hemp inflorescences: From crop-residues to essential oil production. Ind Crop Prod 32:329-337

Bocsa I, Karus M (1998) The cultivation of hemp: botany, varieties, cultivation and harvesting. Hemptech, Sebastopol

Boer MP, Wright D, Feng L, *et al.* (2007) A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics 177:1801-13

Borthwick HA, Scully NJ (1954) Photoperiodic responses in hemp. Bot Gaz 116:14-29

Bouloc P (2006) L'économie agricole du chanvre. In: Bouloc, P. (ed). Le chanvre industriel, production et utilisations. Ed. France Agricole, Paris, France. Pp 185-208

Brown H., Prescott R (2006) Applied mixed models in medicin. 2nd edition. John Wiley and Sons Ltd. Chichester, England

Callaway JC (2004) Hempseed as a nutritional resource: an overview. Euphytica 140, 65-72.

Carberry PS, Muchow RC, Williams R, Sturtz JD, McCown RL (1992) A simulation model of kenaf for assisting fibre industry planning in Northern Australia I. General introduction and phenological model. Aust J Agr Res 43:1501-1513

Carboni A, Paoletti C, Moliterni VMC, Ranalli P, Mandolino G (2000) Molecular markers as genetic tools for hemp characterization. Proceedings of the "Bioresource Hemp 2000", 13-16 September 2000, Wolfsburg, Germany

Chabbert B, Kurek B, Beherec O (2006) Physiologie et botanique du chanvre industriel. In : Bouloc P (ed). Le chanvre industriel, production et utilisations. Ed. France Agricole, Paris, France. Pp 43-74

Chailakhyan MK, Khryanin VN (1978) The influence of growth regulators absorbed by the root on sex expression in hemp plants. Planta 138:181-184

Chailakhyan MK, Khryanin VN (1979) The role of leaves in sex expression in hemp and spinach. Planta 144:205-207

Chanvre Wallon (2012) Procès verbal de l'Assemblée générale ordinaire tenue au Centre d'Economie Rurale, à Marloie, le 4 juin 2012

Chanvre Wallon (2013) Website of the asbl Chanvre Wallon. Retrieved Novembre 7, 2013 from http://www.chanvrewallon.be/site/asbl.html

Chappelle H, Bigaré P (2001) Essais de comportement du chanvre en différentes stations de Wallonie. asbl Sorghal, Ministère de la Région Wallonne - Direction générale de l'agriculture, Ministère de l'emploi et du travail. 148 p

Charles-Edwards DA (1982) Physiological determinants of crop growth. Sydney: Academic Press. 158 p

Charlesworth D (2002) Plant sex determination and sex chromosomes. Heredity 88:94-101

Charlesworth D, Charlesworth B, Marais G (2005) Steps in the evolution of heteromorphic sex chromosomes. Heredity 95:118-128

Clark MS, Parker JS, Ainsworth CC (1993) Repeated DNA and heterochromatin structure in *Rumex acetosa*. Heredity 70:527-536

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-196

Cooper M, van Eeuwijk F a, Hammer GL, *et al.* (2009) Modeling QTL for complex traits: detection and context for plant breeding. Curr Opin Plant Biol 12:231-40

Correia F, Roy DN, Goel K (2001) Chemistry and delignification kinetics of canadian industrial hemp. J Wood Chem Tech 21:97-111

Cosentino SL, Riggi E, Testa G, *et al.* (2013) Evaluation of European developed fibre hemp genotypes (*Cannabis sativa* L.) in semi-arid Mediterranean environment. Ind Crop Prod 50:312-324

Cosentino SL, Testa G, Scordia D, Copani V (2012) Sowing time and prediction of flowering of different hemp (*Cannabis sativa* L.) genotypes in southern Europe. Ind Crop Prod 37:20-33

Costich D, Meagher T, Yurkow E (1991) A rapid means of sex identification in *Silene latifolia* by use of flow cytometry. Plant Mol Biol Rep 9:359-370

Cromack HTH (1998) The effect of cultivar and seed density on the production and fibre content of *Cannabis sativa* in southern England. Ind Crop Prod 7:205-210.

de Meijer EPM (1995) Fibre hemp cultivars: a survey of origin , ancestry , availability and brief agronomic characteristics. J Int Hemp Ass 2:66-73.

de Meijer EPM, Bagatta M, Carboni A, Crucitti P, Moliterni VMC, Ranalli P, Mandolino G (2003) The inheritance of chemical phenotype in *Cannabis sativa* L. Genetics 163:335-346

de Meijer EPM, Keizer LCP (1994) Variation of *Cannabis* for phenological development and stem elongation in relation to stem production. Field Crops Res 38:37-46

de Meijer EPM, Keizer LCP (1996) Patterns of diversity in *Cannabis*. Genet Resour Crop Ev 43:41-52

Dellaporta SL, Calderon-Urrea A (1993) Sex determination in flowering plants. Plant cell 5:1241-51

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via KM algorithm. J Roy Stat Soc Ser B 39:1-38

Desanlis F (2006) Agronomie et culture du chanvre. In: Bouloc, P. (ed). Le chanvre industriel, production et utilisations. Ed. France Agricole, Paris, France. Pp 147-168

Di Stilio VS, Kesseli R V, Mulcahy DL (1998) A pseudoautosomal random amplified polymorphic DNA marker for the sex chromosomes of *Silene dioica*. Genetics 149:2057-2062

Diaz A, Fergany M, Formisano G, *et al.* (2011) A consensus linkage map for molecular markers and quantitative trait loci associated with economically important traits in melon (*Cucumis melo* L.). BMC plant biology 11:111

Doerge R, Zeng Z-B (2011) Module 4: QTL mapping. European Summer institute, Université de Liège, September 2011

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet 3:43-52

Dolezel J, Bartos J (2005) Plant DNA flow cytometry and estimation of nuclear genome size. Ann Bot 95:99-110

Dolezel J, Göhde W (1995) Sex determination in dioecious plants *Melandrium album* and *M. rubrum* by using high-resolution flow cytometry. Cytometry 19:103-106

Dolezel J, Greilhuber J (2010) Nuclear genome size: are we getting closer? Cytometry Part A 77A:635-342

Dolezel J, Greilhuber J, Lucretti S, Meister A, Lysak MA, Nardi L, Obermayer R (1998) Plant genome size estimation by flow cytometry: inter-laboratory comparison. Ann Bot 82 (Suppl. A):17-26

Dolezel J, Sgorbati S, Lucretti S (1992) Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. Physiol Plantarum 85:625-631

Donald CM (1962) In search of yield. J Aust Inst Agric Sci 28:171-178

Donnison IS, Siroky J, Vyskot B, *et al.* (1996) Isolation of Y chromosome-specific sequences from *Silene latifolia* and mapping of male sex-determining genes using representational difference Analysis. Genetics 144:1893-1901

Dufey I, Hakizimana P, Draye X, *et al.* (2009) QTL mapping for biomass and physiological parameters linked to resistance mechanisms to ferrous iron toxicity in rice. Euphytica 167:143-160

Egli DB (1994) Seed growth and development. In: Boote KJ, Bennett JM, Sinclair TR, Paulsen GM (1994) Physiology and determination of crop yield. Ed. ASA, Inc, CSSA, Inc and SSA, Inc, Madison, Wisconsin, USA. Pp 127-148

Egli DB (2004) Seed-fill duration and yield of grain crops. Adv. Agron. 83:243-279

Egli DB, Guffy RD, Heitholt JJ (1987) Factors associated with reduced yields of delayed plantings of soybean. J Agron Crop Sci 159:176-185

EIHA (2011) Position paper of the European Industrial Hemp association (EIHA), 29th April 2011. Retrieved Novembre 7, 2013 from http://www.eiha.org/

Ellen J, Spiertz JHJ (1980) Effects of rate and timing of nitrogen dressings on grain yield formation of winter wheat (*T. aestivum* L.). Fert Res 1:177-190

Emboden WA (1974) *Cannabis* - A polytyoic genus. Econ Botan 28:304-310.

Evrard A (2006) Sorption behaviour of lime-hemp concrete and its relation to indoor comfort and energy demand. Proceedings of the 23rd Conference on passive and low energy Architecture", 6-8 September 2006, Geneva, Switzerland.

Evrard A, De Herde A (2010) Hygrothermal performance of lime-Hemp wall assemblies. J Build Phys 34:5-25

Faeti V, Mandolino G, Ranalli P (1996) Genetic diversity of *Cannabis sativa* L. germplasm based on RAPD markers. Plant Breeding 115:367-370

FAO (2006) World reference base for soil resources 2006. 2nd edition. World Soil Resources Reports No. 103. FAO, Rome.

Faux AM, Berhin A, Dauguet N, Bertin P (2014) Sex chromosomes and quantitative sex expression in monoecious hemp (*Cannabis sativa* L.). Euphytica 196:183–197

Faux AM, Draye X, Lambert R, d'Andrimont R, Raulier P, Bertin P (2013) The relationship of stem and seed yields to flowering phenology and sex expression in monoecious hemp (*Cannabis sativa* L.). Eur J Agron 47:11-22

Fazio G, Staub JE, Stevens MR (2003) Genetic mapping and QTL analysis of horticultural traits in cucumber (*Cucumis sativus* L.) using recombinant inbred lines. Theor Appl Genet 107:864-874

Feeney M, Punja ZK (2003) Tissue culture and Agrobacterium-mediated transformation of hemp (*Cannabis sativa* L.). In Vitro Cell Dev-Pl 39:578-585

Filatov DA, Laporte V, Vitte C, Charlesworth D (2001) DNA diversity in sex-linked and autosomal genes of the plant species *Silene latifolia* and *Silene dioica*. Mol Biol Evol 18:1442-54

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

Flores-Sanchez IJ, Verpoorte R (2008) Secondary metabolism in *Cannabis*. Phytochem Rev 7:615-639

Forapani S, Carboni A, Paoletti C, Moliterni VMC, Ranalli P, Mandolino G (2001) Comparison of hemp varieties using random amplified polymorphic DNA markers. Crop Sci 41:1682-1689

Fournier G, Beherec O (2006) La sélection du chanvre. In: Bouloc P (ed) Le chanvre industriel, production et utilisations. France Agricole, Paris. Pp 89-110

Freeman DC, Harper KT, Charnov EL (1980) Sex change in plants: old and new observations and new hypotheses. Oecologia 47:222-232

Galoch E (1980) The hormonal control of sex differentiation in dioecious plants of hemp *Cannabis sativa* 3. The level of phytohormones in male and female plants in the different stages of their development. Acta Physiol Plant 2:31-40

Garcia AAF, Kido EA, Meza AN, Souza HMB, Pinto LR, Pastina MM, Leite CS, da Silva JAG, Ulian EC, Figueira A, Souza AP (2006) Development of an integrated genetic map of a sugarcane (Saccharum spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. Theor Appl Genet 112:298-314

GC lin et chanvre (2010) Compte-rendu de la reunion du groupe consultative "lin et chanvre" du 15 octobre 2010, Bruxelles. Retrieved November 7, 2013 from http://ec.europa.eu/agriculture/consultations/advisory-groups/flax-hemp/index_en.htm

Gilmour AR (2007) Mixed model regression mapping for QTL detection in experimental crosses. Comput Stat Data An 51:3749-3764

Godin B, Lamaudière S, Agneessens R, *et al.* (2013) Chemical characteristics and biofuel potential of several vegetal biomasses grown under a wide range of environmental conditions. Ind Crop Prod 48:1-12

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137:1121-1137

Groos C, Robert N, Bervas E, Charmet G (2003) Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. Theor Appl Genet 106:1032-40

Haley CS, Knott S (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315-24

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315-24

Haufe J, Karus M (2011) Assessment of life cycle studies on hemp fibre. Biowerkstoff-Report 6:26-27

Haverkort AJ, Boerma M, Velema R, Van de Waart M (1992) The influence of drought and cyst nematodes on potato growth. 4. Effects on crop growth under field conditions of four cultivars differing in tolerance. Neth J Plan Path 98: 179–191

Hennink S (1994) Optimisation of breeding for agronomic traits in fibre hemp (*Cannabis sativa* L .) by study of parent-offspring relationships. Euphytica 69-76

Herrmann D, Boller B, Studer B, *et al.* (2006) QTL analysis of seed yield components in red clover (Trifolium pratense L.). Theor Appl Genet 112:536-45

Heslop -Harrison J (1956) Auxin and sexuality in *Cannabis sativa*. Physiol Plantarum 9:588-597

Hirata K (1924) Cytological basis of the sex determination in *Cannabis sativa* L. Jpn J Genet 4:198-201

Hirose T (2005) Development of the Monsi-Saeki theory on canopy structure and function. Ann Bot 95:483-94

Hoarau JY, Grivet L, Offmann B, D'Hont A, Risterucci AM, Roques D, Glaszmann JC, Grivet L (2001) Genetic dissection of a modern sugarcane cultivar (Saccharum spp.) I. Genome mapping with AFLP markers. Theor Appl Genet 103:84-97

Hobson J, Karus M (2008) Why hemp should be a crop of the future. Retrieved November 7, 2013 from http://www.eiha.org/

Horie T (1994) Crop ontogeny and development. In: Boote KJ, Bennett JM, Sinclair TR, Paulsen GM (1994) Physiology and determination of crop yield. Ed. ASA, Inc, CSSA, Inc and SSA, Inc, Madison, Wisconsin, USA. Pp 153-180

Index Mundi (2013) Retrieved November 7, 2013 from http://www.indexmundi.com/commodities/?commodity=wheat&months=300

IRM (2010) Website of the Institut Royal Météorologique de Belgique. http://www.meteo.be/.

Jacobsen P (1957) The sex chromosomes in *Humulus*. Hereditas 43:357-370

Janick J, Whipkey A (eds) (2007) Issues in New Crops and New Uses. Proceedings of the Sixth National Symposium "Creating Markers for Economic Development of New Crops and New Uses". ASHS Press, Alexandria, VA, USA

Jansen J, De Jong, AG, van Ooijen JW (2001) Constructing dense genetic linkage maps. Theor Appl Genet 102:1113-1122

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. Genetics 135:205-11

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447-55

Jiang C, Zeng Z (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. Genetica 101:47-58

Jiang C, Zeng ZB (1997) Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. Genetica 101:47-58

Johnson DL, Jansen RC, Van Arendonk JAM (1999) Mapping quantitative trait loci in a selectively genotyped outbred population using a mixture model approach. Genet Res 73: 75-83

Kantolic AG, Slafer GA (2001) Photoperiod sensitivity after flowering and seed number determination in indeterminate soybean cultivars. Field Crop Res 72:109-118

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152:1203-16

Karus M, Vogt D (2004) European hemp industry: cultivation , processing and product lines. Euphytica 140:7-12

Kiniry JR, Jones CA, O'toole JC, Blanchet R, Cabelguenne M, Spanel DA (1989) Radiation-use efficiency in biomass accumulation prior to grain-filling for five grain-crop species. Field Crop Res 20:51-64

Knott SA, Haley CS (1992) Maximum likelihood mapping of quantitative trait loci using full-sib families. Genetics 132:1211-1222

Lambinon J, Delvosalle L, Duvigneaud J (eds) (2004) Nouvelle flore de Belgique, du Grand-Duché de Luxembourg, du Nord de la France et des régions voisines. Editions du Patrimoine du Jardin Botanique National de Belgique. Cinquième édition. 1167 p

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199

Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. Proceedings of the National Academy of Sciences of the United States of America 84:2363-7.

Lander ES, Green P, Abrahanson J, Barlow A, Daley M, Lincoln S, Newburg L (1987) MAPMAKER: an interactive computing package for constructing

primary genetic linkages of experimental and natural populations. Genomics 1:174-181

Larkin PJ, Scowcroft WR (1981) Somaclonal variation - a novel source of variability from cell cultures for plant improvement. Theor Appl Genet 60:197-214

Lin M, Lou X, Chang M, Wu R (2003) A general statistical framework for mapping quantitative trait loci in nonmodel systems: issue for characterizing linkage phases. Genetics 165:901-913

Lisson SN, Mendham NJ (2000). Cultivar, sowing date and plant density studies of fibre hemp (*Cannabis sativa* L.) in Tasmania. Aust J Exp Agric 40:975-986

Lisson SN, Mendham NJ, Carberry PS (2000a) Development of a hemp (*Cannabis sativa* L.) simulation model 1. General introduction and the effect of temperature on the pre-emergent development of hemp. Aust J Exp Agric 40:405-411.

Lisson SN, Mendham NJ, Carberry PS (2000b) Development of a hemp (*Cannabis sativa* L.) simulation model. 2. The flowering response of two hemp cultivars to photoperiod. Aust J Exp Agr 40:413-417

Lisson SN, Mendham NJ, Carberry PS (2000c) Development of a hemp (*Cannabis sativa* L.) simulation model 4. Model description and validation. Aust J Exp Agric 40:425-432.

Liu Z, Moore PH, Ma H, Ackerman CM, Ragiba M, Yu Q, Pearl HM, Kim MS, Charlton JW, Stiles JI, Zee FT, Paterson AH, Ming R (2004) A primitive Y chromosome in papaya marks the beginning of sex chromosome evolution. Nature 427:348-352

Loureiro J, Rodriguez E, Dolezel J, Santos C (2006) Comparison of four nuclear isolation buffers for plant DNA flow cytometry. Ann Bot 98:679-689

Lynch M, Walsh B (1998) Genetics and Analysis of Quantitative Traits. Sinauer Associates, Sunderland, MA.

Ma C-X, Casella G, Wu R (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. Genetics 161:1751-62

Major DJ (1980) Photoperiod response characteristics controlling flowering of nine crop species. Can J Plant Sci 60:777-784

Maliepaard C, Jansen J, van Ooijen JW (1997) Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. Genet Res 70:237-250

Malosetti M, Ribaut JM, Vargas M, *et al.* (2007) A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (Zea mays L.). Euphytica 161:241-257

Malosetti M, Visser RGF, Celis-Gamboa C, van Eeuwijk FA (2006) QTL methodology for response curves on the basis of non-linear mixed models, with an illustration to senescence in potato. Theor Appl Genet 113:288-300

Malosetti M, Voltas J, Romagosa I, *et al.* (2004) Mixed models including environmental covariables for studying QTL. Euphytica 37:139-145

Mandolino G, Carboni A (2004) Potential of marker-assisted selection in hemp genetic improvement. Euphytica 140:107-120

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

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

Mandolino G, Ranalli P (2002) The applications of molecular markers in genetics and breeding of hemp. J Ind Hemp 7:7-23

Margarido GR, Mollinari M, Broman K, Garcia G (2012) Package 'onemap': software for constructing genetic maps in experimental crosses: full-sib, RILs, F2 and backcrosses. Version 2.0-3. http://cran.r-project.org/web/packages/onemap/index.html

Margarido GR, Souza P, Garcia F (2007) OneMap: software for genetic mapping in outcrossing species. Hereditas 144:78-79

Marguerit E, Boury C, Manicki A, *et al.* (2009) Genetic dissection of sex determinism, inflorescence morphology and downy mildew resistance in grapevine. Theor Appl Genet 118:1261-78

Martinez O, Curnow RN (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor Appl Genet 85:480-488.

Meagher RB, McLean MD, Arnold J (1988) Recombination within a subclass of restriction fragment length polymorphisms may help link classical and molecular genetics. Genetics 120:809-818

Mediavilla V, Jonquera M, Schmid-Slembrouck I, Soldati A (1998) Decimal code for growth stages of hemp (*Cannabis sativa* L.). J Int Hemp Assoc 5:65,68-74

Mediavilla V, Leupin M, Keller A (2001) Influence of the growth stage of industrial hemp on the yield formation in relation to certain fibre quality traits. Ind Crop Prod 13:49-56

Meijer WJM, van der Werf HMG, Mathijssen EWJM, van den Brink PWM (1995) Constraints to dry matter production in fibre hemp (*Cannabis sativa* L.). Eur J Agron 4:109-117

Menzel MY (1964) Meiotic chromosomes of monoecious Kentucky hemp (*Cannabis sativa*). Bull Torrey Bot Club 91:193-205

Meudt HM, Clarke AC (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances. Trends Plant Sci 12:106-117

Miao H, Zhang S, Wang X, *et al.* (2011) A linkage map of cultivated cucumber (*Cucumis sativus* L.) with 248 microsatellite marker loci and seven genes for horticulturally important traits. Euphytica 182:167-176

Miles C, Wayne M (2008) Quantitative trait locus (QTL) analysis. Nature Education 1(1):208

Mills A, Moot DJ, Jamieson PD (2009) Quantifying the effect of nitrogen on productivity of cocksfoot (*Dactylis glomerata* L.) pastures. Eur J Agron 30:63-69

Ming R, Bendahmane A, Renner SS (2011) Sex chromosomes in land plants. Annu Rev Plant Biol 62:485-514

Mohan Ram HY, Jaiswal VS (1972) Induction of male flowers on female plants of *Cannabis sativa* by gibberellins and its inhibition by abscisic acid. Planta 105:263-266

Mohan Ram HY, Nath R (1964). The morphology and embryology of *Cannabis sativa* Linn.. Phytomorphology 14:414-429

Moliterni VMC, Cattivelli L, Ranalli P, Mandolino G (2004) The sexual differentiation of *Cannabis sativa* L.: a morphological and molecular study. Euphytica 140:95-106

Monteith JL (1972) Solar radiation and productivity in tropical ecosystems. J Appl Eco 9:747-766.

Monteith JL (1977) Climate and the efficiency of crop production in Britain [and Discussion]. Philos Trans R Soc (281) :277-294

Montford S, Small E (1999) Measuring harm and benefit: The biodiversity friendliness of *Cannabis sativa*. Global Biodiv 8: 2–13

Mueller U, Wolfenbarger L (1999) AFLP genotyping and fingerprinting. Trends Ecol Evol 14:389-394

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321-4325

Nandi S, Subudhi PK, Senadhira D, *et al.* (1997) Mapping QTLs for submergence tolerance in rice by AFLP analysis and selective genotyping. Mol Gen Genet 255:1-8

Oliveira E J, Vieira MLC, Garcia AAF, Munhoz CF, Margarido GRA, Consoli L, Matta FP, Moraes MC, Zucchi MI, Fungaro MHP (2008) An integrated molecular map of yellow passion fruit based on simultaneous maximum-likelihood estimation of linkage and linkage phases. J Amer Soc Hort Sci 133:35-41

Ong CK (1983) Response to temperature in a stand of pearl millet (*Pennisetum thyphoides* S. & H.). 1. Vegetative development. J Exp Bot 34:322-336

Otto F (1992) Preparation and staining of cells for high-resolution DNA analysis. In: Radbruch A (ed) Flow cytometry and cell sorting. Springer-Verlag, Berlin. Pp 101-104

Pacifico D, Miselli F, Micheler M, *et al.* (2006) Genetics and Marker-assisted Selection of the Chemotype in *Cannabis sativa* L. Mol Breeding 17:257-268

PAMESEB (2008) Website of the a.s.b.l. PAMESEB (Promotion de l'Agométéorologie dans le Sud-Est de la Belgique). http://www.pameseb.be/

Parker JS, Clark MS (1991) Dosage sex-chromosome systems in plants. Plant Sci 80:79-92

Pastina MM, Malosetti M, Gazaffi R, *et al.* (2012) A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. Theor Appl Genet 124:835-849

Pate DW (1994) Chemical ecology of *Cannabis*. J Int Hemp Assoc 2:32-37

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

Peil A, Schumann E, Flachowsky H, Kriese U, El Ghani M, Riedel M, Weber WE (2000). AFLP markers for male plants of hemp (*Cannabis sativa* L.). Proceedings of the "Bioresource Hemp 2000", 13-18 September 2000, Wolfsburg, Germany

Pertwee RG (2004) Pharmacological and therapeutic targets for Δ 9 - tetrahydrocannabinol and cannabidiol. Euphytica 140:73–82

Picault S (2006) Facteurs influençant les rendements du chanvre industriel: résultats des expérimentations françaises. In: Bouloc P (ed). Le chanvre industriel, production et utilisations. Ed. France Agricole, Paris, France. Pp 111-146

Pigliucci M (2005) Evolution of phenotypic plasticity: where are we going now? Trends Ecol Evol 20:481-6

Piotrowski BS, Carus M (2011) Ecological benefits of hemp and flax cultivation and products 2 . Effects on soil and crop rotations. Nova-Institut 2011-05. 6 p

Praça-Fontes M, Carvalho C, Clarindo W, Cruz C (2011) Revisiting the DNA C-values of the genome size-standards used in plant flow cytometry to choose the "best primary standards". Plant Cell Rep 30:1183-1191

Prothro J, Abdel-Haleem H, Bachlava E, White V, Knapp S, McGregor C (2013) Quantitative trait loci associated with sex expression in an inter-subspecific watermelon population. J Amer Soc Hort Sci 138:125-130

PSPc sa (2008) Etude de faisabilité déterminant les conditions de création d'une production agricole de chanvre dans le contexte de la création d'une filière chanvre en Wallonie. 175 p

Qi X, Lindhout P (1997) Development of AFLP markers in barley. Mol Gen Genet 254:330-6

Qi X, Stam P, Lindhout P (1998) Use of locus-specific AFLP markers to construct a high-density molecular map in barley. Theor Appl Genet 96: 376-384

R Development Core Team (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

Ranalli P (2004) Current status and future scenarios of hemp breeding. Euphytica 140:121-131

Ranalli P, Venturi G (2004) Hemp as a raw material for industrial applications. Euphytica 140:1-6

Reamon-Büttner SM, Jung C (2000) AFLP-derived STS markers for the identification of sex in Asparagus officinalis L.. Theor Appl Genet 100:432-438

Reamon-Büttner SM, Schmidt T, Jung C (1999) AFLPs represent highly repetitive sequences in Asparagus officinalis L. Chromosome Res 7:297-304

Rebai A, Goffinet B, Mangin B (1995) Comparing power of different methods for QTL detection. Biometrics 51:87-99

Reymond M, Muller B, Charcosset A, Tardieu F (2003) Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. Plant Physiol 131:664-675

Rode J, In-Chol K, Saal B, Flachowky H, Kriese U, Weber WE (2005) Sex-linked SSR markers in hemp. Plant Breeding 124:167-170

Rogers SM, Isabel N, Bernatchez L (2007) Linkage maps of the dwarf and normal lake whitefish (*Coregonus clupeaformis*) species complex and their hybrids reveal the genetic architecture of population divergence. Genetics 175:375-98

Rouppe van der Voort J, Draaistra J, Gommers FJ (1997) Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. Mol Gen Genet 255:438-447

Sakamoto K, Abe T, Matsuyama T, Yoshida S, Ohmido N, Fukui K, Satoh S (2005) RAPD markers encoding retrotransposable elements are Linked to the male sex in *Cannabis sativa* L. Genome 48:931-936

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

Sakamoto K, Ohmido N, Fukui K, *et al.* (2000) Site-specific accumulation of a LINE-like retrotransposon in a sex chromosome of the dioecious plant *Cannabis sativa*. Plant Mol Biol 44:723-32

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

Sankari HS (2000) Comparison of bast fibre yield and mechanical fibre properties of hemp (*Cannabis sativa* L .) cultivars. Ind Crop Prod 11:73-84

Sankari HS, Mela TJN (1998) Plant development and stem yield of non-domestic fibre hemp (*Cannabis sativa* L.) cultivars in long-day growth conditions in Finland. J Agron Crop Sci 181:153-159

SAS Institute Inc (2010) SAS OnlineDoc® 9.2. SAS Institute Inc., Cary, NC

SAS Institute Inc (2012) SAS® 9.3 help and documentation. SAS Institute Inc., Cary, NC

Schafer-Pregl R, Salamini F, Gebhardt C (1996) Models for mapping quantitative trait loci (QTL) in progeny of non-inbred parents and their behaviour in presence of distorted segregation rations. Genet Res 67:43-54

Schaffner JH (1921) Influence of environment on sexual expression in hemp. Bot Gaz 71:197-219

Schaffner JH (1923) The Influence of Relative Length of Daylight on the Reversal of Sex in Hemp. Ecology 4:323-334

Schaffner JH (1926) The change of opposite to alternate phyllotaxy and repeated rejuvenations in hemp by means of changed photoperiodicity. Ecology 7:315-325

Scheiner SM (1993) Genetics and evolution of phenotypic plasticity. Annu Rev Ecol Syst 24: 35-68

Schumann E, Peil A, Weber WE (1999) Preliminary results of a German field trial with different hemp (*Cannabis sativa* L .) accessions. Genet Res Crop Evol 46:399-407

Semagn K, Bjornstad A, Xu Y (2010) The genetic dissection of quantitative traits in crops. Electron J Biotechn 13:1-45

Sengbusch RV (1952) Ein weiterer beitrag zur vererbung des geschlechts bei hanf als grundlage für die züchtung eines monözischen hanfes. Z Pflanzenzüchtung 31:319-338

Serquen FC, Bacher J, Staub JE (1997) Mapping and QTL analysis of horticultural traits in a narrow cross in cucumber (*Cucumis sativus* L .) using random-amplified polymorphic DNA markers. Molecular Breeding 3:257-268

Shephard H, Parker J, Darby P, Ainsworth CC (1999) Sex expression in hop (*Humulus lupulus* L. and *H. japonicus* Sieb. Et Zucc.): floral morphology and sex chromosomes. In: Ainsworth CC (ed.) Sex determination in plants. Oxford: BIOS Scientific Publishers. Pp 137-148

Skirvin M (1978) Natural and induced variation. Euphytica 27:241-266

Slafer (2007) Physiology of determination of major wheat yield components. In: Buck HT, Nisi JE, Salomon N (eds) Wheat production in stressed environments. Proceedings of the 7th International Wheat Conference, 27 November–2 December 2005, Mar del Plata, Argentina. Ed. Springer, Dordrecht, Netherlands. Pp 557-566

Small E, Cronquist A (1976a) A practical and natural taxonomy for *Cannabis*. Int Assoc Plant Taxon 25:405-435

Small E, Jui PY, Lefkovitch LP (1976) A numerical taxonomic analysis of *Cannabis* with special reference to species delimitation. Syst Botany 1:67-84

Smith BW (1963) The mechanism of sex determination in *Rumex hastatulus*. Genetics 48: 1265-1288

Soller, M., Brody T (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor Appl Genet 39:35-39

Song JZ, Soller M, Genizi A (1999) The full-sib intercross line (FSIL): a QTL mapping design for outcrossing species. Genet Res 73: 61-73

Spitters CJT (1990) Crop growth models, their usefulness and limitations. Acta Horticulturae 267:349-368

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package. Plant J 3:739-744

Stott CG, Guy GW (2004) Cannabinoids for the pharmaceutical industry. Euphytica 140:83–93

Struik PC, Amaducci S, Bullard MJ, Stutterheim NC, Venturi G, Cromack HTH (2000) Agronomy of fibre hemp (*Cannabis sativa* L.) in Europe. Ind Crop Prod 11:107-118

Summerfield BRJ, Roberts EH, Ellis RH (1991) Towards the reliable prediction of time to flowering in six annual crops. I. The development of simple models for fluctuating field environments. Expl Agric 27:11-31

Summerfield RJ, Asumadu H, Ellis RH, Qi A (1998) Characterization of the photoperiodic response of post-flowering development in maturity isolines of soyabean [*Glycine max* (L.) Merrill] `Clark'. Ann Bot 82:765-771

Taliaferro CM, Hopkins AA, Henthorn JC, Murphy CD, Edwards RM (1997) Use of flow cytometry to estimate ploidy level in *Cynodon* species. Int Turfgrass Soc Res J 8:385-392

Tardieu F, Tuberosa R (2010) Dissection and modelling of abiotic stress tolerance in plants. Curr Opin Plant Biol 13:206-12

Terauchi R, Kahl G (1999) Mapping of the Dioscorea tokoro genome: AFLP markers linked to sex. Genome 42:752-762

Tollenaar M (1999) Duration of the grain-filling period in maize is not affected by the photoperiod and incident PPFD during the vegetative phase. Field Crop Res 62:15-21

Torjek O, Bucherna N, Kiss E, Homoki H, Finta-Korpelova Z, Bocsa I, Nagy I, Heszky L (2002) Novel male-specific molecular markers (MADC5, MADC6) in hemp. Euphytica 127:209-218

Trenbath BR, Angus JF (1975) Leaf inclination and crop production. Field Crop Abstracts 28:231-244

Trenbath BR, Angus JF (1975) Leaf inclination and crop production. Field Crop Abstr 28:231-244

Truta E, Olteanu N, Surdu S, Zamfirache MM, Oprica L (2007) Some aspects of sex determinism in hemp. Analele Stiintifice ale Universitatii « Alexandru Ioan Cuza », Sectiunea Genetica si Biologie Moleculara VIII:31-39

Vagera J, Paulikova D, Dolezel J (1994) The development of male and female regenerants by in vitro androgenesis in dioecious plant *Melandrium album*. Ann Bot 73:455-459

van Bakel H, Stout JM, Cote AG, et al. (2011) The draft genome and transcriptome of *Cannabis sativa*. Genome Biol 12:R102

van der Werf HMG (2004) Life cycle analysis of field production of fibre hemp, the effect of production practices on environmental impacts. Euphytica 140:13-23

van der Werf HMG, Brouwer K, Wijlhuizen H, Withagen JCM (1995a) The effect of temperature on leaf appearance and canopy establishment in fibre hemp (*Cannabis sativa* L.). Annals Applied Biology 126, 551–561

van der Werf HMG, Hassken HJ, Wijlhuizen M (1994) The effect of daylength on yield and quality of fibre hemp (*Cannabis sativa* L.). Eur J Agron 3:117-123

van der Werf HMG, Mathijssen EWJM, Haverkort AJ (1996) The potential of hemp (*Cannabis sativa* L.) for sustainable fibre production: a crop physiological appraisal. Ann Appl Biol 129:109-123

van der Werf HMG, van den Berg W (1995) Nitrogen fertilization and sex expression affect size variability of fibre hemp (*Cannabis sativa* L.). Oecologia 103:462-470

van der Werf HMG, van Geel WCA., van Gils LJC, Haverkort AJ (1995c) Nitrogen fertilization and row width affect self-thinning and productivity of fibre hemp (*Cannabis sativa* L.). Field Crop Res 42:27-37

van der Werf HMG, Wijlhuizen M, De Schutter JAA (1995b) Plant density and self-thinning affect yield and quality of fibre hemp (*Cannabis sativa* L .). Field Crop Res 40:153-164

van Dobben WH (1962) Influence of temperature and light conditions on dry matter distribution, development rate and yield of arable crops. Neth J Agric Sci 10:377-389

van Eeuwijk FA, Bink MC a M, Chenu K, Chapman SC (2010) Detection and use of QTL for complex traits in multiple environments. Curr Opin Plant Biol 13:193-205

van Eeuwijk FA, Malosetti M, Yin X, *et al.* (2005) ANOVA models to eco-physiological QTL models. Aust J Agric Res 56:1-12

van Ooijen (2009) MapQTL 6.0: software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma BV, Wageningen, Netherlands

van Ooijen JW (2011) Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. Genet Res 93:343-9

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522-530

Vera CL, Hanks A (2004) Hemp production in Western Canada. J Ind Hemp 9:79-86

Villalobos FJ, Ritchie JT (1992) The effect of temperature on leaf emergence rates of sunflower genotypes. Field Crop Res 29:37-46

Vinod KK (2006) Mapping of quantitative trait loci (QTL). In: Proceedings of the training programme on "Innovative quantitative traits – Approaches and applications in plant breeding, Tamil Nadu Agricultural University, Coimbatore, India. Pp. 224-242

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23:4407-4414

Vuylsteke M, Mank R, Antonise R, *et al.* (1999) Two high-density AFLP® linkage maps of Zea mays L.: analysis of distribution of AFLP markers. Theor Appl Genet 99:921-935

Warmke HE, Davidson H (1944) Polyploid investigation. Yearbook of the Carnegie Institution of Washington 43:135-139

Waugh R, Bonar N, Baird E, Thomas B, Graner A, Hayes P, Powell W (1997) Homology of AFLP products in three mapping populations of barley. Mol Gen Genet 255:311-321

Westergaard M (1958) The mechanism of sex determination in dieocious flowering plants. Adv Genet 9:217-281

Westerhuis W, Amaducci S, Struik PC, Zatta A, van Dam JEG, Stomph TJ (2009a) Sowing density and harvest time affect fibre content in hemp (*Cannabis sativa*) through their effects on stem weight. Ann Appl Biol 155:225–244

Westerhuis W, Struik PC, van Dam JEG, Stomph TJ (2009b) Postponed sowing does not alter the fibre/wood ratio or fibre extractability of fibre hemp (*Cannabis sativa*). Ann Appl Biol 155:333–348

Whaley JM, Sparkes DL, Foulkes MJ, Spink JH, Semere T, Scott RK (2000) The physiological response of winter wheat to reductions in plant density. Ann appl Biol 137:165-177

Wilson RN, Heckman JW, Somerville CR (1992) Gibberellin is required for flowering in *Arabidopsis thaliana* under short days. Plant Physiol 100:403-408

Wu R, Lin M (2006) Opinion: Functional mapping - how to map and study the genetic architecture of dynamic complex traits. Nat Rev Genet 7: 229-237

Wu R, Ma CX, Casella G (2007) Statistical genetics of quantitative traits: linkage, maps, and QTL. Springer, New York

Wu R, Ma CX, Painter I, Zeng ZB (2002a) Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. Theor Popul Biol 61:349-363

Wu W, Zhou Y, Li W, *et al.* (2002b) Mapping of quantitative trait loci based on growth models. Theor Appl Genet 105:1043-1049

Xu S, Atchley WR (1995) A random model approach to interval mapping of quantitative trait loci. Genetics 141:1189-1197

Xu S, Yi N (2000) Mixed model analysis of quantitative trait loci. Proceedings of the National Academy of Sciences of the United States of America 97:14542-14547

Yan W, Wallace DH (1998) Simulation and prediction of plant phenology for five crops based on photoperiod x temperature interaction. Ann Bot 81:705-716

Yin X, Chasalow SD, Stam P, *et al.* (2002) Use of component analysis in QTL mapping of complex crop traits: a case study on yield in barley. Plant Breeding 121:314-319

Yin X, Goudriaan J, Lantinga EA, *et al.* (2003) A flexible sigmoid function of determinate growth. Ann Bot 91:361-371

Yin X, Stam P, Kropff MJ, Schapendonk AHCM (1999) Crop Modeling, QTL mapping, and their complementary role in Plant Breeding. Agron J 95:90-98

Yin X, Struik PC, Kropff MJ (2004) Role of crop physiology in predicting gene-to-phenotype relationships. Trends Plant Sci 9:426-32

Yin X, Struik PC, Tang J, *et al.* (2005) Model analysis of flowering phenology in recombinant inbred lines of barley. J Exp Bot 56:959-965

Yuan XJ, Pan JS, Cai R, *et al.* (2008) Genetic mapping and QTL analysis of fruit and flower related traits in cucumber (*Cucumis sativus* L.) using recombinant inbred lines. Euphytica 164:473-491

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457-1468