Novel molecular tools to uncover the genetic architecture of hemp fibre quality

Jordi Petit Pedró

Propositions

1. Rhamnogalacturonan type-I pectin from the G-layer improves the strength of hemp bast fibres.
   (this thesis)

2. To improve decortication efficiency, cultivars developed for specific locations are needed.
   (this thesis)

3. The use of nuclear energy slows down climate change.

4. Promoting local tourism is a significant step towards the reduction of greenhouse gas emissions.

5. Frequent flyers should receive air mileage penalties, instead of air mileage rewards.

6. Civil disobedience is fuel for social changes.

Propositions belonging to the thesis, entitled

"Novel molecular tools to uncover the genetic architecture of hemp fibre quality"

Jordi Petit Pedró

Wageningen, 21st February 2020.

Novel molecular tools to uncover the genetic architecture of hemp fibre quality

Jordi Petit Pedró

Thesis committee

Promotor
Prof. Dr L.M Trindade
Personal chair Bioresources Breeding and Genetics
Wageningen University & Research

Co-promotor
Dr E.M.J. Salentijn
Researcher, Plant Breeding
Wageningen University & Research

Other members
Prof. Dr M.E. Schranz, Wageningen University & Research
Prof. Dr P.C. Struik, Wageningen University & Research
Prof. Dr W.E. Vermerris, University of Florida, USA
Dr M.P.L. Calus, Wageningen University & Research

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences

Novel molecular tools to uncover the genetic architecture of hemp fibre quality

Jordi Petit Pedró

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 21 February 2020
at 1:30 p.m. in the Aula.

Jordi Petit Pedró
Novel molecular tools to uncover the genetic architecture of hemp fibre quality, 216 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2020)
With references, with summary in English

ISBN 978-94-6395-267-5
DOI 10.18174/510653

A Tàrrega,
el poble responsable de ser qui i com sóc.

Al pare, a la mare i a l'Aïda,
sense ells no hauria arribat fins aquí.

Table of Contents

Chapter 1

General Introduction

1. Natural plant fibres are useful for renewable and sustainable solutions to cope with global environmental challenges

The human dependence on non-renewable and unsustainable resources has reached a plateau that compromises the needs of future generations. The current widespread use of petroleum-derived products, such as refinery products, and other mineral resources are responsible for large emissions of greenhouse gasses (GHG), driving major environmental challenges (FAO, 2012). By 2050, the world is anticipated to hold 9.6 billion people, who will consume 1.6 times the equivalent of our planet's resources (Zabaniotou and Kamaterou, 2019). Therefore, sustainable alternatives to petroleum or mineral-derived products are needed to reduce GHG emissions while guaranteeing the production of food, feed and non-food commodities for the growing population. Such alternatives are directed towards the development of carbon-neutral bio-based resources, focusing on societal, environmental and economic benefits (Carus and Dammer, 2018). According to the Food and Agriculture Organization (FAO), promoting the use of natural plant fibres, instead of synthetic fibres (petrochemical resource) or non-renewable fibres (mineral resource), contributes to a greener planet (Dam, 2008). This is based on the fact that growing fibre crops results in the fixation in biomass of atmospheric CO_2 through photosynthesis and thus has a positive effect on the CO_2 balance (Dam, 2008).

Natural plant fibres are agglomerations of cells mostly composed of lignocellulosic biomass from plant origin, such as seed hair fibres (e.g. cotton), bast fibres (e.g. flax, hemp, jute, kenaf, sisal, etc.) and woody fibres (eucalyptus, poplar, aspen, etc.) (Dam and Bos, 2004). Traditionally, several types of natural fibres from crops, mainly cotton, hemp and flax, have been used in textile applications. Yet, trade markets of most natural bast fibres (hemp, flax) have seen a decline in the last decades often due to competition with cheaper synthetic fibres and cotton (Dam, 2008). Regardless of the advantages of synthetic fibres over natural fibres, owed to their low cost, homogeneity, guaranteed production and quality and long service life (durability), synthetic fibres have a significant impact on the environment. Several studies have reported that large amount of microplastics are released from home textiles made of synthetic fibres in waste water, polluting freshwater, seas and oceans (Suaria et al., 2016;Henry et al., 2019). Another traditional application of natural fibres from bast fibre crops is the pulping for paper making (de Meijer and van der Werf, 1994) which is cheaper than using wood based pulping, owed to lower energy and chemical requirements (Liu et al., 2018c).

Innovative applications of natural fibres have widened the scope of the use of natural fibres beyond textiles and paper. Fibre bio-composites from bast fibre crops such as flax, hemp or jute have demonstrated to feasibly substitute petrochemical and mineral fibres (glass fibres) in the automotive industry (Holbery and Houston, 2006) and also in building insulation

applications (Wambua et al., 2003;Grohe, 2004;Jones and Brischke, 2017). Additionally, the waste biomass from bast fibre crops is relatively low, as the woody parts of the stems (e.g. hemp) can be used as construction materials (e.g. light-weight concrete (Elfordy et al., 2008)), or for animal bedding (Karus and Vogt, 2004).

Fibre crops, such as flax or hemp, also produce valuable seeds as by-products. These crops can produce high quality seed-oil with a large proportion of omega-3 fatty acids (Leizer et al., 2000). This oil is used in cosmetics (Sapino et al., 2005) and as health promoting functional food (Callaway, 2004). Additionally, cannabinoids are valuable by-products of fibre hemp. Cannabinoids are produced by glandular trichomes and they are accumulated at high concentrations in bracts in female inflorescences and in lower density in the leaves. The most studied cannabinoids from cannabis are delta-9-tetrahydrocannabinolic acid (THCA) and cannabidiol acid (CBDA), but fibre hemp contains only traces of THCA (THC equivalent <0.3%) because higher concentrations are not allowed in field production/agriculture. After heating (decarboxylation) these cannabinoids are transformed to the medical active forms THC and CBD. While THC is mostly used for its psychoactive effects, cannabidiol (CBD) is a non-psychoactive cannabinoid. CBD has become increasingly popular due to its therapeutic potential in a large number of diseases and symptoms without defined side-effects (Baker et al., 2003), when consumed in limited amounts.

In general, comparative studies on the production phase of fibre crops with synthetic products or glass fibres indicate that fibre crops provide environmental benefits. Fibre crops require lower energy inputs on the fibre processing, resulting in reduced CO_2 and other GHG emission levels (Dam, 2008;Peças et al., 2018). Moreover, the production costs of natural fibres are cheaper than the costs of mineral fibres (e.g. glass fibre) in bio-composite applications (Peças et al., 2018). In addition, natural fibre products have an easy end-of-life recycling, without producing undegradable wastes. Yet, if the waste is not recycled but is destroyed to produce energy by incineration, only the CO_2 fraction that was previously fixed by the plant is released to the atmosphere, resulting in CO_2 neutral life cycle (Lynd, 1996).

Actual markets of natural plant fibres are dominated by cotton (Dam, 2008) that accounts for over 70% of the global natural fibre production (FAOSTAT, 2017). Cotton produces high quality fibres, mainly for textile applications. Other natural fibres on the market are derived from bast fibre crops such as hemp, flax and jute. The fibres of these crops have their own properties and applications with only a small share in the total market for natural fibres. Of these crops, hemp is a strong candidate to develop into an economically viable fibre crop that fits into a circular bio-based economy.

2. Hemp as a potential bio-based crop

Hemp (*Cannabis sativa* L.) is a highly productive crop that can reach bast fibre yields up to 3 t ha^{-1} yr^{-1} (Struik et al., 2000;Amaducci et al., 2015;Piotrowski and Carus, 2019). Several studies report the properties of hemp fibres for bio-composites applications (Wambua et al., 2003;Holbery and Houston, 2006;Thygesen et al., 2006;de Bruijn et al., 2009;Placet, 2009;Hughes, 2012;Deyholos and Potter, 2014;Placet et al., 2014;Sorieul et al., 2016;Musio et al., 2018;Peças et al., 2018), indicating the large interest of the bio-composite industry in hemp (Piotrowski and Carus, 2010). Furthermore, the cultivation of hemp has lower environmental impact compared to other fibre crops, as summarized in **Table 1**. In hemp, and also in other bast fibre crops (e.g. flax or jute), the input of fertilizers, agrochemicals for crop protection and water are low (van der Werf and Turunen, 2008), meaning that the fossil fuel consumption needed for the mechanisation of weeding and crop maintenance is generally low (Ebskamp, 2002). In addition, hemp can be used as a phytoremediation crop for cleaning-up heavy metal pollutions (Ahmad et al., 2016) and it also fits well in crop rotation scenarios, increasing the profitability and diversity of the agricultural land (Liu et al., 2012;Amaducci et al., 2015). Hemp can grow in a large diversity of environments, including most European climates (Ebskamp, 2002). As a result, the production of hemp fibres for novel applications might benefit the environment and the European economy (e.g. bio-composites as an alternative to synthetic and mineral fibres).

3. Hemp is a poorly studied crop despite its remarkable interest as a bio-based crop

Due to the strict prohibition of hemp's cultivation and the emergence of markets for synthetic fibres and cotton, the development of hemp was limited during the time when intensive research and breeding programmes drove great improvements in major staple crops (Allegret et al., 2013). Consequently, hemp is a poorly studied crop and there are many gaps in the knowledge of hemp and its fibre quality that hamper the proper exploitation of its breeding potential. These gaps include: the lack of complete characterization of the genetics of hemp, the lack of studies on fibre quality variability, the lack of tools to characterize the cell walls of large amount of hemp accessions, the lack of complete understanding of the influence of genetic-, environment- and genotype-by-environment interactions (G×E) components on traits relevant to hemp fibre quality (including traits such as flowering time and sex determination), the lack of understanding of how the environment affects the fibre quality, the gap of knowledge in the relationships between important traits for fibre quality and finally, the lack of genetic studies that elucidate the genetic/molecular mechanisms controlling these traits.

Table 1. Average fibre yield, fertilizer requirements and water requirements of important natural fibre crops.

Crop	Fibre yield ($t\ ha^{-1}\ yr^{-1}$)	Fertilizer requirements[d] ($kg\ ha^{-1}\ yr^{-1}$)			Water requirements[f] ($mm\ yr^{-1}$)	Reference
		N	P	K		
Cotton	0.6-1.3[ab]	230-250	155-165	100-110	700-1300	(FAO, 2003;Murtaza et al., 2006;Yang et al., 2011)
Hemp	1.2-3[c]	80-100[e]	100[e]	150[e]	200-450	(Struik et al., 2000;Amaducci et al., 2015;Piotrowski and Carus, 2019)
Flax	1.5[c]	60[e]	30-50[e]	70-100[e]	350-450	(Sánchez Vallduví and Sarandón, 2011;Piotrowski and Carus, 2019); https://flaxcouncil.ca)
Jute	2-2.5[c]	35.1	15.4	73.1	500	(Majumdar et al., 2014); http://agritech.tnau.ac.in)

[a] Ginned cotton yield
[b] Seed-hair fibre
[c] Bast fibre
[d] Fertilizer requirements vary depending on soil quality
[e] Fertilizer requirements for fibre end-use
[f] Total water requirements including irrigation and rainfall

3.1 Fibres from hemp: stem structure, fibres and biochemical composition

Fibres from hemp are located in the stem of the plants and are basically cells, and more specifically cell walls. Hemp stems are organized in different tissues that define two different type of fibres, the bast fibres and the woody hemp core (WHC) or shives, as depicted in **Figure 1**. Bast fibres originate from the procambium and correspond to sclerenchyma (phloem) cells (Chernova et al., 2018). Bast fibres encircle the core woody tissue, consisting of xylem cells which are so-called woody fibres.

Two type of bast fibres can be distinguished in hemp, primary and secondary bast fibres. Both bast fibre types are generated from the cambium but secondary fibres are present in very low proportion. Secondary fibres are generated during fibre maturation, after the onset of flowering (Liu et al., 2015). Individual bast fibres, known as elementary fibres, are intimately associated through their middle lamella. One to four dozen of elementary fibres form a bundle of fibres (Thygesen et al., 2006), providing mechanical support to the plant (McDougall et al., 1993). The xylem tissue is composed of individual short (~0.5 mm) xylem cells (Hughes, 2012). Elementary primary bast fibres can reach up to 100 mm in length and 15–25 μm in diameter while elementary secondary bast fibres are generally shorter ~2 mm long and thinner (van der Werf et al., 1994).

From outside to inside the cell lumen, hemp cells have two cell walls: primary and secondary cell walls. The secondary cell wall is divided in three sublayers named as S_1, S_2 and S_3. Bast fibres are characterized by thick secondary cell walls, predominantly sublayer S_2 (~80% of secondary cell wall (Marrot et al., 2013)), that almost fills the lumen of the cell at maturity (Crônier et al., 2005). S_3 layer of cell walls from bast fibres is also referred to as gelatinous or G-layer (Mellerowicz and Gorshkova, 2012;Behr et al., 2016).

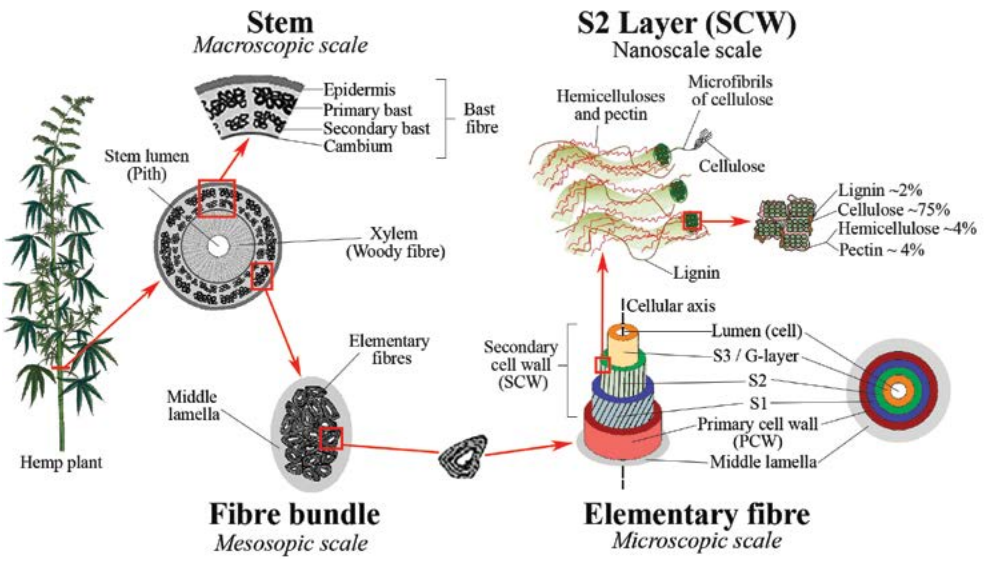


Figure 1. Cross-section and schematic representation of hemp at different scales, from the stem to the microfibrils of cellulose. The schematic representation of cell wall layers is not in scale, as S2 (green layer) represents ~80% of the secondary cell wall (SCW). The vertical lines in S1 and S2 layers of the secondary cell wall indicate the microfibrilar angle (MFA) of the cellulose in each layer. MFA of S2 is almost paralel to the cellular axis. Adaptation of Nykter (2006) and van der Weijde (2016).

The cell walls of plants are basically made up of four building blocks: cellulose, hemicellulose, pectin and lignin. Broadly speaking, the plant cell walls are structures of microfibrils of cellulose, reinforcing a matrix of hemicellulose and either lignin or pectin. The proportions of the different polymers and the orientation of the microfibrils of cellulose vary depending on the plant species, developmental stages, tissues and cell wall layers (Gibson, 2012). The middle lamella is a matrix located in between cells and it is made up of pectin and lignin (Willats et al., 2001). The primary cell walls have generally a higher content of lignin and pectin than the secondary cell walls, which are characterized by a higher content of cellulose (Gibson, 2012).

In hemp, the biochemical structure and spatial organization of the cell wall polymers are different between bast and xylem fibres. Bast fibres of hemp contain a considerable amount of cellulose and low lignin content (van der Werf et al., 1994). Compared to the bast fibres, xylem fibres have a higher content of lignin (as twice as the bast fibres) and hemicellulose-type xylans (Bonatti et al., 2004;van den Broeck et al., 2008). Yet, knowledge of the cell wall composition of xylem fibres is limited, as interest of hemp has traditionally focused on bast fibres. The secondary cell walls of hemp bast fibres are characterized by highly crystalline microfibrils of cellulose, which are deposited almost parallel to the cell axis (Crônier et al., 2005). The microfibril angle (MFA) of crystalline cellulose of the S_2 layer is lower than 10°, which is associated to improve mechanical properties(e.g. fibre strength) (Thygesen et al.,

2006). Previous studies reported that hemp bast fibres contain approximately 4% pectin, 4% hemicellulose, 2% lignin and 75% cellulose (Toonen et al., 2004;Crônier et al., 2005). The content of cellulose continuously increases with the age of the plant, and at the onset of flowering the proportion of polysaccharides decreases, owed to a shift in the carbon partitioning. The nutrients are shifted from the production of stem, leaves and roots towards the production of flowers and seeds around flowering. Notwithstanding this shift, lignification of the bast fibres intensifies at this developmental stage (Liu et al., 2015). These changes in cell wall composition indicate that hemp cell walls are highly dynamic and they undergo massive modifications during plant development (Crônier et al., 2005).

Regarding the G-layer of hemp bast fibres, the pectic polymer rhamnogalacturonan type I has been identified in this gelatinous layer, along with crystalline cellulose (Chernova et al., 2018).

Most cell wall research of hemp has been performed with long and laborious methods whereby the different cell wall polymer fractions are extracted one after the other (sequentially), which is a limiting factor if large number of samples have to be analysed. Consequently, studies on the cell wall composition of many hemp accessions are not yet available, which limits the knowledge on genetic variability, heritability and the influence of the environment on cell wall components.

High-throughput methods to study the cell wall composition are available for grasses, for instance the neutral detergent fibre (NDF) procedure (Goering and van Soest, 1970), which is based on aqueous detergents. Yet, the use of these methods in hemp are limited because of the organization of the hemp stem in different type of fibres, the high crystalline structure of cellulose from bast fibres and the hydrophilic behaviour of certain pectic polymers of hemp, (e.g. rhamnogalacturonan type I) (Huffman and Caballero, 2003;Pettolino et al., 2012). Therefore, novel methods that can phenotype large numbers of hemp accessions in a relatively short period of time are needed to enable the exploration of the potential of hemp fibres. Moreover, it is of great interest to understand the composition of the different types of fibres and how they are organized in mature hemp stems to understand the relation to their functionalities. Notwithstanding this interest, methods to fixate, infiltrate and embed mature and dry hemp stems are not available.

3.2 Fibre quality of hemp

The multiple applications of hemp fibres require high fibre quality to increase the economic value of the end-products and make them more competitive over synthetic and mineral fibres. First of all, the production of excellent fibres requires their separation from the woody core with minimal damage of the bast fibres. Good decortication features depend on the retting (Müssig and Martens, 2003;Müssig et al., 2008;Amaducci et al., 2015;Musio et al., 2018), which is the postharvest processing of the stems to degrade the pectin from the

middle lamella of the vascular cambium (Müssig et al., 2008). These pectins are responsible for cell adhesion in hemp stems and thus for the adhesion of bast to the woody fibres (reviewed in Willats et al. (2001) and in Xiao and Anderson (2013)). The number of cross-links that pectin can make with other cell wall components and the degree of polymerization of pectin side-chains (e.g. arabinan and galactan) influence the strength and elasticity of the cell wall (Willats et al., 2006) and also the resilience to retting. Hence, changes in the contents of monosaccharides derived from pectin are expected to affect the retting and the decortication of the stems. In addition, better decortication efficiencies are expected to produce lower damage to the bast fibres and thus to improve the fibre quality (Ranalli, 2004;Müssig et al., 2008;Salentijn et al., 2015). Minimum postharvest management of the stems to extract the bast fibres is thus desired to improve both the fibre yield and quality.

High cellulose content, large crystalline percentage of cellulose and low microfibrillar angle (MFA) of cellulose in the cell walls of bast fibres are expected to generate fibres of large quality (Frollini et al., 2000;Placet, 2009;Salentijn et al., 2015). Cellulose is the most abundant polymer in the plant cell walls and it provides structural support to the plants. In addition, cellulose content, cellulose crystallinity and low MFA provide mechanical strength to the fibres (Volkmann and Baluška, 2006) and are associated to finer fibres (Thygesen et al., 2006). Large fibre strength and fine fibres are highly appreciated by the bio-composite and textile industry (Linger et al., 2002;Müssig and Martens, 2003;Ranalli, 2004;Holbery and Houston, 2006;Salentijn et al., 2015;Peças et al., 2018). Moreover, fibre strength increases the resistance of the textiles to wear and tear, increasing their durability (Westerhuis, 2016). Likewise, high content of cellulose is appreciated by the paper industry to make large quantity of pulp (Liu et al., 2018c). In contrast, lignin provides rigidity and stiffness to the cell walls and is associated to coarse fibres, which decrease the appeal of the hemp industry (Dam, 2008). In addition, lignin is associated to hinder the fibre extraction because it makes cross-links with pectin (Willats et al., 2001;Kang et al., 2019). Furthermore, hemicellulose type xylans are associated to increase the hydrophobicity to the bast fibres. Large hydrophobic fibres are not desired by the textile industry because colour have difficulties to penetrate the textiles and thus to dye them (Blake et al., 2008;Manabe et al., 2011). Therefore, low contents of lignin and xylans in the bast fibres are expected to increase the quality of hemp fibres for textile applications (reviewed in Ranalli (2004) and in Salentijn et al. (2015)).

Hemp accessions with excellent fibres are also expected to produce a high bast fibre yield (Sankari, 2000). Thus, bast fibre content an important breeding target for the hemp industry (Ranalli, 2004;Salentijn et al., 2015). A previous study reported large heritability for this trait in hemp (Hennink, 1994), meaning that large genetic gains are expected from bast fibre content. In addition, secondary bast fibres are highly lignified (Liu et al., 2015) compared to primary bast fibres. Consequently, excellent bast fibres are expected to have low content of secondary bast fibres (Müssig and Martens, 2003;Chernova and Gorshkova, 2007;Amaducci et al., 2015).

Flowering time triggers the shift in carbon partitioning in hemp and is associated to a change in fibre composition, particularly affecting lignin content, and secondary bast formation (Crônier et al., 2005;Liu et al., 2015). Furthermore, sexual dimorphism between male and female plants of hemp affects plant morphology and fibre composition at plant maturity (further information about sex determination of hemp is available in section 3.5 of this Chapter). Flowering time and sex of hemp are thus other important characteristics influencing the quality of hemp fibres.

Genetic variability studies in hemp have only explored morphological traits and cannabinoids content (de Meijer and van der Werf, 1994;Meijer, 1994;Meijer and Keizer, 1996). The large variation in hemp for morphological traits, such as plant height (e.g. plants from 1 to 4 meters of height), together with the large variability in fibre quality reported for other fibre crops (e.g. poplar, eucalyptus, miscanthus, switchgrass and maize) (Raymond and Schimleck, 2002;Klasnja et al., 2003;Schimleck et al., 2004;McLaughlin et al., 2006;Poke et al., 2006;Boe and K. Lee, 2007;Davis, 2008;Slavov et al., 2014;Torres et al., 2015a;van der Weijde et al., 2017) suggest large variability in traits relevant for hemp fibre quality. Furthermore, large variability in flowering time traits is expected in hemp. This is essentially because significant variability for these traits has been reported in several crops (e.g. almond, apricot, cotton, flax and rice) (Takahashi et al., 2001;Huang et al., 2010;Campoy et al., 2011;Sánchez-Pérez et al., 2014;Soto-Cerda et al., 2014;Kushanov et al., 2017;You et al., 2017). Yet, little research has been conducted on the characterization of the variability of fibre quality traits and flowering time in hemp. Moreover, the complex relationships between these traits have not been extensively studied. As a consequence, the breeding for some traits in hemp may have some unexpected trade-offs to traits related to the quality of the fibres. A better understanding of the variability of fibre quality traits and of the relationships between them in hemp are desired to improve breeding programmes, which seek to develop new hemp cultivars to produce sustainable fibres of high qualities.

3.3 Genes and pathways in cell wall biosynthesis and flowering of hemp

The cell wall is a complex and dynamic structure that involves constant biosynthesis and remodelling of the different components (Houston et al., 2016). It concerns the action of many different genes; about 15% of the genes in each plant species are thought to be involved in the biosynthesis of the cell wall. It is estimated that ~4,000 genes in arabidopsis are protein-encoding genes involved in the cell wall but only ~1,000 of them have been characterized or predicted (Wang et al., 2012). Furthermore, some of these genes can be genetically regulated while the expression of other genes depends on non-genetic mechanisms, such as the environment, increasing the complexity of the cell walls. Van der Broeck and co-workers identified that the gene expression between bast and woody hemp fibres showed differences. Genes encoding proteins involved in lignin metabolism and C1 metabolism were overexpressed in the woody core while genes encoding arabinogalactan related proteins,

lipid transfer proteins, lipoxygenase and endoxyloglucan transferases were found to be overexpressed in the bast fibres (van den Broeck et al., 2008). Lignification pathways include the synthesis of monolignols in the cytoplasm of the cells and export of these components to the apoplast, where they will be incorporated in lignin in the cell wall (reviewed in Boerjan et al. (2003)). Enzymes playing important roles in the lignin biosynthesis and found overexpressed in the woody fibres included: 4-coumarate:CoA ligases (4-CL) (Lewis, 1999), cinnamate-4-hydroxylase (C4'H), p-coumaroyl quinate/shikimate 3'-hydroxylase (C3'H) (Anterola and Lewis, 2002) and phenylcoumaran benzylic ether reductase (PCBER) (Vander Mijnsbrugge et al., 2000). Other enzymes involved in the lignin biosynthesis and also identified overregulated in the woody core included cinnamyl alcohol dehydrogenase, peroxidases, laccases, polyphenol oxidases and coniferyl alcohol oxidase (reviewed in Boerjan et al. (2003)). All these enzymes are involved in the modification of lignin and thus in the fibre quality of hemp. Closely connected to lignin biosynthesis is C1 metabolism, which provides the activated methyl groups for the methylation of several monolignol intermediates, needed to biosynthesise lignin. Key enzymes of the C1 metabolism, and detected in the woody core, included methionine synthase and S-adenosylmethionine-synthase. In addition, the linking enzymes between both metabolic pathways, caffeoyl-CoA o-methyltransferase (CCoAOMT) and acid o-methyltransferase (COMT), were also found overexpressed in the woody core. These enzymes catalyse the transfer of methyl groups from one pathway to the other one (Sterky et al., 1998;van den Broeck et al., 2008). Changes in the expression of the genes that codify for these enzymes of the C1 metabolism might modify lignin content, and as a result also hemp fibre quality.

Flowering is an essential biological process for many plants, as the survival of the species depends on it (Mouradov et al., 2002). Approximately 300 genes are estimated to be involved in flowering in arabidopsis (Bouché et al., 2016). Yet, the genetics and molecular mechanisms of flowering in hemp are poorly studied and little information is currently available.

A review of the genes known to be involved in cell wall composition and flowering time in hemp and in other plant species with common flowering mechanisms or sex determination system with hemp, might provide useful information to get insights into the genetic and molecular mechanisms of these complex traits in hemp.

3.4 Influence of the environment on fibre quality and flowering stages of hemp

Hemp fibre quality and flowering time are strongly influenced by environmental factors, especially by photoperiod and temperature (Amaducci et al., 2012). Hemp is a short-day plant and its flowering is inhibited during long-day photoperiod regimes and is induced when the photoperiod regime shifts above ~10-12 hours of uninterrupted darkness (short-day photoperiod below 12-14 hours day light). Hemp rapidly responds to differences in photoperiod regimes by prolonging the length of the vegetative growth period or by inducing flowering. The

influence of the environment on the biomass yield and quality has resulted in the development of cultivars for specific environments and end-uses (Salentijn et al., 2015). For instance, in the northern European latitudes, where the long-day photoperiod regime lasts longer during hemp's growing season (flowering is inhibited), hemp has been traditionally cultivated for fibre. Meanwhile, in the southern latitudes (shorter daylights) hemp has also been cultivated for seed production as plants flower early and can reach seed maturity.

Regardless of the influence of photoperiod regime and temperature on flowering time and biomass production, little is known about the action mechanisms of environmental factors that affect fibre quality and its phenotypic variation. Understanding the effects of the environment on fibre and flowering might help to develop more efficient breeding programmes.

3.5 Genetics and sex of hemp

Hemp is an open-pollinated species and naturally it is a dioecious crop where male and female flowers are located in different plants. This reproduction behaviour generates a continuous pattern of variation (Sawler et al., 2015). As a consequence, the large heterozygosity of hemp, along with the highly repetitive regions of the genome, hampered any attempt to complete the genome sequence of cannabis (van Bakel et al., 2011;Sawler et al., 2015;Grassa et al., 2018;Laverty et al., 2019).

Sexual dimorphism is characteristic in dioecious hemp. Male and female plants have different morphological, as well as phenological characteristics, and show differences in flowering time, different cell wall composition and different fibre properties (Moliterni et al., 2004;Amaducci and Gusovius, 2010;Faux et al., 2013;Amaducci et al., 2015). Hemp is a diploid species with 9 pairs of homomorphic autosomal chromosomes and a pair of heteromorphic sex chromosomes: X and Y (2n=20). True male plants carry the heterogametic sex (XY) and female plants the homogametic one (XX). Yet, some dioecious hemp plants produce flowers of the opposite sex than the one determined by their genetics (Freeman et al., 1980;Faux et al., 2016). Monoecious genotypes, also termed hermaphrodites, can spontaneously occur. Monoecious plants bear both male and female flowers in the same plant and they carry the homogametic sex (XX). True males have a slender stature, few leaves and hanging inflorescences carrying male flowers. Female plants produce female pistillate flowers in dense panicles heads interspaced with leafy bracts, whereas the morphology of monoecious hemp plants resemble female plants (Faux et al., 2016). Male plants flower mostly preceding female ones (Bócsa and Karus, 1998;Struik et al., 2000) and male plants die after flowering while females remain alive until the seed maturation. As a consequence, fibres from female plants harvested around seed maturity are more lignified than fibres from males (Liu et al., 2015). Male plants are also known to produce finer fibres than females (Amaducci et al., 2015). In addition, the proportion of males and females in

the crop has an effect on the seed yield, as large male proportion is associated to a seed yield reduction. Therefore, dioecious hemp species are characterized by heterogeneity in fibre and seed production. In contrast, monoecious genotypes are more stable in both fibre and seed production and a narrower genetic heterozygosity is characteristic for monoecious plants, owed to its ability to self-pollinate (Bócsa and Karus, 1998;Mandolino and Carboni, 2004). However, monoecious genotypes show unstable sex expression, which tends to lead to a distribution of sex ratios and a gradual return to natural dioecy after few generations (Bócsa and Karus, 1998;Amaducci et al., 2008b). Consequently, this distribution of sex ratios can affect seed propagation.

Despite the presence of specific sex chromosomes, the sex determination in cannabis has been proposed to be based on a X-to-autosome dosage rather than on an active Y chromosome. However, the genetic and molecular mechanisms remain unknown (Westergaard and Demerec, 1958;Grant et al., 1994;Ainsworth, 2000;Moliterni et al., 2004). Deciphering these mechanisms might be helpful to control sex determination and indirectly might benefit the production of fibre of certain quality.

3.6 Genetic studies and breeding of hemp

To date, only three biparental mapping population approaches have performed in hemp. These studies were performed for sex expression and cannabinoid content. Faux and co-workers published the first genetic map made from three F1 segregating population (Faux et al., 2016). Three hemp cultivars were used, one dioecious (Carmagnola) and two monoecious (Uso 31 and Fedora 17). They reported putative recombination between X and Y chromosomes and X chromosomes of monoecious hemp would include fragments homologous with both the X and Y chromosomes of dioecious hemp. In addition, they identified quantitative variation in sex expression using AFLP markers and 5 quantitative trait loci (QTL) for sex expression in hemp. However, the genetic mechanism underlying sex of hemp was not elucidated. Recently, two other groups published simultaneously two genetic maps to study the genetic mechanism behind the synthesis of the two major cannabinoids: THC and CBD. Grassa and co-workers published a genetic map made from an F2 segregating population using a marijuana parental with high content of THC (Skunk#1) and a hemp parental with high content of CBD (Carmen) (Grassa et al., 2018). In addition, Laverty and co-workers published another genetic map made from an F1 segregating population using also a marijuana parental (Purple Kush) and a hemp parental (Finola) (Laverty et al., 2019). Both genetic maps allowed to detect multiple acid synthase loci for both THC and CBD. Furthermore, there are several molecular markers, specific for cannabis, to discriminate between male and female plants. MADC2 marker is a male-associated RAPD marker identified by Mandolino et al. (1999). This sex related marker was characterized by studying the differences in mRNA expression between males and females (Moliterni et al., 2004).

Further research of this marker allowed to develop a SCAR marker, which was used to identify female plants at an early developmental stage (Techen et al., 2010).

Despite these three maps, genetic studies in hemp are limited. The lack of a complete genome sequence and the poor knowledge of genetic variation in hemp has hampered such studies (Sawler et al., 2015). In addition, no genome wide association study (GWAS) has performed in hemp so far. As a result, little is known about the genetic architecture of fibre quality, flowering time and sex determination of hemp and no QTLs have been published about these complex traits.

The lack of association studies has restricted the breeding of hemp to conventional practices. Up to date, breeding methods for hemp commonly consist on the search for natural variation; the generation of varietal parents through mass selection and cross breeding; and the development of experimental cultivars through inbreeding and hybrid breeding (Salentijn et al., 2015). Yet, molecular breeding has not been developed in hemp. The identification of QTLs is of key value for developing new cultivars of hemp through breeding (Mandolino and Carboni, 2004). The use of such genetic markers would speed up the exploration of germplasm diversity and breeding steps.

4. Layout of this thesis

Hemp is a promising bio-based crop, and thus the development of new hemp cultivars with improved fibre properties and adaptations to specific environments is envisioned. The development of improved cultivars is of paramount importance to increase the production of sustainable high quality fibres. This would increase the competitiveness of hemp relatively to other resources. In the views of these prospects, the central objective of this thesis is to increase the knowledge of genetic variability of hemp. Furthermore, we aim to get insights into the genetic and molecular mechanisms and environmental factors controlling fibre traits, cell wall characteristics, flowering time and sex determination relevant to fibre quality.

This study 1) develops high-throughput methods for cell wall phenotyping adapted to special needs of hemp; 2) gets insights into the natural genetic diversity of phenology, cell wall composition, fibre traits, flowering time and sex determination of hemp; 3) assesses the feasibility of a 123 hemp accessions panel in mapping studies for fibre quality; 4) uncovers highly heritable traits aimed at developing novel breeding tools; 5) evaluates the influence of the environment on the phenotypic variation of fibre hemp; 6) develops a genetic association study approach to detect QTLs and candidate genes, although a complete genome sequence is still lacking; 7) identifies genomic regions associated to traits of interest to enable the development of an accelerated genetic improvement of hemp fibre quality, through

molecular breeding and 8) deciphers part of the molecular mechanisms of fibre quality, flowering time and sex expression of hemp. To address these objectives, this thesis has been structured as follows:

This thesis starts with the optimization and development of repeatable and high-throughput methods, to study the cell wall composition and stem morphology of hemp in **Chapter 2**. This chapter also explores the genetic diversity of a small set of hemp accessions and the relationship between the cell wall composition of the different type of fibres and the stem morphology of hemp.

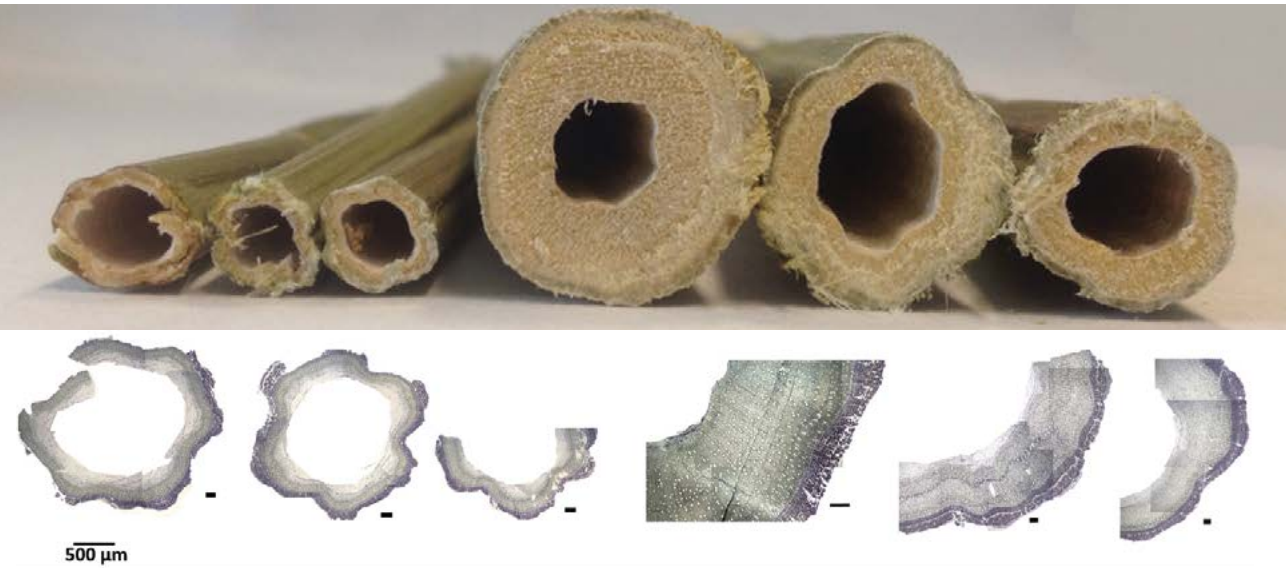
In **Chapter 3** we study the natural genetic variability of phenological traits, fibre measurements, cell wall composition, flowering time traits and sex determination of 123 hemp accessions. The accessions were tested in three European locations with contrasting environments. Moreover, we get insights into the relationships between traits relevant to fibre quality. This chapter also evaluates the potential of the hemp panel for mapping studies and it also evaluates the influence of the environment on hemp fibre traits.

A high-throughput sequencing (RAD-seq) of the hemp panel and the development of genome-wide association studies (GWAS) were performed in **Chapter 4,** to yield insights into the genetic architecture of hemp fibre quality . We develop an approach to identify QTLs and candidate genes without a complete genome sequence.

Chapter 5 reviews the complex interactions between flowering behaviour and fibre quality in hemp. In addition, possible modes of genetic control of flowering time and sex determination have been proposed for hemp by discussing the state-of-art of other plants.

In **Chapter 6** we explore the genetic architecture of flowering time stages in hemp to identify the genetic mechanisms, dependent and independent of the environment, that control the shift between vegetative growth and induction of flowering. In addition, we go one step further by uncovering a genetic mechanism that is a strong candidate for controlling the stability of monoecy in hemp.

Finally, this thesis is culminated with a general discussion in **Chapter 7** on how the knowledge generated in this thesis may contribute to the genetic improvement of hemp for the production of high quality fibres. In this respect, recommendations are also provided to guarantee the success of breeding programmes, seeking to promote hemp in the context of a bio-based alternative crop.



Chapter 2

Phenotypic variation of cell wall composition and stem morphology in hemp (*Cannabis sativa* L.): optimization of methods

Jordi Petit[a], Agata Gulisano[a], Annemarie Dechesne[a], Luisa M. Trindade[a*]

[a] Wageningen UR Plant Breeding, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.
* Corresponding author: Luisa M. Trindade (luisa.trindade@wur.nl)

This chapter has been published (with minor modifications) as:
Petit, J., Gulisano, A., Dechesne, A., and Trindade, L.M. (2019). Phenotypic Variation of Cell Wall Composition and Stem Morphology in Hemp (*Cannabis sativa* L.): Optimization of Methods. *Frontiers in Plant Science* 10. doi: 10.3389/fpls.2019.00959

ABSTRACT

The growing demands for sustainable fibres have stimulated the study of genetic variability in the quality of hemp fibre (*Cannabis sativa* L.). Nevertheless, the lack of high-throughput phenotyping methods that are suited for the analysis of hemp fibre, hampers the analysis of many accessions, and consequently the breeding for this complex trait. In the present report, we developed and optimized the throughput of five methods to study the variability in hemp fibre quality including cell wall extraction, biochemical composition of cell wall polysaccharides, quantification of lignin, quantification of crystalline polysaccharides and morphology of the stems. Six hemp accessions, contrasting for cell wall properties, were used to assess the throughput and suitability of these methods for genetic studies. The methods presented revealed to be highly repeatable, with low coefficients of variation between technical replicates. With these methods we were able to detect significant phenotypic variation in cell wall composition and stem morphology between the six accessions. In addition, the throughput of the methods has been upgraded to a level that enables their use for phenotyping cell wall traits in breeding programmes. The cell wall extraction was optimized to extract enough material for the complete characterization of the cell wall of hemp, while reducing the time for the entire analysis. The throughput of the stem morphological analysis was improved by decreasing the timing of fixation, infiltration and embedding of mature and dry hemp stems. Our methods have the potential to phenotype large number of accessions in a relatively short period of time. The present methods will enable exploration of genetic variability of fibre quality and will contribute to the development of new hemp cultivars with advanced quality of fibres.

1. Introduction

We are entering a circular bioeconomy era and that requires crops that stand for alternative sustainable solutions. Hemp (*Cannabis sativa* L.) is an environmental friendly and multipurpose crop and therefore it presents an attractive candidate for the bio-based economy (Salentijn et al., 2015). Characterization of hemp fibre quality, including stem morphology and cell wall composition and structure, is the first step towards breeding for high yielding cultivars with better fibre quality. Such cultivars, that are more amenable to processing, position hemp as a competitive alternative to poor sustainable fibre crops (e.g. cotton) (Ebskamp, 2002;van der Werf and Turunen, 2008;Amaducci and Gusovius, 2010).

Hemp stems are organized in two types of fibres: bast fibre and woody hemp core (WHC) (Crônier et al., 2005). Bast fibres are collections of one to four dozens phloem cells, known as elementary fibres (Thygesen et al., 2006), while WHC fibres are composed of individual xylem cells (Hughes, 2012). The composition of the cell wall is different in the two types of fibres and varies throughout the different developmental stages (Crônier et al., 2005). Despite the differences, both bast and WHC fibres follow a pattern common in dicot plants (Sarkar et al., 2009). Hemp cells from the bast and WHC fibres are surrounded by a middle lamella, a primary cell wall and a secondary cell wall. Furthermore, cells from the bast fibre also have an additional layer, referred to as gelatinous or G-layer (Mellerowicz and Gorshkova, 2012;Behr et al., 2016). The middle lamella is a network mainly composed of pectin and lignin (Willats et al., 2001). The primary cell wall and the secondary cell wall comprised similar components: mainly cellulose, lignin and a matrix of hemicellulosic polysaccharides and pectins (Morrison, 1988;Bonatti et al., 2004), though in different proportions. The G-layer is characterized by a high content of crystalline cellulose and by the presence of rhamnogalacturonan type I (Chernova et al., 2018). Cell walls from bast fibres are characterized by a high content of cellulose, whereas cell walls from WHC have a larger proportion of lignin and xylans (Bonatti et al., 2004;van den Broeck et al., 2008). In addition, hemp stems are also highly dynamic and they undergo massive modifications during plant development. For instance, lignification of fibres intensifies with flowering and fibres become stiffer (Crônier et al., 2005).

Despite the increasing knowledge of cell wall composition, little is known about the genetics underlying cell wall biosynthesis and stem morphology in hemp. The complex structure of the cell walls and the organization of different types of cells in the stem hinder the development of high-throughput phenotyping methods for the analysis of these traits. In addition, the hydrophilic behaviour of certain pectic polymers of hemp (e.g. rhamnogalacturonan type I) hampers the complete cell wall extraction with the current high-throughput methods, that are based on aqueous detergents (Goering and van Soest, 1970;Huffman and Caballero, 2003;Pettolino et al., 2012). Extraction methods based on alcohols (Alcohol insoluble residue – AIR) would probably be better suited alternatives (Pettolino et al., 2012). Yet, AIR protocols

use low amounts of starting plant material, and consequently several parallel extractions are required before cell walls can be characterized. Thus, the throughput of the hemp cell wall extraction is compromised with the current methods.

The first step to characterize cell wall polysaccharides and lignin comprises the hydrolysis of the carbohydrates into derived monosaccharides (Morrison, 1988). Sulphuric acid (H_2SO_4) is a commonly used acid that hydrolyses the majority of the cell wall structures, except lignin (Morrison, 1988;Sluiter, 2012). Notwithstanding, the current H_2SO_4 methods produce low repeatable data in hemp. The complex stem morphology and the large crystalline polysaccharides from hemp hamper the complete hydrolysis of the cell wall (Crônier et al., 2005;Pu et al., 2013).

The high crystallinity of cellulose and mannan (Millane and Hendrixson, 1994;Crônier et al., 2005) is thought to play an important role in the properties of bast fibres in hemp (e.g. fibre strength) (Marrot et al., 2013;Placet et al., 2014). The current methods to study the crystallinity of cell wall polysaccharides are based on physical methods (Park et al., 2010). These methods involve many handlings, allow to analyse only few samples simultaneously and consequently limit the study of many cultivars. The throughput of these methods is a major limitation to study the variability of crystallinity in hemp bast fibre. Biochemical alternatives adapted to hemp features are a possible alternative to increase the throughput of the analysis. Based on its supramolecular structure, consecutive hydrolysis, using different acids from weak to strong acids release different amounts of glucose and mannose from cellulose and mannan, respectively (Updegraff, 1969;Foster et al., 2010) . The released monosaccharides in each hydrolysis can be used to determine the percentage of crystalline polysaccharide (Foster et al., 2010).

The different types of hemp fibres and the composition of their cell walls are important for both the architecture of the plant and the fibre quality (Crônier et al., 2005). It is therefore of great interest to understand the composition of the different fibres and how they are organized in the stem to understand the relation to their functionalities. The WHC of mature hemp stems is a hard structure and that hampers fixation, infiltration and embedding of the stem and consequently their morphological analysis (Crônier et al., 2005;Behr et al., 2016). Hence, methods to fixate, infiltrate and embed mature hemp stems are of great interest.

In the present study, we developed and optimized the repeatability and the throughput of five methods to assess the quality of hemp cell walls including cell wall extraction, biochemical composition of cell wall carbohydrates, quantification of lignin, quantification of crystalline polysaccharides and morphological analysis of mature stems. The methods were especially optimized to phenotype large number of hemp cultivars and is a first step towards the development of tools to breed for hemp fibre quality.

2. Materials and methods

2.1 Plant material

Six fibre hemp accessions were used in this study: CRA410, CRA412, CRA416, CRA420, FNPC243 and WU101 (**Table 1**). Plants were cultivated in Westerlee (The Netherlands, 53˚N 6˚E) from April to September 2013 and harvested at full flowering.

2.2 Preparation of stem material

Harvested mature stems were naturally dried, in open air under a roof, until the water content was lower than 18% of the total matter. Thereafter, stems were completely dried in the oven at 60˚C for two days to prevent initiation of stem retting. Samples were stored until analysis. Stems needed for cell wall analysis were chopped to 2 cm length pieces using a chopper machine, and re-dried in the oven at 60˚C for 1.5 hours. Subsequently, samples were grinded to 1 mm size using a grinder Peppink 200AN with a 1 mm sifter. Finally, to ensure that cells were completely disrupted, an extra step of grinding was performed. To this purpose, a Retsch Mixer Mill MM400 Retsch was used at maximum frequency (30 Hz) for one minute, using a 13 ml stainless steel jar and two 13 mm steel beads. To avoid an excessive heat up of the samples, grinding jars with the samples were cool down with liquid nitrogen before grinding.

Table 1. Six fibre hemp (*Cannabis sativa* L.) accessions used in this study.

MultiHemp code	Accession name / code	Origin	Accession type	Provider
CRA410	Ermes A	Italy	Fibre	CRA
CRA412	Carmaleonte	Italy	Fibre	CRA
CRA416	Denise	Romania	Fibre	CRA
CRA420	USO 31	Ukraine	Fibre	CRA
FNPC243	A102-111-1	France	Fibre	FNPC
WU101	JSO 16 / 891229	Russia	Fibre	WUR

CRA, FNPC and WUR stand for Research Centre of Industrial Crops, Federation National Producteurs de Chanvre and Wageningen University and Research, respectively.

For the crystalline polysaccharide analysis, only bast fibre was used. Harvested mature stems for bast fibre analysis were naturally dried and then were retted, as described in van den Oever et al. (2003). Briefly, the stems were warm water retted for three days and naturally dried in open air. Thereafter, stems were decorticated with a lab-scaled roller-breaker decortication system, according to Wang et al. (2018), followed by a hand-removal of the remaining shives (WHC). Finally, bast fibre was chopped and double grinded, as described in the previous paragraph.

For the stem morphology analysis, stored dried stems were used to obtain the cross-sections at approximately 70-80 cm from the ground for the further analysis.

2.3 Cell wall extraction

Preparation of cell wall fractions and quantification of the cell wall percentage from the biomass were performed based on the alcohol-insoluble residue (AIR) method from Pettolino et al. (2012) with modifications: sample homogenization, initial biomass amount and drying of cell wall fraction. A summary of the modifications and its implications can be found in **Table 2**.

Milled plant material was mixed using a vortex to ensure a good representation of the sampled tissues, especially for the stem samples. From these samples the AIR fraction was extracted, corresponding to the total cell wall fraction, and used as starting material for the characterization of the cell wall, including monosaccharide composition, quantification of lignin and crystalline polysaccharide analysis. The protocol was scaled up to 1 g starting material and the extraction was performed in 50 ml disposable Nunc tubes suitable for ultracentrifugation. The AIR extraction consisted of two steps: extraction of cell wall and α-amylase digestion.

Cell wall extraction was performed with 36 ml of 80% ethanol (vol/vol) for 30 minutes on ice. Samples were properly mixed every 10 minutes, using a vortex after the ethanol was added. Cell walls were ultracentrifuged for 5 minutes (10.000*g*, 4˚C) using a ultracentrifuge Beckman Avanti with a Fibrelite R14BA – 14x50cy rotor and the supernatant was discarded. This extraction step was repeated three times. Lipids and chlorophyll were removed with 36 ml of absolute acetone (room temperature, 10 minutes). After adding the acetone, samples were mixed using a vortex. Thereafter, samples were ultracentrifuged during 5 minutes (10.000*g*, at room temperature) and the supernatant was discarded. The remaining chlorophyll was removed with 36ml of absolute methanol (room temperature). After a vortexing step, the extraction proceeded for 10 minutes and samples were ultracentrifuged again (5 minutes, at 10.000*g*, at room temperature) and the supernatant was discarded. Finally, the pellet was dried using a RapidVap Vaccum Dry Evaporation System (Labconco, Kansas City, MI, USA) decreasing gradually the pressure with constant shaking.

Extracted cell walls were incubated with α-amylase (porcine pancreas, Megazyme) to remove starch from the AIR fraction. Dried samples were incubated in 6 ml of 10 mM Tris-maleate (Pettolino et al., 2012) buffer for 30 minutes with constant shaking. Starch was gelatinized by boiling the samples in a pot with boiling water (5 minutes), followed by a cool down of the samples to 40˚C with ice. Thereafter, two rounds of α-amylase digestion were conducted. The first one consisted of 2U of enzyme for mg of carbohydrate diluted in 2 ml of 10 mM Tris-maleate buffer for 1 hour at 40˚C in a HLC thermomixer with smooth shaking (300 rpm). The second digestion consisted of half the amount of enzyme for 30 minutes, under the same conditions as the first digestion (40˚C and 300 rpm).

After the digestion, α-amylase was inactivated by adding 36 ml of cold absolute ethanol and precipitated at -20˚C for 1 hour, with a previous vortexing step. Samples were centrifuged in a Multifuge 3S Heraeus with a Sorvall Heraeus rotor (5 minutes, 1.500*g*, room temperature). Supernatant was discarded and three extra washes of 36 ml of absolute ethanol with a vortexing and a centrifuge steps in between were performed (5 minutes, 1.500*g*, room temperature). Finally, pure AIR fraction was dried using a RapidVap system, as described in the previous paragraph of this section.

Quantification of the cell wall percentage was performed following the original protocol (Pettolino et al., 2012) (50 mg as initial weight in a 2 ml microcentrifuge tube) in triplicates with a single modification. The percentage of cell wall was calculated as the difference between the initial weight of the milled plant material and the weight of the sample after the extraction, corrected for dry matter content.

2.4 Monosaccharide composition

The AIR fraction was hydrolysed to analyse the biochemical composition of the cell wall. A two-step sulphuric acid hydrolysis, 72% H_2SO_4 (w/w) at 30˚C followed by 4% H_2SO_4 (w/w) at 121˚C was used to hydrolyse the polysaccharides. The protocol was based on the one described by Sluiter (2012) with modifications (**Table 2**).

Twenty mg of AIR fraction were weighted into a 12 ml glass tube (Schott® culture tube: 18 mm of diameter and 113 mm of length) with a screw leak-proof cap (Schott® red screw cap: DIN thread GL 23 mm of diameter and 20 mm of height) suitable for the autoclave. Tubes were centrifuged for 1 minute at maximum speed in a Multifuge 3S Heraeus with a Sorvall Heraeus rotor to concentrate the AIR fraction at the bottom of the tube. A volume of 0.4 ml of 72% H_2SO_4 was added in the tube and mixed using a vortex. When the AIR fraction was difficult to suspend in the acid, a small chemically inert stir bar was put in the tube and vortexed, followed by a spin down in the centrifuge to concentrate the cell walls and the acid. The stir bar was left in the tube for the entire procedure. The tubes were placed in a New Brunswick Scientific INNOVA42 incubator at 30˚C for 1 hour with constant shaking at 200 rpm. After the first hydrolysis, acid was diluted to 4% by adding 11.42 ml of ultrapure water (milli-Q®). Tubes were properly capped, vortexed and autoclaved (Tuttnauer autoclave 3850EL – D, Breda, NL) at 121˚C for 1 hour. The autoclaving step followed a similar warming up and cooling down procedure between batches to ensure repeatability between them. Upon hydrolysis, samples were centrifuged (10 minutes, 3.500*g*, room temperature). After the second hydrolysis, 2 ml of supernatant were filtered using 0.45 μm PTFE membrane filters and 1x and 10x dilutions of the filtered material were prepared for carbohydrate analyses.

A set of sugar recovery standards (SRS) was prepared to quantify the amount of monosaccharides degraded in the second hydrolysis step and to use them as correction

factors. SRS solution included D-(-)arabinose, D-(+)galactose, D-(+)galacturonic acid, D-(+) glucose, D-glucuronic acid, D-(+)mannose, L-(+)rhamnose and D-(+)xylose in ultrapure water. The concentration was specific for each monosaccharide to resemble their concentration in the hemp cell wall (**Supplementary Table 1**). Two control SRS tubes (no hydrolysis) were prepared adding 11.42 ml of SRS solution with 0.4 ml of ultrapure water and two hydrolysis SRS tubes were prepared adding 0.4 ml of 72% H_2SO_4 instead of ultrapure water. The two hydrolysis SRS tubes were hydrolysed in the same way as the cell wall samples. Finally, all SRS were filtered and diluted 1x and 10x as the hemp samples and they were analysed as follows.

Monosaccharide composition of the cell wall hydrolysates and SRS were analysed using High Performance Anion Exchange Chromatography (HPAEC) on a Dionex™ ICS-5000⁺ DC equipped with a Dionex CarboPac™ PA-100G BioLC™ column (2 x 250 mm) preceded by a similar guard column (2 x 250 mm) (Thermo Fisher Scientific, Sunnyvale, CA, USA). Separation of monosaccharides was performed at a flow rate of 0.25 ml/min at 30˚C. To quantify arabinose, galactose, galacturonic acid, glucuronic acid, mannose, rhamnose and xylose 5 µl of 1x diluted samples were injected in the column using a Dionex AS-AP auto-sampler. To quantify glucose, only 2.5 µl of 10x diluted samples were injected. The separation method consisted of an isocratic elution of 20 mM sodium hydroxide (NaOH) for 25 minutes. Then, a linear gradient was followed by starting at 60 mM sodium acetate (NaOAc) in 100 mM NaOH for 15 minutes and ending at 200 mM of NaOAc in 100 mM NaOH. The column was washed with 1 M NaOAc in 100 mM NaOH for 5 minutes prior to re-equilibration for 30 minutes in 20 mM NaOH. Finally, the eluent was monitored by a 30˚C thermostatted Thermo Scientific ICS-5000⁺ pulsed electrochemical detector (PAD) (Thermo Fisher Scientific, Sunnyvale, CA, USA).

The contents of monosaccharides detected by HPAEC-PAD were corrected for the percentage of monosaccharides degraded during the second step hydrolysis derived from the sugar recovery standards. The percentage of each monosaccharide was calculated relatively to the initial amount of cell wall (AIR fraction) and the cell wall percentage. Monosaccharide composition was analysed in triplicate in different hydrolysis batches.

2.5 Quantification of lignin

The content of lignin was analysed based on an adapted Klason lignin (KL) method essentially as described by van der Weijde et al. (2016) with modifications (**Table 2**). The initial material consisted of AIR fraction instead of Neutral Detergent Fibre (NDF) (Goering and van Soest, 1970). KL consisted of the insoluble fraction of the cell wall to H_2SO_4 after the two step-hydrolysis, previously described in section 2.4; which is considered to be mainly lignin (Sluiter, 2012). After the hydrolysis, samples were cooled down and vacuum-filtered using a glass filtering crucible (30 ml, P4, Klaus Hoffmann, Staudt, Germany) with a pre-washed glass fibre prefilter (EMD Millipore™ AP4004700, MERK) to collect the KL fraction. The residues were dried for 16 hours in the oven at 103˚C and weighed to calculate the percentage of Klason lignin in the AIR fraction. The analysis was performed in triplicate. A

crucial step to increase the repeatability of the analysis was the pre-wash of the glass fibre pre-filters with ultrapure water using the vacuum-filtered system and weighted them after being dried for 16 hours in the oven at 103˚C.

Table 2. Summary of the modifications of the five protocols.

Protocol		Modification	Reason	Improvement
Preparation of stem material	1)	Extra grinding step (Retsch Mixer Mill MM400 Retsch, 30 Hz for 1').	Increase the homogeneity of the samples.	Ensure a good representation of the sampled tissues. Increase the repeatability of the cell wall composition analysis.
Cell wall extraction (AIR)	1)	Milled biomass was properly mixed using a vortex.		
	2)	Initial biomass amount was scaled up from 10-50mg to 1000mg.	Extract enough cell wall necessary for its complete characterization in a single extraction.	Reduction of the extraction time.
	3)	Drying of the cell wall fraction using a RapidVap Vacuum Dry Evaporation System.	Speed up the evaporation of alcohols remaining in the cell wall fraction after extraction.	
Two step sulphuric acid hydrolysis (Monosaccharide composition)	1)	Cell wall content amount was decreased to 20 ± 1 mg.	Increase the amount of samples that can be analysed at the same time, without affecting the repeatability.	Increase the throughput of the cell wall composition analysis (60 samples per batch).
	2)	Centrifuge the tubes after adding the cell wall content.	Concentrate the cell wall on the bottom of the tube to ensure complete hydrolysis.	Increase the repeatability of analysis using low amount of starting cell wall.
	3)	Vortexing step after adding the concentrated acid and centrifuge step.	Increase the mixture of acid with the cell wall to increase the homogeneity of the hydrolysis.	Increase the repeatability of the cell wall composition analysis.
		*Small chemically inert stir bar can be added in the tube.		
	4)	Constant shaping at 200 rpm during the first step of the hydrolysis.	Increase the accessibility of the acid to the entire sample increasing the complete degradation of the crystalline structures.	
	5)	Autoclaving procedure following a similar warming up and cooling down between batches.	Perform the same treatment to all samples.	
	6)	HPAEC: isocratic elution of 20 mM NaOAc in 100 mM NaOH (25') + linear gradient from 60 mM NaOAc in 100 mM NaOH (15') to 200 mM of NaOAc in 100 NaOH.	So far, xylose and mannose eluted at the same minute using the HPAEC and it was not possible to distinguish between them.	Analysis of most (neutral and acid) monosaccharides independently in a single run, including xylose and mannose and speed up the analysis.

Protocol	Modification	Reason	Improvement
Klason lignin	1) Combination of lignin and monosaccharide methods in one procedure using 20 ± 1 mg AIR cell wall fraction.	Reduce the amount of cell wall needed for Klason lignin quantification. Increase the amount of samples that can be analysed without affecting the repeatability.	Increases the throughput of the cell wall composition analysis and decreases the time of preparation.
	2) Pre-wash glass fibre prefilters with ultrapure water using the vacuum-filtered system and dried for 16 hours at 103˚C.	The weight of the glass fibre prefilters might decrease with the filtration of the hydrolysed cell walls. The pre-washing keeps the weight constant. This is especially important when the initial amount of cell wall is really small (≈20 mg).	Increase the repeatability of the analysis.
Crystalline polysaccharides	1) Cell wall content amount was increased to 20 ± 1 mg.	The original protocol (2 mg) was optimized in arabidopsis but when applied in hemp the repeatability was low.	
	2) Third hydrolysis was changed from Saeman hydrolysis (72% H_2SO_4) to the two step sulphuric acid hydrolysis.	Hemp has large crystalline fraction reason why this hydrolysis was changed by the two step sulphuric acid hydrolysis (72% + 4% H_2SO_4).	Ensure the complete hydrolysis of highly crystalline structures and increase the repeatability of the analysis.
Stem morphology	1) Fixation was changed from 72 hours to 48 hours.	Long steps that can be reduced without affecting the process.	Reduction of preparation time. Complete preparation of samples in 4 days while the original protocol was 7 days.
	2) Each step of the dehydration was changed from 2 hour to 30 minutes.		
	3) First step of the infiltration was changed from overnight to 2 hours.		

2.6 Analysis of the crystallinity of cell wall polysaccharides

The percentage of crystalline polysaccharides in the bast fibre was analysed based on an adapted Updegraff method (Foster et al., 2010) with modifications (**Table 2**). The method consists of three consecutive hydrolysis with several washing steps in between, to remove the acid and released monosaccharides. The first hydrolysis uses a weak acid, Trifluoroacetic acid (TFA), that targets mainly non-crystalline polysaccharides, such as xylans and pectins and amorphous cellulose and mannan. The second hydrolysis uses the Updegraff solution, a combination of acetic acid, nitric acid and water in a 8:1:2 ratio (Updegraff, 1969). Updegraff hydrolysis weakens the crystalline structure of cellulose and mannose to

assist the next hydrolysis. The third hydrolysis uses a strong acid (H_2SO_4) to break down the crystalline polysaccharides. Foster et al. (2010) use Seaman hydrolysis consisting of a single step hydrolysis with 72% H_2SO_4. Yet, we exchanged this hydrolysis for the two-step H_2SO_4 hydrolysis described in section 2.4, to ensure the complete hydrolysis of the highly crystalline structures of hemp.

The amount of starting material was scaled up to twenty mg of AIR fraction prepared from the bast fibre, to increase the repeatability of the analysis. AIR fraction was weighted into a 12 ml glass tube (Schott® culture tube: 18 mm of diameter and 113 mm of length) with a screw leak-proof cap (Schott® red screw cap: DIN thread GL 23 mm of diameter and 20 mm of height) suitable for the autoclave. Tubes were centrifuged for 1 minutes at maximum speed. A volume of 2.5 ml of 2M TFA was added to the tubes and incubated (90 minutes, 121˚C) in a HLC Heating thermomixer with smooth shaking. After the first hydrolysis, tubes were cooled down on ice and centrifuged (30 minutes, 3.500*g*, room temperature) using a Multifuge 3S Heraeus with a Sorvall Heraeus rotor. Two ml of supernatant were collected, diluted 10x and stored for further analysis. The remaining TFA and released monosaccharides were washed out with 3 ml of isopropanol. Tubes were centrifuged (30 minutes, 3.500*g*, room temperature) and supernatant was discarded, except for approximately 1.5 ml to avoid disturbing the pellet. The washing step was performed three times. To dry the pellets, tubes were placed in a HLC thermomixer at 40˚C with smooth shaking and a constant stream of nitrogen gas was injected inside the tubes.

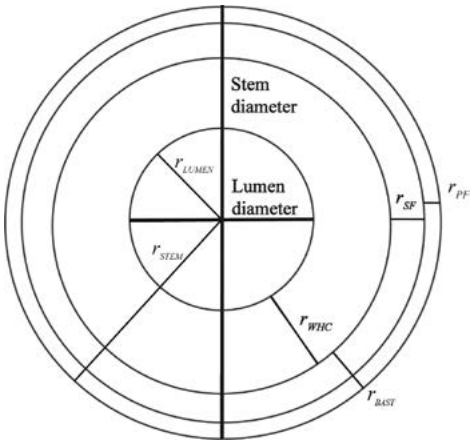


Figure 1. Radius of the different parts of the stem that were used to determine the areas and ratios of different stem structures.

The second hydrolysis consisted of 10 ml of Updegraff solution at 100˚C in a HLC thermomixer for 30 minutes. Tubes were centrifuged (50 minutes, 3.500*g*, room temperature) and supernatant was discarded, except for1.5 ml). Updegraff solution was washed out with 10 ml of ultrapure water, tubes were shake, centrifuged (50 minutes, 3.500*g*, room temperature) and supernatant discarded. Thereafter, the washing step was repeated three more times with 10 ml of acetone. Finally, pellets were dried in a HLC thermomixer at 40˚C

with smooth shaking and constant injection of nitrogen gas. The third hydrolysis with two-step H_2SO_4 was performed as described in section 2.4.
Monosaccharides released upon TFA and H_2SO_4 hydrolysis were determined with a HPAEC-PAD. Total content of cellulose and total content of mannan were the sum of glucose and mannose, respectively released from TFA and H_2SO_4 hydrolysis. Content of cellulose and content of mannan in crystalline forms were determined by the content of glucose and the content of mannose, respectively released from only the H_2SO_4 hydrolysis. The percentage of crystalline cellulose and mannan were calculated as follows:

(1) Percentage of crystalline cellulose, $\%Cryst_{cell} = \frac{Content\ of\ Cellulose\ in\ crystalline\ form}{Total\ content\ of\ Cellulose} * 100$

(2) Percentage of crystalline mannan, $\%Cryst_{man} = \frac{Content\ of\ Mannan\ in\ crystalline\ form}{Total\ content\ of\ Mannan} * 100$

Analyses were performed in triplicate in different batches.

2.7 Morphological analysis of mature hemp stems

Dry mature stems were cut in smalls discs (0.5 cm of length) by using a vertical band saw (FERM SSM1005 Scroll Saw – 90W). To avoid separation between bast and WHC during the cutting process, stems were covered with tissue paper and parafilm® M. Stem sections were fixated, infiltrated and embedded using Technovit® 7100 Kit (Heraeus Kulzer), according to the specifications of the manufacturer with modifications (**Table 2**), in particular incubation times were optimized. Sections were fixed in 0.1 M phosphate buffer with 5% glutaraldehyde under vacuum for 48 hours. To remove the fixative, samples were washed under vacuum with 0.1 M phosphate buffer (4 x 15 minutes) and subsequently with water (2 x 15 minutes). Thereafter, stem sections were dehydrated in a series of 10%, 30%, 50%, 70%, 96% and 100% ethanol under vacuum for 30 minutes each dilution. Samples were pre-infiltrated with a solution of 100% ethanol and Technovit A solution (Technovit liquid 100 ml with 1 g of Hardener I) in a ratio 1:1 under vacuum for 2 hours. Subsequently, samples were infiltrated with Technovit A solution under vacuum for 24 hours. Finally, samples were embedded in the microtome moulds with freshly prepared embedding solution (15 ml of Technovit A solution with 1 ml Hardener II). Once the specimens were solidified, slices of 2 to 5 µm of thickness were cut with a rotary microtome (Reichert-Jung 2055).

For microscopy observations, slices were stained with Toluidine blue, according to O'Brien et al. (1964) with modifications: 0.1% Toluidine blue in 0.1 M phosphate buffer. Dense cellulose fractions (bast fibre) are stained in pink and dense lignin fraction (WHC fibre) in blue. Additionally, the detection of the bast and the WHC could also be performed with immunohistochemistry, as Technovit 7100 is compatible with antibodies. Behr and colleagues reported immunohistochemistry in the hypocotyl of hemp using a set of antibodies on stems embedded with Technovit 7100. LM5, LM10, LM15 and CMB3a from

Plant Probes are antibodies specific to β-1,4-galactan, xylan, xyloglucan and crystalline cellulose, respectively. This set of antibodies was used to differentiate the bast from the WHC (Behr et al., 2016). Observations of the Toluidine blue staining were performed with a light microscope (Axiophot Zeiss) connected to an axioCam digital camera and with a handheld digital microscope (Dino-lite) using Dinocapture software. Morphological analysis of stem included measurements of the radius of bast (primary and secondary bast fibre), WHC, lumen and stem diameter, which were performed using Gimp2 and ImageJ software. **Figure 1** illustrates the measurements of the radius that were taken from each microscopy observations. Measurements were performed in triplicate in each slice and each sample was measured in three slices. The measurements were used to determine the areas and ratios of the different stem structures:

(3) Area of lumen, $A_{Lumen} = \pi * r_{Lumen}^2$

(4) Area of WHC, $A_{WHC} = [\pi * (r_{Lumen} + r_{WHC})^2] - A_{Lumen}$

(5) Area of bast fibre, $A_{Bast} = [\pi * (r_{Lumen} + r_{WHC} + r_{Bast})^2] - (A_{Lumen} + A_{WHC})$

where r_{Lumen} , r_{WHC} and r_{Bast} are the radius of lumen, WHC and bast, respectively.

(6) Bast area% = $\frac{A_{Bast}}{A_{WHC} + A_{Bast}} * 100$

(7) Ratio bast / WHC = $\frac{A_{Bast}}{A_{WHC}}$

(8) Area of primary bast fibre, $A_{PF} = \pi * ((r_{Stem}^2) - (r_{Stem} - r_{PF})^2)$

(9) Area of secondary bast fibre, $A_{SF} = \pi * ((r_{Stem} - r_{PF})^2 - (r_{Stem} - (r_{PF} + r_{SF}))^2)$

Where r_{Stem} , r_{PF} and r_{SF} are the radius of stem, primary and secondary bast fibre, respectively

(10) Ratio primary bast fibre / secondary bast fibre = $\frac{A_{PF}}{A_{SF}}$

2.8 Statistical analyses

Coefficients of variation (CV%) between technical replicates were used to evaluate the repeatability of the methods:

(11) Coefficient of variation = $\frac{Standard\ deviation}{Mean} * 100$

ANOVA was used to determine significant differences between accessions for all the traits and to detect differences between tissues. Statistical analyses were performed using Genstat 19th edition software (VSN International, Hemel Hempstead, UK). Correlation analysis between stem morphological parameters and cell wall traits were performed in R version 3.4.3 statistical software using corrplot function.

3. Results

3.1 Cell wall composition of hemp: bast fibre and stem

Six hemp accessions were analysed for the content and composition of the cell wall in the stem and in the bast fibre. **Figure 2** depicts the overall cell wall composition in the stem and bast fibre reported as the average between accessions. The bast fibre was composed almost exclusively of cell wall (98% AIR%dm), whereas the percentage of cell wall in the whole stem was 92% (**Figure 3**). Glucose, mostly released during hydrolysis of cellulose (Crônier et al., 2005), was the main component of the cell wall. Glucose accounted for 76.24% and 54.4% of the cell wall from the bast and the stem, respectively. The two second main components of the whole stem were xylose, which mainly composes xylan (Pauly et al., 2013), and Klason lignin, accounting for 14% and 14.06% of the total cell wall, respectively. These components only accounted for 1.72% and 2.02% of the bast, respectively. Glucuronic acid, a monosaccharide associated to xylan (Pauly et al., 2013) was also present in higher content in the stem than in the bast: 0.43% versus 0.1%. Mannose, the main component of mannan (Pauly et al., 2013) was present in higher content in the bast fibre than in the stem, 6.88% and 2.28%, respectively. The two monosaccharides in the backbone of pectin –galacturonic acid and rhamnose– (Willats et al., 2001) were found in higher content in the stem than in the bast: 4.77% and 0.82% versus 3.54% and 0.78%. The content of the other pectic monosaccharides –arabinose and galactose– (Willats et al., 2001) was slightly higher in the bast than in the stems (**Figure 2** and **3**). Altogether, the sum of the measured monosaccharides and lignin content reached 93.8% of the total cell wall for the bast and 92.5% for the stem.

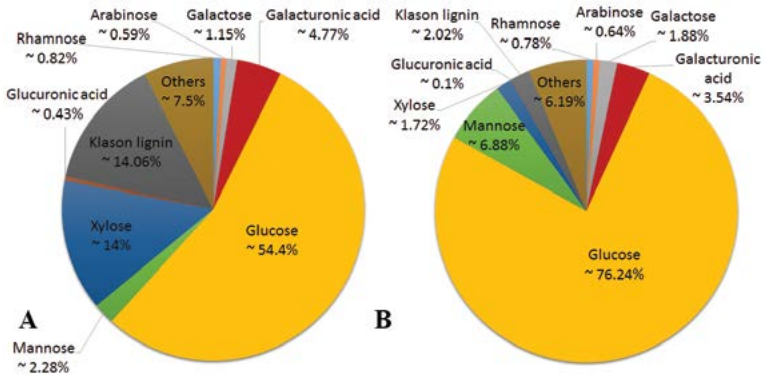


Figure 2. Cell wall composition of hemp stem (**A**) and hemp bast fibre (**B**) as an average of the six hemp accession analysed.

Trait	Description	CRA410		CRA412		CRA416		CRA420		FNPC243		WU101		Differences	
		Stem	Bast	Stem	Bast	Stem	Bast	Stem	Bast	Stem	Bast	Stem	Bast	Stem	Bast
AIR%dm	Alcohol insoluble solid fraction (Percentage of cell wall)	89.9 ± 0.18 (0.2)	96.42 ± 1.57 (1.62)	92.26 ± 0.98 (1.06)	98.69 ± 0.31 (0.31)	92.84 ± 1.05 (1.13)	97.82 ± 0.14 (0.15)	93.2 ± 1.59 (1.7)	99.19 ± 0.42 (0.42)	91.69 ± 1.43 (1.56)	97.13 ± 0.54 (0.55)	92.2 ± 1.19 (1.19)	98.74 ± 1.2 (1.22)	-	-

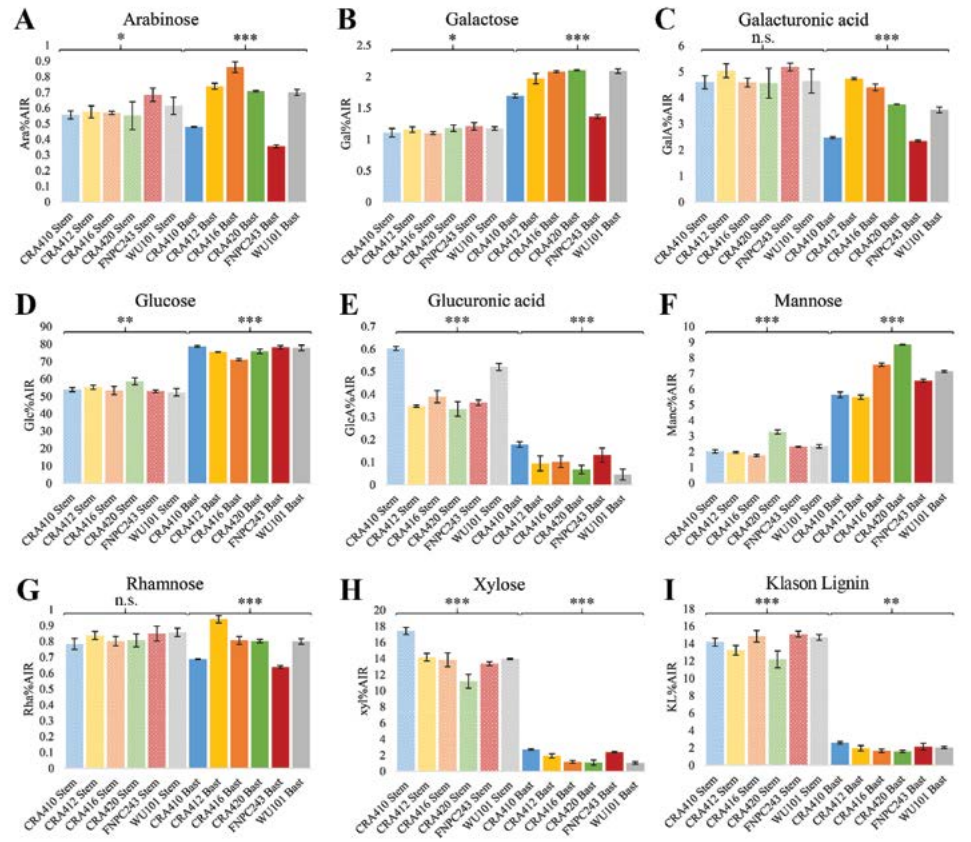


Figure 3. Cell wall composition of the bast fibre and the entire stem of six hemp contrasting accessions. **Table:** The values correspond to the means ± standard deviation. Percentages of coefficient of variation (CV%) between technical replicates are shown between brackets. **Figure:** The columns represent the means and the bar of each column represents the standard deviation. The contents of the nine cell wall components (arabinose (**A**), galactose (**B**), galacturonic acid (**C**), glucose (**D**), glucuronic acid (**E**), mannose (**F**), rhamnose (**G**), xylose (**H**) and Klason lignin (**I**)) were significantly different (p <0.001) between the bast and the stem. Significant levels: *, **, *** and n.s. correspond to significant differences at $p < 0.1$, $p < 0.01$, $p < 0.001$ and no-significant respectively.

3.2 Cell wall composition and content in the six hemp accessions

Phenotyping methods suitable for breeding programmes need to be repeatable and high-throughput. This is needed to distinguish the phenotypes between accessions and to enable analysis of a large number of samples. Six contrasting hemp accessions were used to evaluate the suitability of the methods for breeding for hemp biomass quality. Most parameters showed significant differences between accessions in both tissues, with the exception of cell wall content, as depicted in **Figure 3**. Contents of glucose, glucuronic acid, mannose,

xylose and Klason lignin showed large and significant differences between accessions in the stems. Xylose showed the largest range of variation between the six accessions, CRA410 showed 17% of xylose in the cell wall, while CRA420 showed only 11% of xylose. Arabinose and galactose showed small but significant differences between accessions, while galacturonic acid and rhamnose showed no-significant differences in the stem. In the bast fibre, all cell wall components showed large significant variation and mannose showed the largest range of variation: CRA412 contained 5.5% of mannose in the bast cell wall, while CRA420 had 8.85% of mannose (**Figure 3** and **Supplementary Table 2**).

3.3 Most of the cellulose and mannan in the bast fibre were in a crystalline form

The largest proportion of cellulose from the bast fibre was present in the crystalline form, on average 94% of the cellulose was crystalline (**Table 3** and **Supplementary Figure 1**). Mannan was also highly crystalline, approximately 65% of the total. Both polysaccharides showed significant differences between accessions regarding the proportion of crystalline structures relatively to the total. The range of variation for crystalline cellulose was 1.81% between the six accessions and the range of crystalline mannan was much larger, 17.10%. CRA420 showed the lowest content of crystalline cellulose and mannan, 93.23% and 56.04 respectively, whereas CRA410 showed the largest percentage of crystalline polysaccharides, 95.05% and 73.13%.

Table 3. Percentage of crystalline cellulose and mannan in the bast fibre of six contrasting hemp accession.

Trait	CRA410	CRA412	CRA416	CRA420	FNPC243	WU101	Sign. Level
Percentage of Crystalline cellulose (%Cryst_cell)	95.05 ± 0.588 (0.62)	94.50 ± 0.167 (0.18)	93.66 ± 0.586 (0.63)	93.23 ± 0.296 (0.32)	94.56 ± 0.233 (0.25)	94.30 ± 0.181 (0.19)	**
Percentage of Crystalline mannan (%Cryst_man)	73.13 ± 2.557 (3.50)	65.09 ± 0.242 (0.37)	61.90 ± 3.425 (5.56)	56.04 ± 3.117 (5.56)	69.54 ± 1.784 (2.57)	62.78 ± 1.788 (2.85)	***

Values indicated correspond to mean ± standard deviation. Percentages of coefficient of variation (CV%) between replicates are shown between brackets. Significant levels: ** and ***; significant differences at $p < 0.01$ and $p < 0.001$ respectively.

3.4 Galacturonic acid, rhamnose and xylose were also detected in the crystalline fraction of the bast fibre

Galacturonic acid, rhamnose and xylose were also detected after the two-step sulphuric acid hydrolysis of the crystalline cell wall fraction, as shown in **Table 4** and in **Supplementary Figure 1**. On average 0.5% of xylose content was detected in the crystalline fraction, which represented between 24.26% – 39.46% of the total xylose content, depending on the

accession. A higher content of galacturonic acid content was detected in the same fraction, 0.85% as an average, which represented between 25.9% - 47.6% of the total galacturonic acid content. Both carbohydrates were significantly different between the six accessions but galacturonic acid showed a larger variation. Xylose content in the crystalline fraction ranged from 0.272% to 0.845% whereas galacturonic acid ranged from 0.366% to 1.485%. Small amounts of rhamnose were also identified in the crystalline section, 0.006% as an average, which only represents between 0.76% - 0.9% of the total rhamnose content. However, no significant differences between accessions were found for rhamnose.

Table 4. Content of xylose (Xyl), galacturonic acid (GalA) and rhamnose (Rha) detected in the crystalline fraction of the six hemp accessions.

Trait	CRA410	CRA412	CRA416	CRA420	FNPC243	WU101	Sign. Level
Percentage of Xyl in the crystalline fraction (%)	0.780 ± 0.131 (16.92)	0.34 ± 0.04 (24.26)	0.297 ± 0.054 (18.2)	0.272 ± 0.122 (44.83)	0.84 ± 0.21 (25.1)	0.45 ± 0.19 (42.9)	***
Percentage of Xyl detected to the total Xyl	31.15	12.77	25.6	27.4	35.38	39.46	
Percentage of GalA in the crystalline fraction (%)	0.366 ± 0.036 (9.77)	1.18 ± 0.10 (8.74)	1.485 ± 0.206 (13.87)	0.838 ± 0.083 (9.9)	0.371 ± 0.016 (4.3)	0.868 ± 0.05 (6.22)	***
Percentage of GalA detected to the total GalA	25.9	40.63	47.6	38.66	26.55	39.86	
Percentage of Rha in the crystalline fraction (%)	0.0053 ± 0.00085 (16.09)	0.0067 ± 0.00055 (8.19)	0.0073 ± 0.00052 (7.13)	0.0056 ± 0.0012 (21.87)	0.0045 ± 0.00043 (9.5)	0.0061 ± 0.0011 (18.01)	n.s.
Percentage of Rha detected to the total Rha (%)	0.76	0.71	0.9	0.7	0.71	0.77	

Values indicated correspond to mean ± standard deviation. Percentages of coefficient of variation (CV%) between replicates are shown between brackets. Significant level *** and n.s. stand for significant differences at $p < 0.001$ and no-significant respectively.

3.5 Stem morphological measurements were significantly correlated with most cell wall traits

The morphological structure of the stem was studied in the six hemp accessions to evaluate the organization of the different fibres in the stem and their relationship with the biochemical composition. The morphological study was based on the localization of the polysaccharides and lignin in the stem. Stems from accessions CRA412, FNPC243 and WU101 were the ones with the largest diameters while stems from accessions CRA410, CRA416 and CRA420 were thinner (**Supplementary Figure 2**). Stems from accession WU101 were the thickest with the lowest proportion of bast area, while stems from accession CRA416 were thin with the

largest proportion of bast area, as shown in **Figure 4,** in **Table 5** and in **Supplementary Figure 3**. The ratio bast/WHC and the area of bast fibre were negatively correlated with the stem diameter (r=-0.53 and r=-0.54, respectively) (**Figure 5**). Analysis of the ratio between the primary bast fibre and the secondary bast fibre in the various hemp accessions showed that this ratio was negatively correlated with the stem diameter (r=-0.97), meaning that the accessions with thicker stems showed larger proportion of secondary bast fibres (**Figure 5**, **6**, **Table 5**). Altogether, significant differences were found between stem morphology of distinct accessions.

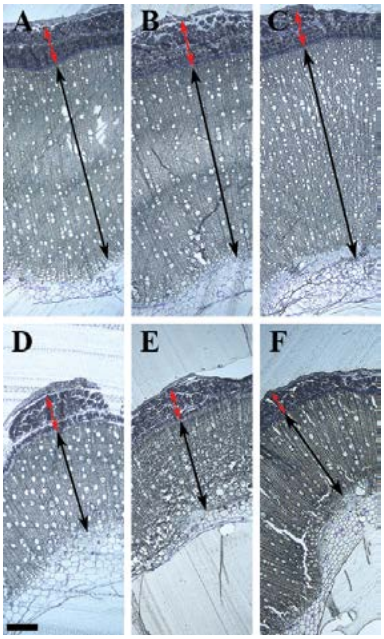


Figure 4. Stem morphology of six contrasting hemp accessions. (**A**) CRA412, (**B**) FNPC243, (**C**) WU101, (**D**) CRA410, (**E**) CRA416 and (**F**) CRA420. Red arrows indicate the bast fibre, black arrows indicate the woody hemp core and scale bars are equivalent to 1000 µm.

Table 5. Stem morphology characteristics of six contrasting hemp accessions.

Trait	CRA410	CRA412	CRA416	CRA420	FNPC243	WU101	Sign. Level
Stem diameter (cm)	0.68 ± 0.029 (4.22)	1.82 ± 0.076 (4.2)	0.9 ± 0.1 (11.11)	0.9 ± 0 (0)	1.73 ± 0.061 (3.33)	2.1 ± 0.1 (4.76)	***
Ratio Bast/ WHC	0.197 ± 0.005 (2.46)	0.218 ± 0.023 (10.48)	0.274 ± 0.024 (8.86)	0.239 ± 0.013 (5.48)	0.21 ± 0.012 (5.73)	0.161 ± 0.013 (8.02)	***
Bast area (%)	16.45 ± 0.338 (2.06)	17.86 ± 1.523 (8.53)	21.51 ± 1.504 (6.99)	19.28 ± 0.85 (4.41)	17.32 ± 0.823 (4.75)	13.86 ± 0.956 (6.90)	***
Ratio Primary/ Secondary bast	3.15 ± 0.53 (16.95)	1.17 ± 0.09 (8.11)	2.58 ± 0.19 (7.43)	2.62 ± 0.38 (14.62)	1.19 ± 0.33 (27.58)	0.84 ± 0.22 (26.19)	***

Values indicated correspond to mean ± standard deviation. Percentages of coefficient of variation (CV%) between technical replicates are shown between brackets. Significant level: *** Significant differences at $p < 0.001$.

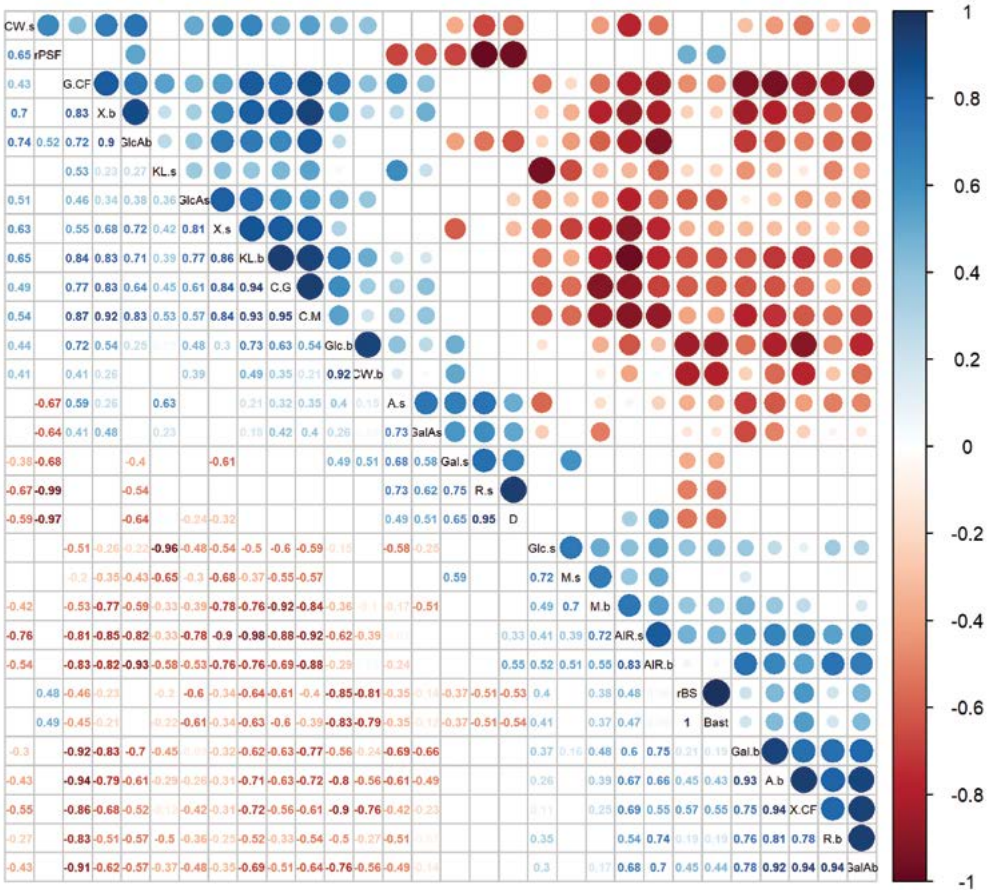


Figure 5. Correlation analysis between morphological parameters and cell wall traits. Significant correlations were set at a confidence level of 0.95 and blank cells represent no significant correlations. The colour and the size of the dots are proportional to the level of significance. **A.s**= arabinose from the stem; **Gal.s**= galactose from the stem; **GalAs**= galacturonic acid from the stem; **Glc.s**= glucose from the stem; **GlcAs**= glucuronic acid from the stem; **M.s**= mannose from the stem; **R.s**= rhamnose from the stem ; **X.s**= xylose from the stem; **KL.s**= Klason lignin from the stem; **TCW.s**= total cell wall components from the stem; **AIR.s**= alcohol insoluble solids fraction from the stem; **A.b**= arabinose from the bast; **Gal.b**= galactose from the bast; **GalAb**= galacturonic acid from the bast; **Glc.b**= glucose from the bast; **GlcAb**= glucuronic acid from the bast; **M.b**= mannose from the bast; **R.b**= rhamnose from the bast; **X.b**= xylose from the bast; **KL.b**= Klason lignin from the bast; **CW.b**= total cell wall components from the bast; **AIR.b**= alcohol insoluble solids fraction from the bast; **C.G**= percentage of crystalline cellulose; **C.M**= percentage of crystalline mannan; **G.CF**= percentage of galacturonic acid in the crystalline cell wall fraction; **X.CF**= percentage of xylose in the crystalline cell wall fraction; **D**= diameter; **rBS**= ratio between bast and WHC; **Bast**= area of bast; **rPSF**= ratio between primary bast and secondary bast fibre.

Relationships between cell wall components and stem morphological measurements were investigated and high significant correlations were identified (**Figure 5**). These included negative correlations between contents of arabinose and galactose in the stem and the ratio between primary bast and secondary bast fibres, the area of bast and the ratio bast/WHC.

Positive correlations were observed between contents of arabinose and galactose in the stem and stem diameter. By contrast, contents of arabinose, galactose, galacturonic acid and rhamnose in the bast fibre were positively correlated with the area of bast and the ratio bast/WHC. Contents of xylose, glucuronic acid and Klason lignin in both stem and bast fibre and the percentage of crystalline structures from the bast were negatively correlated with the area of bast and the ratio bast/WHC. Furthermore, the content of glucose in the stem correlated positively with the area of bast and the ratio bast/WHC, while the content of glucose in the bast fibre correlated negatively with these morphological measurements. Finally, the content of mannose in the bast fibre showed a positive correlation with the area of bast and the ratio bast/WHC.

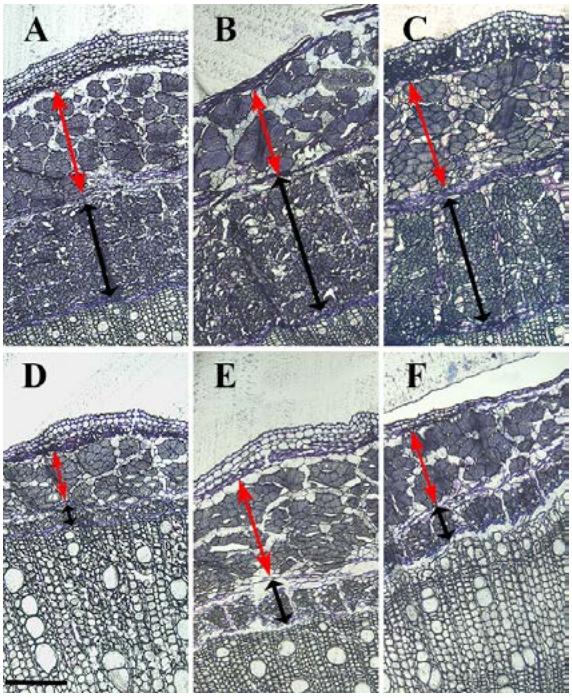


Figure 6. Primary and secondary bast fibre of six contrasting hemp accessions. (**A**) CRA412, (**B**) FNPC243, (**C**) WU101, (**D**) CRA410, (**E**) CRA416 and (**F**) CRA420. Red arrows indicate the primary bast fibre, black arrows indicate the secondary bast fibre and scale bars are equivalent to 200 μm.

Table 6. Radium (in mm) of the primary and secondary bast fibre, woody hemp core (WHC) and pith from the cross-sections of hemp stems.

Radium (mm)	Cross section	Technical replicate	CRA410	CRA412	CRA416	CRA420	FNPC243	WU101
Primary Bast	1	1	109.24	229.17	263.78	231.00	256.58	211.54
		2	159.66	204.17	181.10	194.53	276.32	142.31
		3	142.86	270.83	177.17	234.04	230.26	196.15
	2	1	158.91	195.83	213.38	185.61	222.51	273.03
		2	135.66 (20.01)	241.67 (16.99)	156.05 (20.36)	174.24 (13.99)	240.84 (23.63)	322.37 (32.53)
		3	139.53	250.00	219.75	159.09	251.31	161.18
	3	1	90.00	229.17	267.86	213.59	106.51	116.07
		2	180.00	337.50	187.50	189.32	186.39	285.71
		3	167.50	245.83	160.71	165.05	186.39	205.36
Secondary bast	1	1	33.61	241.67	70.87	94.22	177.63	261.54
		2	46.22	133.33	74.80	97.26	220.39	276.92
		3	84.03	200.00	90.55	94.22	167.76	261.54
	2	1	31.01	191.67	121.02	71.97	178.01	220.39
		2	31.01 (38.09)	241.67 (17.15)	79.62 (37.21)	64.39 (27.43)	164.92 (11.54)	243.42 (9.9)
		3	58.14	225.00	54.14	75.76	180.63	240.13
	3	1	35	250.00	56.55	97.09	177.51	285.71
		2	57.5	212.50	44.64	48.54	224.85	245.54
		3	40	241.67	35.71	43.69	192.31	303.57
WHC	1	1	1025.21	2791.67	881.89	1079.03	2220.39	3934.62
		2	873.95	2641.67	862.20	1091.19	2167.76	4330.77
		3	1067.23	2541.67	870.08	1209.73	2648.03	3880.77
	2	1	883.72	2929.17	1248.41	1007.58	2099.48	3776.32
		2	984.50 (7.09)	2820.83(7.11)	1063.69 (14.64)	814.39 (13.8)	2198.95 (11.71)	3657.89 (8.12)
		3	1003.88	2587.50	894.90	799.24	2651.83	3187.50
	3	1	897.50	2570.83	952.38	1053.40	2026.63	3834.82
		2	992.50	2466.67	1154.76	883.50	1866.86	3700.89
		3	1010.00	2333.33	1148.81	975.73	2292.90	3589.29

Radium (mm)	Cross section	Technical replicate	CRA410		CRA412		CRA416		CRA420		FNPC243		WU101	
Pith	1	1	357.14		0.0		358.27		395.14		305.92		361.54	
		2	386.55		0.0		322.83		395.14		171.05		415.38	
		3	382.35		0.0		385.83		370.82		434.21		330.77	
	2	1	182.17		0.0		286.62		174.24		316.75		384.87	
		2	267.44	(23.18)	0.0	(150.78)	455.41	(26.04)	628.79	(32.97)	253.93	(40.29)	421.05	(16.84)
		3	403.10		0.0		340.76		250.00		204.19		299.34	
	3	1	357.50		191.7		187.50		383.50		443.79		433.04	
		2	257.50		187.5		241.07		388.35		390.53		459.82	
		3	295.00		225.0		252.98		451.46		109.47		522.32	

Percentages of coefficient of variation (CV%) between the replicates are shown between brackets. Pith stands for parenchymatous tissue located between WHC and the lumen.

3.6 The protocols are highly repeatable

Most traits analysed showed small standard deviations and coefficients of variation between technical replicates (**Figure 3**, **Tables 3 – 6** and **Supplementary Table 2** and **Supplementary Figures 2** and **3**). This indicates that the methods used are highly repeatable. Only components that were present in low contents in certain tissues showed larger coefficients of variation in some accessions (e.g. xylose, glucuronic acid and lignin content (KL) in the bast fibre (Crônier et al., 2005)). Generally, the biochemical variation between technical replicates were slightly larger in the stem than in the bast fibre (**Figure 3** and **Supplementary Table 2**).

3.7 The throughput of the protocols for biochemical and morphological analyses of hemp stems was upgraded

The throughput of the protocols for hemp cell wall biochemical characterization was improved by increasing the amount of cell wall extracted in a single round. The starting amount of biomass was scaled up from 10-50 mg of the original protocol (Pettolino et al., 2012) to 1 g. The duration of the AIR preparation remained the same after this modification. Approximately 16 hours were necessary to extract the cell wall. Between 3 to 4 hours were necessary for each part of the protocol: cell wall extract, α-amylase digestion and drying of the pellets after extraction and digestion. A single protocol for monosaccharide characterization and Klason lignin analysis also improved the throughput of the biochemical characterization of hemp cell walls. This protocol also enabled the simultaneous analysis of more samples, 60 samples per run.

The original procedure for the stem morphology (Technovit® 7100 Kit -Heraeus Kulzer-procedure) required 7 days to fixate, infiltrate and embed the samples. The throughput of the protocol developed for analysis of hemp stem morphology was improved by reducing the duration of the procedure to 4 days.

4. Discussion

The growing need towards a circular bioeconomy requires crops that stand for alternative sustainable solutions. Such requirement increases the need for phenotyping methods suitable for breeding programmes for fibre crops, such as hemp. In the present study, we optimized the suitability, repeatability and the throughput of five methods to study the biochemical composition and the morphology of hemp stems.

4.1 Large variability in biochemical composition and morphology of hemp stem

Distinct tissues in the stems of hemp differ in cell wall content and composition, which supports previous findings (Bonatti et al., 2004;Crônier et al., 2005;van den Broeck et al., 2008). The bast fibre showed larger cell wall content than the stem (**Figure 3**), which is in accordance with differences in cell wall thickness between the bast and the WHC. The thicker walls from phloematic cells compared to xylem cells might explain the differences in cell wall content between the bast and the stem (Toonen et al., 2004;Hughes, 2012). Differences in biochemical composition between tissues are essentially in agreement with previous reports. Hemp stems had a higher content of xylose, glucuronic acid and lignin than the bast, which could be explained by the higher contents of xylan and lignin in the WHC than in the bast fibre (Bonatti et al., 2004;van den Broeck et al., 2008). By contrast, bast fibre had higher content of glucose and mannose owed to the higher content of cellulose and mannan in the bast than in the WHC (Crônier et al., 2005). Furthermore, significant differences of galacturonic acid content, the main monosaccharide from pectin (Willats et al., 2001), were found between the bast fibre and the stem. Pectin in the stems of hemp is largely found in the bast fibre and in the middle lamella of the vascular cambium between the bast and the WHC (Crônier et al., 2005). An explanation for the large differences between tissues could be the partial retting of the stems. Retting is a post-harvest treatment of the stems that facilitates the separation of the bast from the WHC. This post-harvest processing mostly degrades the pectin from the middle lamella of the vascular cambium, facilitating this way the detach of bast fibres (Liu et al., 2015). Despite the differences between tissues, the content of monosaccharides derived from pectin from the bast fibre are comparable to those described in previous reports (Crônier et al., 2005).

Cell wall composition of both stem and bast fibre between the six hemp accessions is largely variable (**Figure 3**), which is in line with the large variability in fibre quality, morphology

and biochemical composition of hemp stems described in previous reports (Meijer, 1994;Meijer and Keizer, 1996;Struik et al., 2000;Toonen et al., 2004;Mankowska et al., 2006;Amaducci et al., 2008d;Jankauskiene et al., 2015;Müssig and Amaducci, 2018;Wang et al., 2018). Furthermore, the phenotypic variation of cell wall components was different in the bast and in the stem, suggesting that cell wall composition in different stem tissues is regulated differently (van den Broeck et al., 2008). Xylose and mannose play important roles in the mechanical properties of the fibres (Schönberg et al., 2001;Sorieul et al., 2016), and thus accessions that differ in the composition of these polysaccharides, will most likely have different functionalities. For instance, accessions with larger xylan content could produce extended or stronger polysaccharide matrixes in the WHC, which could enhance the recalcitrance of the stem (Torres et al., 2013).

Crystalline polysaccharides are highly abundant in the bast fibre of hemp supporting previous findings (**Table 3**) (Millane and Hendrixson, 1994;Crônier et al., 2005). In addition, the large variability in crystalline cellulose and mannan between the six hemp accessions could also influence the mechanical properties of hemp fibre. This is because previous reports suggested that crystallinity might affect fibre strength (Bourmaud et al., 2013;Marrot et al., 2013). Remarkably, a fraction of xylose, galacturonic acid and rhamnose were detected in the crystalline cell wall fraction, suggesting that crystalline polysaccharides might protect a fraction of xylan and pectin. These results are in line with previous reports using immunohistochemistry (Gorshkova et al., 2010;Chernova et al., 2018). The large proportion of galacturonic acid found entrapped by crystalline polysaccharides (**Table 4**) could indicate the existence of a pectic polymer with a rhamnogalacturonan I backbone (galacturonic acid and rhamnose) in the G-layer of the bast fibre, as suggested by Chernova et al. (2018). Furthermore, the large variation of galacturonic acid content in this fraction between the six hemp accessions (**Table 4**) could play a role in the regulation of the crystallinity, thus affecting the properties of the crystalline polysaccharides.

The organization of the different fibres in the stem is variable between the six hemp accessions (**Figures 4** and **6**) and is highly related to the biochemical composition of the stem (**Figure 5**), as expected from previous studies (Bonatti et al., 2004;Crônier et al., 2005;van den Broeck et al., 2008;Behr et al., 2016). The bast area and the ratio bast/WHC were positively correlated with the content of glucose and negatively correlated with the contents of xylose and lignin, which support previous findings of the cell wall composition from the bast and the WHC. Bast fibres are characterized by high content of cellulose, while WHC has high contents of xylan and lignin (Bonatti et al., 2004;van den Broeck et al., 2008). The ratio between primary bast and secondary bast fibres was negatively correlated with the contents of arabinose and galactose. Arabinose and galactose are the monosaccharides of the side-chain substitutions of the pectin-type rhamnogalacturonan type I (RGI) (Willats et al., 2001). These relationships suggest a higher content of pectin substitutions in the secondary bast fibre than in the primary bast fibre. The higher content of these substitutions

in the secondary bast fibre suggests stronger polysaccharide matrixes in the cell walls, due to larger amount of cross-links between the side-chains of RGI (Willats et al., 2001). As a consequence, the high content of pectin substitutions might partially explain the large stiffness associated to the secondary bast fibres. Additionally, pectin hampers post-harvest processing of the fibres, such as decortication or scutching (Müssig and Martens, 2003). Therefore, the correlations identified in the present study suggest positive implications for the fibre quality of hemp because large primary bast fibres and low pectin content are highly appreciated, especially for the textile industry (Müssig and Martens, 2003;Chernova and Gorshkova, 2007;Amaducci et al., 2015). Lower content of pectin substitutions in the stem may shorten the retting time of the stems. As a result, the bast fibres are less damaged, which is consistent with previous studies (Liu et al., 2015). Furthermore, the relationships between some stem morphological measurements and the biochemical composition of the bast fibres revealed remarkable results. The bast area and the ratio bast/WHC negatively correlate with the content of glucose and to the percentage of crystalline polysaccharides from the bast, while they positively correlated with the content of mannose and to the monosaccharides derived from pectin in the bast. It seems plausible that these relationships refer especially to an increase of the middle lamella from the bast fibre with the increase of bast area or ratio bast/WHC. Therefore, the increase of bast area might be related to an expansion of the bast in surface, owed to the increase of middle lamella, instead of the increase of total cell wall or the increase of the main components of the bast (e.g. cellulose). These results are beneficial for the breeding for fibre quality, as discussed in previous reports, there is a positive relationship between the middle lamella from the bast fibre bundles and the mechanical properties of the bast fibres (Fuentes et al., 2017).

4.2 Highly repeatable and upgraded high-throughput methods for hemp breeding programmes

The methods developed and optimized to study the biochemical composition and the morphology of hemp stems show large repeatability. This can be concluded from the small differences between technical replicates (**Figure 3** and **Table 3 - 6**), suggesting the suitability of the upgraded methods for the phenotyping and breeding for hemp fibre quality. High quality biochemical and morphological data is essential to study the phenotypic variation between accessions and to be able to identify the genetic architecture underlying quantitative traits (Huang and Han, 2014). Cell wall extraction using the high-throughput methods currently available, such as Neutral Detergent Fibres (NDF) (Goering and van Soest, 1970) are based on aqueous solvents. Those methods are not suitable to extract cell walls comprising important water-soluble components, such as some pectin polysaccharides (Pettolino et al., 2012). The contents of monosaccharides derived from pectin (**Figure 2** and **3**) from the bast fibre in the present study were comparable to those

described in previous reports that used long sequential extraction procedures (Crônier et al., 2005). The alcoholic phobicity properties of the polysaccharides, including pectin, and lignin (Fry, 2003;Huffman and Caballero, 2003;Pettolino et al., 2012;Chen, 2014) can explain the quantification of the almost complete cell wall mass using the optimized methods based on alcohol-based solvents. In addition, the heterogeneous morphology of hemp stems and the large abundance of crystalline polysaccharides in the hemp fibres hinder the complete study of the cell wall composition, which is in line with previous reports (Crônier et al., 2005;Pettolino et al., 2012;Pu et al., 2013). The large repeatability of the monosaccharides and Klason lignin contents between the technical replicates and the almost complete cell wall mass (**Figure 3**) suggest that the optimized two-step sulphuric acid hydrolysis is nearly complete and without degradation of released sugars, as furfural and hydroxymethyl furfural (HMF) were not detectable (data not shown). The modifications of the hydrolysis protocol may allow a better disruption of the cell walls. In turn, this allows a deeper accessibility of the acid through the cell wall structures, allowing a harsher acid performance. This is supported by previous studies (Zhao et al., 2006). Thus, the methods presented in this report combine the high-throughput with the accuracy of the best methods available for cell wall characterization in dicot plants, adapted to hemp limitations. These two features are key for the phenotyping of a large number of plants but maintaining the quality of the phenotypic data needed for quantitative approaches (Huang and Han, 2014).

Altogether, the accuracy of the methods developed for cell wall composition generated high quality data that can be used to develop prediction models with Near-Infrared Spectroscopy (NIRS) (Toonen et al., 2004), and thus increase even further the throughput of hemp cell wall biochemical analysis (Huang and Han, 2014). Notwithstanding, our methods already have the potential to phenotype large number of accessions in a relatively short period of time. Consequently, the methods can already be used to phenotype plants for association studies for fibre quality of hemp

5. Conclusions

Repeatable and high-throughput phenotyping methods to study the genetic variability of fibre hemp are key before we can breed for hemp cultivars with better biomass composition. In the present study, we have developed and optimized five methods to study the phenotypic variation of cell wall composition and stem morphology of hemp. The methods revealed to be repeatable and sensitive to discriminate differential composition between hemp accessions, and thus suitable for characterising the variation in fibre quality of this not extensively breed crop. In addition, the throughput of the methods was significantly upgraded to enable the exploration of the fibre quality and will contribute in the development of breeding tools. The development of such phenotypic methods will have

a valuable impact on the study of hemp and will accelerate the breeding for hemp cultivars with better quality of fibre.

6. Acknowledgements

This research was conducted as part of the MultiHemp project (Multipurpose hemp for industrial bioproducts and biomass) funded by the European Union's Seventh Framework Programme for research, technological developments and demonstration under grant agreement number 311849. The authors are grateful to Isolde Bertram for her technical assistance in the morphological analysis of hemp stems.

Supplementary data

Supplementary Table 1. Concentration of each monosaccharide in the set of carbohydrate recovery standards (SRS).

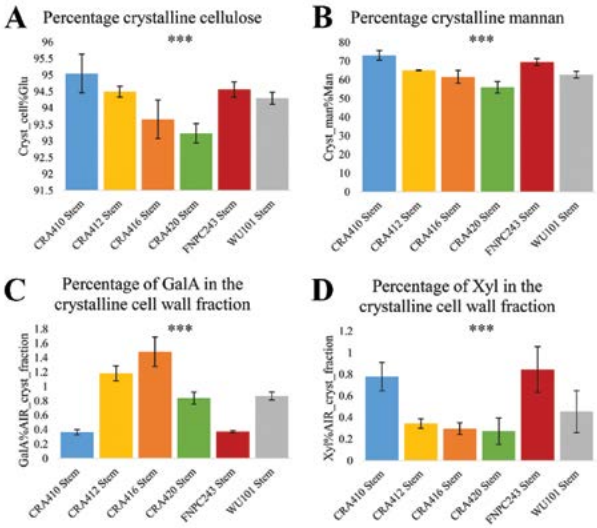
Monosaccharide	Concentration [mg/ml]
Rhamnose	0.016
Arabionse	0.18
Galactose	0.43
Glucose	0.8
Xylose	0.43
Galacturonic acid	0.16
Glucuronic acid	0.011
Mannose	0.07

Supplementary Table 2. Content of monosaccharides and Klason lignin and percentage of AIR fraction (total cell wall) in the stem and in the bast fibre of six contrasting hemp accession.

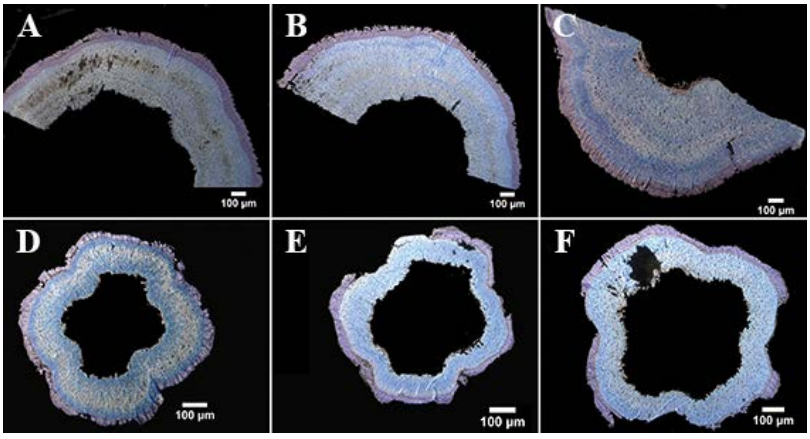
Trait	Description	CRA410		CRA412		CRA416	
		Stem	Bast	Stem	Bast	Stem	Bast
Ara%AIR	Arabinose	0.56 ± 0.025 (4.54)	0.48 ± 0.003 (0.69)	0.58 ± 0.038 (6.68)	0.74 ± 0.019 (2.70)	0.57 ± 0.011 (1.9)	0.86 ± 0.035 (4.02)
Gal%AIR	Galactose	1.11 ± 0.068 (6.14)	1.69 ± 0.03 (1.78)	1.16 ± 0.043 (3.70)	1.97 ± 0.082 (4.18)	1.1 ± 0.024 (2.18)	2.08 ± 0.014 (0.68)
GalA%AIR	Galacturonic acid	4.6 ± 0.252 (5.48)	2.47 ± 0.04 (1.61)	5.05 ± 0.266 (5.27)	4.73 ± 0.039 (0.83)	4.59 ± 0.167 (3.65)	4.4 ± 0.131 (2.98)
Glc%AIR	Glucose	53.89 ± 1.18 (2.19)	78.67 ± 0.53 (0.67)	55.22 ± 1.22 (2.21)	75.5 ± 0.1 (0.13)	53.44 ± 2.33 (4.36)	71.16 ± 0.61 (0.86)
GlcA%AIR	Glucuronic acid	0.603 ± 0.009 (1.52)	0.178 ± 0.012 (6.98)	0.348 ± 0.005 (1.4)	0.095 ± 0.033 (34.78)	0.389 ± 0.027 (6.87)	0.102 ± 0.026 (25.15)
Man%AIR	Mannose	2.03 ± 0.109 (5.35)	5.65 ± 0.175 (3.1)	1.97 ± 0.054 (2.72)	5.49 ± 0.142 (2.58)	1.76 ± 0.065 (3.71)	7.57 ± 0.111 (1.47)
Rha%AIR	Rhamnose	0.78 ± 0.035 (4.41)	0.688 ± 0.003 (0.4)	0.84 ± 0.026 (3.05)	0.939 ± 0.023 (2.45)	0.80 ± 0.03 (3.69)	0.806 ± 0.025 (3.07)
Xyl%AIR	Xylose	17.43 ± 0.44 (2.52)	2.72 ± 0.102 (3.75)	14.17 ± 0.473 (3.34)	1.93 ± 0.260 (13.23)	13.86 ± 0.846 (6.11)	1.18 ± 0.176 (14.92)
KL%AIR	Klason lignin	14.18 ± 0.45 (3.19)	2.6 ± 0.13 (5.15)	13.26 ± 0.54 (4.11)	1.98 ± 0.3 (15.34)	14.86 ± 0.67 (4.48)	1.68 ± 0.2 (11.73)
Total cell wall components	Sum of monosaccahrides and lignin	95.19 ± 2.01 (2.11)	95.15 ± 0.41 (0.43)	92.58 ± 2.44 (2.63)	93.38 ± 0.79 (0.85)	91.37 ± 3.88 (4.25)	89.85 ± 0.35 (0.39)
AIR%DM	Alcohol insoluble solids fraction	89.9 ± 0.18 (0.2)	96.42 ± 1.57 (1.62)	92.26 ± 0.98 (1.06)	98.69 ± 0.31 (0.31)	92.84 ± 1.05 (1.13)	97.82 ± 0.14 (0.15)

The values presented correspond to the means ± standard deviation. Percentages of coefficient of variation (CV%) between technical replicates are shown between brackets. Significant levels: *, **, *** and n.s. correspond to significant differences at $p < 0.1$, $p < 0.01$, $p < 0.001$ and no-significant respectively.

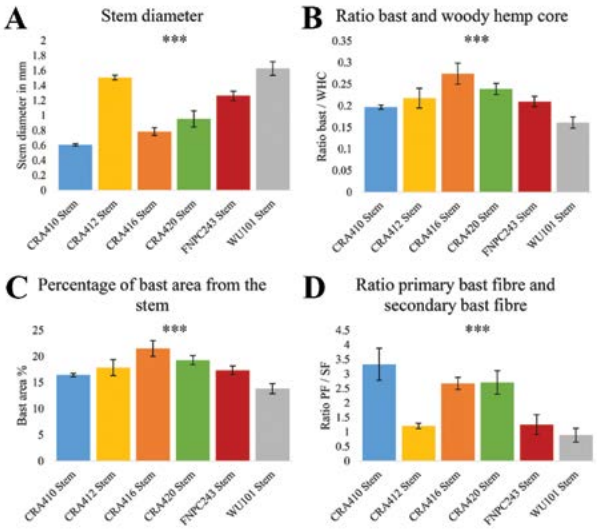
CRA420		FNPC243		WU101		Differences accessions		Differences tissues
Stem	Bast	Stem	Bast	Stem	Bast	Stem	Bast	
0.55 ± 0.088 (15.96)	0.71 ± 0.006 (0.78)	0.68 ± 0.042 (6.19)	0.36 ± 0.009 (2.61)	0.61 ± 0.056 (9.07)	0.7 ± 0.019 (2.72)	*	***	***
1.18 ± 0.05 (4.26)	2.1 ± 0.01 (0.5)	1.21 ± 0.059 (4.92)	1.36 ± 0.03 (2.23)	1.17 ± 0.027 (2.28)	2.09 ± 0.034 (1.64)	*	***	***
4.56 ± 0.572 (12.54)	3.75 ± 0.007 (0.18)	5.18 ± 0.153 (2.94)	2.35 ± 0.033 (1.4)	4.65 ± 0.459 (9.88)	3.54 ± 0.111 (3.12)	n.s.	***	***
58.65 ± 1.95 (3.32)	75.9 ± 1.34 (1.77)	52.1 ± 0.71 (1.35)	78.29 ± 0.95 (1.21)	52.36 ± 2.08 (3.98)	77.91 ± 1.53 (1.97)	**	***	***
0.335 ± 0.032 (9.46)	0.068 ± 0.019 (27.36)	0.363 ± 0.012 (3.27)	0.131 ± 0.031 (23.45)	0.521 ± 0.015 (2.91)	0.045 ± 0.024 (53.5)	***	***	***
3.27 ± 0.135 (4.15)	8.86 ± 0.019 (0.21)	2.32 ± 0.039 (1.67)	6.55 ± 0.086 (1.31)	2.35 ± 0.108 (4.61)	7.16 ± 0.065 (0.91)	***	***	***
0.81 ± 0.041 (5.03)	0.803 ± 0.01 (1.29)	0.85 ± 0.048 (5.61)	0.639 ± 0.008 (1.33)	0.86 ± 0.03 (3.48)	0.799 ± 0.017 (2.18)	n.s.	***	***
11.19 ± 0.852 (7.61)	1.09 ± 0.322 (29.44)	13.37 ± 0.235 (1.76)	2.38 ± 0.092 (3.88)	13.96 ± 0.076 (0.54)	1.06 ± 0.148 (14)	***	***	***
12.2 ± 0.96 (7.85)	1.61 ± 0.14 (8.68)	15.1 ± 0.34 (2.24)	2.16 ± 0.38 (17.38)	14.73 ± 0.33 (2.24)	2.07 ± 0.13 (6.11)	***	**	***
92.73 ± 0.4 (0.43)	94.89 ± 1.01 (1.07)	91.94 ± 0.24 (0.27)	94.22 ± 1.12 (1.19)	91.21 ± 1.45 (1.59)	95.37 ± 1.79 (1.88)	n.s.	***	*
93.2 ± 1.59 (1.7)	99.19 ± 0.42 (0.42)	91.69 ± 1.43 (1.56)	97.13 ± 0.54 (0.55)	92.2 ± 1.19 (1.19)	98.74 ± 1.2 (1.22)	*	*	***



Supplementary Figure 1. Percentage of crystalline cellulose (**A**) and mannan (**B**) in bast fibre of six contrasting hemp accessions. Content of galacturonic acid (GalA) (**C**) and xylose (Xyl) (**D**) detected in the crystalline fraction of the fix hemp accessions. The columns represent the means and the bar of each column represents the standard deviation. Significant levels: **, *** correspond to significant differences at $p < 0.01$ and $p < 0.001$, respectively.



Supplementary Figure 2. Stem morphology of six contrasting hemp accessions. **A**) CRA412, (**B**) FNPC243, (**C**) WU101, (**D**) CRA410, (**E**) CRA416 and (**F**) CRA420.



Supplementary Figure 3. Stem morphology characteristics of six contrasting hemp accessions: stem diameter (**A**) ratio bast and woody hemp core (**B**), percentage of bast area from the stem (**C**) and ratio primary bast fibre and secondary bast fibre (**D**). The columns represent the means and the bar of each column represents the standard deviation. Significant levels: **, *** correspond to significant differences at $p < 0.01$ and $p < 0.001$, respectively.

Chapter 3

Genetic variability of morphological, fibre, cell wall composition and flowering traits in hemp (*Cannabis sativa* L.)

Jordi Petit[a], Elma M. J. Salentijn[a], Maria-João Paulo[b], Claire Thouminot[c], Bert Jan van Dinter[d], Gianmaria Magagnini[e], Jörg-Hans Gusovius[f], Kailei Tang[g], Stefano Amaducci[g], Shaoliang Wang[h], Birgit Uhrlaub[h], Jörg Müssig[h], Luisa M. Trindade[a*]

[a] Wageningen UR Plant Breeding, Wageningen University and Research (WUR), P.O. Box 386, 6700 AJ Wageningen, The Netherlands
[b] Biometris, Wageningen University and Research (WUR), P.O. Box 100, 6700 AJ Wageningen, The Netherlands
[c] Fédération Nationale des Producteurs de Chanvre (FNPC), 72 000 Le Mans, France
[d] VanDinter Semo BV, 9679 ZG Scheemda (VDS), The Netherlands
[e] Centro di ricerca cerealicoltura e colture industriale (CRA), 45100 Rovigo, Italy
[f] Leibniz-Institut für Agrartechnik Potsdam-Bornim e.V. (ATB), 14469 Potsdam-Bornim, Germany
[g] Università Cattolica del Sacro Cuore (UCSC), 29122 Piacenza, Italy
[h] The Biological Materials Group, Biomimetics, City University of Applied Sciences Bremen (HSB), Neustadtswall 30, 28199, Bremen, Germany
[*] Corresponding author: Luisa M. Trindade (luisa.trindade@wur.nl)

This chapter has been accepted (with minor modifications) in *Frontiers in Plant Science.*

ABSTRACT

Hemp (*Cannabis sativa* L.) is a bast-fibre crop well-known for the great potential to produce sustainable fibres. Nevertheless, hemp fibre quality is a complex trait and little is known about the phenotypic variability and heritability of fibre quality relevant traits in hemp. The aim of this study is to gain insights into the variability in fibre quality within the hemp germplasm and to estimate the genetic components, environmental conditions and genotype-by-environment (*G*×E) interactions on fibre quality traits in hemp. To investigate these parameters, a panel of 123 hemp accessions was phenotyped for 30 traits relevant to fibre quality at three locations in Europe, corresponding to climates of northern, central and southern Europe. In general, hemp cultivated in northern latitudes showed a larger plant vigour, while earlier flowering was characteristic of plants cultivated in southern latitudes. Extensive variability between accessions was observed for all traits. Most cell wall components (contents of monosaccharides derived from cellulose and hemicellulose; and lignin content), bast fibre content and flowering traits revealed large genetic components with low *G*×E interactions and high broad-sense heritabilities, making these traits suitable to maximise the genetic gains of fibre quality. In contrast, contents of pectin-related monosaccharides, most agronomic traits and several fibre traits (fineness and decortication efficiency) showed low genetic components with large *G*×E interactions. Thus, large phenotypic differences between accessions in environmental sensitivity were observed, affecting the accession ranking across locations. These results suggest that pectin, agronomic traits and fibre traits are unsuitable targets in breeding programmes of hemp, as their large *G*×E interactions might lead to unexpected phenotypes in untested locations. Furthermore, all environmental effects on the 30 traits were statistically significant, suggesting a strong adaptive behaviour of fibre quality in hemp to specific environments. The high variability in fibre quality observed in the hemp panel, the broad range in heritability and adaptability among all traits prescribe positive prospects for the development of new hemp cultivars with high fibre quality. This can be achieved via both, novel breeding strategies targeted on high heritable traits and/or specific agricultural practices at specific locations to control traits strongly affected by the environment.

1. Introduction

Hemp (*Cannabis sativa* L.) is a well-known bast-fibre crop with evident phenotypic variability in plant morphology between genotypes. For instance, de Meijer and co-workers described large variability in plant height, stem diameter and stem yield between 206 genotypes. They described accessions up to 4 metres of height, while other plants had a dwarf phenotype with less than 1 metre of height (Meijer et al., 1992;Meijer, 1994;Meijer and Keizer, 1996). Extensive variability has also been described in cannabinoid content, in particular for the major cannabinoids: THC (Δ9-tetrathydrocannabinol) and CBD (cannabidiol) (Meijer et al., 1992;Meijer and Keizer, 1996).

The out-crossing behaviour of hemp (Sawler et al., 2015) contributes to the variability whereas another part of the variation in hemp is owed to the dioecious nature of hemp. Male, female and monoecious plants are characterized by large sexual dimorphism affecting plant morphology, flowering time and fibre quality. Male plants have a slender stature, few leaves, flower early and die after flowering. Moreover, male plants have less lignified cell walls, fine fibres and large proportion of primary compared to secondary bast fibres. In contrast, female plants are more leafy, flower later, remain alive until seed maturation, accumulate higher content of lignin in the cell walls and develop larger amount of secondary bast fibres (Amaducci and Gusovius, 2010;Faux et al., 2013). Monoecious plants resemble female plants and are more uniform (Mandolino and Carboni, 2004;Faux et al., 2016). Furthermore, sex determination in cannabis is a quantitative trait. A range of flowers of the opposite sex determined by the genetics can occur in dioecious hemp, and the ratio of female to male flowers in monoecious plants is highly variable (Faux et al., 2013;Faux et al., 2014;Faux et al., 2016). A genetic expression analysis between male and female dioecious plants identified nine mRNA over-expressed in female plants. These mRNA were putatively involved in auxin-related gene expression. The study suggested that the repression of female characteristics in male plants implies the down-regulation of the genes involved in pathways more strictly related to the differentiation of the female sex (Moliterni et al., 2004). In addition, a range of studies revealed that sex determination of hemp is strongly sensitive to external factors, such as to Cu^{++}, Zn^{++} and Pb^{++} ions or to phytohormones (Chailakhyan and Khryanin, 1978;Freeman et al., 1980;Soldatova and Khryanin, 2010;Galoch, 2015;Faux et al., 2016). Such studies suggested that in hemp, non-genetic mechanisms (e.g. epigenetics) might probably affect the control of sex determination (Heikrujam et al., 2014). Consequently, the sexual variation in hemp is expected to be influenced by genetic and environmental components.

Morphological measurements, fibre quality and flowering traits of hemp respond strongly to environmental factors, particularly to photoperiod and temperature but also to soil composition and crop management (Faux et al., 2013;Amaducci et al., 2015;Sawler et al., 2015). Hemp is a short day plant and its flowering time is influenced by changes in the photoperiod regime (Amaducci et al., 2012). In locations where the shift from long-day

towards short-day photoperiod regimes occurs early, hemp plants flower early. However, in locations where the shift occurs later, the critical photoperiod for flowering is reached later (reviewed in Salentijn et al. (2019)). This behaviour affects plant development because plants accumulate biomass during the vegetative growing period but nutrients are shifted from the production of stem, leaves and roots towards the production of flowers and seeds, around the onset of flowering. In addition, lignification of cell walls intensifies after flowering, along with secondary bast fibre formation (van der Werf and Turunen, 2008;Liu et al., 2015). Crop management features (e.g. plant density, irrigation and harvesting time) are also reported to generate differences in phenological traits,(e.g. plant height and stem diameter) (Amaducci et al., 2015). Therefore, hemp accessions cultivated under specific environmental conditions are expected to have specific fibre composition and properties.

Hemp is a sustainable fibre crop with great potential for the production of a plethora of bio-based products. Yet, hemp cultivars with improved fibre properties are needed to promote hemp in the emerging bio-based economy. The first step in a breeding programme is to characterize the genetic variability for the traits of interest, and that can be done by characterizing them in a wide range of accessions. Understanding the contribution of the genetic (*G*), environment (*E*) and genotype-by-environment (*G*×E) interactions components in fibre quality traits is essential to study the stability of fibre quality across different environments and thus, improve the success of breeding programmes.

To date, little research has been conducted on the variability of hemp traits relevant to fibre quality (e.g. cell wall composition, stem decortication, bast fibre content after decortication or fineness of extracted fibre bundles). The objectives of this study are to evaluate the genetic variability, *G*×E interactions and heritability of 30 traits relevant for fibre quality of hemp, and identify which traits are worth to be further investigated with mapping studies. The relationships between the 30 traits will also be investigated.

2. Materials and methods

2.1 Plant material

A test panel of 123 hemp accessions was used in this study to investigate the phenotypic variability of fibre quality in hemp. This panel included mainly fibre accessions, one oil accession, cultivar Finola, one ornamental and few accessions with other uses (**Table 1**).

2.2 Field Experimental Design

The effects of the environment and the genotype-by-environment (*G*×E) interactions on the phenotypic variation of fibre quality were assessed on the basis of three locations across

Europe at respectively high-, mid- and low- latitude. The environments mostly differed in photoperiod and temperature regimes and water availability, as shown in **Table 2**. The 123 hemp accessions were grown in: Rovigo, at CRA (Centro di ricerca cerealicoltura e colture industriale) in Italy (45˚N 11˚E); Chèvrenolles, Neuville-sur-Sarthe, at FNPC (Fédération Nationale des Producteurs de Chanvre) in France (48˚N 0.2˚E) and Westerlee, at VDS (VanDinter Semo BV) in the Netherlands (53˚N 6˚E). Field trials were performed between April and September 2013. Each field trial had a randomized complete block design with three biological replicates (plots) per accession and location. The experimental units were plots of 1 m^2 in Italy and the Netherlands and of 1.5 m^2 in France. In all three locations the same sowing density were used to aim a density of 100 plants/m^2. Plants in the three middle rows were used for phenotyping. In dioecious accessions, phenotyping was performed only in female plants given the difference in fibre quality due to sex dimorphism in hemp. Field trials were harvested at temperature degree days (Σ˚C, the cumulated Celsius degree day over a period at a base temperature of 1˚C) of 1740.25˚C, 1421.1˚C, and 1843.3˚C in CRA, FNPC and VDS respectively, corresponding to full flowering for most accessions in each location.

Table 1. Panel of 123 hemp (*Cannabis sativa* L.) accessions.

MultiHemp code	Accession name / Code	Origin	Accession type	Population type	Provider
MH-AGM-701	Fibrol / Other	Hungary	Fibre	B	AGM
MH-AGM-702	Tiborszallasi / Other	Hungary	Fibre	B	AGM
MH-AGM-703	Tisza / Other	Hungary	Fibre	B	AGM
MH-AGM-704	KC Dora / Other	Hungary	Fibre	B	AGM
MH-AGM-705	Monoica / Other	Hungary	Fibre	B	AGM
MH-CAAS-601	CYM171 / Other	China	Fibre	B	CAAS
MH-CAAS-602	CYM28 / Other	China	Fibre	B	CAAS
MH-CAAS-603	Yunma 5 / Other	China	Fibre	B	CAAS
MH-CAAS-604	CYM49 / Other	China	Fibre	B	CAAS
MH-CAAS-605	CYM273 / Other	China	Fibre	B	CAAS
MH-CRA-401	CRA_1 / Other	Italy	Fibre	B	CRA
MH-CRA-402	CRA_2 / Other	Italy	Fibre	B	CRA
MH-CRA-404	Delta llosa / Other	Spain	Fibre	B	CRA
MH-CRA-405	CRA_4 / Other	Italy	Fibre	B	CRA
MH-CRA-406	Carma Monoica/ Other	Italy	Fibre	B	CRA
MH-CRA-407	Supermono / Other	Italy	Fibre	B	CRA
MH-CRA-408	Fibranova (CRA_5) / Other	Italy	Fibre	B	CRA
MH-CRA-409	Carmagnola / Other	Italy	Fibre	B	CRA
MH-CRA-410	Ermes A / Other	Italy	Fibre	B	CRA
MH-CRA-411	CS (CRA_6) / Other	Italy	Fibre	B	CRA
MH-CRA-412	Carmaleonte / Other	Italy	Fibre	B	CRA
MH-CRA-413	CRA_7 / Other	Italy	Fibre	B	CRA
MH-CRA-414	W-1 / Other	Italy	Fibre	B	CRA
MH-CRA-415	Zenit / Other	Romania	Fibre	B	CRA
MH-CRA-416	Denise / Other	Romania	Fibre	B	CRA
MH-CRA-417	CRA_8 / Other	Italy	Fibre	B	CRA

MultiHemp code	Accession name / Code	Origin	Accession type	Population type	Provider
MH-CRA-418	SVGB-10611 / Other	Italy	Fibre	B	CRA
MH-CRA-419	USO 14 Monoica / Other	Ukraine	Fibre	B	CRA
MH-CRA-420	USO 31 / Other	Ukraine	Fibre	B	CRA
MH-FNPC-201	Other / A11-121-1	France	Fibre	B	FNPC
MH-FNPC-202	Other / A11-121-2	France	Fibre	B	FNPC
MH-FNPC-203	Other / A11-121-3	France	Fibre	B	FNPC
MH-FNPC-204	Other / A11-121-4	France	Fibre	B	FNPC
MH-FNPC-205	Other / A11-121-5	France	Fibre	B	FNPC
MH-FNPC-206	Other / A11-121-6	France	Fibre	B	FNPC
MH-FNPC-207	Other / A11-121-7	France	Fibre	B	FNPC
MH-FNPC-209	Other / A11-121-9	France	Fibre	B	FNPC
MH-FNPC-210	Other / A11-121-10	France	Fibre	B	FNPC
MH-FNPC-211	Other / A11-121-11	France	Fibre	B	FNPC
MH-FNPC-212	Other / A11-121-12	France	Fibre	B	FNPC
MH-FNPC-213	Other / A11-121-13	France	Fibre	B	FNPC
MH-FNPC-214	Other / A11-121-14	France	Fibre	B	FNPC
MH-FNPC-215	Other / A11-121-15	France	Fibre	B	FNPC
MH-FNPC-216	Other / A11-121-16	France	Fibre	B	FNPC
MH-FNPC-217	Other / A11-121-17	France	Fibre	B	FNPC
MH-FNPC-218	Other / A11-121-18	France	Fibre	B	FNPC
MH-FNPC-219	Other / A11-121-19	France	Fibre	B	FNPC
MH-FNPC-220	Other / A11-121-20	France	Fibre	B	FNPC
MH-FNPC-221	Other / A11-121-21	France	Fibre	B	FNPC
MH-FNPC-222	Other / A11-121-22	France	Fibre	B	FNPC
MH-FNPC-223	Other / A11-121-23	France	Fibre	B	FNPC
MH-FNPC-224	Other / A11-121-24	France	Fibre	B	FNPC
MH-FNPC-225	Other / A10-122-1	France	Fibre	B	FNPC
MH-FNPC-226	Other / A10-122-2	France	Fibre	B	FNPC
MH-FNPC-227	Other / A10-122-4	France	Fibre	B	FNPC
MH-FNPC-228	Other / A103-122-1	France	Fibre	B	FNPC
MH-FNPC-229	Other / A103-122-2	France	Fibre	B	FNPC
MH-FNPC-230	Other / A103-122-3	France	Fibre	B	FNPC
MH-FNPC-231	Other / A103-122-4	France	Fibre	B	FNPC
MH-FNPC-232	Other / A103-122-6	France	Fibre	B	FNPC
MH-FNPC-233	Other / A103-122-8	France	Fibre	B	FNPC
MH-FNPC-234	Other / A103-122-10	France	Fibre	B	FNPC
MH-FNPC-235	Other / A9-122-1	France	Fibre	B	FNPC
MH-FNPC-236	Other / A9-122-2	France	Fibre	B	FNPC
MH-FNPC-237	Other / A9-122-3	France	Fibre	B	FNPC
MH-FNPC-238	Other / A9-122-4	France	Fibre	B	FNPC
MH-FNPC-239	Other / A102-122-1	France	Fibre	B	FNPC
MH-FNPC-240	Other / A102-122-2	France	Fibre	B	FNPC
MH-FNPC-241	Other / A102-122-3	France	Fibre	B	FNPC
MH-FNPC-242	Other / A102-122-4	France	Fibre	B	FNPC
MH-FNPC-243	Other / A102-111-1	France	Fibre	B	FNPC
MH-FNPC-244	Other / A102-111-2	France	Fibre	B	FNPC

MultiHemp code	Accession name / Code	Origin	Accession type	Population type	Provider
MH-FNPC-245	Other / A7-104-1	France	Fibre	B	FNPC
MH-FNPC-246	Other / A7-105-4	France	Fibre	B	FNPC
MH-FNPC-248	Other / B6-093-3	France	Fibre	B	FNPC
MH-FNPC-250	Other / B6-093-17	France	Fibre	B	FNPC
MH-FNPC-251	Férimon / Other	France	Fibre	B	FNPC
MH-FNPC-252	Fédora 17 / Other	France	Fibre	B	FNPC
MH-FNPC-253	Félina 32 / Other	France	Fibre	B	FNPC
MH-FNPC-254	Epsilon 68 / Other	France	Fibre	B	FNPC
MH-FNPC-255	Futura 75 / Other	France	Fibre	B	FNPC
MH-FNPC-256	Santhica 27 / Other	France	Fibre	B	FNPC
MH-IWNRZ-901	Bialobrzeskie / Other	Poland	Fibre	B	IWNRZ
MH-IWNRZ-902	Beniko / Other	Poland	Fibre	B	IWNRZ
MH-IWNRZ-903	Tygra / Other	Poland	Fibre	B	IWNRZ
MH-LARC-501	Katlakalna / Other	Latvia	Fibre	B	LARC
MH-UOY-801	Finola / Other	Finland	Seed	B	UOY
MH-VDS-301	Chameleon / Other	Netherlands	Fibre	B	VDS
MH-VDS-302	Marcello / Other	Netherlands	Fibre	B	VDS
MH-VDS-303	Markant / Other	Netherlands	Fibre	B	VDS
MH-VDS-304	Ivory / Other	Netherlands	Fibre	B	VDS
MH-WU-101	JSO 16 / 891229	Russia	Fibre	B	WUR
MH-WU-102	Ajkai-A-TF / 891054	Hungary	Fibre	L	WUR
MH-WU-103	Fibrimon 56 / 880828	France	Fibre	B	WUR
MH-WU-104	Rastislavicke / 880816	Slovakia	Fibre	B	WUR
MH-WU-105	Krasnodarskaja 56 / 891333	Ukraine	Fibre	B	WUR
MH-WU-106	Dneprovskaja 84 / 921054	Russia	Fibre	L	WUR
MH-WU-107	Other / 883290	Russia	Fibre	L	WUR
MH-WU-108	Lovrin 110 / 883173	Romania	Fibre	B	WUR
MH-WU-109	Bialobrzeskie / 891223	Poland	Fibre	B	WUR
MH-WU-110	Other / 880973	Spain	Other	Other	WUR
MH-WU-111	Kompolti Sargászáru / 883049	Hungary	Fibre	B	WUR
MH-WU-112	Other / 883262	Spain	Other	Other	WUR
MH-WU-113	Kompolti hybrid TC / 891070	Hungary	Fibre	B	WUR
MH-WU-114	Fibrimon 56 / 891158	France	Fibre	B	WUR
MH-WU-115	Other / 921203	Canada	Other	W	WUR
MH-WU-116	Panorama var. globosa / 910914	Hungary	Ornamental	B	WUR
MH-WU-117	Silistrenski / 901107	Bulgaria	Fibre	B	WUR
MH-WU-118	Csehslovák-A-TF / 891068	Slovakia	Fibre	Other	WUR
MH-WU-119	Other / 891288	Poland	Fibre	Other	WUR
MH-WU-120	Other / 891090	Turkey	Other	L	WUR
MH-WU-121	Komoroi-A-TF / 891046	Hungary	Fibre	L	WUR
MH-WU-122	Other / 883289	Russia	Fibre	L	WUR
MH-WU-123	Juznaja Odnovremenno / 883293	Russia	Fibre	B	WUR
MH-WU-124	Other / 891240	Spain	Other	Other	WUR
MH-WU-125	Orosi-A-TF / 891059	Hungary	Fibre	Other	WUR
MH-WU-126	Kompolti / 883048	Hungary	Fibre	Other	WUR
MH-WU-127	Dneprovskaja odnodomnaja 6 / 891326	Ukraine	Fibre	B	WUR

MultiHemp code	Accession name / Code	Origin	Accession type	Population type	Provider
MH-WU-128	Other / 891327	Other	Fibre	L	WUR
MH-WU-129	Superfibra / 883040	Italy	Fibre	B	WUR
MH-WU-130	Other / 891057	Hungary	Other	L	WUR
MH-WU-131	Other / 891094	Turkey	Other	L	WUR
MH-WU-132	Other / 880817	Germany	Other	Other	WUR

Population types B, L and W stand for breeding material, landraces and wild material, respectively. Accession type is based on use. The provider refers to the institution that provided the accessions: AGM, Agromag Kft. from Hungary; CAAS, Bast Fibre Crops - Chinese Academy of Agricultural Science from China; CRA (also known as CREA), Centro di ricerca cerealicoltura e colture industriale from Italy; FNPC, Federation National Producteurs de Chanvre form France; IWNRZ, Institute of Natural Fibres and Medicinal Plants from Poland, LARC; Latgale Agricultural Research Centre from Latvia; UoY, University of York from United Kingdom; VDS, VanDinter Semo from the Netherlands and WU, Wageningen University from the Netherlands. Other stands for no specific accession name, code, origin, accessions type or population type for those accessions.

Table 2. Environmental characteristics of the three field trial locations (CRA, FNPC and VDS) during the growing season of the MultiHemp project in 2013.

	CRA	FNPC	VDS
Sow dates	18-19 April	15 May	6 May
Harvest dates	15-17 July	29-31 July	6-9 September
Days of growing season	90	77	126
Daylight 19th April (hours)	13:41	13:51	14:14
Daylight 15th May (hours)	14:51	15:10	15:51
Daylight 15th June (hours)	15:37	16:01	16:57
Daylight 31st July (hours)	14:45	15:01	15:43
Daylight 31st August (hours)	13:20	13:28	13:45
Σ°C	1691.2 - 1740.25	1369.25 - 1421.1	1800.05 - 1843.3
Average Σ°C/day	19.12	18.22	14.51
Σrainfall (mm)	184.8	195.7	363
Days without rain (%)	43.9	62.5	51.49
Average rainfall/day	2.031	2.509	2.858
Average min RH%/day	45.23	44.9	59.56
Average max RH%/day	93.02	94.39	97.85

Daylight regimes were attained from https://www.timeanddate.com/sun/; Σ°C= accumulated Celsius degrees during the growing season; Σrainfall= the total amount of rainfall in mm during the growing season; and RH%= percentage of relative humidity.

2.3 Phenotypic data analysis

In total 30 parameters were phenotyped, including seven agronomic traits, four different flowering traits including sex determination, nine fibre measurements (morphological and processing-related properties) and ten parameters of cell wall composition (**Table 3**).

2.3.1 Agronomic measurements

Growth was monitored in 3 plants per plot by measuring plant height, weekly until flowering. Heights of the plants at different Σ°C (Σ°C=580 and Σ°C=900) were used to calculate the

growing rate of the plants. Growth rate (GR) over the period of Σ°C=580 and Σ°C=900 (GR in percentage per 10 Σ°C) was calculated as follows:

$$GR = \left(\frac{Height\,900^{\frac{1}{(900-580)}}}{Height\,580} - 1 \right) x\,100\,x\,10\,, \quad (1)$$

where Height 580 and Height 900 are the heights of the plants at Σ°C=580 and Σ°C=900, respectively.

At harvesting time, plants were cut at the base of the stem and agronomic traits were measured. Stem height (H in cm) was measured after harvest. Stem diameter (D in mm) was measured at 10 cm above the ground. Stem volume (stem_volume in cm^3) of the whole stem was calculated as follows:

$$Stem\_volume = \pi r^2 h\,, \quad (2)$$

where r is the radium (half diameter) and h is the stem height. GR, H, D and Stem_volume were phenotyped in 3 plants per plot and data was provided as the mean per plot.

Table 3. Summary statistics of 30 traits in 123 hemp accessions

Trait	Abbreviation	Trait group	Cell wall	Mean	Min.	Max.	Range	Standard deviation	Coefficient of variation (CV%)
Stem diameter, after harvest (mm)	D	Agronomic	-	8.776	3.267	24.02	20.76	4.076	46.44
Total DW of 5 plants as fraction of the FW of 5 plants (%)	DW(%)	Agronomic	-	0.358	0.218	0.729	0.511	0.0917	25.64
Dry weight 5 stems (%)	DW_S(%)	Agronomic	-	0.703	0.332	0.824	0.492	0.0711	10.11
Dry weight 5 stems (g)	DW_S5(g)	Agronomic	-	146.3	5.4	844.1	838.7	146.6	100.2
Grow rate 10Cd (%)	GR_10Cd	Agronomic	-	3	0.938	6.912	5.974	0.826	27.52
Stem height (cm)	H	Agronomic	-	198.5	56.13	324.8	268.7	63.91	32.19
Stem volume (cm3)	Stem_volume	Agronomic	-	188.8	4.811	1812	1807	225.8	119.6
Acid Detergent Lignin (%)	ADL%dm	Cell wall	Lignin	9.071	7.249	13.83	6.58	0.912	10.05
Arabinose (%)	Ara%dm	Cell wall	Pectin	0.797	0.412	1.207	0.794	0.22	27.61
Galactose (%)	Gal%dm	Cell wall	Pectin	1.555	0.924	2.156	1.231	0.311	20.01
Galacturonic acid (%)	GalA%dm	Cell wall	Pectin	4.739	3.372	6.685	3.313	0.781	16.48
Glucose (%)	Glc%dm	Cell wall	Cellulose	48.89	41.73	56.52	14.8	2.584	5.285
Glucuronic acid (%)	GlcA%dm	Cell wall	Xylan	0.367	0.205	0.554	0.349	0.0696	18.98
Klasson Lignin (%)	KL%dm	Cell wall	Lignin	14.7	10.71	19.34	8.627	1.374	9.342
Mannose (%)	Man%dm	Cell wall	Mannan	2.738	1.826	3.773	1.946	0.332	12.12
Rhamnose (%)	Rha%dm	Cell wall	Pectin	0.746	0.619	0.903	0.284	0.0565	7.571
Xylose (%)	Xyl%dm	Cell wall	Xylan	13.63	10.74	17.5	6.761	1.453	10.67
SHIVES% (%)	(χ)	Fibre	-	11.76	0	38.68	39.2	5.806	49.39
Bast content after decortication (%)	BCD%	Fibre	-	29.07	11.61	51.05	39.44	6.431	22.12

Trait	Abbreviation	Trait group	Cell wall	Mean	Min.	Max.	Range	Standard deviation	Coefficient of variation (CV%)
The average stem weight (g)	M_o	Fibre	-	10.3	0.89	38.99	38.99	6.566	63.76
Fibre weight before the separation (g)	MF0	Fibre	-	24.31	0.111	50.7	50.59	8.923	36.71
Fibre weight after the separation (g)	MF1	Fibre	-	21.96	0.374	57.94	57.56	8.587	39.11
Fibre fineness parameter, high compression (mm water)	PH	Fibre	-	8.468	2.979	21.33	18.35	3.074	36.3
Fibre fineness parameter, low compression (mm water)	PL	Fibre	-	14.1	7.819	22.33	14.51	2.799	19.85
Decortication index (%)	ηDec_1	Fibre	-	80.58	38.79	97.37	58.58	11.04	13.7
Decorticability (%)	ηDec_2	Fibre	-	94	78.77	101.1	22.28	3.503	3.727
Beginning flowering time (Σ˚C)	FL_BEGIN	Flowering	-	1178	204.8	2329	2124	338	28.7
Full flowering time (Σ˚C)	FL_FULL	Flowering	-	1552	416.8	3466	3049	497.9	32.07
Begin of flowering relative to day of first emergence (days)	VEG	Flowering	-	69.75	12	143	131	20.58	29.51
Sex determination	Sex_det	Flowering	-	2.222	1	3.077	2.077	0.86	38.71

Fresh weight was measured in 5 entire plants per plot including stem and leaves. Thereafter, leaves and stems were separated and dried at 60˚C for 48 hours. Different dry weights were calculated: dry weight of 5 stems (DW_5S in g), dry weight as fraction of the fresh weight of 5 stems (DW_S% in percentage) and dry weight as fraction of the fresh weight of 5 plants (DW% in percentage).

2.3.2 Flowering parameters

Emergence of the plants was scored as the accumulated Celsius degree days or temperature sum (Σ˚C) at the day of first emergence. Emergence was scored in one row per plot at day=N, N+2, N+4 and N+7, where N is the day of sowing. Flowering time traits were also measured in Σ˚C at 10 plants per plot. Beginning of flowering (FL_Begin in Σ˚C) and full flowering (FL_Full in Σ˚C) were calculated relative to the emergence as:

$$FL\_Begin = \Sigma^{\circ}C_{Beginning\ flowering} - \Sigma^{\circ}C_{Emergence}, \quad (3)$$

$$FL\_Full = \Sigma^{\circ}C_{Full\ flowering} - \Sigma^{\circ}C_{Emergence}, \quad (4)$$

where $\Sigma^{\circ}C_{\text{Beginning flowering}}$, $\Sigma^{\circ}C_{\text{Full flowering}}$ and $\Sigma^{\circ}C_{\text{Emergence}}$ are the accumulated Celsius degree days respectively at the beginning of flowering, full flowering and at the day of first emergence. The length of vegetative growth period (VEG in days) is the growing period of the plants in days, as measured from the day of first emergence until FL_Begin. Sex determination was phenotyped assessing "1" for predominantly dioecious plants, "2" for the mix of dioecious and monoecious plants and "3" for predominantly monoecious.

2.3.3 Fibre traits (morphological and processing-related properties)

The measurements of the processing related properties were performed on stem segments of at least 100 cm. The first 20 cm from the base of the plant and the 30 cm from the top were discarded. Stem portions were naturally dried, in open air under a roof, until the water content was less than 18% of the mass. Thereafter, stems were warm water retted for 3 days at an average temperature of 23°C, according to van den Oever et al. (2003). After water retting, stems were naturally dried again and stored at 20 ± 3°C and relative humidity of 60 ± 5%. All stem weight measurements were calculated as an average in grams of 10 stems per plot. The first measurement of stem weight, $M_{0,}$ was performed straight before the decortication. Each specimen was decorticated individually with a lab-scale roller-breaker decortication system, according to Wang et al. (2018). Stem portions passed through all decortication steps six times. The weight of each decorticated specimen was measured and recorded after each passage through the decorticator, M_i in grams where "i" is the passage number from 1 to 6. After the sixth passage, the remaining shives (also known as woody hemp core, WHC) were removed manually from the bast and the shives-free bast was weighed (M_7). The fibre bundles of the shive-free bast were separated using a Worthmann coarse separator unit (Worthmann Maschinenbau GmbH, Barßel-Harkebrügge, Germany). The weight of the separated fibre bundles were measured and recorded before (MF0 in g) and after (MF1 in g) the separation.

The stem weight (M_0) as well remaining weights after respective decortication steps (M_2, M_6 and M_7) were used to calculate the bast content and the decortication efficiency parameters, according to Wang et al. (2018). Bast content after decortication (BCD in percentage) was calculated as the fraction between the mass of shives-free bast (M_7) and the mass of the initial non-decorticated stems (M_0):

$$BCD\% = \frac{M_7}{M_0} \text{x} 100 \,. \tag{5}$$

The initial decortication efficiency ($\eta_{\text{Dec\_1}}$ in percentage) describes the efficiency of the initial stage of the decortication process. It was calculated by using the following formula:

$$n_{Dec\_1} = \frac{M_0 - M_2}{\text{M}_0 - M_7} \text{x} 100 \,. \tag{6}$$

The ultimate decortication efficiency ($\eta_{\text{Dec\_2}}$ in percentage) estimates the efficiency of the overall decortication process known as decorticability. The decorticability indicates the

difference between the weight of the bast fibre after the final removal of the remaining shives after the decortication (M_7) and the weight of the bast after the last round of the decortication process (M_6):

$$n_{Dec\_2} = \frac{M_0 - M_6}{M_0 - M_7} \text{x} 100 . \quad (7)$$

Shives content after decortication (χ in percentage) describes the ratio of the shives that remained stuck to the bast after the decortication:

$$(\chi) = \frac{M_6 - M_7}{M_6} \text{x} 100 . \quad (8)$$

Fineness of extracted fibre bundles was indirectly characterized by measuring the permeability of air flow injected in the bast fibre bundles with a defined mass (Müssig and Amaducci, 2018). The permeability of air is an indicator of the fibre bundle surface. Fineness was measured using a Shirley IIC Fineness and Maturity Tester (Shirley FMT), according to Müssig (2001) and Müssig et al. (2010). Twelve technical replicates of 4 ± 0.005 g of separated bast fibre bundles were weighed for the analysis after 24 hours of acclimatization at 20°C and 65% relative humidity of air for sample standardization. Two different air compressions were injected in each sample: low compression of air at a flow rate of 4 litres of air per minute and high compression at a flow rate of 1 litre of air per minute (Montalvo and Faught, 1999). Two different Shirley values were obtained: P_L and P_H. P_L (in mm water) is the pressure of the air injected at a low compression and P_H (in mm water) is the pressure of the air injected at a high compression of air. Both measurements were calculated as the mean of the twelve specimens per sample.

Table 4. Summary of internal cross-validation statistics of mPLS models used for the prediction of cell wall composition. See **Table 3** for abbreviations.

Trait	Number of samples	Biochemical data			NIRS predicted data			r^2	SEC	SECV
		Mean	Min	Max	Mean	Min	Max			
ADL%dm	112	9.23	6.55	14.13	9.19	5.07	13.30	0.90	0.43	0.53
Ara%dm	110	0.74	0.27	1.41	0.72	0.00	1.45	0.94	0.06	0.08
Gal%dm	112	1.45	0.74	2.55	1.45	0.33	2.57	0.89	0.12	0.13
GalA%dm	112	5.15	3.89	8.61	5.07	2.26	7.88	0.91	0.27	0.38
GlcA%dm	115	0.39	0.17	0.75	0.39	0.08	0.70	0.71	0.06	0.07
Glc%dm	114	49.34	42.38	58.56	49.22	37.83	60.61	0.75	1.89	2.06
KL%dm	114	14.97	9.812	20.53	14.96	8.77	21.15	0.87	0.79	0.87
Man%dm	116	2.63	1.69	4.09	2.63	1.16	4.10	0.80	0.22	0.28
Rha%dm	112	0.82	0.62	1.13	0.82	0.56	1.08	0.76	0.04	0.05
Xyl%dm	112	14.10	9.53	18.70	14.10	6.99	19.70	0.91	0.55	0.69

r^2 = squared Pearson coefficient of correlation. SEC = standard error of the calibration. SECV = standard error of the cross-validation

2.3.4 Biochemical analysis of hemp cell walls

Hemp cell walls are mostly composed of polysaccharides and lignin and this was therefore the main target of the biochemical analysis. Polysaccharide composition was measured based on the content of the monosaccharides that are specific for each polysaccharide. In total 10 cell wall parameters were measured: the monosaccharide glucose (Glc%dm) that is mostly derived from cellulose; mannose (Man%dm) derived from mannan; xylose (Xyl%dm) and glucuronic acid (GlcA%dm) derived from xylan; arabinose (Ara%dm), galactose (Gal%dm), galacturonic acid (GalA%dm) and rhamnose (Rha%dm) derived from pectin, and furthermore two measurements for lignin, Klason lignin (KL%dm) and acid detergent lignin (ADL%dm). All parameters were calculated as percentage of the dry matter. All cell wall traits were measured with multivariate prediction models based on near-infrared spectroscopy (NIRS), after a calibration curve for each cell wall trait was developed. In details, 5 stems of each plot were harvested, after which the un-retted stem were dried, pooled and grinded, according to Petit et al. (2019a). The ground samples were scanned using a Foss DS2500 near-infrared spectrometer (Foss, Hillerød, Denmark) to obtain the NIR spectra of stem samples (details in van der Weijde et al. (2016)). A subset of 114 samples was selected based on the variation of the NIR spectra and biochemically analysed to develop the prediction models. The models were externally validated with 24 randomly selected samples. Biochemical analysis of monosaccharide composition and content of Klason lignin were performed according to Petit et al. (2019). Acid detergent lignin was analysed according to the protocol developed by Ankom Technology (ANKOM Technology Corporation, Fairpoint, NY), based on the previous established protocols (Goering and van Soest, 1970). Biochemical and NIRS data from model and validation samples were used to develop the prediction models and to validate them, using WinISI version 4.9 statistical software (Foss, Hillerød, Denmark). The prediction equations were generated using modified partial least square regression analysis (Shenk and Westerhaus, 1991) with 1:4:4:1 as parameters for derivation and smoothing of NIRS. The first 7 principal components were manually selected for modelling the equations, which included 98.76% of the NIRS variation. The quality of the prediction models was evaluated with an internal cross-validation and with an external validation. The cross-validation of the models was evaluated by calculating the squared Pearson coefficient of correlation (r^2) and the standard error of the cross-validation (SECV) between the biochemical and the predicted data of the model samples. The external validation was evaluated by calculating the r^2 between the biochemical and the predicted data of the external validation samples. Models were also evaluated by comparing the standard error of the prediction (SEP) and the standard error of laboratory (SEL) of external samples. **Tables 4** and **5** detail internal cross-validation and external validation statistics of the prediction models. The models were used to predict the cell wall composition of 1034 stem samples.

2.4 Statistical analyses

In order to study the variability of fibre quality in hemp, an ANOVA model was used to determine the significant differences of each variance component in the 30 traits: genotype (*G*), environment (*E*), blocks within environment (*B*), genotype-by-environment interactions (*G*×E) and residual variance (ε).

The analysis was performed following the model:

$$y = \mu + E + B + G + G{\times}E + \varepsilon, \quad (9)$$

where *y* is the trait, μ is the grand mean, *E* is the effect due to the environment, *B* is the effect of block within environment, *G* is the genotypic effect, *G*×E is the genotype-by-location interactions and ε is the residual error.

Table 5. Summary of external validation statistics of mPLS models used for the prediction of cell wall composition. See **Table 3** for abbreviations.

Trait	Number of samples	Biochemical data			NIRS predicted data			Slope	Intercept	r^2	SEP	SEL
		Mean	Min	Max	Mean	Min	Max					
ADL%dm	24	8.79	6.84	12.64	9.06	7.74	12.35	1.14	-0.26	0.91	0.47	0.25
Ara%dm	24	0.83	0.42	1.18	0.80	0.45	0.93	1.05	0.02	0.89	0.07	0.04
Gal%dm	24	1.66	0.91	2.31	1.53	0.98	1.67	1.45	0.12	0.81	0.23	0.08
GalA%dm	24	5.23	3.99	6.08	5.17	4.06	5.60	0.96	0.06	0.81	0.35	0.16
GlcA%dm	24	0.33	0.52	0.35	0.35	0.50	0.32	0.94	-0.02	0.75	0.05	0.02
Glc%dm	24	50.91	47.96	50.4	49.30	46.80	48.71	0.62	1.64	0.22	3.54	0.70
KL%dm	24	14.51	11.68	17.44	14.50	14.90	17.33	0.86	0.05	0.60	0.91	0.29
Man%dm	24	2.90	2.15	2.52	2.86	2.30	2.68	1.14	0.05	0.74	0.25	0.10
Rha%dm	24	0.87	0.71	1.03	0.84	0.72	0.90	1.38	0.02	0.82	0.05	0.02
Xyl%dm	24	13.93	17.07	14.17	13.90	16.68	13.91	0.94	0.03	0.71	0.76	0.30

r^2 = squared Pearson coefficient of correlation. SEP = standard error of the prediction. SEL = Standard error laboratory

In addition, a random effects model was used to determine the estimates of variance components of the phenotypic variation following the model:

$$\sigma_y^2 = \sigma_E^2 + \sigma_{B(E)}^2 + \sigma_G^2 + \sigma_{G\times E}^2 + \sigma_\varepsilon^2, \quad (10)$$

where σ_y^2, σ_E^2, σ_B^2, σ_G^2, $\sigma_{G\times E}^2$, σ_ε^2 are the variances for *Y*, *E*, *B*, *G*, *G*×E and ε, respectively. The variance components were reported as the percentage of each component to the total phenotypic variation. Both ANOVA and random effects models were performed using a REML (Restricted Maximum Likelihood) algorithm. For each trait, the stability of the accessions across locations was determined with the size of the variation due to genotype-by-environment interactions relative to the main genotypic component, as in Gitonga et al. (2014):

$$Ratio\, G\times E / G = \frac{\sigma_{G\times E}^2}{\sigma_G^2}. \quad (11)$$

The broad-sense heritabilities (H^2) were calculated across the three environments, as the fraction of the genetic component (σ_G^2) to the total genotypic effect (σ_G^2, $\sigma_{G\times E}^2$ and σ_ε^2), including the genotype-by-environment interactions and the residual variance corrected by the number of blocks and environments, as in Renaud et al. (2014):

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{G\times E}^2}{n\cdot E} + \frac{\sigma_\varepsilon^2}{n\cdot B\times n\cdot E}}. \quad (12)$$

where $n\cdot E$ is the number of environments and $n\cdot B$ is the number of blocks. REML and broad-sense heritability (H^2) analyses were performed using Genstat 19th edition software (VSN International, Hemel Hempstead, UK).

Summary statistics of the 30 traits and the accessions was performed in Genstat 19th edition software. Correlation analysis between the 30 traits was performed in R (http://www.r-project.org/) version 3.4.3 statistical software using corrplot function. The adjusted mean of the phenotypic values across the three locations was used for each trait to study the main correlations independently of the effect of the environment. The most important correlations were represented with a network analysis based on graphical lasso performed in R, using the High-Dimensional Undirected Graph Estimation (HUGE) package version 1.2.7 (Zhao et al., 2012). The best solution, using the StARS criterion at $\alpha = 1$ (Liu et al., 2010), was non-informative. Several lambdas were used and an intermediate solution was found at $\lambda = 0.51$. Boxplots were also graphed in R using ggplot2 package.

3. Results

3.1 Fibre quality variability of the hemp accession panel

Significant differences between the averaged performance across the three environments for all accessions were found for all traits ($p < 0.001$). Most traits showed extensive variation among the accessions of the hemp panel. This was revealed by the wide range and the large coefficients of variation for each trait presented in **Table 3**. For most traits, low variation between individual plants of the same accession was observed, except for most agronomic, monosaccharides derived from pectic polysaccharides and fibre traits (**Supplementary Table 1**). Traits with wide variation between accessions included Glc%dm, Man%dm, Xyl%dm, ADL%dm, KL%dm, BCD%, P_H, P_L and four flowering traits (**Table 3**). **Tables 6**, **7** and **8** show the averaged phenotypic values and the coefficients of variation across all environments of these 12 traits, for the accessions that displayed the most contrasting phenotypic values.

Accessions IWNRZ-902 (Beniko) and WU-111 (Kompolti Sargászáru) showed the largest contents of Glc%dm, Man%dm and BCD%, while LARC-501 (Katlakalna) and UOY-801 (Finola) showed the opposite phenotypic characteristics. In contrast, the opposite patterns were found for Xyl%dm, ADL%dm and KL%dm, where IWNRZ-902 and WU-111 showed the

lowest phenotypic values and LARC-501 and UOY-801 showed the largest values (**Table 6** and **7**). CAAS-601 showed the finest fibre bundles, while IWNRZ-902 showed the coarsest ones, as presented in **Table 7**. Chinese accessions (CAAS) were the latest to flower and showed the longest vegetative growth period (VEG). In addition, some Chinese accessions (CAAS-601, CAAS-602, CAAS-603, and CAAS-605) did not reach full flowering, before the end of the field trials, in the Netherlands and France but they did in Italy (data not shown). In contrast, LARC-501, UOY-801 and WU-122 were the earliest accession to flower and to reach full flowering and they showed the shortest VEG (**Table 8**). Finally, contrasting accessions for sex determination can be found in **Table 8**. Sex determination highlighted large range of variation between predominantly dioecious and predominantly monoecious. For instance, Chinese accessions and UOY-801 showed only dioecious plants (score=1) and IWNRZ-901, IWNRZ-902 and IWNRZ-903 showed only monoecious plants (score=3). Other accessions showed dioecious plants mixed with monoecious plants in different proportions. LARC-501 showed larger number of dioecious than monoecious plants (score=1.22), WU-119 showed approximately equal amount of dioecious and monoecious plants while CRA-415 showed more monoecious than dioecious plants (score=2.44).

Table 6. Summary statistics of hemp accessions with extreme phenotypes for 5 cell wall traits.

Accession	ADL%dm		Glc%dm		KL%dm		Man%dm		Xyl%dm	
	Mean ± SD	CV%	Mean ± SD	CV%	Mean ± SD	CV%	Mean ± SD	CV%	Mean ± SD	CV%
MH_IWNRZ_902	7.72 ± 0.20	(2.63)	54.50 ± 2.32	(4.26)	11.93 ± 0.48	(4.02)	3.29 ± 0.42	(12.74)	11.36 ± 0.65	(5.52)
MH_LARC_501	10.89 ± 0.83	(7.61)	44.57 ± 0.64	(1.43)	16.93 ± 1.48	(8.75)	2.34 ± 0.35	(14.88)	15.83 ± 1.45	(9.17)
MH_UOY_801	12.81 ± 0.89	(6.92)	43.00 ± 1.33	(3.08)	18.18 ± 0.34	(1.88)	2.19 ± 0.26	(11.94)	15.87 ± 1.09	(6.84)
MH_WU_111	8.03 ± 0.73	(9.05)	53.58 ± 2.67	(5.01)	12.08 ± 1.44	(11.92)	2.65 ± 0.20	(7.63)	11.68 ± 0.99	(8.48)
MH_WU_122	11.78 ± 0.71	(6.02)	43.95 ± 0.80	(1.83)	17.63 ± 0.71	(4.04)	2.28 ± 0.35	(15.47)	15.64 ± 1.62	(10.34)

Phenotypic values indicate the averages across the three environments. Promising accessions across all three locations are highlighted in bold. See **Table 3** for abbreviations.

Table 7. Summary statistics of hemp accessions with extreme phenotypes for 3 fibre traits.

Accession	BCD%		PH		PL	
	Mean ± SD	CV%	Mean ± SD	CV%	Mean ± SD	CV%
MH_CAAS_601	16.68 ± 4.42	(26.52)	14.62 ± 6.21	(42.46)	19.69 ± 3.90	(19.82)
MH_IWNRZ_902	41.50 ± 10.89	(26.24)	3.40 ± 0.60	(17.48)	8.69 ± 1.23	(14.14)
MH_LARC_501	21.74 ± 3.76	(17.32)	13.81 ± 2.70	(19.45)	17.41 ± 0.77	(4.39)
MH_UOY_801	17.89 ± *	(*)	*	(*)	*	(*)
MH_WU_111	40.70 ± 2.36	(5.80)	6.70 ± 1.68	(25.90)	12.76 ± 1.38	(10.81)

Phenotypic values indicate the averages across the three environments. Promising accessions across all three locations are highlighted in bold. (*) stands for missing value. See **Table 3** for abbreviations.

Table 8. Summary statistics of hemp accessions with extreme phenotypes for flowering traits.

Accession	FL_Begin		FL_Full		VEG		Sex_det	
	Mean ± SD	CV%	Mean ± SD	CV%	Mean ± SD	CV%	Mean	CV%
MH_CAAS_601	2057.40 ± 470.42	(22.87)	3428.50 ± 64.95	(1.89)	119.00 ± 35.68	(29.98)	1 ± 0.00	(0.00)
MH_CAAS_602	2071.60 ± 445.83	(21.52)	3442.53 ± 40.65	(1.18)	119.57 ± 34.70	(29.02)	1 ± 0.00	(0.00)
MH_CAAS_603	2087.42 ± 389.66	(18.67)	3440.87 ± 43.53	(1.26)	117.78 ± 30.76	(26.12)	1 ± 0.00	(0.00)
MH_CAAS_604	1928.48 ± 418.76	(21.71)	2663.01 ± 705.98	(26.51)	106.92 ± 33.36	(31.20)	1 ± 0.00	(0.00)
MH_CAAS_605	2090.60 ± 411.36	(19.68)	3440.87 ± 43.53	(1.26)	115.92 ± 31.00	(26.75)	1 ± 0.00	(0.00)
MH_CRA_415	784.87 ± 131.15	(16.71)	1059.09 ± 124.30	(11.74)	50.51 ± 12.30	(24.34)	2.44 ± 0.51	(20.83)
MH_IWNRZ_901	929.90 ± 190.08	(20.44)	1246.96 ± 129.77	(10.41)	57.95 ± 17.05	(29.41)	3 ± 0.00	(0.00)
MH_IWNRZ_902	997.02 ± 214.51	(21.52)	1293.58 ± 97.08	(7.50)	61.23 ± 18.01	(29.42)	3 ± 0.00	(0.00)
MH_IWNRZ_903	951.50 ± 146.35	(15.38)	1287.25 ± 31.05	(2.41)	58.50 ± 12.55	(21.46)	3 ± 0.00	(0.00)
MH_LARC_501	620.85 ± 236.13	(38.03)	840.68 ± 248.70	(29.58)	41.15 ± 20.45	(49.71)	1.22 ± 0.38	(31.49)
MH_UOY_801	469.03 ± 239.07	(50.97)	639.83 ± 204.32	(31.93)	32.63 ± 20.16	(61.78)	1 ± 0.00	(0.00)
MH_WU119	685.38 ± 347.01	(50.63)	965.48 ± 342.91	(35.52)	47.68 ± 29.53	(61.93)	1.83 ± 0.24	(12.86)
MH_WU_122	440.12 ± 165.34	(37.58)	615.45 ± 148.60	(24.14)	30.34 ± 14.38	(47.40)	1 ± 0.00	(0.00)

Phenotypic values indicate the averages across the three environments. Promising accessions across all three locations are highlighted in bold. See **Table 3** for abbreviations.

3.2 Elucidating the key components of fibre quality variability in the hemp panel

The ANOVA model highlighted significant differences ($p < 0.001$) for all variance components in all traits. In addition, as shown in **Figure 1** and **Table 9**, the random effects model revealed traits with phenotypic variations strongly influenced by the genetic component, other traits mostly influenced by the environment component and a reduced group of traits that was mainly influenced by the residual variance (ε%).

Traits with extensive influence of the genetic component (>40%) comprised: flowering traits, cell wall traits (including contents of monosaccharides derived from cellulose and hemicelluloses and lignin content) and the fibre trait BCD% (**Figure 1** and **Table 9**). The variation in flowering traits: FL_Begin, FL_Full, VEG and Sex_det showed genetic components of respectively 74%, 79%, 46% and 69%; the content of the monosaccharide from cellulose, Glc%dm 68% and the contents of monosaccharides from hemicelluloses were calculated respectively 55% (for Xyl%dm) and 42% (for Man%dm). The two measurements of lignin displayed genetic components of 44% (for ADL%dm) and 48% (for KL%dm), respectively. BCD% showed a genetic component of 67%. All these ten traits showed larger genetic component than *G*×E interactions. Ratios *G*×E/*G* close to zero were detected for all these traits. These results indicatedlarge stability of the accession ordering across environments. Consequently, all these traits displayed high H^2, ranging from 0.88 to 0.96, as detailed in **Table 9**. Furthermore, the contents of lignin (KL%dm) and VEG revealed a large influence of the environment component, respectively 34% and 42%. The other flowering traits (FL_Begin, FL_Full and Sex_det) showed small environment components, respectively 9, 5 and 3%. Glc%dm, Xyl%dm, Man%dm, ADL%dm and BCD% showed environment components ranging from 11 to 24% (**Table 9**).

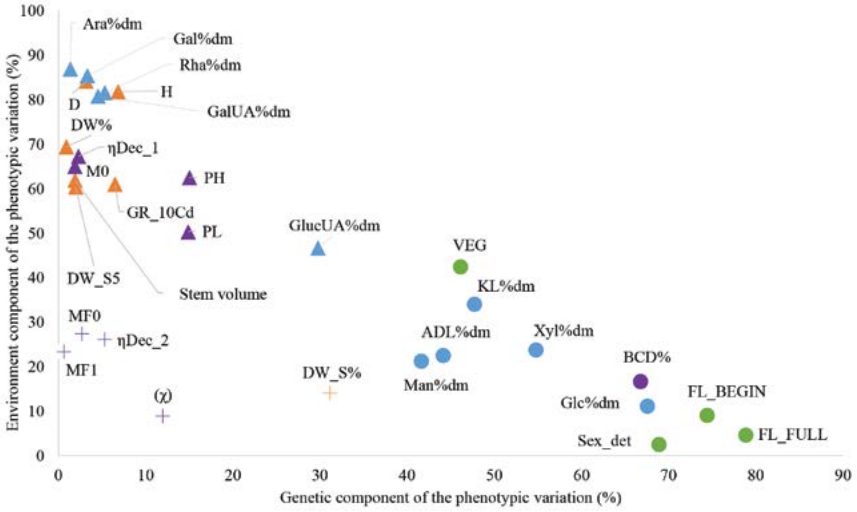


Figure 1. Scatterplot of the phenotypic variation of 30 traits explained by the percentage of genetic (x-axis) and environment (y-axis) components. The colour of the markers indicates the type of trait: orange, green, purple and blue represent agronomic parameters, flowering/sex stages, fibre measurements and cell wall traits respectively. The shape of the markers indicates the pattern of the phenotypic variation predominant in each trait: circle, triangle and cross represent genetic, environment and residual main component respectively. See **Table 3** for abbreviations.

Traits with large influence of the environment component (>30%) comprised: several agronomic traits, cell wall traits (e.g. the content of monosaccharides derived from pectin and GlcA%dm) and most fibre traits (**Figure 1** and **Table 9**). Agronomic traits (e.g. D, H, Stem_volume, GR_10Cd, DW(%) and DW_S5(g)) showed environment components larger than 60%, whereas the genetic components ranged from 1 to 7%. The composition of pectin, reflected by the contents of Ara%dm, Gal%dm, GalA%dm and Rha%dm was extensively influenced by the environment component (>80%) and only 1.3 to 5.3% by the genetic component. Glucuronic acid, a component of xylan (hemicellulose), expressed as GlcA%dm, showed a strong influence of the environment (47%) but also highlighted a substantial genetic component (30%). In addition, fibre traits such as, η_{Dec_1}, P_H, P_L and M_o showed environment components of respectively 67%, 62%, 50% and 65%. η_{Dec_1} and M_o showed low genetic components of 2% and fineness traits showed genetic components of 15%. The *G*×E/*G* ratios showed different performances in these agronomic measurements, fibre traits and pectin related monosaccharides. DW(%), Stem_volume, Ara%dm, M_o and η_{Dec_1} showed large *G*×E/*G* ratios ranging from 1.42 to 5.41. These traits showed large differences between accessions in environmental sensitivity, indicating alteration of the accession ordering across environments. These differences in sensitivity were reflected by low H^2, ranging from 0.21 to 0.45. In contrast, D, DW_S5(g), GR_10Cd, H, Gal%dm, GalA%dm, Rha%dm, P_H and P_L showed interaction ratios ranging from 0.32 to 0.85. These results may indicate that, despite the significant genetic component of the phenotypic variation, the small effects of

the genetic component in some traits (particularly D, DW_S5(g), GR_10Cd, H, Gal%dm, GalA%dm and Rha%dm) hampered the assessment of the *G*×E interactions and thus the ratios *G*×E/*G* are small. As a consequence, considering the definition of H^2 (Renaud et al., 2014), the ratios *G*×E/*G* below 1 can explain unexpected H^2 (Gitonga et al., 2014), ranging from 0.33 to 0.8, from traits with mostly environment component.

Dissection of the phenotypic variation revealed that a few fibre traits and one agronomic parameter showed large unexplained variance (ε%), as depicted in **Figure 1**. MF0, MF1, (χ), η_{Dec_2} and DW_S% showed residual variances of respectively 38%, 51%, 46%, 41% and 40% (**Table 9**). Furthermore, MF0, MF1, (χ) and η_{Dec_2} traits showed genetic components ranging from 1 to 12% and large *G*×E/*G* ratios, ranging from 1.61 to 30.70. As a consequence, low H^2 were detected for these traits.

Table 9. Variance components and broad-sense heritability (H^2) of 30 traits. The variances explained by each component are shown as the proportion of total variance (%). See **Table 3** for abbreviations.

Trait	Location (*E*%)	Block within Location (*B*%)	Genotype (*G*%)	Genotype x Location (*G*×*E*%)	Error (ε%)	Ratio *G*×*E*/*G*	H^2
D	84.13	1.72	3.16	1.00	9.98	0.32	0.69
DW(%)	69.41	4.43	0.89	2.70	22.57	3.03	0.21
DW_S(%)	14.08	1.45	31.15	13.01	40.30	0.42	0.78
DW_S5(g)	60.37	4.63	1.96	1.30	31.74	0.66	0.33
GR_10Cd	60.99	0.95	6.48	5.81	25.77	0.90	0.57
H	81.81	1.33	6.84	2.31	7.71	0.34	0.81
Stem_volume	61.88	3.91	1.89	4.30	28.01	2.28	0.29
ADL%dm	22.52	0.98	44.14	8.72	23.64	0.20	0.89
Ara%dm	86.87	0.59	1.35	1.92	9.26	1.42	0.45
Gal%dm	85.38	1.16	3.30	2.16	8.00	0.65	0.67
GalA%dm	80.77	0.38	4.54	2.88	11.44	0.63	0.67
Glc%dm	11.14	1.52	67.57	4.30	15.48	0.06	0.96
GlcA%dm	46.64	1.96	29.77	2.97	18.66	0.10	0.91
KL%dm	34.04	0.45	47.73	3.31	14.47	0.07	0.95
Man%dm	21.24	0.53	41.62	8.24	28.36	0.20	0.88
Rha%dm	81.54	0.66	5.31	2.71	9.78	0.51	0.73
Xyl%dm	23.77	0.29	54.79	4.86	16.29	0.09	0.94
(χ)	8.91	13.73	11.97	19.22	46.16	1.61	0.51
BCD%	16.66	0.89	66.79	5.36	10.29	0.08	0.96
M_o	64.98	0.00	1.88	7.60	25.54	4.04	0.26
MF0	27.39	9.67	2.68	22.59	37.66	8.43	0.19
MF1	23.37	6.51	0.61	18.73	50.78	30.70	0.05
PH	62.47	0.80	15.03	5.61	16.09	0.37	0.80
PL	50.25	1.18	14.89	12.59	21.09	0.85	0.69
ηDec_1	67.25	1.28	2.30	12.45	16.71	5.41	0.28
ηDec_2	26.14	11.10	5.29	16.61	40.87	3.14	0.34
FL_Begin	9.06	0.40	74.43	8.97	7.15	0.12	0.95
FL_Full	4.62	0.05	78.86	14.66	1.81	0.19	0.94
VEG	42.44	0.30	46.13	6.16	4.98	0.13	0.95
Sex_det	2.53	0.91	68.90	10.52	17.13	0.15	0.93

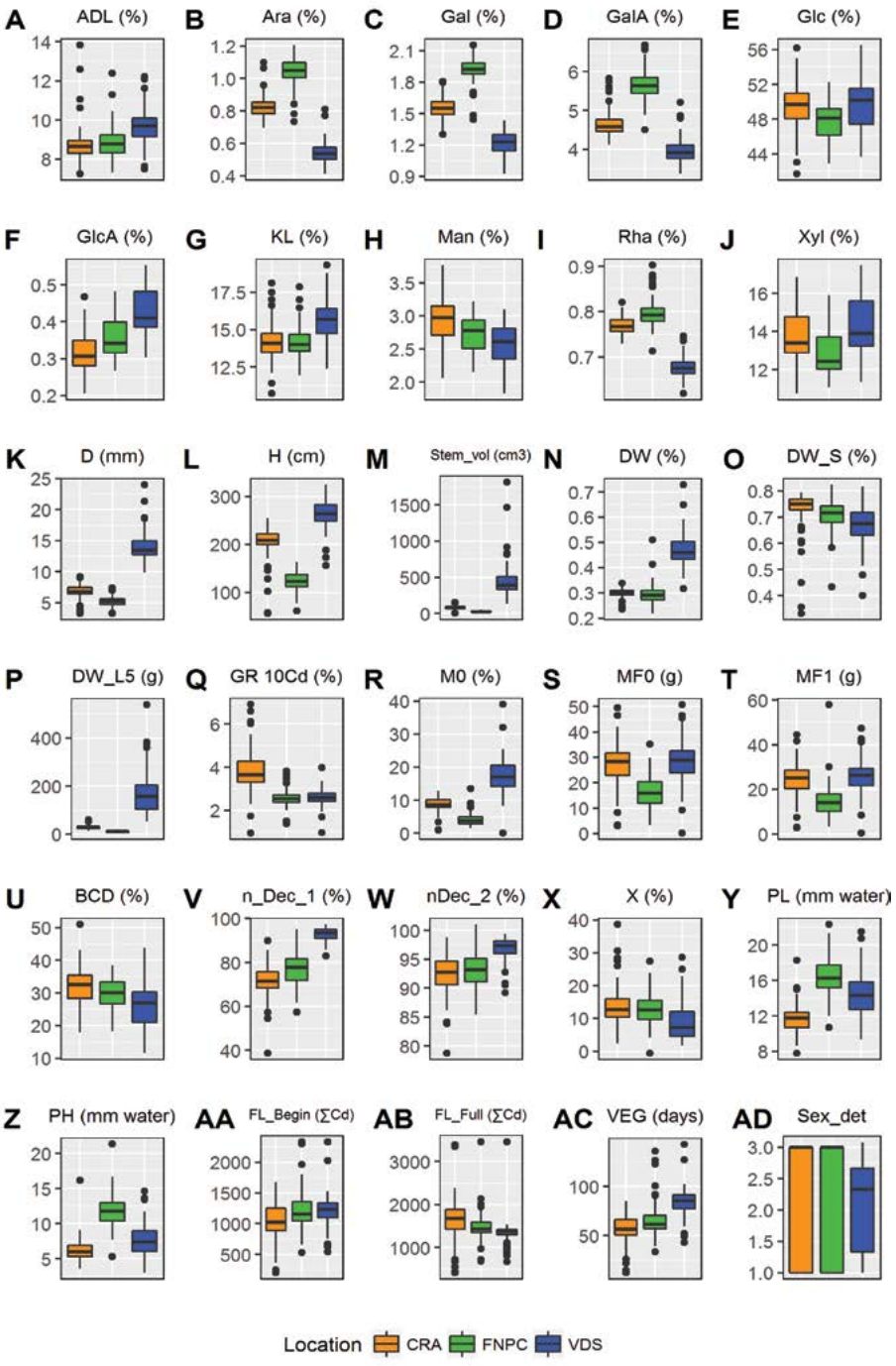


Figure 2. *Box*-plots summarizing the variation of a hemp panel for 30 fibre quality traits in three locations with contrasting environments. For every *box*-plot, the horizontal line represents the median of the trait, the box represents the interquartile range, the bars outside the box represent the extremes and the crosses indicate the outliers. In every panel, the x-axis indicates the location, as specified in the legend. See **Table 3** for abbreviations.

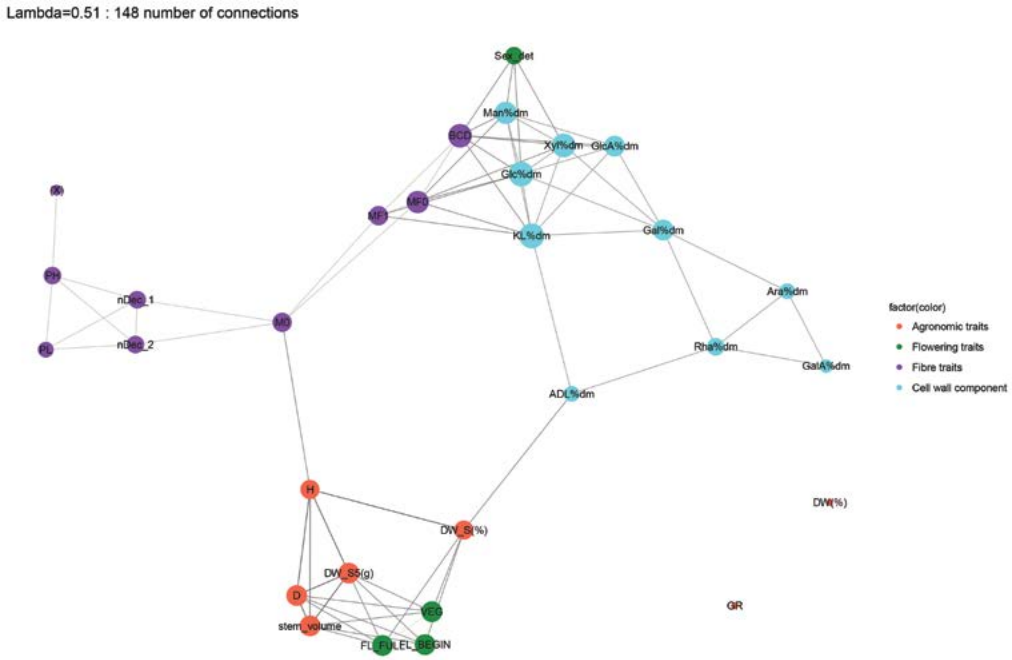


Figure 3. Network analysis based on graphical lasso of 30 hemp traits including agronomic parameters, flowering stages, fibre traits and cell wall components at λ = 0.51. See **Table 3** for abbreviations.

3.3 Large adaptive behaviour of hemp fibre quality under specific environments

The significant effect of the environment component of all traits suggested strong adaptability of hemp fibre quality to different environmental conditions. **Figure 2** shows environmental specific responses or adaptations of these traits in different locations. Plants grown in the Netherlands were quite different from plants grown in the other two locations. They produced larger biomass (DW(%), DW_S5(g), lignin content and GlcA%dm), thicker stems, taller plants and plants flowered later and over a shorter period than in the other locations. In addition, the decortication parameters (ηDec_1, ηDec_2 and (χ)) showed larger efficiencies in stems from plants grown in this location. In contrast, monosaccharides derived from pectin showed the largest contents in plants grown in France, while the lowest contents were found in plants grown in the Netherlands. Fineness properties followed the same pattern, as monosaccharides derived from pectin. Finally, plants grown in Italy flowered earlier and over a longer period of time.

3.4 Elucidating the relationships between traits relevant to hemp fibre quality

A network analysis was performed between the 30 traits to identify the most relevant correlations for fibre quality, as shown in **Figure 3** and **4**. To avoid relationships associated only to specific environments, correlation coefficients were calculated from the averaged phenotypic values across the three environments. The results revealed 148 main connections at $\lambda = 0.51$. The network highlighted that most connections were established between same type of traits. Furthermore, several important connections were identified between traits of different nature. Fibre traits, such as the amount of fibre bast (BCD%, MF0 and MF1) showed connections with cell wall components and with Sex_det. Most cell wall components also showed significant correlations with Sex_det. Only one connection between cell wall components and agronomic traits was detected (ADL%dm and DW_S(%)). In addition, flowering time traits (FL_Begin, FL_Full and VEG) showed significant relationships with most agronomic traits and ADL%dm.

The fibre trait BCD% showed strong correlations with cell wall components. BCD% was positively correlated with Glc%dm and Man%dm (respectively r^2=0.94 and r^2=0.82) and negatively correlated with Xyl%dm and GlcA%dm (respectively r^2=-0.91 and r^2=-0.87). In addition, the decortication trait (χ) showed a small but significant positive correlation with GalA%dm (r^2=0.33). The contents of lignin (KL%dm) and glucose highlighted the largest negative correlation (r^2=-0.93). Lignin measurement ADL%dm was negatively correlated with DW_S(%) and with flowering time traits. Flowering time traits were positively correlated with Stem_volume, D and DW_S5(g). Finally, sex determination was positively correlated with the contents of bast, Glc%dm and Man%dm, while it was negatively correlated with the contents of lignin (KL%dm), Xyl%dm and GlcA%dm, agronomic traits (D, Stem_volume) and flowering time traits. The positive correlations with sex determination were associated to monoecious accessions or accessions with a larger fraction of monoecious plants, while the negative correlations were associated to dioecious accessions or accessions with a larger fraction of dioecious plants.

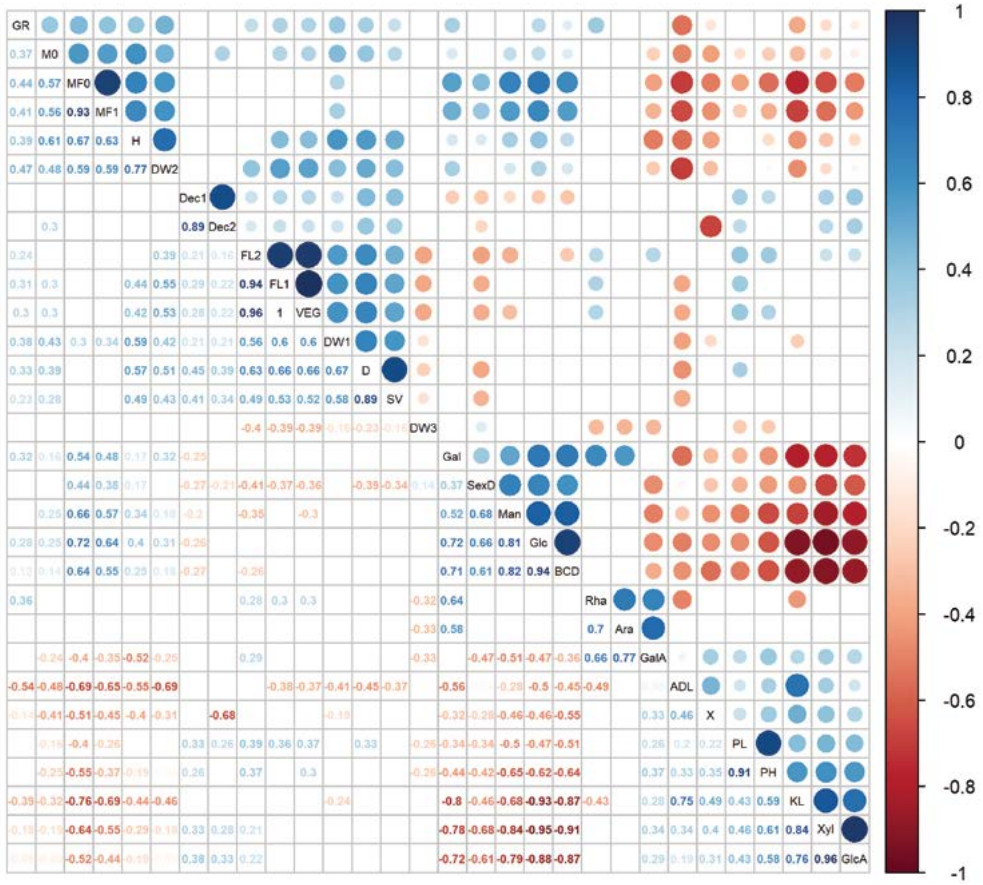


Figure 4. Correlation analysis between 30 agronomic measurements, flowering stages, fibre traits and cell wall components. Significant correlations were set at a confidence level of 0.95 and blank cells represent no significant correlations. The colour and the size of the dots are proportional to the level of significance. **Rha**= Rha%dm; **Ara**= Ara%dm; **Gal**= Gal%dm; **GalA**= GalA%dm; **Glc**= Glc%dm; **Man**= Man%dm; **Xyl**= Xyl%dm; **GlcA**= GlcA%dm; **KL**= KL%dm; **ADL**= ADL%dm; **FL1**= FL_Begin; **FL2**= FL_Full; **SexD**= Sex determination; **DW1**= DW_S5(g); **DW2**= DW_S(%); **DW3**= DW(%); **GR**= GR_10Cd; **SV**= Stem_volume; **BCD**= BCD%; **X**= (χ); **Dec1**= ηDec_1; **Dec2**= ηDec_2. See **Table 3** for abbreviations.

4. Discussion

4.1 Fibre quality traits are extensively variable and heritable but also adaptable to specific environments

In the present study, a panel of 123 hemp accessions was used to study the variability in hemp fibre quality and to enlighten key components of this variability. The analysis of the hemp panel revealed extensive variation in 30 fibre quality traits among the hemp accessions. In addition, some accessions displayed characteristics that are highly appreciated by the hemp industry. Such traits included high contents of bast fibre and cellulose, low contents of lignin and pectin, fine fibre bundles (high P_L and P_H values) and late flowering time (Ranalli, 2004;Salentijn et al., 2015). Accessions IWRNZ-902 and WU-111, exhibited several of these phenotypes and thus have a considerably higher quality fibre compared to many other accessions. These accessions indicate valuable germplasm to include in breeding programmes, aiming to improve hemp fibre quality.

We observed that important fibre quality traits have a large fraction of heritable phenotypic variation, as indicated by the large *G*% and H^2. As in the present study, extensive genetic studies have shown large heritabilities for cell wall components (cellulose, hemicellulose and lignin) in other fibre crops, such as poplar, eucalyptus, miscanthus, switchgrass and maize. Several studies reported large heritabilities for lignin and cellulose content in different species of poplar and eucalyptus, ranging from 0.4 to 0.9 (Raymond and Schimleck, 2002;Klasnja et al., 2003;Schimleck et al., 2004;Poke et al., 2006;Davis, 2008). Two other studies reported large heritabilities for contents of cellulose, hemicellulose and lignin in miscanthus, ranging from 0.4 to 0.72 (Slavov et al., 2014;van der Weijde et al., 2017). Studies on switchgrass also reported large heritabilities for biomass traits, ~0.6 (McLaughlin et al., 2006;Boe and K. Lee, 2007). Similarly, Torres and co-workers reported large heritability estimates for biomass traits in maize, >0.6 (Torres et al., 2015a). In addition, a previous study in hemp reported a large heritability for the bast fibre content of 0.84 (Hennink, 1994). Furthermore, similar heritabilities for flowering time were reported in several plant species (e.g. almond (reviewed in Sánchez-Pérez et al. (2014)), apricot (Campoy et al., 2011), adabidopsis (Sasaki et al., 2015), cotton (Kushanov et al., 2017), flax (Soto-Cerda et al., 2014;You et al., 2017) and rice (Takahashi et al., 2001;Huang et al., 2010)) ranging from 0.66 to 0.93. In addition, Hennink also reported in hemp a heritability estimate for date of flowering of ~1 (Hennink, 1994). It seems plausible that large fraction of the phenotypic variation of biomass and flowering traits might be controlled by highly "robust genetic systems", although they are highly complex and polygenic traits. This is because respectively ~4000 (Wang et al., 2012) and ~300 genes are estimated to be involved in cell wall synthesis and flowering in arabidopsis (Wang et al., 2012;Bouché et al., 2016). The robust genetic systems might work to control the performance of these traits so that

they are less sensitive to environmental differences. This guarantees important functions, such as fibre production and reproduction regardless of the environment.

The crucial functions of these traits might support such high heritabilities controlled by robust genetic mechanisms. From an evolution point of view, the cell wall performs a structural function in shaping the cells and consequently to plant bodies (Sarkar et al., 2009). Particularly, cellulose and lignin can withstand mechanical pressure exerted by the gravitational pull and the load of the plant body. Therefore, cellulose and lignin provide mechanical strength to the plant (Volkmann and Baluška, 2006). In addition, lignin provides protection functions against UV radiation and against pathogens (e.g. microbes, fungi and animals) that allowed plants to conquer terrestrial habitats (Iiyama et al., 1994;Popper et al., 2011). Hemicellulosic polysaccharides also provide structural rigidity to the cell walls (Pauly et al., 2013). Hemicellulose and lignin create a matrix around microfibrils of cellulose, which affects the recalcitrance of the cell walls (Torres et al., 2015a). Furthermore, flowering is an essential biological process for many plants, as the survival of the species depends on it (Mouradov et al., 2002). Consequently, these biomass and flowering traits seem to perform essential roles that cannot be widely modified. This is essentially because the consequences might be lethal for the plant.

Furthermore, we observed that hemp fibre quality traits are not strongly stable across environments. This is indicated by the significant genotype-by-environment (*G*×E) interactions. The results showed that some accessions were more sensitive in some environments than in others. It seems likely that the phenotypic variation associated to the *G*×E interactions might be controlled by "plastic genetic systems", where certain genes are expressed when combined with specific environmental conditions. This is the first study describing significant *G*×E interactions in fibre and flowering traits in hemp. Studying these interactions is important owed to their implications for the set-up of selection experiments. This is because the ranking of accessions is dependent on the environment (van der Weijde et al., 2017). Selection for traits with large *G*×E interactions in breeding programmes might lead to biased selection decisions, owed to the unknown effects in untested environments. Nonetheless, *G*×E interactions were much small in most important cell wall components (e.g. contents of glucose, mannose, xylose, ADL and KL), fibre content and flowering traits in hemp. Similar small *G*×E interactions have been reported for biomass traits, especially contents of cellulose, hemicellulose and lignin, in several fibre crops (e.g. alfalfa (Sheaffer et al., 1998), maize (Dolstra et al., 1992;Cox et al., 1994;Argillier et al., 1997;Barrière et al., 2008;Torres et al., 2015a), miscanthus (van der Weijde et al., 2017) and switchgrass (Hopkins et al., 1995)). In addition, Sasaki and co-workers also reported small *G*×E interactions in flowering time in arabidopsis (Sasaki et al., 2015). Therefore, the extent of the *G*×E interactions on these fibre quality related traits might not strongly affect the ordering of hemp accessions across environments and might not interfere in selection decisions.

Interesting examples of large *G*×E interactions in hemp are the contents of some pectin related monosaccharides. The evolution and the functions of pectin in plants might explain these results. In the stem of plants, pectins are mostly present in the middle lamella between cells and are involved in the intracellular adhesion, providing integrity and rigidity to plant tissues and to the stem. They also play important roles in the defence mechanisms against pathogens. In addition, they are involved in the regulation of the ion transport and in the water holding capacity (Voragen et al., 2009). Yet, pectic polysaccharides are highly dynamic structures and their content dramatically changes across tissues and plant species (Willats et al., 2001). Pectin has almost disappeared in the stems from several modern plants, such as grasses (Carpita and Gibeaut, 1993;Carpita, 1996;Sarkar et al., 2009;Voragen et al., 2009). This suggests that important functions of pectin might be evolutionary replaced by other cell wall components. Lignin is the newest cell wall component to appear in plants and has some parallel functions with pectin, such as structural support and defence functions (Sarkar et al., 2009). As a consequence, dramatic changes of pectin content might not be lethal to plants, owed to a putative partial compensation from other cell wall components. This might allow larger plasticity in sensitivity of certain accessions in different environments. In addition, further research to understand the *G*×E interactions might get insights into the evolution of pectin in plants. Several genetic studies suggested that the *G*×E interactions are crucial in the genetic local adaptations (Via and Lande, 1985;Fournier-Level et al., 2011;Sasaki et al., 2015). This is because these interactions account for the genotypic variation in the phenotypic response to the environment (El-Soda et al., 2014).

Fibre quality related traits were strongly influenced by the differences of the environments across the trial locations. Previous studies have reported large sensitivity of hemp to the environment, particularly to the photoperiod and temperature regimes, that affected the vegetative growth and flowering of the plants (Faux et al., 2013;Amaducci et al., 2015;Sawler et al., 2015). This large sensitivity can be understood as a strong general response of hemp accessions to adapt to the environment, independently of the heritable genetic control of the traits (described in previous paragraphs). This behaviour might be the result of the optimization of the plant fitness under specific environmental conditions. Consequently, the environment of the growing locations should be taken into account when selecting the cultivation purpose of hemp. Consequently, breeding should be done for use in a specific environment.

Examples of the adaptive behaviour of hemp are the differences in biomass production and flowering time across environments. Plants of the same accession grown in the Netherlands produced larger biomass and flowered later than plants grown in the other two locations. These variations can be explained by differences in photoperiod regimes across locations. Hemp is a short-day plant and the length of the vegetative growth period depends on the shift from long- to short-day photoperiod regimes (Amaducci et al., 2012). The vegetative growth period is characterized by biomass production, after which this behaviour shifts towards

fibre maturation (secondary fibre formation and lignification) and plant reproduction during flowering development (van der Werf and Turunen, 2008;Liu et al., 2015). At northern latitudes the shift from long-day towards short-day photoperiod regimes occurs later. Therefore, the critical photoperiod for flowering (~14 – 16 hours) is reached later in the growing season of hemp (Struik et al., 2000;Hall et al., 2012). As a result, the cultivation of hemp focused on the production of fibres may be better in northern latitudes. Meanwhile, the cultivation of hemp for seeds or dual purpose seed/fibre in southern latitudes may be more adapted to the environment (Amaducci et al., 2015). The selection of the cultivation purposes based on the environment might increase the profitability of hemp cultivation. This could complement the high fibre quality achieved by the breeding programmes, using the heritable phenotypic variation.

Another example of the adaptive behaviour of hemp biomass is the difference in production of monosaccharides derived from pectin from plants across environments. Plants grown in France showed the largest content of monosaccharides derived from pectin, while plants of the same accessions in the Netherlands showed the lowest contents. It has been previously reported that pectin plays a role in modulating cell wall architecture in response to low availability of water (reviewed in Le Gall et al. (2015)). This is essentially because of the water holding capacity function of pectin (Voragen et al., 2009). As shown in **Table 2**, France had lower rainfall and larger days without rain than the Netherlands. Based on the environmental differences between the locations and the functions of pectin, it seems plausible to hypothesize that hemp plants may increase the content of pectin in the stem, partially as a response to the changes in water availability. This relationship may have important implications in the improvement of the fibre quality of hemp. This is because the contents of monosaccharides derived from pectin are poorly heritable traits and pectin plays a key role in the fibre quality. Pectin has been associated to difficulties in decortication, which results in increased fibre damage (Ranalli, 2004;Müssig et al., 2008;Salentijn et al., 2015) after fibre decortication. Furthermore, plants grown in the Netherlands showed larger decortication efficiencies than in the other two locations. In addition, the content of galacturonic acid, the main component of pectin (Willats et al., 2001), was positively correlated with the shive content after decortication (χ). These results indicate that lower contents of galacturonic acid in the stems results in easier decortication. Moreover, it suggests that water availability may play a role in the pectin content. If that is the case, this knowledge could be used to improve fibre quality associated to poor heritable traits. Crop management, such as irrigation could contribute to decrease the pectin content and thus improve the quality. The present study provides interesting results to further investigate the influence of water availability on pectin content of hemp. The use of a wider range of locations (contrasting and more detailed environmental conditions), and the use of controlled experiments may be useful to get insights into the role of specific environmental factors in hemp fibre quality.

The extensive phenotypic variability of the 30 traits in hemp enables the mapping of relationships between several traits in the context of fibre quality. The identification of such correlations is of great relevance for the development of breeding programmes for hemp fibre quality. The correlations between flowering time traits, agronomic measurements and the content of lignin (ADL%dm) suggest significant implications in the fibre quality. Accessions with long vegetative growth period (VEG) and late flowering time showed larger plant vigour, larger biomass accumulation and lower content of lignin, which is in accordance with the lignification of stems that intensifies with flowering (Liu et al., 2015). As a result, breeding for several independent traits (e.g. content of glucose (cellulose) and flowering time traits) may have trade-offs to other traits that may complementarily increase the hemp fibre quality. Furthermore, sex determination revealed significant relationships with cell wall components and fibre traits. These results may be explained by the sexual dimorphism of hemp (Mandolino and Carboni, 2004;Amaducci and Gusovius, 2010;Faux et al., 2013;Faux et al., 2014;Faux et al., 2016). Monoecious accessions were related to high contents of glucose, mannose, amount of fibre (BCD%, MF0 and MF1) and early flowering. Meanwhile dioecious plants were related to large plant vigour, high content of lignin (KL%dm) and xylans (Xyl%dm and GlcA%dm) and late flowering. The results of the hemp panel analysis confirm that monoecious accessions have larger fibre qualities than dioecious. These differences may be explained by the larger uniformity in fibre production common in monoecious accessions compared to dioecious accessions (Mandolino and Carboni, 2004;Salentijn et al., 2015). Finally, the sex determination of hemp is another key element that should be taken into account when selecting the germplasm for breeding programmes, as it has important implications in fibre quality (Amaducci and Gusovius, 2010;Amaducci et al., 2015).

4.2 Implications of the fibre quality variability in the development of new hemp cultivars with improved fibre properties

In the present study, the extent of fibre quality variation amongst accessions reveal a good hemp panel. This panel can be used to further study the genetic architecture of fibre quality, flowering and sex characteristics of hemp. As Davis described in a previous study, in order to genetically improve some traits, they must be heritable (Davis, 2008). The contents of most cell wall components (glucose, mannose, xylose, glucuronic acid and lignin (ADL and KL)), content of bast fibre, flowering time measurements and sex determination of hemp have high heritabilities. This means that these traits are good candidates for genetic association studies. The selection of plants harbouring favourable alleles for these traits would maximize the genetic gains expected from the breeding programmes, that aim to increase the quality of the bast fibre. However, traits with low genetic components and relatively large *G*×E interactions, such as monosaccharides derived from pectin, are not appropriate candidates for mapping studies. This is because the statistical tools currently

available have low power when the genetic components are small. As a result, the statistical power of the association for these traits would drop, leading to high false-positives and/or false-negative rates (Huang and Han, 2014;Bernardo, 2016). Yet, the large adaptive behaviour of these traits suggest that crop management practices may be a good alternative to breeding for traits poorly heritable. Therefore, the combination of breeding programmes for traits with large genetic components and crop management for traits with small genetic component may be a good strategy to improve the potential of hemp as a high yielding, sustainable crop of excellent fibres.

5. Conclusions

The results of this study prescribe positive prospects for the development of new hemp cultivars with improved fibre quality properties. In particular, the hemp accession panel reveal to be a good dataset for mapping studies. This is because of the extensive phenotypic variability of 30 fibre quality traits. The content of most cell wall components (cellulose, hemicellulose and lignin), bast fibre content, flowering time traits and sex determination showed large heritable variationthat can be used in breeding programmes. In addition, all traits showed statistically significant genotype-by-environment interactions, in different percentages depending on the traits. These results suggest that the phenotypic variation in fibre quality of hemp has a fraction of heritable variation sensitive to the environment. These *G*×E interactions might play a crucial role in the genetic local adaptation of the plants.

In addition, fibre quality traits were strongly affected by the environment, such as photoperiod and temperature regimes and probably water availability. These sensitivities can be understood as adaptations to the environment, independently of the heritable genetic variation. The adaptive behaviour of poorly heritable traits such as pectin, might be used to develop strategies(e.g crop management practices) to increase fibre quality alternatively to breeding programmes. Furthermore, the relationships between traits indicate that breeding for one trait may have potential trade-offs in other traits, which may be beneficial for the final hemp product. Finally, the correlation analysis revealed that monoecious have larger fibre quality than dioecious hemp. This is probably because of the uniformity in fibre production common in monoecious accessions. Altogether, we advocate for novel hemp breeding programmes that breed for highly heritable traits, takes into account the environmental sensitivity of fibre quality, considers the sex determination of the germplasm in the breeding schemes and also consider potential trade-offs between traits.

6. Acknowledgements

This work was conducted as part of the MultiHemp project (Multipurpose hemp for industrial bioproducts and biomass) funded by the European Union's Seventh Framework Programme for research, technological developments and demonstration under grant agreement number 311849. The authors are grateful to Annemarie Dechesne, Wen Fang, Dominik Ewald, Wei Xiong and Li Gu from WUR for their technical assistance during laboratory analysis. We also gratefully acknowledge the help of Salvatore Musio from UCSC, Carsten Lühr from ATB and Felix Beck, Christopher Graf, Kevin Janke, Niklas Lehmensieck, Anne-Kristin Lenz, Sebastian Mulot, Fabian Plum, Simon Riechel, Tobias Riphaus, Fabian Uhrlaub and Timo Wunsch from HSB for their help during harvest of the field trials and decortication of stems.

Chapter 4

Elucidating the genetic architecture of fibre quality in hemp (*Cannabis sativa* L.) using a Genome-Wide Association Study

Jordi Petit[a], Elma M. J. Salentijn[a], Maria-João Paulo[b], Christel Denneboom[a], Eibertus N. van Loo[a], Luisa M. Trindade[a*]

[a] Wageningen UR Plant Breeding, Wageningen University and Research, P.O. Box 386, 6700 AJ Wageningen, The Netherlands
[b] Biometris, Wageningen University and Research, P.O. Box 100, 6700 AJ Wageningen, The Netherlands
[*] Corresponding author: Luisa M. Trindade (luisa.trindade@wur.nl)

To be submitted as:
Jordi Petit, Elma M. J. Salentijn, Maria-João Paulo, Christel Denneboom, Eibertus N. van Loo and Luisa M. Trindade. Elucidating the genetic architecture of fibre quality in hemp (*Cannabis sativa* L.) using a Genome-Wide Association Study

ABSTRACT

Hemp (*Cannabis sativa* L.) is a bast-fibre crop with a great potential in the emergent bio-based economy. Yet, breeding of hemp for fibre quality is restricted and that is mainly due to the limited knowledge of the genetic architecture of its fibre quality. To better understand the genetic components determining the quality of fibre, a panel of 123 hemp accessions, with large phenotypic variability, was used to study the genetic basis of seven cell wall and bast fibre traits relevant to fibre quality. These traits have large genetic variance components and broad-sense heritabilities in this hemp panel, as concluded from the phenotypic evaluation across three test locations with contrasting environments. The hemp panel was genotyped using Restriction site Associated DNA Sequencing (RAD-seq). Subsequently, a large set (>600.000) of selected genome-wide SNP markers was used for a Genome-Wide Association Study (GWAS) approach to investigate the large phenotypic variation of the hemp panel and to get insights into quantitative trait loci (QTL) controlling fibre quality traits. In absence of a complete hemp genome sequence, identification of QTLs was based on the following characteristics: 1) association level to traits, 2) fraction of explained trait variance, 3) collinearity between QTLs and 4) detection across different environments. Using this approach, 16 QTLs were identified across locations for different fibre quality traits, including contents of glucose, glucuronic acid, mannose, xylose, lignin and bast fibre content. Among them, six were found across the three environments. The markers composing the QTLs across locations are useful tools to develop novel genotypes of hemp with improved fibre quality. Twelve putative candidate genes for the biosynthesis and modification of monosaccharides, polysaccharides and lignin were detected in direct vicinity of QTLs. These candidate genes were suggested to be involved in the regulation of fibre quality in hemp. This study provides new insights into the genetic architecture of fibre traits, identifies QTLs and putative candidate genes that form the basis for breeding for fibre quality traits in hemp.

1. Introduction

Hemp (*Cannabis sativa* L.) is a bast-fibre crop with a great potential in the emergent bio-based economy. It is an environmentally friendly crop that fits well into crop rotation scenarios and sustainable agriculture. Besides, it can also be used for bio-remediation purposes of polluted lands (Struik et al., 2000;Amaducci et al., 2008b;van den Broeck et al., 2008;Amaducci and Gusovius, 2010). This multi-purpose crop is, apart from being a valuable source of cannabinoids and oils (Salentijn et al., 2015), an alternative source of fibres relative to poor sustainable crops, (e.g. cotton) and to non-renewable glass and fossil-based fibres (Ebskamp, 2002;van der Werf and Turunen, 2008;Amaducci and Gusovius, 2010). Despite the large interest in hemp, it is a relatively poorly developed crop. This is because of the strong decline in hemp production in the last century, when intensive breeding programmes drove great improvements in major staple crops (Salentijn et al., 2015). Consequently, hemp breeding was restricted and to date little is known about the genetic architecture that underlies hemp fibre quality.

Recently, large phenotypic variability in 30 traits related to fibre quality of hemp has been reported in a diverse panel of 123 hemp accessions native to different regions in the world, and cultivated in three test locations with contrasting environments (Petit et al., 2019c). The study reported that fibre quality in hemp is a clear example of quantitative trait, in which cumulative effects of genetic and environmental factors, as well as genotype and environment interactions (*G*×E), play important roles. Yet, the mode of regulation of the different traits was found to differ, ranging from traits with large genetic components, large heritabilities and low *G*×E interactions to traits that are largely controlled by environmental components, with small contributions of genetic factors.

Traits appreciated by the hemp industry are associated to fibres with high quality properties. These traits include high bast fibre content, high content of cellulose and low contents of lignin and pectin in the cell walls, fine bast fibres and efficient decortication properties of the stems (Ranalli, 2004;Salentijn et al., 2015). For the contents of bast fibre, cellulose and lignin large genetic components were described in the hemp accession panel from Petit et al. (2019c). This suggests that the panel is a promising dataset to further study the genetic basis of these traits. In contrast, the variation of pectin, bast fibre fineness and decortication efficiency were only to a low level controlled by genetic factors, which would hamper the study of their genetic components. Nonetheless, Petit and co-workers suggested agronomic practices to control traits with low genetic components, as they showed adaptive behaviour under certain environmental conditions (Petit et al., 2019c).

Despite the large variability of hemp fibre quality and the large influence of genetics on important traits of fibre quality, the genetic mechanisms controlling these traits remain mostly unknown in hemp. Different genetic approaches can be performed to study its genetic basis, such as reverse genetics (candidate gene approaches) and forward genetics

(genetic mapping studies). Reverse approaches determine the function of a specific candidate gene by analysing the phenotypic effect associated to its altered gene sequence or expression (Zhu and Zhao, 2007). The identification of candidate genes can be based on mapping studies, or on transgenic approaches to knock out genes. Since not much of such knowledge is available for hemp, putative candidate genes can be identified by studying the available functional knowledge of genes controlling a trait in other plant species. Salentijn and colleagues reviewed the state of art of candidate gene studies for fibre quality in hemp. Briefly, most of the reported genes have a function in the lignin metabolism or code for phytohormones involved in plant development with a possible effect on lignin (Salentijn et al., 2019). However, many different metabolic pathways are involved in the biosynthesis and regulation of the cell walls. As a consequence, the function of an altered gene might be replaced by another one, whereby the genes are functionally redundant (Boerjan et al., 2003;Pauly et al., 2013;Kumar et al., 2016). Forward genetic approaches determine the genetic basis associated to a trait by studying the variation of the phenotype and link this to specific genetic variation (Zhou and Stephens, 2012). Forward approaches do not focus on a single gene but on the complexity of loci associated to the trait to identify quantitative trait loci (QTLs). Therefore, forward genetics would be a good approach to get insights into the genetic architecture of fibre quality, which is a complex trait, governed by multiple genetic loci. To date, only three genetic association studies have been performed in hemp and they focused on sex expression and cannabinoid content (Faux et al., 2016;Grassa et al., 2018;Laverty et al., 2019). These studies were based on biparental mapping population approaches to detect QTLs in a genetic map. Despite the interest, to the best of our knowledge, no association studies for fibre quality of hemp and no Genome-Wide Association Studies (GWAS) have been reported for hemp. The lack of a complete genome sequence (Sawler et al., 2015;Laverty et al., 2019) and the lack of panels of hemp accessions harbouring large variability in fibre quality have hampered such studies on this crop (Sawler et al., 2015). As a result, no QTLs have been reported for hemp fibre quality. Furthermore, hemp breeding programmes are currently based on conventional breeding, while molecular breeding has not yet been developed in hemp (Salentijn et al., 2015). Molecular breeding would accelerate the development of new cultivars with improved fibre properties. Marker assisted selection would allow a selection of promising individuals at early developmental stages and would decrease the phenotyping costs of the breeding programmes. Genetic association studies are thus of great value to upgrade breeding programmes toward molecular approaches.

Insights in the genetic architecture of hemp fibre quality are essential to develop new breeding strategies, which seek for novel hemp cultivars with improved fibre properties. The objective of this study was to identify QTLs associated to fibre quality. The highly variable panel of 123 diploid hemp accessions (2n=20) was used for this purpose (Petit et al., 2019c). GWAS analyses were performed for fibre quality traits, with a large genetic components. The study was performed in three test locations with contrasting environments.

2. Materials and methods

2.1 Plant material

A panel of 123 hemp accessions was used in this study (**Table 1**, **Chapter 3**) (Petit et al., 2019c). The panel included mostly fibre type accessions but also an oil seed cultivar, landraces and breeding material. The panel originated from different countries in Europe, Asia (China) and North America (Canada). Plants were cultivated in three locations across Europe to assess different environments: Rovigo (CRA - Centro di ricerca cerealicoltura e colture industriale, Italy, 45˚N 11˚E), Chèvrenolles, Neuville-sur-Sarthe (FNPC - Fédération Nationale des Producteurs de Chanvre, France, 48˚N 0.2˚E), and Westerlee (VDS - VanDinter Semo BV, The Netherlands, 53˚N 6˚E). The multi-location trial was established between April and September of 2013. Each location had a randomized complete block design with three biological replicates per accession. The experimental units were plots of 1m² in Italy and the Netherlands and of 1.5m² in France.

2.2 Phenotyping and data collection

Phenotyping of cell wall traits was performed essentially as described in Petit et al. (2019a) and in Petit et al. (2019c). Briefly, six cell wall traits (contents of glucose (Glc%dm), glucuronic acid (GlcA%dm), mannose (Man%dm), xylose (Xyl%dm), acid detergent lignin (ADL%dm) and Klason lignin (KL%dm)) and one fibre parameter (BCD%) were measured after harvesting the stems. Plants were cut from the based (10 cm above the ground) when the accumulative temperature degree days (Σ˚C, the accumulated Celsius degree day over a period at a base temperature of 1˚C) were 1740.25Σ˚C, 1421.1Σ˚C, and 1843.3Σ˚C in CRA, FNPC and VDS, respectively. These temperatures corresponded to the Σ˚C when most accessions reached full flowering in each location. The biochemical composition of the cell walls from stems of the 123 accessions was measured with multivariate prediction models based on near-infrared spectroscopy (NIRS). Data were predicted for each accession in each block and in each location. The bast fibre content after decortication (BCD%) was measured as an average of 10 stems per plot after warm water retting and decortication, using a lab-scale roller-breaker decortication system, according to Wang et al. (2018) and Petit et al. (2019c).

2.3 DNA extraction

Genomic DNA was isolated from young grinded hemp leaves (~20 to 400 mg, lyophilized material) using a CTAB method (Doyle and Doyle, 1987) with additional steps to remove proteins, polysaccharides and RNA. First, the leave material was treated with proteinase K in 500µl TES buffer (100 mM TRIS pH=8.0, 10 mM EDTA pH=8.0, 2% SDS, 200µg/ml proteinase K) for 1 hour at 60°C. Thereafter, the CTAB method was performed using a final

concentration of 1% CTAB and 1.4N NaCl. Polysaccharides were removed by incubation for 1 hour on ice in 1.2M NH4Ac, centrifugation for 10 minutes at 13.000 rpm and discarding the pellet. The supernatant was then treated with RNAseA (100µg RNAseA per 700µl supernatant, 30 minutes at 37°C) and extracted with chloroform:IAA. The DNA was precipitated and dissolved in 550µl ultra-pure water. The CTAB method was repeated on this sample and the resulting DNA was suspended in 30 µl ultra-pure water. DNA samples were further purified over a column (Genomic DNA Clean and concentrator-10, Zymo Research) and controlled for quality and DNA concentration on agarose gel and by Qubit™ Fluorometric quantitation to provide high quality genomic DNA for massive sequencing.

Hemp is an outcrossing species and each accession used in the GWAS panel might have a certain degree of genetic heterogeneity, despite being phenotypically homogeneous (Sawler et al., 2015). Therefore, to cover all allelic variation within accessions, the genomic DNA of eight individual plants per accession was isolated and pooled in equimolar amounts, resulting in 123 samples for genotyping by RAD-seq.

2.4 Restriction-site-associated DNA sequencing (RAD-seq)

The GWAS mapping panel was sequenced with RAD-seq to identify Single Nucleotide Polymorphism (SNP) distributed over the genome to be used as molecular markers. High quality genomic DNA (2.5 to 5 µg at a concentration ≥ 25ng/µl) was digested using the restriction enzyme *Eco*RI. Then, RAD libraries with insert sizes of 300 to 550bp, were prepared for each sample, as described by Baird et al. (2008). The 123 samples were paired end sequenced on an Illumina platform (PE150) in two rounds to provide 2 x 1 Gbp genomic data per sample. RAD library preparation and sequencing were performed by Beijing Genomics Institute (BGI, Hongkong).

2.5 RAD-seq data analysis

Adaptors from the sequences were trimmed and low quality reads were removed. Low quality reads comprised reads with >50% of the bases Q≤12, unknown bases >3%, reads that lack a part of the multiplexing barcode, and could not be identified, and reads lacking the key sequence of the enzyme used. The clean sequence reads of each sample were mapped to the *C. sativa* 'Purple Kush' (canSat3 version GCA_000230575.1) genome reference (van Bakel et al., 2011), using Burrows-Wheeler Alignment Tool (Li and Durbin, 2009) (specific BWA parameters: (o) max number of or fraction of gap opens =1; (e) max number of or fraction of gap extensions=50; (m) maximum entries in the queue=100000). Picard-tools (v1.118) was used to sort the Sequence Alignment Map (SAM) files by coordinate and convert them to Binary Alignment Map (BAM) files and to mark duplicate reads. The average mapping rate was 55.54% (range 50.3% to 85.7%). Subsequently, SOAPsnp was used to call SNPs in each sample (Li et al., 2009).

2.6 SNP marker selection

Samples for each accession consisted of pools of eight diploid plants to cover all allelic variation present in the accessions. Each sample therefore harbours DNA from 16 alleles, represented by mostly expected 2 different nucleotides but occasionally 3 and rarely 4 different nucleotides (A, G, C and T) at a position. For each polymorphic site, the possible allele frequencies (%A, %G, %C and %T) were calculated per accession and in the GWAS panel.

Quality SNP marker selection was performed based on a 100% call rate of the SNPs in the 123 hemp accessions. Markers with a minor allele frequency below 2% and with a major allele frequency above 98% in the mapping panel were removed. Only biallelic markers were selected, the frequency sum of the two major alleles was equal or above 95%. In addition, to ensure allelic variation in the mapping panel, the markers with a standard deviation in the frequency of the major allele below 0.1 were removed. In total a set of 612.452 SNPs was selected for the genetic analysis. Each SNP was scored as the proportion of the major allele in the pool sample of plants from the same accession. Quality SNP marker selection was performed in R version 3.4.3 statistical software.

2.7 Analysis of population structure

A kinship matrix was used to study the genetic relatedness between the accessions. The VanRaden (2008) method, following the same approach as in Kruijer et al. (2015), was used to calculate the kinship matrix by using the entire set of selected markers. To investigate the patterns of population structure, a Principal Coordinate Analysis (PCoA) of the kinship matrix was performed using Genstat 19th edition. A dendrogram of the kinship was also calculated in R version 3.4.3 statistical software, using *APE* package (Paradis et al., 2004). In addition, clusters of accessions inferred from the kinship matrix were further analysed to study the level of population structure in the GWAS panel. Pairwise comparisons of coefficient of population differentiation (*Fst*) between clusters were calculated using the Wright's formula (Wright, 1969):

$$F_{ST} = \frac{\bar{p}(1-\bar{p}) - \sum c_i p_i (1-p_i)}{\bar{p}(1-\bar{p})} = \frac{\bar{p}(1-\bar{p}) - p(1-p)}{\bar{p}(1-\bar{p})}, \quad (1)$$

where the allele frequency in the i^{th} population is p_i, the relative size of the i^{th} population is c_i and the average allele frequency between the two populations is $\bar{p}$.

2.8 Analysis of pairwise correlation between markers

The identification of the boundaries of QTLs with a fragmented genome sequence in scaffolds is difficult. The study of the pairwise correlation between markers in certain genomic regions and by mapping groups of collinear markers (MultiQTL modelling, see section 2.9) can overcome this limitation.

To estimate the threshold to discriminate between collinear and non-collinear markers, a study of pairwise correlation-distance between markers from the same scaffold was performed. The study was performed on the 24 largest scaffolds (~130kbp – ~600kbp) of the cannabis genome canSat3 version GCA_000230575.1 (van Bakel et al., 2011), that were covered with SNP markers. The analysis used the marker allele frequencies of the 123 hemp accessions to calculate the squared-allele frequency correlation (r^2) between pairs of bi-allelic markers. The correlation values (r^2) were compared with the physical distance between the marker pairs. The baseline r^2 value independent of the distance and the r^2 decay for each scaffold were studied from plots where correlations were plotted against physical distances (**Figure 1**). Correlation analyses were performed using Genstat 19th edition and plots were generated in R version 3.4.3 statistical software.

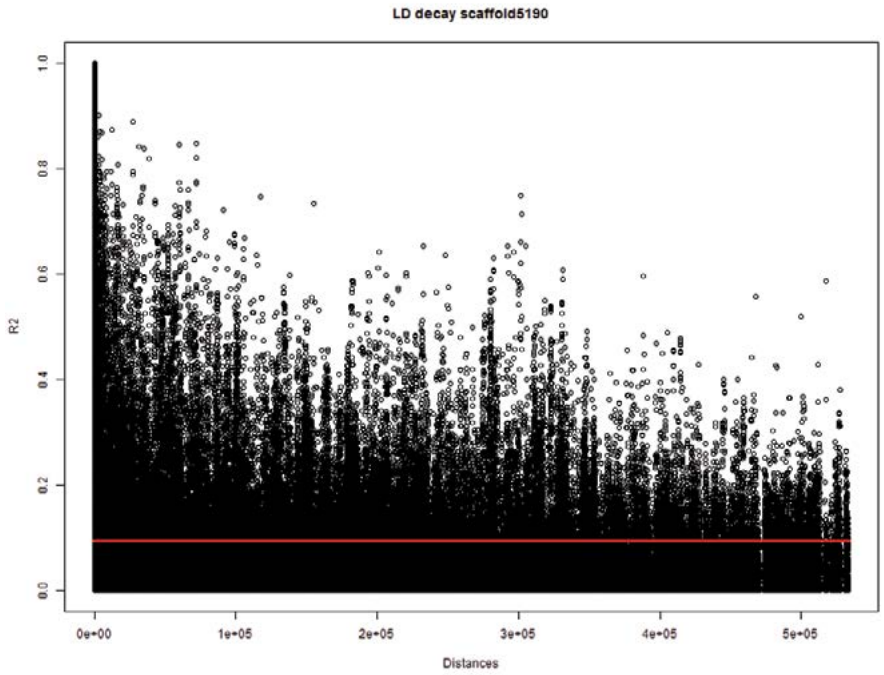


Figure 1. Estimation of the linkage disequilibrium decay of the largest scaffold (scaffold5190; 565.9 kbp) covered with SNP markers. The X-axis indicates the physical distances and the Y-axis indicates the pairwise correlations (r^2) between all pairs of markers from the scaffold. Red line indicates the critical value of r^2.

In most tested genomic scaffolds, the r^2 decayed with the distance. Nevertheless, a baseline value of $r^2 \approx 0.1$ was detected in all scaffolds independently of the distance between markers. This baseline value of r^2 was set as a threshold to identify non-random associated markers. When markers non affected by population structure and were correlated with r^2 equal or above 0.1 ($\sim r \geq 0.3$) they were hypothesized to be physically close on the same chromosome

and therefore $r^2 \geq 0.1$ was used as threshold for collinearity of markers in the multiQTL models (see section 2.9). For those markers involved in genetic relatedness processes, genetic drift and selection can create linkages between markers, even if they are located in different chromosomes. The correlation study between marker frequencies also detected large correlations between markers that are far apart (distances up to ~500kbp and correlations up to $r^2=0.6$). This indicated that a fraction of the QTLs can be located in alleles that comprise non-random correlation between markers over long physical distances. Thus, QTLs can be composed of different genomic scaffolds.

2.9 Genome-Wide Association Study (GWAS) analysis

A Linear Mixed Model (LMM) was used to identify significant associations between SNP markers and fibre quality traits, using a kinship correction (VanRaden, 2008) and following the same approach as in Yu et al. (2006), Huang et al. (2010), Malosetti et al. (2011), Zhou and Stephens (2012) and Kruijer et al. (2015). To account for multiple testing, we performed a Bonferroni correction based on the number of independent markers (Li and Ji, 2005). The cumulative distributions of observed and expected *p*-values were inspected for 3000 randomly selected SNP markers, to assess the correction for population structure (Yu et al., 2006). The expected *p*-values were the significance level of the associations between genotypes- and phenotypes under the assumptions that markers were unlinked to the polymorphism controlling the variation of the traits. An independent analysis was performed separately for seven fibre quality traits of hemp and over three locations.

A MultiQTL model was performed following a forward selection procedure to identify QTLs associated to a trait. Here, a QTL is defined as a group of significant and collinear QTL-markers, represented by the marker that explains the largest phenotypic variance (representative QTL-marker). The MultiQTL model is the combination of non-collinear QTLs, whereby each QTL explains a specific part of the phenotypic variation in the population. First, the best representative QTL-marker was selected into the model. To avoid multicollinearity in the MultiQTL models, correlations between the selected marker and remaining candidate markers were determined. All significant markers correlated at $r \geq 0.3$ ($\sim r^2 \geq 0.1$) were considered collinear to the first one. The remaining significant markers ($r<0.3$) were considered candidates to add to the model and forward selection was continued with those. The explained variance (r^2) of the MultiQTL model was calculated by the correlation between the fitted trait values and the observed traits values.

The same value of threshold for collinearity described in section 2.8 was confirmed using a different approach. Different conditions of correlation (r value) were used in the MultiQTL models, ranging from r=0.1 to r=0.9. At a threshold of $r \leq 0.3$ ($\sim r^2 \leq 0.1$) the number of QTLs and explained variance remained mostly steady, while at $r>0.3$ ($\sim r^2>0.1$) the number of QTLs and the explained variance dramatically increased (**Supplementary Table 1**).

In order to find groups of significant markers, a Principal Component Analysis (PCA) was performed. The PC1, PC2 and the $-\log_{10}P$ of the association between genotypes- and phenotypes were plotted to detect the clusters.

To further characterize the genetics of fibre quality traits across locations, a correlation analysis was performed between the QTLs of the three MultiQTL models for each trait, corresponding to the three test locations. For a trait, a common QTL region can be composed of different genomic scaffolds, that harbour correlated ($r \geq 0.3; r^2 \geq 0.1$) representative QTL-markers from MultiQTL models of two or three locations.

GWAS, PCA and correlation analyses were performed in Genstat 19th edition. The 3D scatter plots were performed in Excel version 14.0, using the macro Excel 3D Scatter Plot version 2.1 (Doka, 2013).

2.10 Transcriptome annotation

The transcriptome of the hemp cultivar *C. sativa* 'Finola' (finola1), aligned to the *C. sativa* 'Purple Kush' (canSat3 version GAC 000230575.1) assembly, was downloaded from the *C. sativa* Genome Browser Gateway at http://genome.ccbr.utoronto.ca/cgi-bin/hgGateway (van Bakel et al., 2011). The transcriptome of 'Finola' was annotated using Blast2go (blastx 2.8.0; E value cut of 0.001) (**Supplementary File 1**) (Conesa et al., 2005;Conesa and Gotz, 2008). Genomic scaffolds associated to the QTLs were analysed for transcripts. Special focus was given to genes with predicted functions related to the biosynthesis and modification of the cell wall, based on gene description and relevant research papers.

3. Results

3.1 Development of markers for genome-wide association studies

RAD-seq generated 3.717,57 million clean reads (557.4Gb) with an average of 29.7 million reads per sample (range 15.1-49.2 millions) and an average read length of 149.9bp. Our genotyping approach resulted in the detection of 2.852.901 SNPs with a missing rate < 50% in the population (number of samples with missing data divide by total number of samples). These SNPs were used for genotyping by scoring nucleotide frequencies in each accession. In total 612.452 SNPs were selected as informative markers for the genetic analysis. Selected markers covered 35.590 scaffolds of the draft canSat3 genome (van Bakel et al., 2011), including the largest scaffolds. In total, a cumulative length of ~503 Mb was covered by the scaffolds, which corresponds to ~59.7% of the haploid genome sequence of hemp (843Mb).

3.2 Low level of population structure in the hemp accession panel

The PCoA of the kinship matrix revealed the presence of some structure among the 123 hemp accessions. PCo1 divided most French accessions from the others with some levels of admixture and PCo2 divided French accessions in two groups. Nonetheless, PCo1 and PCo2 only explained 23.18% of the total genetic variance (**Figure 2a**). The dendrogram of the kinship matrix revealed a similar structure (**Figure 2b**). Five groups were identified, three of them showed large level of admixture, including accessions from different origin, whereas the remaining two clusters virtually showed only French accessions. Besides the patterns of structure inferred from the kinship matrix, coefficients of population differentiation (F_{st}) between the five groups showed low levels of differentiation between them, ranging from 0.019 to 0.058 (**Table 1**).

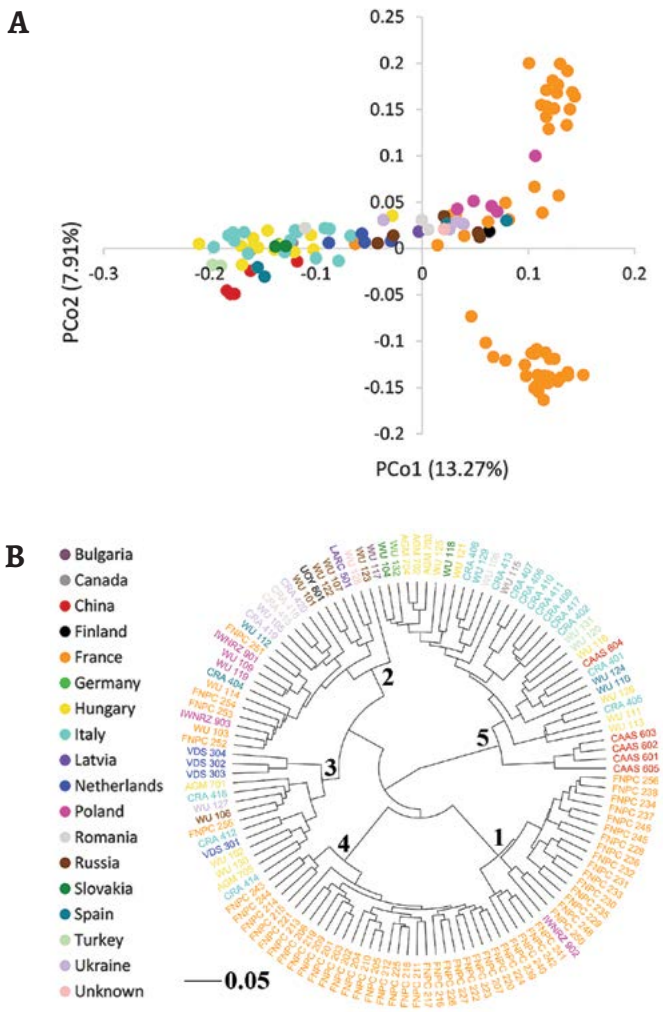


Figure 2. Population structure of the 123 hemp accessions. (**A**) Principal Coordinates Analysis (PCoA) and (**B**) dendrogram of the kinship matrix with the number of the sub-cluster in the node. Colour palette indicates the geographic origin of the accessions.

3.3 Effective control of population structure in highly heritable fibre quality traits

The extensive phenotypic variation and the large genetic components of several cell wall traits (contents of glucose, mannose, xylose, glucuronic acid, Klason lignin and acid detergent lignin) and bast fibre content makes them excellent tools to study the genetic architecture of fibre quality of hemp (**Table 3** and **9**, **Chapter 3**). The high heritabilities of these seven traits allow the study of the additive control of fibre quality traits. In addition, the small interactions between genotypes and environments (*G*×E) for these traits allow to study the heritable variation in fibre quality sensitive to the environment (Petit et al., 2019c).

To control putative effects of population structure in the phenotypic variation of highly heritable traits, the efficiency of the kinship matrix was assessed in the seven fibre quality traits, as depicted in **Figure 3**. Without the kinship matrix, observed $-\log_{10}P$ values were higher than the expected ones, suggesting large number of false-positive associations. In contrast, when the effects of population structure were controlled by the kinship matrix, observed and expected $-\log_{10}P$ values were similar in the seven fibre quality traits. Therefore, the kinship matrix properly controlled the effects of population structure on the seven traits to identify true-positive associations. The phenotypic data of the seven traits can thus be used to get insights into the genetic control of fibre quality.

Table 1. Pairwise coefficients of population differentiation (Fst) between the five clusters of the GWAS panel inferred from the dendrogram from the kinship matrix (**Figure 1**).

Clusters	1	2	3	4	5
1	-				
2	0.034296	-			
3	0.05169	0.019437	-		
4	0.04746	0.033938	0.043606	-	
5	0.058453	0.034621	0.015365	0.054827	-

3.4 Significant associations between markers and fibre quality traits in hemp

GWAS analyses resulted in the identification of over 2500 markers, mapping to single loci on 1515 different genomic scaffolds. These associations were found significantly associated ($-\log_{10}P \geq 4.047$) to at least one trait and one location (**Supplementary File 2**). These results indicated that the number of significant markers in each genomic scaffold was mostly one to three markers per scaffold. Yet, some scaffolds showed higher number of significant markers, such as scaffold4465 (length of 14.465bp) that showed 12 significant markers (**Supplementary File 2**). Moreover, the distribution of significant markers within the scaffolds showed differences between different scaffolds. For instance, the 12 significant markers from scaffold4465 were spread in a genomic sequence of ~9000bp, while the 11 significant markers from scaffold34707 clustered in a shorter sequence of ~600bp. Moreover, markers that clustered close to one edge of the scaffold sequence (i.e. significant markers from scaffold4465) indicated that the putative flanking genomic sequence of that

scaffold could also have significant markers associated to the same trait/s. Furthermore, several markers were found significantly associated to several fibre quality traits. For example, many markers from scaffold4465 were found significantly associated to contents of glucose, xylose, lignin (ADL, KL) and bast fibre content (**Supplementary File 2**). These results suggest a common genetic control between different traits.

3.5 Identification of QTLs for trait and location: MultiQTL models

Altogether, 90 QTLs were detected in the models among all traits and locations. The number of QTLs detected for the same trait generally differed across locations. The explained variance of the models for the same traits also differed across locations. In total, only 8 of the QTLs were detected in more than one location for the same trait or were detected common in different traits (**Table 2**). These differences were likely to be explained by the small *G*×E interactions on the traits (**Table 9**, **Chapter 3**).

Table 2. QTLs for fibre and cell wall traits of hemp.

Trait	Type of trait	Location	QTLs (n)	Explained variance	Representative QTL-markers
ADL%dm	Cell wall	CRA	3	60.32	scaffold73277_3647; scaffold74027_10718; scaffold4782_157368
ADL%dm	Cell wall	FNPC	5	40.2	scaffold34707_3473; scaffold80551_6348; scaffold2608_892; scaffold46943_424; scaffold60230_6140
ADL%dm	Cell wall	VDS	4	56.15	scaffold12000_76415; scaffold4268_1985; scaffold12558_4900
Glc%dm	Cell wall	CRA	3	27.75	scaffold45492_8629; scaffold15962_51386; scaffold71896_5178
Glc%dm	Cell wall	FNPC	5	63.08	scaffold55006_23494; scaffold118257_4014; scaffold25092_982; scaffold137175_654; scaffold53823_15026
Glc%dm	Cell wall	VDS	3	55.78	scaffold69515_8050; scaffold18086_2629; scaffold8221_7904
GlcA%dm	Cell wall	CRA	13	60.87	scaffold82406_5533; scaffold66904_776; scaffold111279_1125; scaffold16765_1949; scaffold9389_35935; scaffold39478_12834; scaffold11938_20472; scaffold2003_8850; scaffold2733_12157; scaffold869_24280; C32032885_328; scaffold29101_9917; scaffold1838_41611
GlcA%dm	Cell wall	FNPC	3	47.19	scaffold90847_19970; scaffold15717_124591; scaffold76130_8395
GlcA%dm	Cell wall	VDS	6	57.23	scaffold55265_18829; scaffold68246_496; scaffold5954_52112; scaffold20722_6013; scaffold79841_7656; scaffold28137_3788
KL%dm	Cell wall	CRA	2	40.11	scaffold45492_8629; scaffold10943_380
KL%dm	Cell wall	FNPC	6	40.2	scaffold53823_15026; scaffold63415_1601; scaffold69322_5540; scaffold47933_3213; scaffold68105_6992; scaffold6550_126605
KL%dm	Cell wall	VDS	3	48.53	scaffold12000_76415; scaffold51533_356; scaffold13365_3919
Man%dm	Cell wall	CRA	6	39.93	scaffold17542_9887; scaffold2731_11599; scaffold9919_11591; scaffold47229_16597; C32103569_1584; scaffold17032_37572

Trait	Type of trait	Location	QTLs (n)	Explained variance	Representative QTL-markers
Man%dm	Cell wall	FNPC	7	66.64	scaffold34580_21508; scaffold12140_37109; scaffold4854_60; scaffold71367_190; scaffold34081_8008; scaffold900_4618; scaffold48672_475
Man%dm	Cell wall	VDS	2	46.59	scaffold90847_19970; scaffold32213_3500
Xyl%dm	Cell wall	CRA	4	23.38	scaffold103588_1151; scaffold49748_10229; scaffold30777_901; scaffold14559_5718
Xyl%dm	Cell wall	FNPC	4	47.43	scaffold42215_3622; scaffold14590_17087; scaffold25801_1953; scaffold3268_34793
Xyl%dm	Cell wall	VDS	3	32.92	scaffold55265_18829; scaffold56299_8302; scaffold32213_3500
BCD	Fibre	CRA	3	41.8	scaffold56299_8302; scaffold13466_34599; scaffold35378_13209
BCD	Fibre	FNPC	3	61.08	scaffold26831_57993; scaffold25080_10706; scaffold62704_12566
BCD	Fibre	VDS	2	45.71	scaffold56299_8302; scaffold3696_33201

Number of QTLs (n) identified in the respective MultiQTL models per trait and per test location (CRA=Italy; FNPC=France and VDS=The Netherlands), with explained variance of the full model. Trait: ADL%dm = Acid Detergent Lignin content, BCD= Bast fibre content after decortication, Glc%dm= Glucose content, GlcA%dm= Glucuronic acid content, KL%dm= Klason lignin content, Man%dm= Mannose content and Xyl%dm= Xylose content; Representative QTL-markers= QTL-markers that explain the largest variance in a MultiQTL model. Markers are indicated as the scaffold number and the position of the SNP within the scaffold (canSat3 assembly GAC_000230575.1).

The 3D scatter plots (of x=PC1, y=PC2 and z=-$\log_{10}$P) highlighted clusters of significant markers that resembled the peak tops of Manhattan plot distributions, despite the lack of physical positions on a chromosome (**Figure 4**). It was observed that all clusters included significant markers from all the three trial locations. Two clusters of markers were identified for respectively contents of glucose, xylose, lignin (ADL, KL) and bast fibre content, while only a single cluster of markers was identified for respectively contents of mannose and glucuronic acid.

3.6 QTL regions across locations for fibre quality

The correlation analyses across locations highlighted 16 common QTLs across two or three locations, as detailed in **Table 3**. Eight QTLs were identified across two test locations. Among them, one for respectively contents of ADL (QTL_{ADL2}), KL (QTL_{KL3}), mannose (QTL_{Man2}) and bast fibre content (QTL_{BCD2}); and two for respectively contents of glucose (QTL_{Glc2}, QTL_{Glc3},) and glucuronic acid (QTL_{GlcA2}, QTL_{GlcA3}). In total, eight QTLs were identified across the three locations, one for respectively contents of ADL (QTL_{ADL1}), glucose (QTL_{Glc1}), glucuronic acid (QTL_{GlcA1}), mannose (QTL_{Man1}), xylose (QTL_{Xyl1}) and bast fibre content (QTL_{BCD1}); and two for content of KL (QTL_{KL1} and QTL_{KL2}).

Table 3. Identification of QTL regions across locations for seven fibre quality traits of hemp. QTLs common across locations were identified by correlation analyses ($r \geq 0.3$, $\sim r^2 \geq 0.1$).

QTL across locations	Trait	Correlated representative QTL-markers across MultiQTL models					
		CRA		FNPC		VDS	
QTL_{ADL1}	ADL%dm	scaffold4782_157368		scaffold80551_6348		scaffold4268_1985	
QTL_{ADL2}	ADL%dm	scaffold73277_3647		scaffold34707_3473		-	
QTL_{BCD1}	BCD%dm	scaffold56299_8302		scaffold25080_10706		scaffold56299_8302	
QTL_{BCD2}	BCD%dm	scaffold13466_34599		-		scaffold3696_33201	
QTL_{Glc1}	Glc%dm	scaffold45492_8629		scaffold53823_15026		scaffold18086_2629	
QTL_{Glc2}	Glc%dm	scaffold15962_51386		scaffold118257_4014		-	
QTL_{Glc3}	Glc%dm	-		scaffold137175_654		scaffold69515_8050	
QTL_{GlcA1}	GlcA%dm	scaffold39478_12834		scaffold90847_19970		scaffold55265_18829	scaffold79841_7656
QTL_{GlcA2}	GlcA%dm	scaffold111279_1125		-		scaffold28137_3788	
QTL_{GlcA3}	GlcA%dm	scaffold1838_41611		scaffold15717_124591		-	
QTL_{KL1}	KL%dm	scaffold45492_8629		scaffold69322_5540	scaffold53823_15026	scaffold51533_356	
QTL_{KL2}	KL%dm	scaffold10943_380		scaffold63415_1601		scaffold13365_3919	
QTL_{KL3}	KL%dm	-		scaffold6550_126605		scaffold12000_76415	
QTL_{Man1}	Man%dm	scaffold47229_16597	scaffold2731_11599	scaffold34580_21508		scaffold90847_19970	
QTL_{Man2}	Man%dm	C32103569_1584		-		scaffold32213_3500	
QTL_{Xyl1}	Xyl%dm	scaffold30777_901		scaffold42215_3622		scaffold55265_18829	

QTL across locations= name given to the QTL common across locations; Trait: ADL%dm= Acid Detergent Lignin content, BCD= Bast fibre content after decortication, Glc%dm= Glucose content, GlcA%dm= Glucuronic acid content, KL%dm= Klason lignin content, Man%dm= Mannose content and Xyl%dm= Xylose content; CRA= representative QTL-markers from MultiQTL models from CRA, Italy; FNPC= representative QTL-markers from MultiQTL models from FNPC, France and VDS= representative QTL-markers from MultiQTL models from VDS, The Netherlands.

Table 4. Putative candidate genes in QTLs across locations associated to fibre quality traits.

QTL across locations	Trait	Scaffold	Candidate gene (Transcript)	Description	Class of metabolism	UniProt code	References
QTL_{ADL1}	ADL%dm	scaffold4782	FN14269	*Polygalacturonase*	Polysaccharide	B9SPH9 and A0A1J3JMW1	(Chan et al., 2010;Blande et al., 2017)
			FN26643	*2-phytyl-1,4-beta-naphtho-quinone methyltransferase*	Lignin	W9RCM0 and A0A1R3I6R5	(He et al., 2013)
QTL_{Glc1}	Glc%dm	scaffold53823	FN16068	*Cytochrom* b_5	Lignin	A0A1R3IKW6, W9R7N2 and W9R7N2	(He et al., 2013)
QTL_{Glc3}	Glc%dm	scaffold137175	FN29173	*Phosphoenolpyruvate carboxylase (pepc), housekeeping isozyme*	Monosaccharide	W9RZ79	(He et al., 2013)
QTL_{GlcA1}	GlcA%dm	scaffold55265	FN02726	*p-coumaroyl shikimate 3-hydroxylase (c3h1)*	Lignin	A0A068FPP3	(Kamdee et al., 2014)
QTL_{GlcA3}	GlcA%dm	scaffold15717	FN00563	*Glyceraldehyde-3-phosphate dehydrogenase (gapdh)*	Monosaccharide	W9R8N5	(He et al., 2013)
			FN19953	*Chalcone synthase (chs)*	Lignin	A0A068PX04, Q1G6S4, Q94LW8 and R9WQY5	(Lihova et al., 2006)
			FN08380	*Stilbene synthase (sts)*	Lignin	Q6BAR5	(Richter et al., 2005)
			FN15157	*Alpha-mannosidase*	Polysaccharide	A0A087G6Y0 and A0A1S2XQU0	(Parween et al., 2015)
QTL_{KL1}	KL%dm	scaffold53823	FN16068	*Cytochrome* b_5	Lignin	A0A1R3IKW6 and W9R7N2	(He et al., 2013)
QTL_{KL2}	KL%dm	scaffold10943	FN05922	*Methyltransferase*	Lignin/Polysaccharide	W9S3D7	(He et al., 2013)
QTL_{KL3}	KL%dm	scaffold6550	FN11795	*Glucan endo-1,3-beta-glucosidase 3 / F-box* protein SKIP28	Polysaccharide	W9SPC5, A0A1S2YDV5 and A0A1S3UV78	(He et al., 2013;Kang et al., 2014;Parween et al., 2015)
QTL_{Man1}	Man%dm	scaffold34580	FN16350	*Glucose-1-phosphate adenylyltransferase (glgc)*	Monosaccharide	W9R328	(He et al., 2013)
QTL_{Xyl1}	Xyl%dm	scaffold55265	FN02726	*p-coumaroyl shikimate 3-hydroxylase (c3h1)*	Lignin	A0A068FPP3	(Kamdee et al., 2014)

QTL across locations= name given to the common QTL region across locations; Trait: ADL%dm = Acid Detergent Lignin content, BCD= Bast fibre content after decortication, Glc%dm = Glucose content, GlcA%dm = Glucuronic acid content, KL%dm = Klason lignin content, Man%dm = Mannose content and Xyl%dm = Xylose content; Scaffold= genomic scaffold of canSat3, GAC_000230575.1 that composes the QTL across locations (**Table 3**); Candidate gene= transcript of hemp cultivar 'Finola' (transcriptome assembly; Finola) located in a QTL across locations, annotated for a putative function in cell wall metabolism and/or fibre quality (Blast2go; gene description, relevant research papers); Description= description of hemp transcript, as found by Blast2Go annotation, based on homology to sequence from GenBank with known or predicted functions; UniProt code= protein code of the gene.

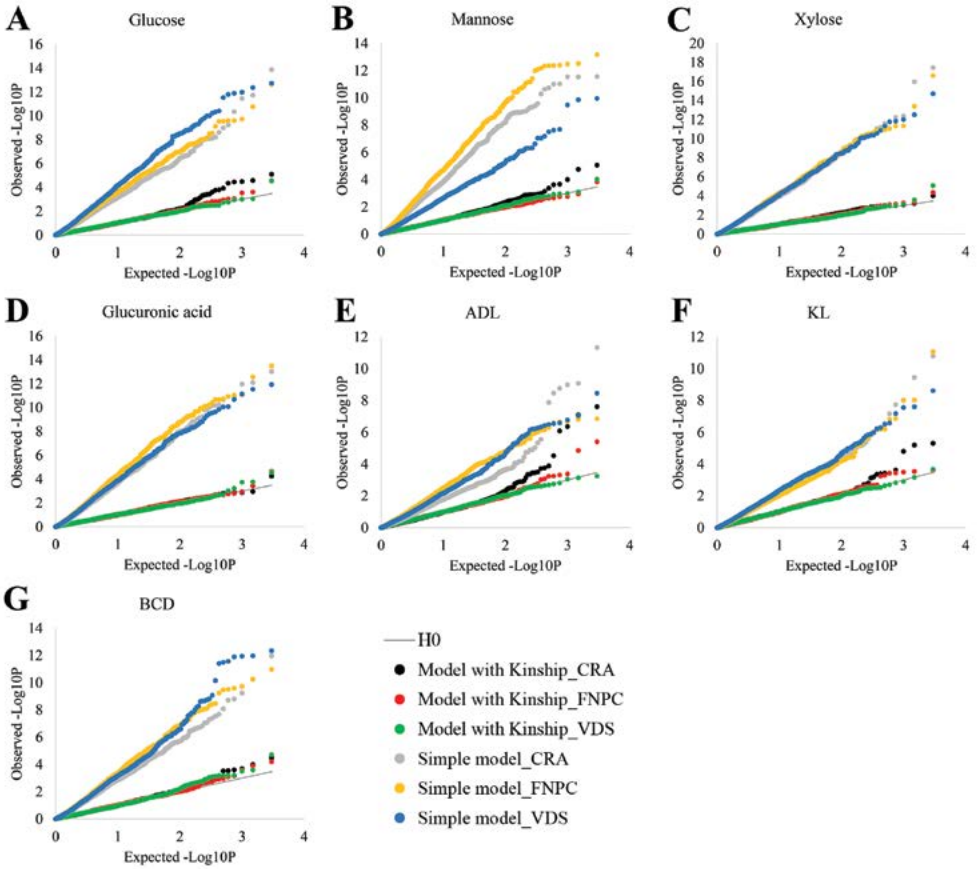


Figure 3. Cumulative distributions of *p*-values of contents of glucose (**A**), mannose (**B**), xylose (**C**), glucuronic acid (**D**), ADL (**E**), KL (**F**) and BCD (**G**). The cumulative distributions of the observed $-\log_{10}P$ for the simple model (no controlled for population structure) and the kinship models (controlled for population structure) under the expectation that random SNP markers are unlinked to the polymorphism controlling the traits.

Furthermore, several QTLs across locations for different traits shared some representative QTL-markers. This is the case of QTL_{Glc1} for glucose content and QTL_{KL1} for lignin content that shared the representative QTL-marker scaffold53823_15026. QTL_{Xyl1} for xylose content and QTL_{GlcA1} for glucuronic acid also shared a representative QTL-marker (scaffold55265_18829) (**Table 2** and **3**). The common scaffolds between QTLs for several traits indicated co-localization of some QTLs in tightly-close genomic regions.

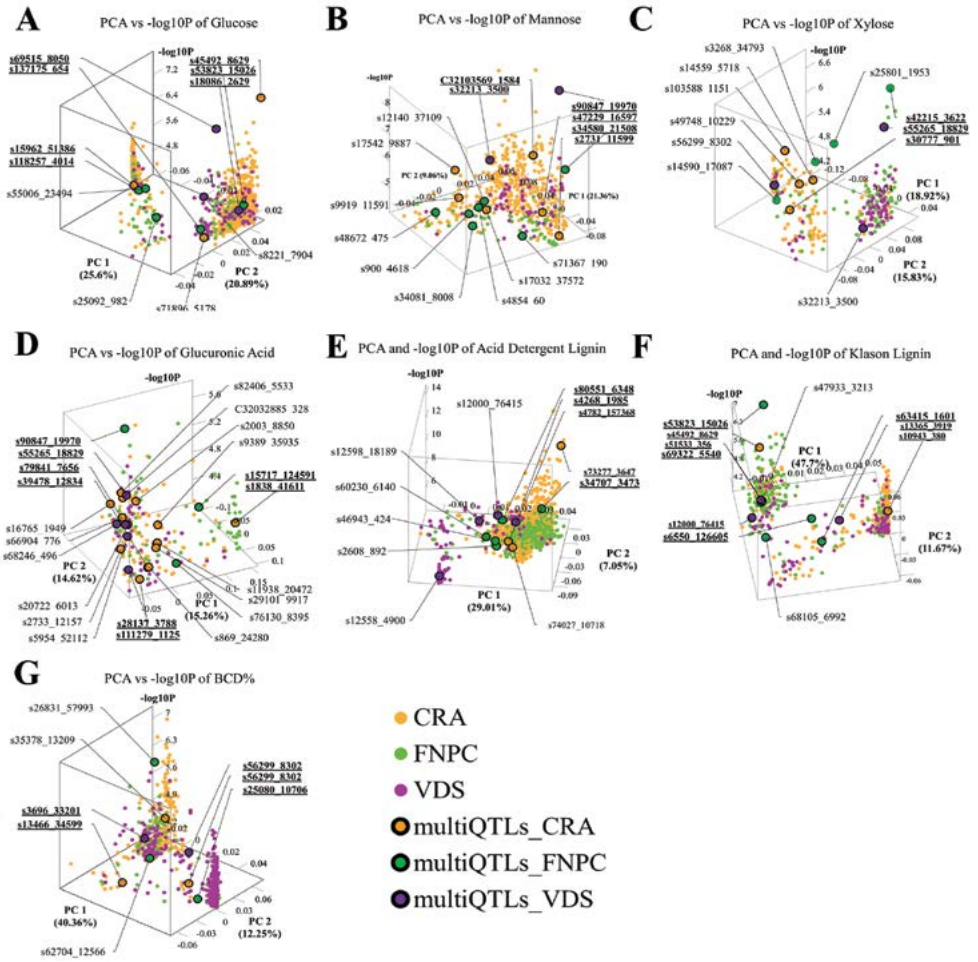


Figure 4. 3D scatter plots of PCAs depicting variation in the allele frequency profiles of significant QTLs for respectively contents of glucose (**A**), mannose (**B**), xylose (**C**), glucuronic acid (**D**), ADL (**E**), KL (**F**) and BCD (**G**). X-axes indicate Principal Component 1, Y-axes indicate Principal Component 2 and Z-axes indicate the significant level of the GWAS association ($-\log_{10}P$). Each plot is shown in the angle that represents better the results. Each dot represents a significant QTL-marker. Orange, green and purple dots indicate the significant QTL-markers detected in CRA, FNPC and VDS, respectively. Orange, green and purple dots with a black circle represent the representative QTL-markers of the MultiQTL models from CRA, FNPC and VDS, respectively. QTL-names in bold and underlined indicate representative QTL-markers from different MultiQTLs models that belong to a QTL region across locations.

3.7 Candidate genes for fibre quality of hemp

In total, 166 transcripts were identified in scaffolds associated to the QTLs and 102 of them showed sequence homology to annotated structural genes from other species (**Supplementary Table 4**). These genes were involved in several plant physiological

processes, including metabolism of proteins, lipids, carbohydrates (monosaccharides and polysaccharides) and lignin. Housekeeping genes, transcription factors and genes involved in nuclear processes (e.g. transport and regulation of chromatin and DNA) were also identified. Moreover, among the annotated transcripts were identified genes involved in plant defence mechanisms against pathogens, genes involved in redox reactions and genes involved in perception of light, such as phytochromes. Finally, transporters of magnesium, potassium and calcium; genes involved in plasmodesmata and genes involved in the transport of vesicles in the cytoplasm were also identified. Among the 102 transcripts, 12 genes were identified as candidate genes for six QTLs across locations, as detailed in **Table 4**.

A *polygalacturonase* gene was identified in QTL_{ADL1} for ADL content. This gene degrades pectins. Two genes of the metabolism of lignin and monosaccharides were identified in QTLs associated to glucose content: a *cytochrome b5* in QTL_{Glc1} and a *phosphoenolpyruvate carboxylase* (*pepc*) in QTL_{Glc3}. In addition, *cytochrome b5* was also identified in QTL_{KL1} for lignin content. This gene mapped in the scaffold53823 which underneath both QTL_{Glc1} and QTL_{KL1}. The lignin-related gene *p-coumaroyl shikimate 3-hydroxylase (c3h1)* was also identified in a scaffold (scaffold55265) covered by QTLs for two different traits. *C3h1* was identified in QTL_{GlcA1} for glucuronic acid content and in QTL_{Xyl1} for xylose content. A gene of the metabolism of monosaccharides (*glyceraldehyde-3-phosphate dehydrogenase* (*gapdh*)), two genes of the metabolism of lignin (*chalcone synthase* (*chs*) and *stilbene synthase* (*sts*)) and a gene of the metabolism of polysaccharides (α-mannosidase) were identified in QTL_{GlcA3} for glucuronic acid content. A *methyltransferase* gene involved in the biosynthesis of many cell wall components was identified in QTL_{KL2} and a *glucan endo-1,3-beta-glucosidase 3* gene was identified in QTL_{KL3}. Finally, a *glucose-1-phosphate adenylyltransferase* (*glgc*) involved in the metabolism of monosaccharides was identified in QTL_{Man1} for mannose content.

4. Discussion

The study of the genetic architecture of hemp fibre quality is fundamental to develop molecular strategies to breed for new hemp cultivars with improved fibre properties. The genetics of hemp fibre quality have not been widely studied and no Genome-Wide Association Studies (GWAS) have yet been reported for hemp. Nonetheless, GWAS analyses have proved to be a powerful approach to detect genetic components controlling quantitative traits. GWAS analyses were used to detect QTLs for agronomic traits, biomass content and cell wall composition in several fibre crops, such as cotton (Liu et al., 2018b) and flax (Xie et al., 2018).

In the present study, we developed a modified GWAS approach to identify and map QTLs in hemp. In the absence of a complete genome sequence it is difficult to map the boundaries of

QTLs and to identify independent QTLs (Sawler et al., 2015). To overcome this limitations, we developed a MultiQTL approach based on the levels of explained variance and the correlations between significant markers. This approach was applied in hemp and several QTLs for fibre quality traits were successfully identified. This method enables genome-wide association analyses for hemp, and it can be extended to genetic studies in other orphan species, for which no complete and assembled genome is available (Graham, 2013).

All fibre quality traits assessed in this study were largely heritable across different environments (H^2=0.88-0.96) and the rankings of the accessions were similar at the three locations for all traits (**Table 9, Chapter 3**) (Petit et al., 2019c). Genetic studies on cell wall composition in different crops also showed large heritabilities. In maize, heritabilities for biomass traits, such as contents of cellulose, hemicellulose and lignin were reported above 0.6 (Torres et al., 2015a). In miscanthus, the heritabilities of these traits ranged from 0.4 to 0.72 (Slavov et al., 2014;van der Weijde et al., 2017). These large heritabilities indicate that the genetic gains of fibre quality can be maximized in several crops.

In this report, sixteen QTLs across locations were found to be associated to fibre quality in hemp. For six QTLs across locations, cell wall candidate genes were identified. Among those were genes involved in the metabolism of monosaccharides, genes involved in the metabolism of polysaccharides, genes involved in the metabolism of glycoproteins and genes involved in the metabolism of lignin. *Phosphoenolpyruvate carboxylase* (*pepc*), *glyceraldehyde-3-phosphate dehydrogenase* (*gapdh*) and g*lucose-1-phosphate adenylyltransferase* (*glgc*) are genes involved in the metabolism of monosaccharides that were detected respectively in QTL_{Glc3}, QTL_{GlcA3} and QTL_{Man1}. *Phosphoenolpyruvate carboxylase* (*pepc*) codifies for an enzyme that catalyses the carboxylation of phosphoenolpyruvate to oxaloacetate to produce malate. The activity of this enzyme has been reported to play multiple roles, among them fibre elongation, carbon storage and energy production (Zieher, 2010). The accumulation of malate and sugars, is thought to play an important role in the fibre elongation, through osmotic regulation and charge balance. For instance, Li and co-workers reported in cotton that in periods of rapid elongation phase, *pepc* genes were highly expressed but weakly at slow elongation periods (Li et al., 2010). In addition, the activity of *pepc* is involved in the carbon storage and energy production by consuming phosphoenolpyruvate (Jeanneau et al., 2002). Phosphoenolpyruvate results from the glycolysis of glucose and consequently reduces the source of monosaccharides for other pathways, such as to biosynthetise cell wall polysaccharides. *Glyceraldehyde-3-phosphate dehydrogenase* (*gapdh*) is another gene involved in glycolysis (Piattoni et al., 2017) and its overexpression can also decrease the monosaccharide availability. G*lucose-1-phosphate adenylyltransferase* (*glgc*) is a gene involved in the partitioning of α-D-glucose-1P for starch and cell wall polysaccharide biosynthesis. *Glucose-1-phosphate adenylyltransferase* codifies for an enzyme that catalizes the conversion of α-D-glucose-1P to ADP-glucose. ADP-glucose will be further used for starch biosynthesis

(de Setta et al., 2014). Therefore, this enzyme competes for sustrate with enzymes involved in the conversion of α-D-glucose-1P to UDP-glucose (Lozovaya et al., 1996), that will be further used for the biosynthesis of cell wall polysaccharides. UDP-glucose serves as a direct source of UDP-galactose, UDP-rhamnose and UDP-glucuronic acid; and an indirect source of UDP-xylose, UDP-galacturonic acid and UDP-apiose (Wierzbicki et al., 2019).

Alpha-mannosidase, *glucan endo-1,3-β-glucosidase 3* and *polygalacturonase* are genes involved in the metabolism of glycoproteins and polysaccharides, that were detected in QTLs for fibre quality traits (**Table 4**). *Alpha-mannosidase* codifies for an enzyme involved in early N-glycan processing. The N-glycans are oligosaccharides ($Glc_3Man_9GlcNAc_2$) linked to nitrogen (N) and involved in the biosynthesis of N-glycoprotein (Liebminger et al., 2009). Cell wall N-glycoproteins are important structural components of the cell walls (Roberts et al., 1985;Nguema-Ona et al., 2014). They are a little understood group of proteins that play a diversity of functions, including signalling and interacting with the surrounding environment; and plant defence. Moreover, they are essential for the plant development and responses to stress (Nguema-Ona et al., 2014;Strasser, 2014). In arabidopsis, three α-mannosidase genes have been identified, named as *mns* genes. Arabidopsis mutants for *mns* genes resulted in the formation of aberrant N-glycan. N-glycoproteins are incorrectly folded, affecting the function of the proteins (Nguema-Ona et al., 2014). The mutant plants displayed short, swollen roots and altered cell walls. Arabidopsis mutant for one, two or three *mns* genes displayed lower amount of homogalacturonan-type pectin. Consequently, Nguema-Ona and co-workers suggested that N-glycosylation of glycoproteins plays a role in the correct targeting, assembly, or stability of cell wall biosynthetic or remodelling enzymes (Nguema-Ona et al., 2014). Furthermore, the cell wall is a highly dynamic structure with constant synthesis and degradation of cell wall components throughout plant development (Houston et al., 2016). Degradation of cell wall components is an important step during pathogenesis. It is known that pathogens trigger the expression of endogenous plant genes that induce a degradation of the cell wall. Therefore, genes involved in cell wall associated defence can modify cell wall composition (reviewed in Underwood (2012)). *Glucan endo-1,3-β-glucosidase 3* codifies for an enzyme involved in degradation of cell wall polysaccharides and plant defence against pathogens. In arabidopsis, this enzyme has been implicated in pathogenesis when the plants were infected with tobamo virus (Zavaliev et al., 2013). The infection triggered the endogenous secretion of this enzyme in the cell wall, which degraded callose at plasmodesmata. Consequently, the damage in the cell wall enhanced the virus spread. Arabidopsis mutants for this gene retained the enzyme in the endoplasmic reticulum and was not secreted. Zavaliev and co-workers observed that the cell walls of the mutant plants were unaffected. As a result, alterations in expression or function of *glucan endo-1,3-β-glucosidase 3* gene is likely to affect the dynamism and architecture of the cell walls. *Polygalacturonases* are another group of genes codifying for enzymes involved in degradation of the cell wall, precisely homogalacturonan pectin. These enzymes function

in a wide range of developmental processes, including cell separation and cell elongation to control the shape of the plant body (Gonzalez-Carranza et al., 2007;Babu and Bayer, 2014). They play important roles in the primary cell wall of meristematic and elongating cells, before secondary cell walls can fortify the fibres (Zhong and Ye, 2007). In arabidopsis, 69 *polygalacturonases* has been identified and the alteration of several of those was shown to produce aberrations in plant developmental processes (reviewed in Babu and Bayer (2014)). For instance, *polygalacturonase involved in expansion1 (pgx1)* was shown to be involved in hypocotyl elongation. Arabidopsis mutants for *pgx1* gene showed strong cell-elongation defects, alterations in pectin molecular mass and cell wall composition (Xiao et al., 2014).

Cytochrome b_5, *p-coumaroyl shikimate 3-hydroxylase* (*c3h1*), *chalcone synthase* (*chs*), *stilbene synthase* (*sts*) and *2-phytyl-1,4-β-naphthoquinone methyltransferase* are genes involved directly or indirectly to the metabolism of lignin, that were detected in QTLs for lignin content and contents of monosaccharides (**Table 4**). A recent study in arabidopsis described the product of *cytochrome b_5* as an obligate electron shuttle protein specific for the biosynthesis of syringyl lignin subunit (Gou et al., 2019). The study revealed that the arabidopsis mutant for *cytochrome b_5* suppressed the catalytic activity of the enzymes cinnamic acid 4-hydroxylase (C4H) and ferulate 5-hydroxylase (F5H). These two enzymes are essential in the biosynthesis of lignin and the suppression of their activity affects lignin content (Boerjan et al., 2003;Sykes et al., 2015). *P-coumaroyl shikimate 3-hydroxylase* is another essential gene in the lignin biosynthetic pathway (Boerjan et al., 2003). In eucalyptus and alfalfa, the down-regulation of *c3h1* expression displayed lower lignin content (Ralph et al., 2006;Ralph et al., 2012) In addition, in poplar, the repression of *c3h1* was associated to alterations in the cell wall and in improved sugar release (Sykes et al., 2015). Moreover, these studies showed that the repression of *c3h1* did not lead to deleterious impacts on plant growth. These results indicate positive implications for increasing fibre quality without affecting plant fitness. Furthermore, external treatment of chalcone in soybean was reported to inhibit lignin biosynthesis (Chen et al., 2011). This suggests that variation in key genes of chalcone (*chs*) and stilbene (*sts*) biosynthesis are responsible for variation in lignin content (Vannozzi et al., 2012).

Moreover, the identification of the same genomic scaffolds in QTLs for different traits is an indication of co-localization of QTLs for fibre quality in tightly-close genomic regions. For instance, QTL_{Glc1} and QTL_{KL1} for respectively glucose and lignin content shared a genomic scaffold. Co-localization of scaffolds was also found in QTL_{Xyl1} and QTL_{GlcA1} for respectively xylose and glucuronic acid content (**Table 3**). Cell wall composition is a complex trait and many cell wall components are highly interconnected. Petit and co-workers reported in hemp large correlations between cell wall components, such as contents of glucose and lignin (r^2=-0.93) and contents of xylose and glucuronic acid (r^2=0.96) (Petit et al., 2019c). The large correlation between different cell wall components is an indication that QTLs affecting

a trait, will automatically affect other correlated cell wall traits. The similar phenotypic variation between correlated traits is thus associated in the same genomic regions. For instance, the gene *cytochrome* b_5, identified in the genomic scaffold common for glucose and lignin content, can directly affect lignin content. The alteration in lignin content can then indirectly affect the content of glucose. Novaes and co-workers reviewed evidences that the negative correlation between lignin and cellulose (glucose) is partially mediated at the level of transcription (Novaes et al., 2010). The overproduction of a cell wall component can inhibit the production of another component, through transcription factors mediating different pathways (e.g. NYB46, MYB83 and secondary wall NAC (SWN)) (Zhong and Ye, 2012). Another example of co-localization is the QTLs for contents of xylose and glucuronic acid. These monosaccharides belong to the same polysaccharide, xylans in eudicots plants (Pauly et al., 2013). Therefore, the variation of one component affects the variation of the other one. Moreover, *c3h1* gene for lignin biosynthesis was identified in the common QTL between xylose and glucuronic acid. Lignin and hemicellulose type xylans are cross-linked (Pauly et al., 2013) and highly correlated (Petit et al., 2019c). Allelic variation of *c3h1* can directly lead to phenotypic variation of lignin content, as reported in other crops (Ralph et al., 2006;Ralph et al., 2012;Sykes et al., 2015). Thereafter, alteration of lignin content can affect the cross-links and indirectly the contents of xylose and glucuronic acid. An evidence for this are the studies that report modifications of xylans (e.g. methylations of glucuronic acid) affecting lignin composition (reviewed in Wierzbicki et al. (2019)).

Finally, the candidate genes identified in this study are of great significance for further studies, including reverse genetic approaches, such as candidate gene functional studies. Such studies would be very valuable for understanding the molecular mechanisms responsible for hemp fibre quality.

Implications of the QTLs for fibre quality in the context of genetic improvement of hemp

The identification of QTLs across locations for hemp fibre quality will have relevant implications for the breeding programmes of the crop. Breeding for QTLs identified across two locations will lead to genotypes that perform well under certain environments but not necessarily under other environments. Furthermore, when breeding for QTLs identified across the three locations, the advantage is that the resulting improved genotypes will perform well under more environments. The two or three markers for each QTL across locations would be the first candidates to include in marker assisted selection (MAS) schemes of hemp. The implementation of molecular markers in hemp breeding programmes is expected to speed up the development of new hemp cultivars with improved fibre properties. The use of molecular markers have several advantages, including the early stage selection of promising plants. In addition, the use of markers avoids to phenotype large number of samples for fibre quality, which usually involves large costs and time in breeding programmes.

Furthermore, the identification of QTLs and candidate genes for fibre quality, using the hemp panel, indicates that this panel was large enough to include an extensive range of genotypic and phenotypic variation for genetic studies in hemp. The large variability of the hemp panel enables further mapping studies for hemp traits that still remain poorly studied. The hemp panel and the methodology developed in this study are a great value to extend the genetic basis of many traits beyond cell wall traits, such as flowering time and sex determination (Salentijn et al., 2019) and understand their interaction with fibre quality traits. Those studies would also provide more molecular breeding tools to accelerate the development of new hemp cultivars with improved fibre quality.

5. Conclusions

The present study provides valuable insights into the genetic and molecular architecture of the fibre quality in hemp, which advocates for positive prospects to modernise breeding programmes of hemp towards molecular approaches. The sixteen QTLs for fibre quality are the first candidates to include in marker assisted selection (MAS) schemes of hemp. The implementation of these QTLs in molecular breeding programmes in hemp will accelerate the selection of interesting individuals and will avoid the phenotyping of large number of samples, leading to important cost reductions. In addition, the identification of QTLs and candidate genes for fibre quality enables the use of the hemp panel in further studies to extend the genetic architecture of other important traits of hemp. Furthermore, the correlation method used in this study to identify non-collinear QTLs can be extended to genetic studies in other orphan species, for which no complete genome is available.

6. Acknowledgements

This work was conducted as part of the MultiHemp project (Multipurpose hemp for industrial bioproducts and biomass) funded by the European Union's Seventh Framework Programme for research, technological developments and demonstration under grant agreement number 311849. The authors are grateful to Richard Finkers for his technical assistance in the annotation of the transcriptome.

Supplementary data

Supplementary Table 1. Number of QTLs and explained variances of the multiQTL models for the content of glucose in the three locations under different putative conditions of independence (r).

Trait	Location	Correlation (r)	Number of QTLs	Explained variance
Glucose	CRA	0.1	2	30.93
Glucose	CRA	0.2	2	29.18
Glucose	CRA	0.3	3	29.22
Glucose	CRA	0.4	2	29.36
Glucose	CRA	0.5	6	47.67
Glucose	CRA	0.6	5	65.27
Glucose	CRA	0.7	19	84.22
Glucose	FNPC	0.1	2	37.42
Glucose	FNPC	0.2	2	39.07
Glucose	FNPC	0.3	5	64.63
Glucose	FNPC	0.4	5	68.22
Glucose	FNPC	0.5	14	80.07
Glucose	FNPC	0.6	13	84.38
Glucose	FNPC	0.7	6	74.78
Glucose	VDS	0.1	3	53.23
Glucose	VDS	0.2	3	56.17
Glucose	VDS	0.3	3	55.72
Glucose	VDS	0.4	6	79.21
Glucose	VDS	0.5	6	66.4
Glucose	VDS	0.6	14	79.26
Glucose	VDS	0.7	8	75.62

Supplementary File 1, **Supplementary File 2** and **Supplementary Table 4** can be found in the following link: https://wageningenur4-my.sharepoint.com/:f:/g/personal/jordi_petitpedro_wur_nl/ElnmN8O9jItCqh4XMK5dVkQBWcMXrvV3WxIV-cvDcdZ2TA?e=O3Dq0K

Chapter 5

The complex interactions between flowering behaviour and fibre quality in hemp

Elma M. J. Salentijn[a], Jordi Petit[a], Luisa M. Trindade[a*]

[a] Wageningen UR Plant Breeding, Wageningen University and Research (WUR), P.O. Box 386, 6700 AJ Wageningen, The Netherlands

[*] Corresponding author: Luisa M. Trindade (luisa.trindade@wur.nl)

This chapter has been published (with minor modifications) as:
Salentijn, E.M.J., Petit, J., and Trindade, L.M. (2019). The Complex Interactions Between Flowering Behaviour and Fiber Quality in Hemp. *Frontiers in Plant Science* 10.
doi: 10.3389/fpls.2019.00614

ABSTRACT

Hemp, *Cannabis sativa* L., is a sustainable multipurpose fibre crop with high nutrient and water use efficiency, and with biomass of excellent quality for textile fibres and construction materials. The yield and quality of hemp biomass are largely determined by the genetic background of the hemp cultivar but are also strongly affected by environmental factors, such as temperature and photoperiod. Hemp is a facultative short-day plant, characterized by a strong adaptation to photoperiod and a great influence of environmental factors on important agronomic traits, such as "flowering time" and "sex determination". This sensitivity of hemp can cause a considerable degree of heterogeneity, leading to unforeseen yield reductions. Fibre quality for instance is influenced by the developmental stage of hemp at harvest. Also, male and female plants are differing in stature and produce fibres with different properties and quality. Next to these causes, there is evidence for specific genotypic variation in fibre quality among hemp accessions. Before improved hemp cultivars can be developed, with specific flowering times and fibre quality, and adapted to different geographical regions, a better understanding of the molecular mechanisms controlling important phenological traits such as "flowering time" and "sex determination" in relation to fibre quality in hemp is required. It is well known that genetic factors play a major role in the outcome of both phenological traits but, the major molecular factors involved in this mechanism are not characterized in hemp. Genome sequences and transcriptome data are available but their analysis mainly focussed on the cannabinoid pathway for medical purposes. Herein, we review the current knowledge of phenotypic- and genetic data available for "flowering time", "sex determination" and "fibre quality" in short-day and dioecious crops, respectively, and compare them with the situation in hemp. A picture emerges for several controlling key genes, for which natural genetic variation may lead to desired flowering behaviour, including examples of pleiotropic effects on yield quality and on carbon partitioning. Finally, we discuss the prospects for using this knowledge for the molecular breeding of this sustainable crop via a candidate gene approach.

1. Introduction

Hemp (*Cannabis sativa* L. spp.) is increasingly attractive as multipurpose crop for the sustainable production of fibres, oils and cannabinoids ((van der Werf et al., 1996;Struik et al., 2000;Callaway, 2004;Karus and Vogt, 2004;van der Werf, 2004;Barth and Carus, 2015;Andre et al., 2016) and references therein). As a quantitative short-day plant, photoperiod and temperature input are the key factors that determine the timing of flowering. Adaptation to the latitude of growth, characterized by a specific photoperiod and temperature regime, is very important for hemp production. Hemp flowering is inhibited in a regime of long-day photoperiod (LD) and is induced when a number of short-day photoperiods (SD) have passed (threshold ~10-12 hours uninterrupted darkness; critical photoperiod ~14-12 hrs daylight). If the critical short-day is not reached within the growing season or if cultivars are very late flowering, the plants remain vegetative. Another important aspect of hemp phenology is its sexual dimorphism. Hemp is naturally dioecious, with the unisexual male and female flowers located in separate plants. Male and female plants do not flower and age simultaneously, with the male plants usually flower earlier (protandry) and age earlier. Since male flowers don't produce seed, a high frequency of male plants in the crop will reduce the seed yield. Male plants are also more susceptible to pests but have a finer fibre, which is an advantage for application in textile manufacture. The development of stable monoecious cultivars (hermaphrodite plants, carrying male and female flowers on the same plant) is an important breeding goal. Compared to dioecious cultivars, the monoecious cultivars are more uniform in plant height, stem- and seed production (Borthwick and Scully, 1954;Lisson et al., 2000a;b;Amaducci et al., 2008a;Amaducci et al., 2008b;Amaducci et al., 2012;Salentijn et al., 2015;Small, 2015). However, while monoecious hemp is better for dual harvest of fibre and seed, it is considered that dioecious genotypes are superior for fibre production. There is evidence for genotypic variation in fibre quality but, due to the large variability in fibre characteristics created by environment and large influence of fibre extraction methods the identification of cultivars with specific fibre qualities is very difficult (Berenji et al., 2013;Amaducci et al., 2015).

Knowledge on the typical phenology of hemp is a prerequisite for successful hemp production and breeding for optimal combinations of flowering time and fibre quality in a specific environment. Many aspects of hemp flowering are reviewed in detail by Hall et al. (2012). First of all, the development stage of hemp must be carefully monitored to determine the right moment for harvesting biomass for different products, and flowering time is a main indicator for this. Around the onset of flowering, nutrient flow and carbon partitioning is shifted to the development of flowers and seeds, and less to the development of stems, leaves and roots. Regarding fibre hemp, stem yield shows the highest increase before the onset of flowering and is positively correlated with the duration of the vegetative phase, and this is a reason why late flowering cultivars with a prolonged vegetative phase,

are producing the highest stem biomass (Höppner and Menge-Hartmann, 2007;Faux et al., 2013;Hall et al., 2014b;Tang et al., 2016). The time of maximal stem, bark and fibre yield ('technical maturity') is reached at full (male) flowering (Mediavilla et al., 2001). Flowering time also marks the time-point of secondary fibre formation, from the bottom-upwards in the stem. At the flowering stage the lignification process continuous and intensifies (Keller et al., 2001;Liu et al., 2015), accompanied by a decrease in cellulose and pectin deposition with plant maturity (Liu et al., 2015). This situation results in a proportional decrease in the primary bast fibre layer and increase in secondary bast fibre fraction along the stem. Thus, the quality of the fibres is influenced by the developmental stage at harvest and it differs between different sections of the stem. Based on fibre quality measurements the best quality fibres are obtained from the middle part of stems, harvested around flowering (Keller et al., 2001;Mediavilla et al., 2001;Li et al., 2013;Liu et al., 2018b). In addition to this, the after harvest process "field retting" is important for a good separation of bast fibres from the woody core (shives). This process is also influencing the fibre properties such as the colour, cellulose content and crystallinity of the fibres (Mazian, 2018). The decision of harvest date, should therefore also be determined on the basis of the optimal weather conditions for field retting (not extremely wet or dry). For the dual production of seed and fibre, harvest takes place at seed maturity, resulting in an increased proportion of more lignified and higher content of secondary bast fibres (Amaducci et al., 2015).

A great wish of breeders is to gain more control over the phenology and fibre quality of hemp in order to breed for cultivars with specific combinations of flowering time and seed- and fibre quality properties. Current hemp cultivars still contain levels of genotypic- and phenotypic heterogeneity in sensitivity to photoperiod and instability of monoecy, of which the outcome in different environments is hard to predict. In this respect, the complex phenological traits "flowering time" and "sex determination" are very important traits to consider in fibre hemp breeding. However, currently there is only little molecular information available for hemp. Here, we evaluate current knowledge on genetic components for fibre quality, flowering time control and sex determination in hemp and in other crops, relevant to hemp. Candidate genes for phenology and fibre quality in hemp are proposed and the prospects for using this knowledge for hemp breeding is discussed.

2. Genetic factors involved in "flowering-time": learning from other crops

How are plants capable to sense changes in their living environment, and how is the plant capable to respond to changing environments to ensure the most efficient timing of flowering?

Many studies, on multiple crops, have focused on the results of exhaustive analysis of the numerous genes involved in complex flowering time gene networks in the model crops

Arabidopsis thaliana (long-day (LD) plant) and rice (short-day (SD) plant), that are accessible via interactive databases ((Bouché et al., 2016), http://www.flor-id.org, wikipathways: WP2312 (arabidopsis) & WP2178 (rice)). From these studies a complex picture emerged, whereby flowering is precisely controlled by cross-talk between multiple signalling pathways combining environmental and endogenous factors. This regulation network enables the plant to reproduce in changing environments.

The knowledge of flowering time control is continuously expanding to other species, a broader range of flowering-related traits, such as yield- and stress components, and specific allelic functionalisation of flowering time regulatory genes (Jung et al., 2017). The main pathways regulating flowering time are: 1) the photoperiodic pathway, induced by variation in day length, 2) the gibberellic acid (GA)-dependent pathway, 3) the autonomous pathway, governed by the plant's physiology status, independent of day length, 4) the vernalization and temperature pathway, induced by cold/ambient temperatures, 5) the aging pathway, induced by developmental factors that render the plant competent to flowering and 6) the sugar pathway, in which the sugar status of the plant plays a role (e.g. Cho et al. (2018)).

From studies in model crops, it is known that specific key genes in the regulation of flowering time are acting at distinct stages. The first stage is the perception and transduction of external signals; secondly, environmental- and endogenous signals are transferred to special nodes in the signalling pathway and; thirdly, downstream "integrator genes" confer the capacity to flower to the meristems by the activation of floral meristem identity genes.

Hemp flowering is extremely sensitive to changes in photoperiod and temperature, and therefore the "photoperiodic pathway" and the "temperature pathway" seem to play prominent roles in regulating flowering time in hemp.

The first stages in the "photoperiodic pathway" are the perception of light by photoreceptors (Smith, 2000;Christie et al., 2014;Jenkins, 2014;Galvao and Fankhauser, 2015;Xu et al., 2015;Kong and Okajima, 2016) and signal transduction to the central node of the "photoperiodic pathway" (the transcription factors "GI-CO-FT signalling cascade, also written "*gi-co-ft*" when referring to the genes). The nuclear transcription factor CONSTANS (CO) acting in this node is essential for the induction of expression of *flowering locus T* (*ft*) gene, coding for the mobile flower-promoting signal (FT). Expression of *gene constans* (*co*), is temporal regulated by the "circadian clock" gene *gigantea* (*gi*), which has many other regulating functions in plant development, synchronizing genes in a 24 hours daily rhythm (e.g. Valverde et al. (2004) (*co* gene); Wenkel et al. (2006) (*co* gene) and Mishra and Panigrahi (2015) (*gi* gene)). The next steps are performed by downstream acting "integrator genes" such as *flowering locus T (ft), flowering locus C (flc) and suppressor of overexpression of constans1 (soc1)* (for references see Immink et al. (2012)).

One of the most intriguing integrator genes involved in flowering is *flowering locus T* (*ft* or florigen; member of the *cets* gene family). *Flowering locus T* is the mobile flower-promoting signal that is transported over long distance from the leaf to meristems elsewhere in the plant to induce flowering (for references see e.g. Notaguchi et al. (2008 and Wigge (2011)). It is now clear that specific members of the *cets* gene family have antagonistic functions in maintaining vegetative growth (indeterminate growth) or promote flowering (determinate growth), indicating that flowering is regulated by balancing between family members (McGarry and Ayre, 2012;Lifschitz et al., 2014). *Flowering locus T* codifies for protein FT that originates from specific phloem companion cells of leaf veins (Chen et al., 2018). After loading to the phloem, FT protein is transported to meristems where it promotes the transition from vegetative growth to flowering, together with other "integrator genes", by regulating floral meristem identity genes, which are coding for transcription factors (e.g. LEAFY, APETALA1, AGAMOUS-LIKE 24, and FRUITFULL) (e.g. Lee and Lee (2010)).

An example of a successful approach to model flowering is the specific control of *ft* gene expression by transgenic ectopic overexpression or inactivation of *ft* gene. This approach was applied in angiosperm trees such as poplar, apple, citrus and in a variety of woody and herbaceous species to inducing precocious flowering (early flowering, shortening the juvenile stage), to facilitate an intermediate steps that accelerates the breeding process (for references see Klocko et al. (2016)). However, it has to be considered that in most crops *ft* gene is encoded by small gene-families, with different functionalities and expression profiles among the members, within and among species. While some *ft* family members stimulate flowering, others may have the opposite function and inhibit flowering. Examples are reviewed by Wigge (2011), such as the case of sugar beet (*B. vulgaris*), where two *ft-like* genes are expressed, an activator and a repressor of flowering, and the case of tomato (*L. esculentum*) where one of the *ft* genes was not functioning on flowering but influenced leaf maturation, stem growth and the formation of abscission zones. In such a situation careful selection of specific candidate *ft*-loci is required. An example for the targeted inactivation of a specific *ft* gene, leading to delayed flowering, is in soybean. In soybean, ten *ft* homologs have been found and two of them are confirmed to control flowering (*Gmft2a* and *Gmft5a*). Cai et al. (2018) used the CRISPR/Cas 9 system to specifically knock out *Gmft2a* in soybean resulting in truncated non-functional proteins. The mutations were stably inherited to the next generations and several homozygous, transgene clean lines, without signs of off-target activity were selected. These induced mutants displayed, as expected, late flowering under both SD and LD conditions. However besides *ft*, also other genes in the flowering time pathway could affect growth as it was observed a in late flowering, double loss of function mutant of *soc1* and *fruitfull* genes that displayed a woody phenotype (Melzer et al., 2008).

FLOWERING LOCUS C (FLC) transcription factor mainly integrates signals from ambient temperatures, vernalization and autonomous signals. In many plant species a period

of cold temperature is required in order to promote flowering in the following spring (vernalization). FLC is a floral repressor and, in arabidopsis, a period of low temperature is needed to release this inhibition (Michaels and Amasino, 1999;Bouché et al., 2015;Cheng et al., 2017). *Flowering locus C* genes appeared to have been lost in several lineages of flowering plants (Ruelens et al., 2013). In hemp, the influence of temperature is especially important during the juvenile stage (basic vegetative stage). Hypothetically, it is possible that *flc-like* genes, if expressed in hemp, are involved in timing the completion of the basic vegetative phase that requires a certain temperature input, before entering the photoperiod sensitive phase. Initiation of hemp growth requires a base air temperature of around 1°C, with an optimal temperature for growth of 29°C and a ceiling temperature 41°C. For completion of the basic vegetative stage and floral initiation, temperature degree days in the range 306×Cd to 636×Cd are required (Amaducci et al., 2008a).

The third main integrator gene, *suppressor of overexpression for constans1* (*soc1*) codifies for a MADS-box transcription factor (SOC1) that, together with another transcription factor (AGL24), directly can activate the floral meristem identity gene *leafy* (*lfy*). *Suppressor of overexpression for constans1* gene is acting in response to signals from multiple pathways. FLC transcription factor (acting in the vernalization pathway) can bind to the promoters of *soc1* and *ft* genes to repress their expression (Liu et al., 2008). *Suppressor of overexpression for constans1* gene is indirectly regulated by *constans* gene (*co*), via the levels of FT transcription factor (photoperiod pathway). Expression of *soc1* gene is also regulated via the "aging pathway", via a squamosa promoter-binding-like transcription factor (SPL9) and microRNA156 (see below). Finally, SOC1 transcription factor integrates the "gibberellic dependent pathway" via a yet unclarified mechanism (for more information see Lee and Lee (2010), Immink et al. (2012) and Hyun et al. (2016)).

3. Natural genetic variation in flowering-time

Hemp adapts strongly to the growing season of a given region and therefore it is important to grow the right cultivar for the desired yield, fibre, seed or both. Also, the maturity of the crop at harvest has a strong influence on the fibre quality. As a general rule, late dioecious cultivars are good for fibre production and early cultivars better for seed production. For dual harvest of fibre and seeds, monoecious early- or mid-early flowering cultivars are advised (e.g. Faux et al. (2013), Amaducci et al. (2015) and Tang et al. (2016)). A range of cultivars is already available but there is still need for improvement, for instance by breeding cultivars with specific fibre quality, lower lignin content, lower pectin content, altered architecture or higher seed yield and quality, in combination with specific flowering time (early-, mid-, late-flowering for specific environments) and sex determination (monoecious or dioecious) (Hall et al., 2012;Salentijn et al., 2015).

Early flowering, under non-inductive (LD) conditions and reduced sensitivity to photoperiod is interesting for latitudinal adaptation of hemp and pivotal for reproductive success and good fibre and seed yields in Northern latitudes. Late flowering, with a prolonged vegetative stage is used for adaptation to low latitudes to obtain more yield.

In many crops, genetic variation in specific genes has been employed for the development of selection markers to speed up breeding of cultivars adapted to more extreme environments. Because flowering time is a complex trait, characterization of flowering time pathways and the specific natural allelic variation underlying genetic loci correlated with quantitative traits (QTLs) for flowering time, the development of molecular markers for flowering behavior in hemp is needed. If we consider short-day crops like hemp, examples for the molecular control of flowering time can be found in the monocot rice (*Oryza sativa*) and the dicot legume soybean (*Glycine max* (L.) Merr.).

In the short-day plant rice, natural variation in *ft* homologs is a characteristic for more day neutral flowering. The main *ft* homolog in rice, *Heading date 3a* (*Hd3a*) (Kojima et al., 2002), is upregulated in SD photoperiods and inhibited in LD, to allow flowering only in a SD regime and the central regulation node of flowering Arabidopsis (*gi-co-ft*) is conserved in rice (*Osgi-Hd1-Hd3a*). *Hd3a* and its close homolog *rice flowering locus T 1* (*rft1*) are essential for flowering under SD conditions (Komiya et al., 2008). In specific rice cultivars, that were selected for early flowering in LD conditions *rft1* gene is the major floral activator LD (Komiya et al., 2009). A rice orthologue of *Atsoc1*, *Osmads50* and *Ehd1* are positive regulators of *rft1* gene, while *Hd1* , *Ghd7* , *Ghd8* and *prr37* genes delay flowering under long day conditions resulting in increased plant height, grain number per panicle and grain yields (Yano et al., 2001;Hayama et al., 2003;Doi et al., 2004;Xue et al., 2008;Komiya et al., 2009;Ryu et al., 2009;Zhang et al., 2015;Hill and Li, 2016). Combinations of functional- and non-functional alleles of these floral suppressors, contributes to early flowering in LD and adaptation of rice to specific climates (Xue et al., 2008;Zhai et al., 2015).

In soybean, several natural variants controlling flowering and seed maturity time have been used in breeding for adaptation to the more Northern regions (longer periods of LD). These are mainly variants in the *e* genes (*e1* to *e10*), the *juvenile* (*j*) gene and *ft* genes (*Gmft2a*, *Gmft5a* and *ft4*). The *e1* gene (specific for legume) (Xia et al., 2012) and the *ft4* gene (Samanfar et al., 2017) codify for transcription factors that act as floral repressors by down-regulating *Gmft2a* and *Gmft5a* in LD conditions. *E2* gene is an orthologue of *gigantea* gene (Watanabe et al., 2011), whereas *e3* and *e4* genes are encoding *phytochrome A* genes (respectively *GmphyA3* and *GmphyA2*) (Liu et al., 2008;Watanabe et al., 2009) that integrate red to far-red ratios. The genes underlying loci e5 to e8 have not been identified yet, e9 gene is *Gmft2a* (Zhao, et al., 2016), and *ft4* gene is a candidate gene for the *e10* locus (Samanfar et al. 2017). Plants carrying loss of function alleles for *e1* to *e4* genes lead to photoperiod insensitivity by allowing higher expression levels of the *ft* genes and promoting flowering under long-day conditions (for more references see Zhai et al. (2015), Zhao et al. (2016) and Copley et al. (2018)).

4. Transition to the reproductive stage and bast fibre quality of hemp

Transition to the adult, reproductive phase is an important moment. Around the onset of flowering, nutrient flow and carbon partitioning is shifted to the development of flowers and seeds, and less to the development of stems, leaves and roots. In hemp, the transition from the vegetative to the adult phase can be recognized by a change in the leave arrangement from opposite to alternate, the formation of inflorescences (Hall et al., 2012), a reduction in stem growth (Faux et al., 2013;Tang et al., 2018), and formation of secondary more lignified bast fibres from the bottom-upwards in the stem (Keller et al., 2001;Mediavilla et al., 2001;Li et al., 2013;Liu et al., 2015;Liu et al., 2018a). Furthermore, Liu and co-workers observed a reduction in bast content and thickness of the primary bast fibre layer in stems with plant maturity, that was related to the development and ripening of the seeds (Liu et al., 2015). Bast fibres of hemp are bundles of phloem cells derived from the vascular bundles of stems, with primary bast fibres derived from the procambium and secondary bast fibres from the vascular cambium (Gorshkova et al., 2012). One way in which the quality of hemp fibres can be improved is to reduce the proportion of more lignified and shorter secondary bast fibres relative to primary bast fibres (Amaducci et al., 2015).

With the aim to identify key genes related to hemp fibre quality Van den Broeck et al. (2008) studied differential expression profiles of over 1000 unique hemp genes in bast tissue versus the more lignified core tissue during development. They found that hemp genes acting in five interconnected metabolic pathways (pentose phosphate pathway, shikimate pathway, aromatic amino, acid biosynthesis, lignin biosynthesis, and one-carbon metabolism) were upregulated in the lignified core tissue, suggesting a direct or indirect link with lignin. Most of these genes were also found to be expressed in the bast fibre tissue but at a much lower level. The relative expression levels of some lignin related genes increased during development. For instance, a gene with homology to *caffeoyl-CoA O-methyltransferase*, involved in lignin biosynthesis peaked in bast tissue at 91 days after sowing. More recently, Guerriero and co-workers performed annotation and transcriptional profiling (RNAseq) of over 3000 transcript assemblies in bast fibre tissues derived from different hemp stem sections (Guerriero et al., 2017). It was observed that in comparison to other stem sections, the transcriptome of the older internodes showed enrichment for phytohormone-related genes (e.g. genes involved in auxin metabolism, gibberellic acid-, abscisic acid- and jasmonic acid biosynthesis), together with genes involved in non-cellulosic polysaccharide deposition and lignification. This is in accordance with a high degree of lignification in more mature fibre tissues at the bottom of the stems and gives many leads to candidate genes for further functional analysis (**Table 1**). The important role of jasmonic acid in the stimulation of secondary growth was strengthened by the observations of Behr and co-workers, who showed that exogenous application of jasmonic acid on young hemp plantlets stimulated the formation of additional secondary phloem fibres and enhanced the lignin content

(Behr et al., 2018a). Putative candidate hemp genes for the biosynthesis of monolignols, their oxidative coupling (laccases and class III peroxidases), lignin deposition (dirigent-like proteins) and stereo-conformation of lignans (dirigent proteins) were studied in more detail (on the gene expression and protein level) underpinning their putative functional relation to lignification of hemp bast fibres (Behr et al., 2018b).

In arabidopsis, specific miRNAs (small non-protein coding RNA molecules; microRNA; miRNAs) in the leaves are functioning as a signalling system that allow the plant to monitor the progress in development to adulthood and adequately help to time, and to induce the flowering stage. This signalling operates via a balance between the amounts of two miRNAs, miR156 and miR172. From the juvenile to adult stage of arabidopsis, a decrease in miR156 and an increase in miR172 in the leaves is observed. Expression of *miR156* gene can repress adult leave traits and flowering by binding and inhibiting the genes coding for squamosa promoter-binding-like (SPLs) transcription factors that allow floral transition by activating *miR172* gene. In contrast, miR172 promotes flowering and adult leave traits (Wu et al., 2009;Matsoukas, 2014). In the short-day crop soybean, *miR156* and *miR172* genes were shown to be regulated by photoperiod. Indeed, in soybean, lower miR156 levels (repressor of flowering) and higher miR172 (inducer of flowering) levels were observed under SD- than LD- photoperiods (Liu et al., 2015;Sánchez-Retuerta et al., 2018). Interestingly, *miR156* gene also functions in the phenylpropanoid biosynthesis pathway, where miR156 has been shown to regulate the metabolic flux during flavonoid biosynthesis (Gou et al., 2011;Gupta et al., 2017). The expression of miRNA families (*csa-miR156*, *csa-miR159a*, *csa-miR171b*, *csa-miR172a*, *csa-miR5021a*, *csa-miR6034*) is *in silico* predicted in hemp (Das et al., 2015;Hasan et al., 2016). However, to date no further information is available on the regulatory patterns of specific miRNAs in hemp. An interesting research question would be if lignin and secondary cell wall deposition that occurs around flowering can be attributed to a specific miRNAs. Due to the presence of lignification and high variation in flower characters, hemp seems to be an ideal model crop to study interactions between signalling pathways. This is expected to result in the identification of efficient molecular tools to improve the fibre quality in hemp.

Table 1. A selection of candidate genes for bast fibre quality in hemp.

Candidate gene	Protein description / ortholog	Function; Species	Reference
Wat1	WALLS ARE THIN	Auxin efflux transporter required for secondary wall formation in fibres (Arabidopsis). Upregulated in bast fibres of older, thicker and more lignified stem sections (Hemp)	(Ranocha et al., 2010;Guerriero et al., 2017)
Omt1	Flavone 3'-O-methyltransferase 1	Catalyzes the methylation of monolignols, the lignin precursors. Upregulated in bast fibres of older, thicker and more lignified hemp stem sections (Arabidopsis; Hemp)	(van den Broeck et al., 2008;Moinuddin et al., 2010;Guerriero et al., 2017)
CCoAomt	Caffeoyl-CoA O-methyltransferase 1	Synthesis of feruloylated polysaccharides. Upregulated in bast fibres of older, thicker and more lignified hemp stem sections (Arabidopsis; Hemp)	(Do et al., 2007;Guerriero et al., 2017)
Nac Myb4	MYB- & NAC domain containing protein	Involved in lignin biosynthesis. Several are upregulated in bast fibres of older, thicker and more lignified hemp stem sections (Hemp)	(Zhao and Dixon, 2011;Guerriero et al., 2017;Behr et al., 2018b)
Dlp4 Dlp5	Dirigent-like proteins	Putatively involved in lignin deposition (Hemp)	(Behr et al., 2018b)
Irx12	Laccase4	Oxidative coupling of monolignols (H, G,S-units) (Arabidopsis). Upregulated in bast fibres of older, thicker and more lignified hemp stem sections (Hemp)	(Brown et al., 2005;Zhao et al., 2013;Guerriero et al., 2017;Behr et al., 2018b)
Lox2 4cll7	lipoxygenase 2 and 4-coumarate-CoA ligase-like 7	Jasmonic acid biosynthesis (Arabidopsis; Hemp)	(Bell and Mullet, 1993;Schneider et al., 2005;Guerriero et al., 2017;Behr et al., 2018a)

Genes involved in gibberellic acid signalling (GA) and DELLA transcription factors (see **Table 3**) are interesting candidate genes for all three hemp traits; flowering-time, sex determination and fibre quality, depending on the specific developmental stage- and/or tissues were they are expressed.

5. The diverse roles of gibberellic acid and other phytohormones

Another player in the regulation of flowering time is the phytohormone gibberellic acid (GA). Gibberellins are known to be required for normal growth and development in several species. In arabidopsis, gibberellins are known to be involved in the "gibberellic acid (GA)-dependent flowering pathway" by regulating gene *Atsoc1* and the floral meristem identity gene *leafy* (*Atlfy*) (Hedden and Phillips, 2000;Moon et al., 2003;Mutasa-Göttgens and Hedden, 2009). The GA dependent growth regulated pathway is connected with the light regulatory pathway and the circadian clock via proteins PHYTOCHROME INTERACTING FACTORS (PIFs). *Pifs* genes are regulated by light through phytochrome, and the interaction with the GA-regulated DELLA proteins can block their activity, and thereby suppress the ability of PIFs to promote gene expression and growth (Mutasa-Göttgens and Hedden, 2009;Leivar and Monte, 2014). Also, the biosynthesis pathway of gibberellins is regulated by several endogenous and environmental factors including light, developmental stage, and hormone balance (Hedden and Phillips, 2000). Changes in active GA levels that occur for instance in response to altering light intensities, are influencing plant cell development and the cell wall composition (e.g. Falcioni et al. (2018)). The enzyme GIBBERELLIN 20 OXIDASE (GA20OX) is a key enzyme in the formation of bioactive GAs, whereas GIBBERELLIN 2 OXIDASE (GA2OX) acts in the opposite way by inactivating bioactive GAs. Both enzymes were shown to modulate plant growth when genetically modified. Overexpression of *ga20ox* gene in various plant species resulted in e.g. increased seed yields, biomass increase, longer xylem fibres, longer and larger leaves, whereas knockdown of *ga2ox* gene resulted in increased (tobacco) growth- and fibre production (reviewed in Lima et al. (2017)). Transgenic tobacco plants expressing the arabidopsis genes *ga20ox* or *ga2ox* showed, high and low GA levels, respectively, resulting in elongated and stunted pants, respectively. The effects on dry matter accumulation that were found among these transgenic tobacco plants were most likely due to changes in lignin deposition. This was the result of the upregulation of genes acting in lignin biosynthesis at increased GA levels (Biemelt et al., 2004). Overexpression of *ga2ox* in jatropha and arabidopsis induced dwarfs with smaller leaves, flowers and fruits, with a late flowering effect observed only in the latter (Hu et al., 2017). In hemp, exogenous application of GA is used to induce male flowers on female plants but, male plants showed no change in sex determination when treated with GAs (Ram and Jaiswal, 1972). Exogenous application of GA on leaves increased the growth of hemp and the treated plants showed a greater number of fibres compared to controls. The individual fibres were larger in diameter, more lignified and up to ten times as long as the fibres from the untreated plants (Atal, 1961). Application of the phytohormone jasmonic acid to young hemp plantlets resulted in an increased secondary growth, as well as the formation of additional secondary phloem fibres, increase in lignin deposition and upregulation of lignin-related genes (Behr et al., 2018a). Also, fibres in the bottom parts of hemp stems were enriched for the expression of genes involved in GA biosynthesis and the biosynthesis of other phytohormones (Guerriero et al., 2017) pointing at the involvement of phytohormones in the regulation of secondary fibre growth.

6. Genetic components of sex determination

Hemp has a diploid genome (2n = 20) composed of 9 pairs of autosomal chromosomes and one pair of sex chromosomes. Like in human, the gender of hemp is known to be influenced by a XY chromosome system. The hemp males are always XY, while females carry the XX karyotype (Ainsworth, 2000;Moliterni et al., 2004;Ming et al., 2011). Monoecious hemp, with the female and male flowers located in the same plant, has generally the female XX karyotype (Faux et al., 2014). The key factors that are driving this sexual dimorphism are still unknown (Westergaard and Demerec, 1958;Ainsworth, 2000;Matsunaga and Kawano, 2001;Ming et al., 2011). The sex determination seems to be more stable and definite in the male XY karyotype, showing the typical male morphology. However, the ability to develop male flowers on monoecious XX karyotypes shows that the male determining and/or female suppressing factors are not necessarily located in the Y chromosome (Faux et al., 2016). To identify sex linked genomic sequences in hemp, linkage mapping has been performed (Mandolino and Ranalli, 2002;Peil et al., 2003;Faux et al., 2016). Faux and co-workers used populations, segregating for male and female plants, to map several sex linked QTL loci, putatively located in sex chromosomes. Furthermore, groups of markers co-segregating with sex- and with stability of sex determination were found (Faux et al., 2016). Comparison of gene expression (cDNA-AFLP) in early male and female apices resulted in the identification of several differentially expressed fragments, with homology to genes coding for a permease, a ubiquitin (SMT3-like protein), heavy chain of a kinesin 9 protein, and a Rac-GTP binding protein, which may be involved in auxin regulated gene expression (Moliterni et al., 2004).

Regarding sex determination an obvious similarity is found between spinach (*Spinacea oleracea*) and hemp. Like hemp, spinach is dioecious with occasionally monoecious plants in specific lines and crosses, but, in contrast to hemp no heteromorphic sex chromosomes are observed (Ramanna, 1976). Sex determination in spinach is determined by a locus for sex determination carrying the Y and X alleles, whereas monoecy is controlled by a single, incomplete dominant gene on the M locus that is closely linked to the X/Y locus, as was determined in a specific breeding line (Yamamoto et al., 2014). In the presence of the incomplete dominant M allele, female plants (XmXm and XXm) are monoecious, whereby the homozygous XmXm plants show a higher degree of male flowers compared to the XXm plants (M masks X) and, because Y is dominant over X and M, YX and YXm plants are male. So, M is a male promoting, female suppressing factor but is less effective than the Y allele. In spinach monecious lines are used for breeding because of the high degree of homozygosity and high-male monoecious lines are wanted therefore as male parents in breeding programmes (Yamamoto et al., 2014). Also, in both spinach and hemp, gibberellins promote masculinization. Recently, West and Goldberg studied the role of gibberellic acid signalling (GA) in sex determination of spinach and came up with an interesting model

for the action of genes underlying sex determination in spinach (West and Golenberg, 2018). They observed differential expression of the *gibberellic acid insensitive* gene (*Spgai*), which codifies for a transcription factor of the DELLA family, among female and male inflorescences, with a high expression of *Spgai* gene observed in female inflorescences. Based on gene function analysis studies, a signalling pathway towards sex determination was proposed in reaction to GA application. In short: high levels of GA inhibit *Spgai* gene. GAI transcription factor inhibits the expression of spinach *B-class* homeotic genes, which are masculinizing factors that stimulate male organ formation and at the same time suppress the development of female organs in flower primordia. So, in conditions of high GA levels (external GA application) the GAI transcription factor content is reduced, resulting in the release of the inhibition on the *B-class* homeotic genes, formation of male organs and inhibition of female organ development. Indeed, in female inflorescences a two-fold higher expression of *Spgai* was observed compared to male plants, which is in agreement with a higher GAI transcription factor content leading to female organ development.

7. Pleiotropic effects and carbon partitioning

Variation in flowering time is often linked to variation in developmental traits, such as plant height, ear height (in maize), seed yield, seed quality traits, leave number, cell wall composition and secondary growth (e.g. Melzer et al. (2008) in arabidopsis, Durand et al. (2012) and Vanous et al. (2018) in maize, Cober and Morrison (2010) and Copley et al. (2018) in soybean, Shen et al. (2018) in *Brassica napus* and Petit et al. (2019c) in hemp). Members of the *ft* gene family may be involved in these pleiotropic effects (see above), but also other genes operating in signalling networks may connect flowering traits with development and growth. For instance, based on expression network profiling using a late flowering, woody double mutant (*soc1* and *ful* genes) of arabidopsis (Melzer et al., 2008), three genes with dual function in growth and flowering were indicated as potential candidates for the link between the flowering pathway and growth (*xal1*, *an3* and *rem1*) of which one, *an3*, has *ft*-like properties (Davin et al., 2016).

This correlation of traits complicates selection procedures, since negative co-effects on traits have to be considered. An example is soybean, where earliness is often accompanied by a loss in seed yield and quality. To examine these pleiotropic side-effects of early flowering, a series of isogenic soybean lines carrying "photoperiod insensitive alleles" (at genes *e1*, *e2*, *e3*, *e4*, and *e7* for early flowering, under LD, see above) was monitored for multiple agronomic traits. The whole series of isogenic lines, including lines with mutations in multiple loci, provided a range of flowering times, maturities, and yields. For isogenic lines with a single mutant locus, early flowering was often associated with shorter plants, reduced lodging and early maturity but unfortunately also with reduced seed yields. Among the lines with multiple mutations some interesting lines with zero yield reduction were found, which

might be due to additive- or epistatic effects of combined alleles (Cober and Morrison, 2010). In a search for novel loci and genes for photoperiod insensitivity and maturity in soybean, Copley and co-workers performed genome-wide association studies (GWAS) and identified several novel loci for maturity traits (Copley et al., 2018). However, as most traits were correlated, also most QTLs were co-localized. This correlation of phenotypes can be explained by either a clustering of several genes in a locus or by a "pleiotropic effect" of a single gene on several traits.

Such correlated changes in phenotypic patterns, may reflect the shifts in carbon-partitioning that take place during development, which affect the overall plant morphology. In earlier cultivars, less biomass is accumulated in stem and leaves and therefore less carbon is available for seed production. Interestingly, overexpression of an *agamous-like* MADS-box gene (*Gmagl1*) codifying for a transcription factor induced early flowering in soybean, but without negative effects on seed production or on oil and protein content in seeds (Zeng et al., 2018). The only pleiotropic effect of earliness in these transgenic lines was that they had smaller petals and shortened inflorescences. Based on this, it was hypothesized that the transgenic plants may compensate for the energy required for developing fruiting organs by reducing a further allocation to vegetative organs (shortened inflorescences and slightly reduced growth of petals).

In arabidopsis, it was shown that the shift in carbon partitioning during development is tightly controlled and involves the action of sucrose transporters (SUTs), hexose transporters (STPs) that function in uptake to a cell, and transporters for export out of the cell (SWEET), as well as sucrose cleavage enzymes, such as cell wall invertases (CINs), vacuolar invertases (VINs), and sucrose synthases (SUSs). Besides their multiple roles, among which acting as energy resource (sugars), storage molecules (starches), and structural components (fibres), carbohydrates can act as signal molecules (Cho et al. (2018) and references therein). Trehalose-6-phosphate and hexokinase 1 are such important signalling metabolites, regulating carbon assimilation and sugar status in plants. At flower induction, sugar consumption for growth is reducing and the remaining glucose that is accumulating in the phloem of leaves can eventually promote expression of florigens, while trehalose-6-phosphate functions in the shoot apical meristem to promote the flowering signal pathway downstream of those florigens (Ponnu et al., 2011;Matsoukas, 2014;Cho et al., 2018).

In hemp, the upregulation of genes acting in lignin biosynthesis in older bast fibres (Guerriero et al., 2017) may reflect carbon partitioning towards lignin biosynthesis in phloem tissues around flowering.

8. A candidate gene approach towards genetic control of hemp phenology and fibre quality

Flowering time, sex determination and fibre quality of hemp are quantitative traits that are governed by many genetic loci, each with a certain effect on the phenotype in a specific environment or at a certain developmental stage, or in general. A "candidate gene approach" can contribute to increase the knowledge about these traits. The identification of biosynthesis- and signal routes that play a role in the traits enable the identification of candidate genes with the greatest effects on the downstream phenotype, and to predict pleiotropic effects on other traits. A selection of genes that are hypnotized to have profound effects on phenology and bast fibre quality in fibre hemp are shown in **Table 1**, **2** and **3**.

Hemp orthologues for genes acting in flowering time signalling pathways are putative candidate genes for the regulation of flowering time in the SD plant hemp. A selection of promising candidates is shown in **Table 2**.
The most obvious candidates are orthologues of the "*gi-co-ft*" core genes of the "photoperiodic pathway" and genes coding for the phytochrome receptors. In other SD crops, early flowering was often observed in plants carrying nonsense mutations in genes that are repressors of flowering in LD conditions (e.g. orthologues to soybean *e1* to *e4* genes, (Langewisch et al., 2017) and orthologues of rice *Ghd7*, *Ghd8*, *prr37* and *phyB* (e.g. Xue et al. (2008)). Later flowering can for instance be found in plants with nonsense mutations in the florigenes or other genes that stimulate flowering (Cai et al., 2018).

When considering a candidate gene one has to take into account that many genes, acting in the flowering time signalling networks, have pleiotropic effects on other traits. In addition, these genesmay belong to a gene family of which the individual family members have different functions. For instance, the "circadian clock" genes (e.g. *gi* gene) and other "circadian" genes involved in regulation of flowering time are involved in many biological processes, and can result in pleiotropic effects on for instance floral transition, leaf movement, stomata opening, seed germination, and hypocotyl elongation (e.g. Ding et al. (2007), Kolmos et al. (2009), Wenden et al. (2011), Mishra and Panigrahi (2015) and Shim and Imaizumi (2015)). Also, squamosa promoter-binding-like proteins (SPLs) belong to a family of functionally specialized transcription factors with multiple roles in plant phase transition, flower and fruit development, plant architecture, gibberellins signalling, sporogenesis, and response to copper and fungal toxins (Preston and Hileman, 2013).

Table 2. A selection of candidate genes for controlling flowering-time in the short-day crop hemp.

Candidate gene	Protein description/ortholog	Function; Species	Reference
E2	GIGANTEA	Photoperiod sensitivity (Soybean)	(Watanabe et al., 2011)
GmphyA2	*Phytochrome A*	Photoperiod sensitivity (Soybean)	(Liu et al., 2008)
GmphyA3			(Watanabe et al., 2009)
Gmft2a, Gmft5a	FLOWERING LOCUS T	Promoting flowering (Soybean)	(Kong et al., 2010)
GmFT4 &e1	FLOWERING LOCUS T	Repressors of *GmFT2a* and *GmFT5a* in LD (Soybean)	(Samanfar et al., 2017)
			(Xia et al., 2012)
J	EARLY FLOWERING 3	Relieving the suppression of FT expression by E1; loss of function alleles show delayed flowering (Soybean)	(Lu et al., 2017)
Hd3a	HEADING DATE 3A	Promotes flowering in SD (Rice)	(Monna et al., 2002)
	FLOWERING LOCUS T		(Kojima et al., 2002)
Rft1	RICE FLOWERING LOCUS T 1	Promotes flowering in SD & LD (Rice)	(Komiya et al., 2008;Komiya et al., 2009)
Osmads50	MADS-box transcription factor 50/*Atsoc1*	Promotes flowering in LD (Rice)	(Ryu et al., 2009)
			(Komiya et al., 2009)
Ehd1	Two-component response regulator ORR30	Promotes flowering in SD (Rice)	(Doi et al., 2004)
Hd1	Zinc finger protein HD1 / *constans*	Inhibition of flowering in LD (Rice)	(Hayama et al., 2003)
Ghd7	Transcription factor GHD7		(Xue et al., 2008)
Ghd8/Hd5	Nuclear transcription factor Y subunit B-11		(Yano et al., 2001)
Prr37	Two-component response regulator-like PRR37		(Yano et al., 2001)
Flc	MADS-box protein FLOWERING LOCUS C, AGAMOUS-LIKE 25	Temperature dependent flowering. Repressor of flowering (Arabidopsis)	(Michaels and Amasino, 1999)

LD=long-day photoperiod; SD=short-day photoperiod

Table 3. A selection of candidate genes for sex determination, growth and development in hemp. Genes involved in gibberellic acid signalling (GA) and DELLA transcription factors are interesting candidate genes for all three hemp traits; flowering-time, sex determination and fibre quality, depending on the specific developmental stage- and/or tissue were they are expressed.

Candidate gene	Protein description/ortholog	Function; Species	Reference
Gai-like	DELLA protein / *gai* gene	May inhibit B-class homeobox genes, that promote male organ development. Upregulated in female inflorescences of spinach (Spinach).	(Peng et al., 1997) (Dill et al., 2001) (West and Golenberg, 2018)
Gid1	Gibberellin receptor GID1	Gibberellin (GA) receptor; Interacts with DELLA proteins in the presence of GA4 (Rice, Arabidopsis).	(Nakajima et al., 2006) (Griffiths et al., 2006) (Ueguchi-Tanaka et al., 2005)
Ga20ox	Gibberellin 20 oxidase	Key oxidase enzymes in the biosynthesis of gibberellin (Rice, Arabidopsis).	(Phillips et al., 1995) (Rieu et al., 2008b)
Ga2ox	Gibberellin 2-beta-dioxygenase	Catabolism of biologically active gibberellins; GA homeostasis (Rice, Arabidopsis).	(Thomas et al., 1999) (Rieu et al., 2008a)
Spl	Squamosa promoter-binding like transcription factor	A family of plant-specific transcript factors that play crucial roles in the regulation of plant growth and development.	(Klein et al., 1996) (Preston and Hileman, 2013) (Liu et al., 2016)

Regarding sex determination genes involved in gibberellic acid signalling (GA) and DELLA transcription factors are interesting candidate genes that may also have a side-effect on fibre quality (**Table 1** and **3**).

In a situation, where different quantitative characteristics have to be combined, the ability to select in an early stage for plants with specific flowering characteristics would already be an important step for breeding. Genetic variation at candidate gene loci can be utilized to select specific haplotypes via "haplotype tagging SNPs" (htSNPs). These htSNPs improve the efficiency of association studies performed for the selection of alleles in the population that are associated with phenotypic variation in the trait (Ehrenreich et al., 2009). In short day crops, such as soybean and rice, molecular markers for maturity and flowering time based on genetic variation in candidate genes for flowering time are already used (e.g. Langewisch et al. (2017) in soybean and Shabir et al. (2017) in rice). In hemp molecular markers have mostly been developed for forensic studies to differentiate drug-type cannabis from hemp or for the early detection of male-plants ((Mandolino et al., 2002); reviewed in Onofri and Mandolino (2017)).

It should also be stressed that finding "candidate genes" across species has limitations because the function of candidate genes across species may be similar but often not identical (e.g. Salentijn et al., (2007) and Wong et al., (2014)). So, a prerequisite for a successful application of the "candidate gene approach" is functional knowledge of candidate genes in hemp. At present, the knowledge of gene function and gene expression in *Cannabis sativa* is still limited, and mainly focused on genes acting in cannabinoid biosynthesis. Several molecular technologies such as "genome editing" and "targeted mutagenesis" contribute to gene functional analysis and to the generation of plants with specific mutations. Genome editing by the CRISPR/Cas9 system appears to be a very precise and efficient tool for functional analysis of specific genes and the development of useful mutants in several crops (e.g. Hille et al. (2018) and Schindele et al. (2018). The system yet requires genetic transformation and regeneration of transgenic CRISPR/Cas9 plants from undifferentiated cells (callus tissue) or protoplasts. Regarding hemp, protocols for shoot regeneration from callus are known, but these work only efficient for specific hemp accessions (Andre et al., 2016;Chaohua et al., 2016) and, to our knowledge, cases of efficient production of transgenic hemp plants, produced via *Agrobacterium* mediated genetic transformation have not been published yet. As such, hemp is still considered a recalcitrant plant for genetic modification and thus for CRISPR/Cas9.

If mutants can be obtained in a less recalcitrant hemp cultivar, the specific mutations in such transgenic lines can be delivered to breeding lines, via cross breeding. However, this will introduce also unwanted traits and the Cas9-gRNA cassette in the receiving parent, and many subsequent breeding steps are required to restore original traits. Recently, Kelliher, Wang and co-workers published a new approach (Haploid-Inducer Mediated Genome Editing) to overcome such problems in maize cultivars (Kelliher et al., 2019;Wang et al., 2019). This approach combines the technology of haploid induction with CRISPR/Cas9 genome editing and requires specific haploid inducer lines (e.g. carrying homozygous mutations in *cenh3* gene for dicots) that are stably transformed with constructs expressing the CRISPR Cas9-gRNA editing tools. The gametes of such lines can transfer the editing tools to recalcitrant cultivars (via cross breeding instead of genetic transformation). Due to the haploid inducer, the genome carrying the CRISPR sequences and the haploid inducer, is eliminated short after fertilization and haploid embryos are formed that, upon chromosome doubling, can grow into plants that yield 100% inbred seed. It appeared that the short time of interaction of the two genomes after fertilization was enough to induce specific mutations in the recipient genome. Above all, the CRISPR genes are not present in genome of the resulting crop which can be advantageous in connection with GMO regulations. Application of this system in hemp is not to be expected in the short term since transformable hemp accessions, together with a haploid induction system are not immediately available.

Targeted mutagenesis or TILLING (McCallum et al., 2000) is another way to select for plants with mutations in specific genes. For this strategy seeds or pollen are treated with specific chemicals that make at random point-mutations throughout the genome. Large populations of mutated plants are then screened for the presence of mutations in specific genes using high throughput sequencing, or other screening technologies. This strategy was used in hemp to find plants with specific induced knock-out and missense mutations in *Csfad2* and *Csfad3* genes leading to altered seed-oil composition in the seed hemp cultivar Finola (Bielecka et al., 2014). Such a strategy requires facilities to grow large mutant populations and for seed storage and breeding steps to obtain homozygous mutations or combine different mutations.

A very useful tool for hemp genomics is the draft genome sequence of hemp (covering 534 Mb of the haploid hemp genome that is 818 to 843 Mb in size) published by Van Bakel and co-workers, including more than 30,000 transcripts assemblies (NCBI TSA: JP449145.1 to JP482359.1; PK00001.1 to PK29878.1), and the "*in silico*" gene expression profiles of these genes (Massimino, 2017) (van Bakel et al., 2011). Two initiatives to improve the hemp genome were undertaken, that independently resulted in the assembly of the hemp genome in 10 pseudomolecules (scaffolds, separated by gaps) representing the 10 different chromosomes of hemp (2n=2x=20) (Grassa et al., 2018;Laverty et al., 2019). It was experienced that the assembly of the hemp genome was complicated by the presence of large quantities of repetitive DNA (~73% of the hemp genome), the heterozygous character of hemp (van Bakel et al., 2011;Sawler et al., 2015), and an expected high degree of karyotype polymorphisms among hemp cultivars (Razumova et al., 2016). This situation was approached by using long read sequencing technologies (PacBio SMRT, Nanopore sequencing) next to the standard Illumina sequencing technology to span large stretches of repetitive DNA. For the assembly of the genome a combination of physical- and genetic mapping was applied (Laverty et al., 2019). It was found that most recombination events occurred in the gene rich regions near the chromosome ends. Furthermore, three pseudomolecules appeared to have recombination only on a single arm of the chromosome (telocentric) and one of these may represent the sex chromosome, whereas the other two may represent the chromosomes that harbour 5SrDNA and 45SrDNA (Laverty et al., 2019). The map is still not completed (see NCBI assembly no. GCA_003417725.2 & GCA_000230575.4; GCA_900626175.1) and not all known transcripts and male-specific markers could be mapped. Dedicated genetic mapping and sequencing strategies may further unravel the complex genetic structure of the hemp genome, and may detect hemp lines that accommodate specific genetic variation.

9. Concluding remarks

Here we review aspects of the traits "flowering time" ,"sex determination" and "fibre quality" that are relevant to hemp. This information can be utilized to predict putative candidate genes, which can serve as targets for the development of molecular markers for these traits. For the development of such breeding tools, it is important to know the allelic variation underlying candidate genes that is responsible for the phenotypic variation. A big advantage for hemp is the presence of a high level of natural genotypic- and phenotypic variation, which makes it possible to perform efficient GWAS studies to validate the putative biological function of candidate genes, and to discover novel genomic regions involved. Furthermore, we like to point to the importance of high throughput phenotyping protocols, which are needed to map QTL loci, including small effect loci.

10. Acknowledgements

This research was conducted as part of the MultiHemp project (Multipurpose hemp for industrial bioproducts and biomass) funded by the European Union's Seventh Framework Programme for research, technological developments and demonstration under grant agreement number 311849, and co-funded by the Dutch ministry of Agriculture (LVN) as part of the research theme "Total use of Resources for Food and Materials".

Chapter 6

Genetic architecture of flowering time and sex determination in hemp (*Cannabis sativa L.*): a Genome-Wide Association Study

Jordi Petit[a], Elma M. J. Salentijn[a], Maria-João Paulo[b], Christel Denneboom[a] and Luisa M. Trindade[a*]

[a] Wageningen UR Plant Breeding, Wageningen University and Research, P.O. Box 386, 6700 AJ Wageningen, The Netherlands
[b] Biometris, Wageningen University and Research, P.O. Box 100, 6700 AJ Wageningen, The Netherlands
[*] Corresponding author: Luisa M. Trindade (luisa.trindade@wur.nl)

To be submitted as:
Jordi Petit, Elma M. J. Salentijn, Maria-João Paulo, Christel Denneboom, and Luisa M. Trindade. Genetic architecture of flowering time and sex determination in hemp (*Cannabis sativa* L.): a Genome-Wide Association Study.

ABSTRACT

Flowering time and sex determination of hemp (*Cannabis sativa* L.) strongly influence fibre quality and seed production of the crop. The control of these traits is paramount for the breeding of new cultivars. Yet, little is known about the genetics underlying such complex traits and a better understanding requires in depth knowledge of the molecular mechanisms responsible for these traits. In this report, the genetic architecture of flowering time and sex determination in hemp was studied using a Genome-Wide Association Studies (GWAS) approach. Association studies were performed on a 123 hemp accession panel, tested in three contrasting environments, using a set of 600K SNP markers. Altogether, eight QTLs were identified across environments; six for flowering time traits and two for sex determination. These QTLs covered genomic regions with 33 transcripts predicted to be involved in flowering and sex determination as well as a microRNA, *miR156*. Genes related to perception and transduction of light and transcription factors well-known to regulate flowering were identified in QTLs for flowering time traits. Transcription factors and genes involved in the balance of phytohormones, specially auxins and gibberellic acid, were identified in QTLs for sex determination. Sex determination QTLs were associated with the development of male flowers in female plants and thus with the stability of sex determination in monoecious plants. The present study elucidates relevant knowledge on the genetic mechanism of flowering and sex determination traits in hemp, and provides new tools for breeding hemp.

1. Introduction

Hemp (*Cannabis sativa* L.) is naturally a dioecious species with male and female flowers in different individuals. Yet, monoecious genotypes can spontaneously occur bearing both male and female flowers in the same plants. Sexual dimorphism is characteristic between male and female hemp plants. True males can be recognized by their typical morphology characterised by slender stature, few leaves and hanging inflorescences carrying male flowers. Female plants produce their female, pistillate flowers in dense panicles heads interspaced with leafy bracts, whereas the morphology of monoecious hemp resembles female plants (Faux et al., 2016). Male plants flower mostly preceding female ones (Bócsa and Karus, 1998;Struik et al., 2000) and male plants die after flowering, while females remain alive until seed maturation. Lignification intensifies around the onset of flowering and continues until seed maturation. As a consequence, female plants are more lignified than males (Liu et al., 2015). Male plants are also known to produce finer fibres than females but males are more susceptible to pests (Amaducci et al., 2015). In addition, the proportion of males and females has an effect on the seed yield, as large male proportion is associated to a seed yield reduction. A recent study reported large differences in the content and quality of fibres between dioecious and monoecious plants. The study reported that monoecious were associated to larger bast fibre content, and higher contents of cellulose and mannan, while dioecious showed larger vigour plants with later flowering and higher contents of xylans and lignin (Petit et al., 2019c). Furthermore, dioecious hemp species are characterized by heterogeneity in fibre and seed production. In contrast, monoecious genotypes are more stable in both fibre and seed production because plants are more uniform (Mandolino and Carboni, 2004;Salentijn et al., 2015). The differences in properties between dioecious and monoecious plants have traditionally led to a specialization of their applications, as sex determination of hemp strongly influences fibre quality and seed yield (reviewed in Salentijn, et al. (2019)). As a consequence, dioecious are mostly used for fibre production, while monoecious are produced for dual purpose: fibre and seed production (Amaducci and Gusovius, 2010;Amaducci et al., 2015).

Hemp is a diploid species with 9 pairs of homomorphic autosomal chromosomes and a pair of heteromorphic sex chromosomes: X and Y (2n=20) (Moliterni et al., 2004). The haploid genome size is 818Mbp for female plants and 843Mbp for male plants (van Bakel et al., 2011). Sex determination system in dioecious hemp has been well studied and scientists have agreed that male plants carry the heterogametic sex (XY) and female plants the homogametic one (XX). However, despite the presence of specific sex chromosomes, some dioecious hemp plants produce flowers of the opposite sex than the one determined by their genetics. Monoecious hemp plants carry the homogametic sex (XX) and the ratio of female to male flowers in a single monoecious plant is highly variable. This variation ranges from monoecious plants that have predominantly male flowers to predominantly female flowers

(Faux et al., 2013;Faux et al., 2014;Faux et al., 2016). Dioecious hemp species abundantly exist in nature, while monoecious plants have been selected during the domestication of the plant. Yet, the monoecious trait is particularly recessive and shows unstable sex expression, which tends to lead to a distribution of sex ratios, including unisexual plants, and a gradual return to natural dioecy after few generations (Bócsa and Karus, 1998;Amaducci et al., 2008b). Constant strict selection of monoecious plants is thus necessary to maintain the monoecy of some cultivars (Faux et al., 2014).

Some studies suggested that in hemp, besides genetic mechanisms, epigenetic mechanisms might probably also affect the control of sex determination (Heikrujam et al., 2014). Soldatava and Khryanin suggested that sex determination in hemp might be affected by the switch of genes at the transcriptional level or post-transcriptional level, without changing the DNA sequences (Soldatova and Khryanin, 2010). They observed that accumulation of Cu^{++} and Zn^{++} ions with zeatin induced feminization, whereas accumulation of Pb^{++} ions and phytohormone GA favoured a masculinization effect. Thus, the sex determination in hemp is affected by external factors, which also include hormonal treatments (Chailakhyan and Khryanin, 1978;Freeman et al., 1980;Galoch, 2015;Faux et al., 2016).

Flowering time is another phenological trait with great relevance for hemp breeding. Hemp is a short-day plant strongly sensitive to changes in the photoperiod regime, temperature input, rainfall and stress factors (Amaducci et al., 2008b;Amaducci et al., 2012;Salentijn et al., 2015;Petit et al., 2019c;Salentijn et al., 2019). Flowering time of hemp rapidly responds to these environmental variations to synchronise the flowering with the environment, especially with the day-length of the growing season (Amaducci et al., 2012;Salentijn et al., 2019). Hemp flowering is inhibited during long-day photoperiod regimes and is induced when the photoperiod regime shifts to ~10-12 hours of uninterrupted darkness. Temperature input is of special importance for development in the juvenile stage of hemp (references in Amaducci et al. (2008b), Amaducci et al. (2012) and Salentijn et al. (2019). Transition to flowering has large consequences on the vegetative growth of hemp. Around the onset of flowering, the nutrient flow and the carbon partitioning is shifted from the stems, leaves and roots towards the development of flowers and seeds. Moreover, flowering time also marks the time-point of secondary bast fibre formation, which is characterized by intense lignification (Liu et al., 2015). Consequently, flowering time also has important implications for the fibre quality of hemp (Salentijn et al., 2019).

A breeding goal of paramount importance for hemp is to gain control of flowering time and sex determination. This will help to breed for cultivars better adapted to specific photoperiod regimes and with desired seed and/or fibre yields and qualities. For instance, less sensitivity to photoperiod for specific production areas, and more stability to monoecy are a desired breeding goals for seed production. Meanwhile, late flowering, more dioecy control and a constant fibre quality are important breeding goals for developing fibre type

cultivars (Hall et al., 2012;Amaducci et al., 2015;Salentijn et al., 2015). Notwithstanding the interest in controlling sex and flowering in hemp, the genetic and molecular knowledge of this crop is limited (Salentijn et al., 2015). Previous studies described large quantitative variation in sex expression in monoecious hemp (Faux et al., 2013;Faux et al., 2014). In addition, Petit and co-workers reported large range of variation in flowering time traits and sex determination in a panel of 123 hemp accessions (Petit et al., 2019c). The study described strong influence of genetic components with large heritability estimates for flowering time and sex determination. In addition, small but significant genotype-by-environment interactions (*G*×E) were reported. These interactions indicated significant heritable variation of flowering time sensitive to the environment. Besides these pioneering studies, in hemp little is known about the genetic mechanisms that control sex determination, the origin of high plasticity of sex expression, sexual dimorphism and flowering time.

The present study describes a GWAS approach to characterize the genetic architecture underlying the length of the vegetative period, flowering time, and sex determination. A panel of 123 hemp accessions was used to measure these traits in three locations across Europe with contrasting photoperiod regimes. The hemp panel included a large variation of monoecious and dioecious genotypes and early and late flowering accessions (Petit et al., 2019c). Here, we present QTLs for flowering time traits and sex determination, and characterize whether these QTLs are present in all locations, or present only in particular environments. Furthermore, candidate genes for the QTLs across locations were identified and will contribute to a better understanding of the molecular mechanisms behind flowering and sex determination in hemp.

2. Materials and Methods

2.1 Plant material

A panel of 123 hemp accessions was used in this study. The panel included a large diversity of accessions from various origins and primarily used for different purposes (**Table 1**, **Chapter 3**). The accessions of the panel included large variation in time of flowering and in type of sex, namely dioecious, monoecious and accessions with large range of variation in between (Petit et al., 2019c). Plants were grown in three locations across Europe with different photoperiod regimes; in Rovigo (CRA - Centro di ricerca cerealicoltura e colture industriale, Italy, 45˚N 11˚E), in Chèvrenolles, Neuville-sur-Sarthe (FNPC - Fédération Nationale des Producteurs de Chanvre, France, 48˚N 0.2˚E), and in Westerlee (VDS - VanDinter Semo BV, The Netherlands, 53˚N 6˚E). Field testing was performed between April and September 2013. Three biological replicates were grown per accession and location in a randomized complete block design and the experimental units were plots of $1m^2$ in CRA and VDS and of $1.5m^2$ in FNPC.

2.2 Phenotyping of the GWAS panel

Flowering time traits were measured in 10 plant per plot and data were provided as the mean per plot. Emergence of the plants was scored in one row per plot at day=N, N+2, N+4, N+7, where N is the day of sowing. Beginning of flowering (FL_Begin in Σ˚C (the accumulated Celsius degree day over a period at a temperature of 1˚C)) and full flowering (FL_Full in Σ˚C) were calculated relative to the day of emergence as follows:

$$FL\_Begin = \Sigma^{\circ}C_{Beginning\ flowering} - \Sigma^{\circ}C_{Emergence}, \quad (1)$$

$$FL\_Full = \Sigma^{\circ}C_{Full\ flowering} - \Sigma^{\circ}C_{Emergence}, \quad (2)$$

where $\Sigma^{\circ}C_{Beginning\ flowering}$, $\Sigma^{\circ}C_{Full\ flowering}$ and $\Sigma^{\circ}C_{Emergence}$ are the accumulated Celsius degree days respectively at the beginning of flowering, at full flowering and at the day of first emergence. The length of vegetative growth period (VEG in days) is the period of vegetative growth of the plants in days, as measured from the day of first emergence until FL_Begin. Sex determination was phenotyped per plot assessing "1" when the plants were dioecious, "2" when the plants were a mix of dioecious and monoecious and "3" when the plants were monoecious.

2.3 Genotyping of the GWAS panel

The genotyping of the GWAS panel was performed using restriction-site associated DNA sequencing (RAD-seq), essentially as previously described by Petit et al. (2019b). Briefly, high quality genomic DNA (2.5 to 5 µg at a concentration ≥25 ng/µl) of the plants was extracted following a modified CTAB method (Petit et al., 2019b). To cover all allelic variation within accessions, the genomic DNA of eight plants per accessions were pooled in equimolar amounts, resulting in 123 samples. RAD libraries with insert sizes of 300 to 550bp, were prepared for each sample using the restriction enzyme *Eco*RI, as described by Baird et al. (2008). The 123 samples were paired end sequenced on an Illumina platform (PE150) in two rounds to provide 2 x 1 Gbp genomic data per sample.

After sequencing, quality check adaptors were trimmed from the sequences and low quality reads were removed. Low quality reads comprised reads with >50% of the bases Q≤12, unknown bases >3%, reads that lack a part of the multiplexing barcode and could not be identified, and reads lacking the key sequence of the enzyme used. Subsequently, clean reads were aligned to the reference *C. sativa* 'Purple Kush' assembly, (canSat3 version GCA_000230575.1) (van Bakel et al., 2011), using Burrows-Wheeler Alignment Tool (specific BWA parameters: (o) max number of or fraction of gap opens=1; (e) max number of or fraction of gap extensions=50; (m) maximum entries in the queue=100000) (Li and Durbin, 2009). Picard-tools (v1.118) was used to sort the Sequence Alignment Map (SAM) files by coordinate and convert them to Binary Alignment Map (BAM) files and to mark duplicate reads. The average mapping rate was 55.54% (range 50.3% to 85.7%). Next, SOAPsnp (Li

et al., 2009) was used to call SNPs in each sample (SOAP options description: -u (enable rank sum test), -t (set transition/transversion ratio to 2:1 in prior probability); -L [45]; -Q [40]). RAD library preparation, sequencing, SNP calling and genotyping were performed by Beijing Genomics Institute (BGI, Hongkong).

For each polymorphic site, the allele frequencies (%A, %G, %C and %T) were calculated per accession and each SNP was scored as the proportion of the major allele per sample. Subsequently, averaged frequencies of major alleles were calculated per SNP over all 123 samples of the GWAS panel. Then, quality SNPs for genotyping were selected using the following selection criteria: 100% call rate in the mapping panel; SNPs with a minor allele frequency below 2% and with a major allele frequency above 98% in the mapping panel were removed; only biallelic SNPs selected with the frequency sum of the two major alleles in the mapping panel equal or above 95%; SNPs with a standard deviation in the frequency of the major allele below 0.1 were removed. Quality SNP marker selection was performed in R version 3.4.3 statistical software. In total, 621.452 SNP markers were selected for the genetic analysis (Petit et al., 2019b). The genotypic data of these SNP markers from the GWAS panel were used in the present study.

2.4 Genome-wide association studies for flowering stages and sex determination of hemp

In the absence of a complete physical map of the hemp genome, a specific approach was followed for QTL analyses that included: 1) selection of significant markers by association analyses (GWAS), 2) MultiQTL modelling of significant markers to determine QTLs per trait and per location, 3) PCA analysis of the significant markers to identify clusters of markers and 4) determination of QTLs across locations.

First, a Linear Mixed Model (LMM) was used to identify significant associations between genotypes and flowering/sex traits (significant QTL-markers), using a kinship correction (VanRaden, 2008) and following the same approach as in Yu et al. (2006), Huang et al. (2010), Malosetti et al. (2011), Zhou and Stephens (2012), Kruijer et al. (2015) and Petit et al. (2019b). To account for multiple testing, a Bonferroni correction was performed based on the number of independent markers (Li and Ji, 2005). To assess the correction for population structure, the cumulative distributions of observed and expected *p*-values were investigated for 3000 randomly selected SNP markers (Yu et al., 2006). The expected *p*-values were the significance level of the associations under the assumptions that markers were unlinked to the polymorphism controlling the variation of the traits. An independent analysis was performed separately for three flowering time traits and sex determination across three locations.

A MultiQTL model was performed on the significant QTL-markers, following a forward selection procedure to identify QTLs associated to a trait. Here, a QTL is defined as a

cluster of significant and collinear QTL-markers represented by the best QTL-marker that explains the largest phenotypic variance (representative QTL-marker). The MultiQTL model is the combination of non-collinear QTLs, whereby each QTL explains a specific part of the phenotypic variation in the population. Therefore, first the representative QTL-marker was selected into the model. Then, to avoid multicollinearity between QTLs from the same MultiQTL model, pairwise correlations between the selected QTL-marker and the remaining candidate QTL-markers were assessed. Significant QTL-markers correlated at $r \geq 0.3$ ($\sim r^2 \geq 0.1$) were considered collinear to the first one. Significant QTL-markers at $r < 0.3$ ($\sim r^2 < 0.1$) were candidates to add to the model and forward selection was continued with those. The explained variance of the model was calculated by linear regression (r^2) between the fitted trait values and the observed trait values.

A Principal Component Analysis (PCA) was performed to detect groups of significant colinear markers per trait. The PCA included the representative QTL-markers selected by the three MultiQTL models (three locations) and the ones that were removed to avoid multicollinearity in the models. The PC1, PC2 and the $-\log_{10}P$ of the association were plotted in 3D scatter plots to detect the clusters.

A further correlation analysis was performed to identify QTLs across locations for flowering time traits and sex determination. The analysis was performed for each trait separately, using the QTLs from each of the three MultiQTL location models. A QTL common across two or three locations was partially composed of the scaffolds with correlated representative QTL-markers ($r \geq 0.3$).

The threshold to assess multicollinearity between significant markers was based on the relationship pairwise correlation–physical distance between pairs of markers along the largest scaffolds for the *Cannabis sativa* genome (canSat3 version GCA_000230575.1) (van Bakel et al., 2011), as described in Petit et al. (2019b).

GWAS, PCA and correlation analyses were performed in Genstat 19th edition and the 3D scatter plots were performed in Excel version 14.0, using the macro Excel 3D Scatter Plot version 2.1 (Doka, 2013).

2.5 Transcriptome annotation and candidate gene identification

The representative transcriptome of the drug-type *C. sativa* 'Purple Kush' (canSat3) (transcriptome sequencing project: bioproject PRJNA 74271; NCBI GI:351590686 to GI:351629476) was downloaded from the Cannabis Genome Browser Gateway at http://genome.ccbr.utoronto.ca/cgi-bin/hgGateway (van Bakel et al., 2011). The transcriptome was annotated using Blast2go (Conesa et al., 2005;Conesa and Gotz, 2008) (**Supplementary File 1**). A selection of Purple Kush transcripts with sequence homology (blastx 2.8.0; E value cut of 0.001) to well-known flowering and sex related genes was performed. The selection was

based on gene description, gene ontology and relevant research papers on gene functions. Examples of selected genes ontology (GO) terms included P:regulation of photoperiodism, flowering, P:response to red light; P:flower development; P:short-day photoperiodism, flowering; P:photoperiodism; F:photoreceptor activity; F:phosphatidylethanolamine binding; F:DNA photolyase activity; P:positive/negative regulation of gibberellic acid mediated signalling pathway, P:floral organ morphogenesis and P:floral organ development. Finally, genomic scaffolds with significant QTL-markers were analysed for the presence of the selected transcripts associated to flowering and sex.

2.6 miRNA analysis

The six microRNA (*csa-miR156*, *csa-miR159a*, *csa-miR171b*, *csa-miR172a*, *csa-miR5021a* and *csa-miR603*), predicted and validated in *Cannabis sativa* by Das et al. (2015) and Hasan et al. (2016), were subjected to BLASTn (Altschul et al., 1997) search against the genome of *C. sativa* 'Purple Kush' assembly (van Bakel et al., 2011). The nucleotide sequences of the miRNA can be found in **Table 1**.

Table 1. Nucleotide sequences of 6 miRNA in silico predicted and validated in *Cannabis sativa* by (Das et al., 2015) and their positions in the cannabis genome (van Bakel et al., 2011).

miRNA	Sequence	Genomic scaffold	Start position	End position
csa-miR156	TGACAGAAGAGAGAGAGCAT	scaffold4343	146352	146372
		scaffold123704	2010	2030
csa-miR159a	TTTGGATTGAAGGGAGCTCTA	scaffold11004	3	24
csa-miR171b	TGATTGAGCCGTGCCAATATC	scaffold900	36428	36449
		scaffold76882	41343	41364
		scaffold113749	6009	6030
csa-miR172a	AGAATCTTGATGATGCTGCAT	scaffold140375	1454	1475
csa-miR5021a	TGAGAAGAAGAAGAAGAAAA	scaffold5250	22028	22048
		scaffold5438	42938	42958
		scaffold24090	217	237
		scaffold28679	4250	4270
csa-miR6034	TCTGATGTATATAGCTTTGGG	-	-	-

3. Results

3.1 Large phenotypic variation in flowering time and sex determination of hemp

The extensive phenotypic variation and the large heritable behaviour of beginning of flowering, full flowering, length of the vegetative growth period (VEG) and sex determination enabled the study of the genetics underlying flowering and sex in hemp, through a GWAS approach (**Table 3** and **9**, **Chapter 3**) (Petit et al., 2019c).

To identify genetic markers associated to flowering time and sex determination of hemp, the effects of the population structure on the associations between the genotypes- and phenotypes were studied. As shown in **Figure 1**, the GWAS analyses without considering the kinship correction showed large differences between expected and observed *p*-values of the associations for both, flowering time and sex determination traits. Yet, when using the kinship correction, observed and expected *p*-values showed similar distributions of *p*-values. Therefore, with the kinship correction, the population structure affecting flowering and sex traits was mostly controlled and allowed the identification of molecular markers associated to the traits.

3.2 Significant markers and MultiQTL models for flowering stages and sex determination of hemp

GWAS analyses resulted in the identification of 7875 significant associations ($-\log_{10}P \geq 4.047$) between genotypes and phenotypes for the four traits (**Supplementary File 2**). These associations corresponded to 4225 SNP markers. Among them, 2008 markers were associated to more than one trait and/or location, while 2217 markers were associated to one trait in a single location. Significant markers mapped to single loci on 2195 different genomic scaffolds. The amount of significant markers in each genomic scaffold was generally larger than one marker in more than 50% of the scaffolds, ranging from 2 to 25 significant markers. The density of significant markers per scaffold was not related to the length of the scaffold sequence. For instance, scaffold925 harboured 25 significant markers and had a sequence length of ~42kbp, while scaffold13362 also harboured 25 significant markers but with a length of ~130kbp. In addition, the distribution of markers within each scaffold showed large variation between scaffolds. For instance, significant markers in scaffold925 were condensed in a region of ~18kbp, while the markers in scaffold13362 were spread along a larger sequence (~40kbp). Furthermore, some significant markers were found close to one edge of the scaffold, such as significant markers from scaffold13362. This indicated that the flanking genomic sequences of that scaffold can also have significant markers associated to the same trait/s. Therefore, these results suggest that a QTL can span to different genomic scaffolds.

The MultiQTL model analyses identified 63 QTLs (named with the representative QTL-marker) from all MultiQTL models for flowering time and sex determination traits, as summarized in **Table 2**. The three MultiQTL models for each trait showed different number of QTLs and different explained variances. The differences between models were likely to be explained by the small *G*×E interactions affecting flowering and sex traits of hemp (**Table 3**, **Chapter 3**). In addition, models for traits in FNPC and VDS showed similar explained variances. These results suggested similar genetic control of flowering time and sex determination in plants cultivated in FNPC and VDS, while this control might partially differ in plant cultivated in CRA.

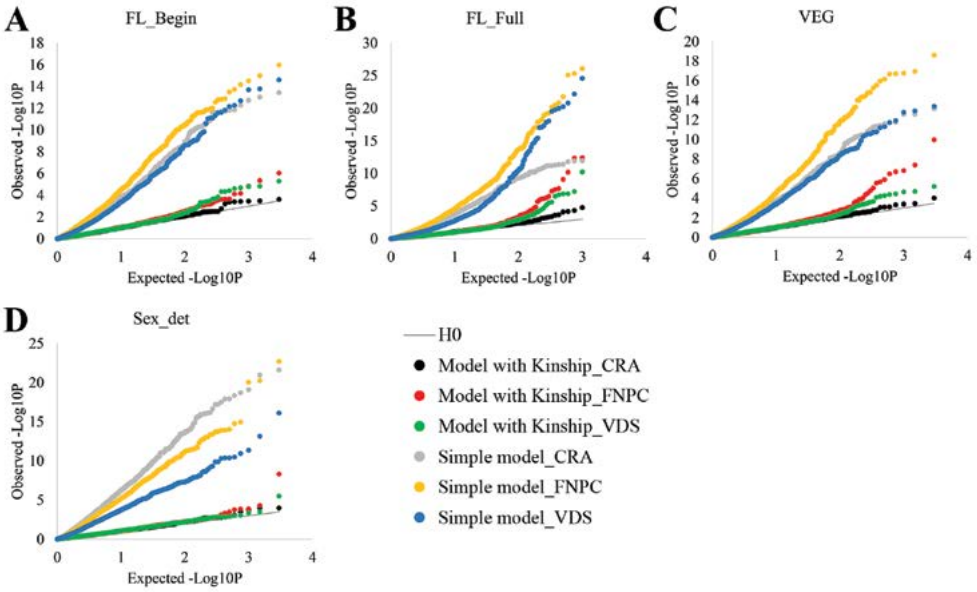


Figure 1. Cumulative plot distributions of beginning of flowering (**A**), full flowering (**B**), length of the vegetative growth period (VEG) (**C**) and sex determination (**D**). The cumulative distributions of the observed $-\log_{10}P$ for the simple model (no corrected for population structure) and the kinship models (corrected for population structure) under the expectation that random SNP markers are unlinked to the polymorphism controlling the traits.

Table 2. QTLs for flowering time traits and sex determination of hemp.

Trait	Location	QTLs (n)	Explained variance	Representative QTL-marker
FL_Begin	CRA	2	31.05	scaffold82158_8091; scaffold73915_9581
FL_Begin	FNPC	4	62.01	scaffold3201_6178; scaffold53106_2049; scaffold1534_18526; scaffold26621_95222
FL_Begin	VDS	7	78.31	scaffold2427_91742; scaffold10176_16016; C32110397_2120; scaffold47558_14031; scaffold31281_81403; scaffold21952_1421; scaffold17677_7005
FL_Full	CRA	1	21.9	scaffold15398_26963
FL_Full	FNPC	1	71.05	scaffold112592_6178
FL_Full	VDS	10	88.87	C32042317_287; scaffold15151_24297; scaffold137236_3543; scaffold27165_77253; scaffold4636_30422; scaffold7440_111147; scaffold24246_6032; scaffold4468_47061; scaffold16027_72705; scaffold55664_1689
VEG	CRA	4	39.82	scaffold82158_8091; scaffold2814_39823; scaffold118639_3805; C32101075_450
VEG	FNPC	3	76.45	scaffold27873_3085; scaffold35523_33322; scaffold925_39325
VEG	VDS	9	79.77	scaffold2427_91742; scaffold2696_4663; scaffold33154_78667; scaffold49174_45; scaffold3375_2450; scaffold29523_2165; scaffold63042_7506; scaffold38131_3503; scaffold27843_6598
Sex_det	CRA	2	63.78	scaffold21799_5192; scaffold95298_5540
Sex_det	FNPC	6	81.06	scaffold32801_3642; scaffold9506_1277; scaffold116376_5115; scaffold9283_52650; scaffold3545_18840; scaffold15087_19344
Sex_det	VDS	14	77.5	scaffold46170_31046; scaffold10107_1735; scaffold8222_31139; scaffold7247_5227; scaffold156_223; scaffold23473_1894; scaffold158807_3378; scaffold1968_8392; scaffold41933_1745; scaffold8391_22059; scaffold10301_73345; scaffold40564_1913; scaffold28242_13131; scaffold65111_1444

The number of QTLs (n= number of groups of collinear markers; at r≥0.3) in MultiQTL models generated per trait (FL_Begin= Beginning of flowering, FL_Full= Full flowering, VEG= Length of the vegetative growth period and Sex_det= Sex determination) and location (CRA, Italy; FNPC, France and VDS, The Netherlands), with the percentage explained phenotypic variance of the full model. Each QTL is named with the marker that explained the largest variance of the QTL (representative QTL-marker). Markers are indicated as the scaffold number and the position of the SNP within the scaffold (canSat3 assembly GAC_000230575.1).

3.3 QTLs across locations for flowering time and sex traits of hemp

3D scatter plots of the PCAs showed that markers associated to the three flowering time traits clustered in two groups (**Figure 2**). Yet, the distribution of the markers showed differences across locations. For beginning of flowering (FL_Begin) and length of the vegetative growth (VEG), most markers detected in CRA clustered in a single group, while markers detected in FNPC and VDS clustered in two groups, one of them co-localized with markers from CRA.

These results indicated that the genetic regulation for both traits (FL_Begin and VEG) is partially common across the three locations. In contrast, for full flowering, most markers from VDS clustered in a single group that only co-localized with markers from FNPC, as shown in **Figure 3**. Therefore, the results suggested a common genetic regulation of full flowering between FNPC and VDS. The 3D scatter plot from sex determination revealed a single dense cluster of markers from the three locations, alluding to a common genetic regulation of sex determination across all tested locations. Moreover, each cluster included at least one representative QTL-marker from the location (MultiQTL model) where most markers of the cluster were identified.

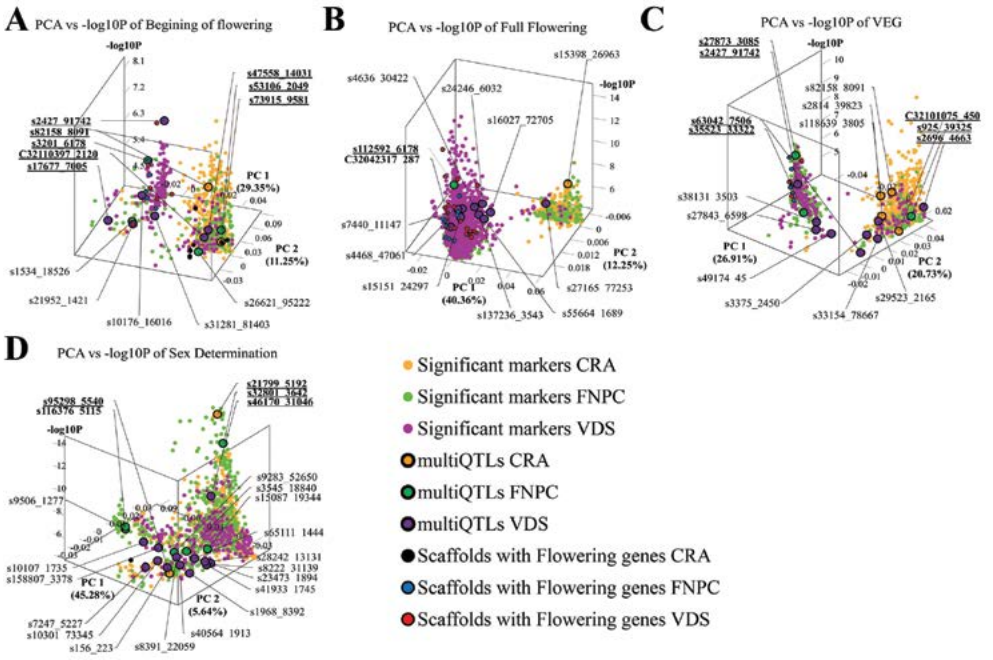


Figure 2. 3D scatter plots of PCAs depicting variation in the allele frequency profiles of significant markers for respectively beginning of flowering (**A**), full flowering (**B**), length of the vegetative growth period (VEG) (**C**) and sex determination (**D**). X-axes indicate Principal Component 1, Y-axes indicate Principal Component 2 and Z-axes indicate the significant level of the GWAS association (-$\log_{10}$P). Each plot is shown in the angle that represents better the results. Each dot represents a significant QTL-marker from the GWAS analysis. Orange, green and purple dots indicate significant QTL-markers detected in CRA, FNPC and VDS, respectively. Orange, green and purple dots with a black circle represent the representative QTL-markers of QTLs of the MultiQTL models from CRA, FNPC and VDS, respectively. Black, blue and red dots with a black circle represent scaffolds with significant QTL-markers associated to the trait in the GWAS with flowering/sex related genes in CRA, FNPC and VDS respectively (**Table 3**). QTL-names in bold and underlined indicate representative QTL-markers from different MultiQTLs models that belong to QTL regions across locations.

Finally, the correlation analyses between the representative QTL-markers of the three MultiQTL models per trait revealed six QTL across locations for flowering traits and two

for sex determination, as detailed in **Table 3**. Among them, four QTLs were identified across two locations, two for length of the vegetative growth (QTL_{VEG2} and QTL_{VEG3}) and one respectively for full flowering (QTL_{FL_Full1}) and for sex determination (QTL_{Sex_det2}). QTL_{VEG2}, QTL_{VEG3} and QTL_{FL_Full1} were identified across FNPC and VDS, while QTL_{Sex_det2} was identified for both CRA and FNPC locations. Furthermore, four QTLs were identified across three locations, two for beginning of flowering (QTL_{FL_Beg1} and QTL_{FL_Beg2}) and one for respectively length of the vegetative growth (QTL_{VEG1}) and sex determination (QTL_{Sex_det1}).

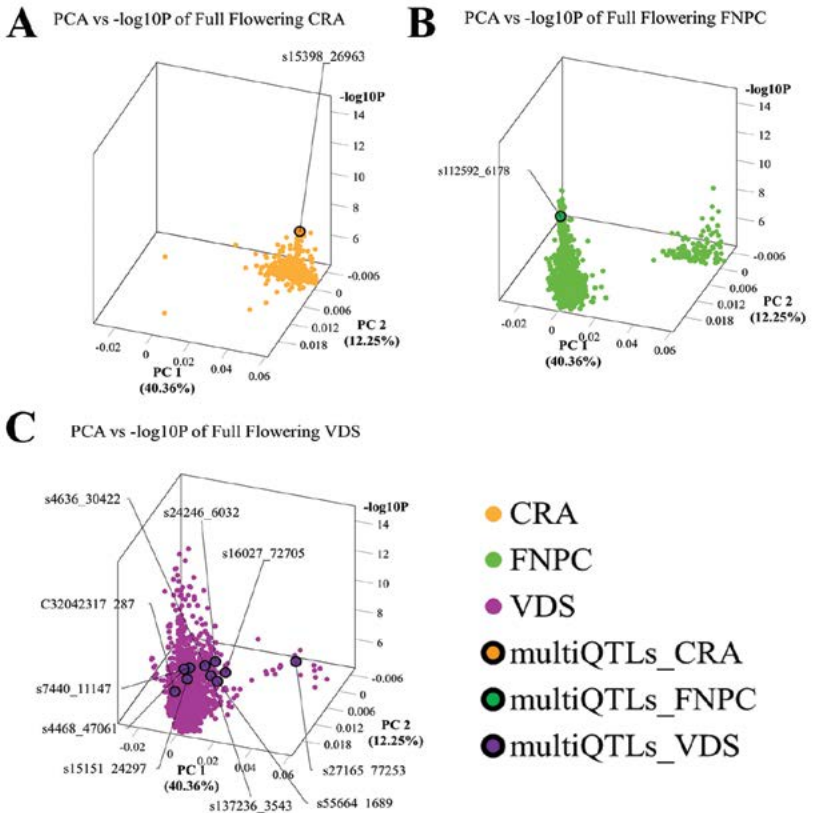


Figure 3. 3D scatter plots of PCAs depicting clustering of significant markers on basis of variation in the allele frequency profiles in the mapping population and significance in GWAS, for respectively full flowering in CRA (**A**), full flowering in FNPC (**B**) and full flowering in VDS (**C**). X-axes indicate Principal Component 1, Y-axes indicate Principal Component 2 and Z-axes indicate the significant level of the association ($-\log_{10}P$). Orange, green and purple dots indicates the significant markers detected in CRA, FNPC and VDS, respectively. Orange, green and purple dots with a black circle represent the representative QTL-markers of the MultiQTL models from CRA, FNPC and VDS, respectively.

3.4 Candidate genes for flowering stages and sex determination of hemp

In the transcriptome of cannabis 'Purple Kush' (van Bakel et al., 2011), 252 transcripts predicted to be involved in flowering were identified (**Supplementary Table 2**). Among them, 33 transcripts were found in 30 genomic scaffolds with significant QTL-markers collinear to the QTLs across locations (**Table 4**). Of these 33, 20 candidates were found to be specific for flowering time traits, five specific for sex determination and eight for both type of traits.

The candidate genes included genes of the perception and transduction of light in the flowering pathway. Among them were identified cryptochromes (*cry1*), phytochrome A (*phyA*) and E (*phyE*), suppressor of PHYA-105 (*spa1*), ultraviolet-B receptors (*uvr8*) and the *xap5* circadian timekeeper gene, which coordinates light signals for proper timing of photomorphogenesis. In addition, several genes that codify for transcription factors well-known to regulate flowering were also identified, namely *agamous*, *bed*, *bZIP*, *constans*, *floricaula*, *leafy*, *flowering locus D*, *flowering locus T*, MAD-box transcription factors, *squamosa* and *vrn1*. Twelve of the above mentioned candidates were detected in scaffolds associated only to QTL_{FL_FULL} for full flowering (**Figure 2**, **Table 4** and **Supplementary Table 1** and **2**). Furthermore, the regulatory element of flowering genes, *miR156* was found in scaffold4343 associated to the three flowering time traits (**Table 4**).

Table 3. Identification of QTL regions across locations for flowering time traits and sex determination of hemp. QTLs common across locations were identified by correlation analysis ($r \geq 0.3$, $\sim r^2 \geq 0.1$).

QTL across locations	Trait	Correlated representative QTL-markers across MultiQTL models		
		CRA	FNPC	VDS
QTL_{FL_Beg1}	FL_Begin	scaffold82158_8091	scaffold3201_6178	C32110397_2120 scaffold17677_7005 scaffold2427_91742
QTL_{FL_Beg2}	FL_Begin	scaffold73915_9581	scaffold53106_2049	scaffold47558_14031
QTL_{FL_Full1}	FL_Full	-	scaffold112592_6178	C32042317_287
QTL_{VEG1}	VEG	C32101075_450	scaffold925_39325	scaffold2696_4663
QTL_{VEG2}	VEG	-	scaffold27873_3085	scaffold2427_91742
QTL_{VEG3}	VEG	-	scaffold35523_33322	scaffold63042_7506
QTL_{Sex_det1}	Sex det	scaffold21799_5192	scaffold32801_3642	scaffold46170_31046
QTL_{Sex_det2}	Sex det	scaffold95298_5540	scaffold116376_5115	-

QTL across locations= name given to a QTL common across locations; Trait: FL_Begin= Beginning of flowering, FL_Full= Full flowering, VEG= Length of the vegetative growth period and Sex_det= Sex determination; CRA= representative QTL-markers from MultiQTL models from CRA, Italy; FNPC= representative QTL-markers from MultiQTL models from FNPC, France and VDS= representative QTL-markers from MultiQTL models from VDS, The Netherlands.

The candidate genes putatively involved in sex determination were identified across ten scaffolds underlying the QTLs. Among them, five scaffolds were co-localized with QTLs for flowering traits, sharing some candidate genes. Most of the common candidate genes between flowering and sex determination traits coded for transcription factors, such as *bZIP*, *bed*, *agamous* and *flowering locus T*. The candidate genes specific for sex determination coded for transcripts related to the metabolism and regulation of phytohormones, such as *auxin response factors* genes (*arf2* and *arf5*), and *gibberellin acid insensitive* gene (*gai* gene – DELLA proteins). The candidate genes specific for sex determination were mostly identified in QTL_{Sex_det1} across the three environments (**Figure 2**, **Table 4** and **Supplementary Table 1** and **2**).

4. Discussion

Flowering time and sex determination of *Cannabis sativa* L. strongly influence fibre quality and seed yield (Mandolino and Carboni, 2004;Amaducci and Gusovius, 2010;Faux et al., 2013;Amaducci et al., 2015;Liu et al., 2015;Salentijn et al., 2015;Petit et al., 2019c;Salentijn et al., 2019). Thus, understanding the genetic mechanisms underlying such complex traits will help to develop molecular markers for breeding for different type of hemp cultivars, such as for fibre and oil production. Such knowledge will greatly contribute to improve the stability of sex determination across locations, the predictability of flowering times and the uniformity of the crop, with positive effects on the quality of end-products.

Here, association studies were performed for flowering time and sex determination using a panel of 123 hemp accessions. The phenotyping was performed in three locations with contrasting environments across Europe. The extensive and heritable variation of the flowering and sex traits and the small genotype-by-environment (*G*×E) interactions of these traits (Petit et al., 2019b) made possible to study the genetics of flowering and sex determination in hemp. In addition, the genomic regions harbouring identified QTLs were further characterized to identify putative candidate genes explaining the molecular basis of these traits in hemp.

4.1 Elucidating molecular mechanisms of hemp flowering time

The six QTLs and the 28 candidate genes detected here, provide key regulators of hemp flowering time involved in different flowering pathways, as depicted in **Figure 4**. Four candidate genes involved in the perception and transduction of environmental signals (photoperiod flowering pathway), such as *cryptochrome 1* (*cry1*), *phytochrome A* (*phyA*) and *E* (*phyE*) and *suppressor of phytochrome A* (*spa1*) were identified in the QTL_{FL_FULL}, for full flowering, across two environments. These genes codify for photoreceptors that regulate light responses under different light conditions, such as light quantity, quality and timing (e.g. *cry1*, *phyA*, and *phyE*) and for proteins that inhibit photoreceptor's activity (e.g. *spa1*) (Lin, 2000). The

activity of these genes triggers or inhibits a signalling pathway that regulates flowering (Lin, 2000) (**Figure 4**). Previous research in arabidopsis and tobacco concluded that mutations in cryptochromes (*cry1* (Lin et al., 1995) and *cry2* (Guo et al., 1998)) and phytochromes (*phyA* (Johnson et al., 1994;Reed et al., 1994)) result in alterations in the flowering time. QTL_{FL_FULL} was identified across FNPC and VDS, the two Northern European locations with a longer photoperiod regime compared to CRA. The differences in photoperiod regimes across locations and the role of the photoreceptors and their suppressors contribute to understand the behaviour of this QTL. The expression of the arabidopsis *cry1* gene in transgenic tabacco has been reported to increase the hypersensitivity to blue, UV-A and green light (Lin et al., 1995). Therefore, the quantitative genetic variation of QTL_{FL_FULL} in hemp can affect the function of the genes and generate a hypersensitivity to specific light conditions (e.g. specific environment from Northern Europe), triggering a response in flowering time.

Furthermore, the candidate gene *vrn1*, also identified in $QTL_{FL\_FULL,}$ can provide deeper insights into flowering time of hemp. *Vrn1* is a vernalization related gene identified in arabidopsis (Michaels and Amasino, 1999;Levy et al., 2002;Bouché et al., 2016;Cheng et al., 2017). Hemp has no vernalization requirements but temperature is known to be a factor affecting the length of the juvenile stage (Amaducci et al., 2012;Salentijn et al., 2019). Therefore, flowering in hemp is likely to involve a temperature related pathway. In arabidopsis, the gene *vrn1* codes for a protein that functions in stable repression of the major actor in the vernalization pathway: the floral repressor transcription factor, FLOWERING LOCUS C (FLC). FLC represses the expression of flowering integrator genes, thereby inhibiting the plants to flower (**Figure 4**). A period of low temperatures is needed to trigger the expression of *vrn1*, which in turn leads to flowering (Cheng et al., 2017). As hemp has no vernalization requirements, quantitative variation affecting *vrn1* in association to temperature sensitivity are likely to affect flowering. Therefore, the lower temperatures during the growing season of hemp at Northern European latitudes are likely to affect *vrn1* and lead to an acceleration of the juvenile state and of flowering.

Table 4. Candidate genes in QTL regions across locations.

Scaffold	Markers (n)	Trait/s	QTLs across locations collinear to the markers	Candidate gene (Transcript)	Gene description
scaffold21952	2	FL_BEGIN / VEG	QTL_{FL_Beg1} / QTL_{VEG2}	PK17906.1	*Agamous-like* MADS-box (AGL93 Transcription factor)
scaffold16869	4	Sex_det	QTL_{Sex_det1}	PK04762.1	*Auxin response factor 2B-like isoform X1* (*arf2* gene)
scaffold21255	1	Sex_det	-	PK19328.1	*Auxin response factor 5* (*arf5*)
scaffold2205	1	FL_FULL	QTL_{FL_Full}	PK10791.1	*Vrn1 gene* - B3 domain-containing transcription factor VRN1-like
scaffold16340	3	Sex_det	QTL_{Sex_det1}	PK23964.1	bZIP transcription factor 16-like
scaffold2448	4	FL_BEGIN / FLL_FULL / Sex_det / VEG	QTL_{FL_Beg2} / QTL_{VEG1} / QTL_{Sex_det2}	PK08841.1	bZIP transcription factor 27-like
scaffold11691	3	FL_BEGIN / FL_FULL / VEG	QTL_{FL_Full}	PK19567.1	*Cryptochrome 1 family* (*cry1* gene)
scaffold31176	3	Sex_det	QTL_{Sex_det1}	PK09859.1	*Gai* gene - DELLA RGL1-like protein (Repressor of the gibberellin (GA) signalling pathway. Regulates the floral development)
scaffold102797	3	FL_FULL	QTL_{FL_Full}	PK10917.1	*Floricaula leafy* (flower and leaf development)
scaffold19844	3	VEG	QTL_{VEG1}	PK11476.2	*Flowering locus D* (*fld* gene)
scaffold4564	17	FL_BEGIN / FL_FULL / VEG	QTL_{FL_Beg1} / QTL_{FL_Full} / QTL_{VEG2}	PK13716.1	Phosphatidylethanolamine-binding PEBP (Regulation of flower development; *Florigen – Flowering locus T-like1 –ft* gene)
scaffold3201	4	FL_BEGIN / FL_FULL / VEG / Sex_det	QTL_{FL_Beg1} / QTL_{FL_Full} / QTL_{VEG2}	PK16758.1	*Heading date 3a-like* (*Flowering locus t-like3* or *hd3a* gene)
scaffold24080	2	FL_FULL	QTL_{FL_Full}	PK08698.1	*Flowering locus T-like4* or *ft-like4* gene
scaffold10005	9	FL_FULL / Sex_det	QTL_{FL_Full} / QTL_{Sex_det1}	PK12609.1	*Flowering time control partial* (Autonomous flowering pathway; inhibiting FLC transcription factor)
scaffold5190	7	FL_BEGIN / FL_FULL / VEG / Sex_det	QTL_{FL_Beg1} / QTL_{FL_Full} / QTL_{VEG2}	PK04407.1 PK28902.1 PK14825.2	*Gibberellin-20 oxidase* (*ga20ox* gene) *Gibberellin 2-beta-dioxygenase 1* (*ga2ox* gene) *Agamous-like* MADS-box AGL6 transcription factor
				PK20658.1	*Probable UDP-N-acetylglucosamine-peptide N-acetylglucosaminyltransferase* SPINDLY protein – *spy* gene
scaffold9737	2	FL_FULL	QTL_{FL_Full}	PK20658.1	*Probable UDP-N-acetylglucosamine-peptide N-acetylglucosaminyltransferase* SPINDLY protein – *spy* gene

scaffold75641	1	FL_FULL	QTL_{FL_Full1}	PK05121.1	*Leucine-rich repeat receptor-like serine threonine- kinase* BAM3 isoform X1 transcription factor
scaffold112970	1	VEG	QTL_{VEG1}	PK15717.1	MADS-box transcription factor (*soc1* gene)
scaffold42686	1	FL_FULL	QTL_{FL_Full1}	PK25677.1	*Phytochrome A* (*phyA* gene)
scaffold33564	1	FL_FULL	QTL_{FL_Full1}	PK18424.1	*Phytochrome E* (*phyE* gene)
scaffold35877	1	sex_det	-	PK12852.1	*Probable lysine-specific demethylase* - ELF6 transcription factor (Circadian clock, photoperiodism, flowering)
scaffold4343	7	FL_BEGIN / FL_FULL / VEG	QTL_{FL_Beg2} / QTL_{VEG1}	*csa-miR156* PK22320.1	*miRNA156* and *squamosa promoter-binding 1-like* (*spl1*-like gene)
scaffold1953	3	FL_FULL	QTL_{FL_Full1}	PK13755.1	*Suppressor of* PHYA-105 1 (*spa1* gene)
scaffold46774	1	FL_BEGIN / FL_FULL / VEG	QTL_{FL_Beg2} / QTL_{VEG1}	PK07299.1	*Ultraviolet-B receptor* (*uvr8* gene)
scaffold1342	1	FL_FULL	QTL_{FL_Full1}	PK10982.1	*Xap5* gene *circadian timekeeper*
scaffold64764	1	FL_FULL	QTL_{FL_Full1}	PK05670.1	*Bed* gene - domain-containing RICESLEEPER 2-like
scaffold10189	3	FL_FULL / Sex_det	QTL_{FL_Full1} / QTL_{Sex_det2}	PK10457.1	*Bed* gene - domain-containing RICESLEEPER 2-like
scaffold29022	1	VEG	QTL_{VEG1}	PK12682.1	*Bed* gene - domain-containing RICESLEEPER 2-like
scaffold12036	3	FL_FULL	QTL_{FL_Full1}	PK18183.1	*Zinc finger constans-like 16* (*zinc finger constans-like 6*)
scaffold13708	1	FL_FULL	QTL_{FL_Full1}	PK23361.1	*Zinc finger constans-like 2*

Candidate genes located in genomic regions of markers collinear to representative QTL-marker from MultiQTL models (QTL= representative QTL-markers + collinear markers) for flowering time traits (FL_begin, VEG, FL_full) and/or sex determination (Sex_det). Scaffold= genomic scaffold of canSat3 genome that harbours significant markers; Makers (n)= number of significant markers present on the scaffold; Candidate gene= transcript of hemp cultivar 'Purple Kush' located in a QTL across locations, annotated for a putative function in flowering time/sex determination; Description= description of putative candidate genes. See **Supplementary Table 4** for the Genbank accession codes.

Candidate genes located in genomic regions of markers collinear to representative QTL-marker from MultiQTL models (QTL= representative QTL-markers + collinear markers) for flowering time traits (FL_begin, VEG, FL_full) and/or sex determination (Sex_det). Scaffold= genomic scaffold of canSat3 genome that harbours significant markers; Makers (n)= number of significant markers present on the scaffold; Candidate gene= transcript of hemp cultivar 'Purple Kush' located in a QTL across locations, annotated for a putative function in flowering time/sex determination; Description= description of putative candidate genes. See **Supplementary Table 4** for the Genbank accession codes.

Genes coding for transcription factors involved in flowering, such as *agamous*, *bZIP*, *soc1* (MADS-box transcription factor), *flowering locus D*, *flowering locus T* and *squamosa* and the flowering regulatory element, *miRNA156* were identified in QTLs for beginning of flowering and length of vegetative growth. All these genes are involved in endogenous flowering pathways independently of the photoperiod and vernalization/temperature pathways (**Figure 4**). For instance, *squamosa* and *miRNA156* are involved in the regulation of the flowering gene *suppressor of overexpression of constans1* (*soc1*) (Wu et al., 2009;Hyun et al., 2016). *Squamosa* gene codes for a transcription factor that promotes the expression of *soc1*, while *miR156* represses its expression. In arabidopsis, *soc1* codes for a MADS-box transcription factor that can directly activate the floral meristem identity gene *leafy*, which directly triggers the flowering meristems of the plants. In addition, *soc1* can generate fluctuations in the balance of the phytohormone gibberellic acid (GA) and activate flowering (Lee and Lee, 2010;Immink et al., 2012;Hyun et al., 2016). *Flowering locus T*, also known as *ft* or *florigen*, is a mobile flower-promoting signal transported from the leaves to the meristems elsewhere in the plant. *Flowering locus T* induces the expression of other flowering identity genes (e.g. *agamous*), which induce meristems to flower. *Flowering locus D* codes for another transcription factor (FLD) that, as *vrn1* gene, controls the floral repressor FLC. However, the mode of action of FLD is independent of the temperature. This transcription factor probably mediates histone demethylation at FLC locus, inactivating the expression of FLC and consequently FLD triggers flowering (He et al., 2003;Jiang et al., 2007;Liu et al., 2007). Therefore, genetic variation in these genes reported to affect flowering, through endogenous flowering pathways, can lead to flowering with less sensitivity to the environment.

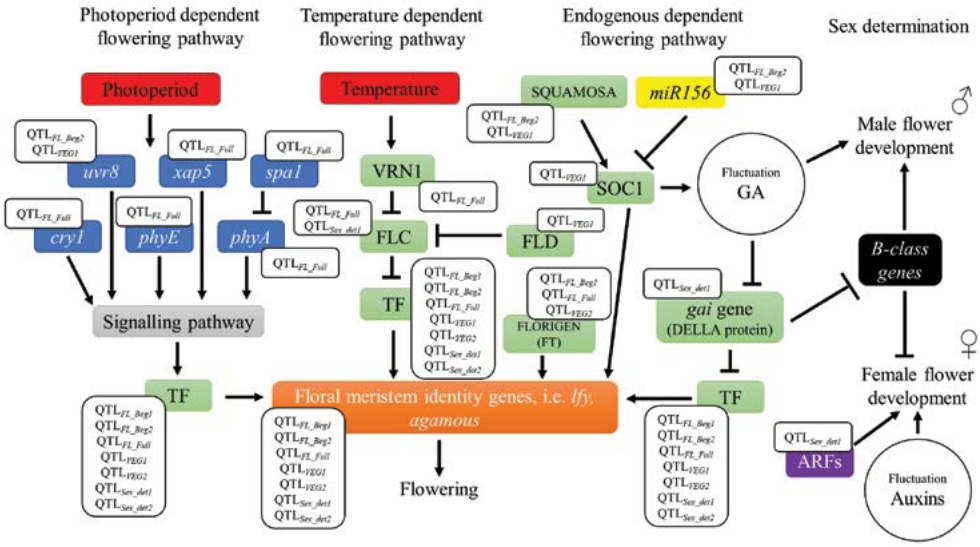


Figure 4. Flowering and sex determination metabolic pathways: identification of candidate genes underneath the QTLs. The QTLs for flowering time are found in different flowering dependent pathways (photoperiod, temperature and endogenous flowering pathways). Photoperiod pathway involves genes of the perception and transduction of light signals (ultraviolet-B receptors (*uvr8*), circadian timekeeper (*xap5*), suppressor of PHYA-105 (*spa1*), cryptochromes (*cry1*), phytochrome A (*phyA*) and phytochrome E (*phyE*)). Temperature pathway involves *vrn1*, a vernalization dependent transcription factor. Both photoperiod and temperature pathways activate signalling pathways and/or transcription factors involved in the endogenous flowering pathway to regulate floral meristem identity genes, such as *leafy* (*lfy*). Genes that code for these transcription factors include *flowering locus C* (FLC), *flowering locus D* (FLD), *flowering locus T* (FLORIGEN or FT), *suppressor of overexpression of constans1* (*soc1*) and *gibberellic acid insensitive* (*gai* gene – DELLA protein), among others. TF is used to summarize all transcription factors inducing floral meristem identity genes. Endogenous pathway also include the regulatory element of flowering genes, *miR156*. The QTLs for sex determination are found in metabolic pathways involved in regulation of phytohormones gibberellic acid (GA) or auxins. These pathways include *B-class* homeotic genes involved in the development of male flower organs and *auxin response factors* genes (ARFs) involved in female flower development.

4.2 Molecular mechanisms of hemp sex determination

The identification of the two QTLs for sex determination across locations is a significant step towards understanding the sex determination system of cannabis. Two auxin response factors genes (*arf2* and *arf5*), *bZIP* transcription factor 16-like and gene *gibberellic acid insensitive* (*gai*) that codes for the DELLA RGL1-like protein were identified in QTL_{Sex_det1} for sex determination. These genes are involved in the balance of the phytohormones auxins and gibberellic acid, which are known to play an active role in the sex expression in many crops, such as hemp or spinach ((Heslop-Harrison, 1956;Chailakhyan and Khryanin, 1978;Galoch, 2015) and reviewed in Salentijn et al. (2019). For instance, West and Golenberg reported that in spinach the *gibberellic acid insensitive* (*Spgai*) gene was two-fold higher expressed in inflorescences from females compared to male inflorescences (West and Golenberg, 2018).

Both, the exogenous GA application and the reduction of *Spgai* gene expression in female plants, induced the development of male flowers in spinach (**Figure 4**). Based on their research they proposed a model for the sex determination of monoecious plants in spinach, named M locus, which could apply to hemp. In conditions of high expression of *Spgai* there was an inhibition of the spinach *B-class* homeotic genes. These *B-class* homeotic genes are masculinizing factors of flower development, promoting the formation of male organs and at the same time they inhibit female organ formation in flowers (**Figure 4**). When applying GA to spinach, *Spgai* gene was inhibited and thus the inhibition on the masculinizing *B-class* genes was released, resulting in the formation of male organs. Soldatova and Kryanin also reported a masculinization effect of female hemp plants owed to external treatment with GA (Soldatova and Khryanin, 2010).

Furthermore, Heslap-Harrison provided evidences that the treatment of male dioecious hemp with auxins showed the development of female flowers (Heslop-Harrison, 1956) (**Figure 4**). Auxins are phytohormones known to be involved in flower development (Hardtke and Berleth, 1998). Auxins regulate this process by controlling gene expression via transcription factors AUXIN RESPONSE FACTORS (ARF). These transcription factors control the expression of genes involved in female flower development (Li et al., 2016) (**Figure 4**). Arabidopsis mutant for the *auxin response factor 8* (*arf8*) gene developed seedless fruits (Goetz et al., 2006). Goetz and co-workers reported that transcription factor ARF8 acts as an inhibitor to stop further development of carpel in the absence of fertilization and the generation of signals to induce seed development (Goetz et al., 2006). In addition, Liu and co-workers reported floral development defects and female sterility in tomato where *arf6* and *arf8* genes were down-regulated with microRNAs (Liu et al., 2014). Therefore, it is likely that QTL_{Sex_det1} contributes to regulate sex expression, for example through downregulation of *gai* gene and/or downregulation of *auxin response factors* genes. Both effects are likely to promote the development of male flower organs in female plants, leading to a monoecious hemp. This is essentially because the positive effect alleles of QTL_{Sex_det1} are associated to promote monoecy in hemp plants. Thus, variation in expression of these genes, due to allelic variation, is likely to be responsible for the phenotypic variation of monoecy in the hemp panel.

4.3 Implications of the QTLs for hemp breeding programmes

The identification of the six QTLs for flowering time will have positive implications for breeding for hemp fibre quality and seed yield. Flowering time is a well-known heritable trait, as large heritabilities have been previously reported for hemp and for several plant species (Takahashi et al., 2001;Huang et al., 2010;Campoy et al., 2011;Sánchez-Pérez et al., 2014;Soto-Cerda et al., 2014;Sasaki et al., 2015;Kushanov et al., 2017;Petit et al., 2019c). However, flowering of hemp is also known to be strongly sensitive to environmental factors (Faux et al., 2013;Hall et al., 2014b;Amaducci et al., 2015;Sawler et al., 2015;Petit et al., 2019c).

The identification of QTLs common across environments with contrasting photoperiod regimes reveals the genetic basis of flowering control that is poorly sensitive to environmental variation. Breeding for these QTLs will develop genotypes with more specific flowering times across locations. For instance, the QTLs for beginning of flowering and for vegetative growth period can be used to select accessions with a high frequency of respectively "early", "medium" or "late" QTL alleles. Late flowering can have positive implications for plants cultivated for fibre at Southern European Latitudes. At Northern European Latitudes both medium and late flowering times are preferred for fibre production (Amaducci et al., 2015). Plants that do not shift the carbon partitioning from biomass to flower and seed production directly after the critical photoperiod is reached (~14h of daylight), are likely to produce fibre of large quality. This is because lignification intensifies after flowering (Liu et al., 2015) and lignin is negatively associated to fibre quality (Petit et al., 2019c).

Furthermore, the QTL for full flowering (QTL_{FL_FULL}) can be used to breed for cultivars with accelerated flowering (short flowering period). This acceleration is likely to have good implications for fibre and seed yield of dual-purpose hemp, especially in the short growing season of Northern Europe. This is essentially because the short flowering duration favours a uniform crop development and increases fibre quality (Amaducci et al., 2008b).

Finally, the identification of the two QTLs for sex determination across locations will allow molecular breeding to control sex determination of hemp across generations. This is important because monoecious plants show particular unstable sex expression, which tends to a gradual return to natural dioecy after few generation (Bócsa and Karus, 1998;Amaducci et al., 2008b). In fact, the large occurrence of males in monoecious cultivars is one of the major bottlenecks in the production of large seed yields. Therefore, breeding for these QTLs will develop genotypes that will have a more stable monoecy sex determination. Moreover, the production of fibre and seed from these genotypes, will be more uniform and consequently the quality of these traits is likely to increase (Mandolino and Carboni, 2004;Salentijn et al., 2015;Petit et al., 2019c).

5. Conclusions

The results of this study prescribe positive prospects to gain control over flowering time and stability over sex determination. The development of hemp cultivars with specific flowering properties and sex conditions will increase the commercial value of the crop and reduce risks of heterogeneous quantity and quality of seed and fibre. QTLs that are identified across locations are of particular interest to develop genotypes with specific flowering times. To our knowledge this is the first report about the identification of QTLs for sex determination of hemp. Finally, this report provides new insights into the molecular mechanisms underlying flowering time and monoecy determination in cannabis.

6. Acknowledgements

This work was conducted as part of the MultiHemp project (Multipurpose hemp for industrial bioproducts and biomass) funded by the European Union's Seventh Framework Programme for research, technological developments and demonstration under grant agreement number 311849.

Supplementary data

Supplementary File 1, **Supplementary File 2**, **Supplementary Tables 1** and **Supplementary Table 2** can be found in the following link:
https://wageningenur4-my.sharepoint.com/:f:/g/personal/jordi_petitpedro_wur_nl/ElnmN8O9jItCqh4XMK5dVkQBWcMXrvV3WxIV-cvDcdZ2TA?e=O3Dq0K

Chapter 7

General Discussion

1. Embedding of the research

With the projected increase in the world's population, along with the undoubtedly increase in climate instability, a more efficient production of food and non-food commodities is needed. Therefore, we are called to produce more outputs with less inputs. In addition, the consumer awareness towards products from renewable and/or sustainable sources and the number of governmental guidelines towards sustainability are increasing. An example of this is the European Guideline 2000/53/EC, that has set the goal of improving material recyclability in automobiles to 95% (Peças et al., 2018). The production of natural fibres requires less energy input than the production of synthetic and mineral fibres (Dam, 2008;Peças et al., 2018). Therefore, this situation has increased the interest in natural fibres, in particular for bio-composites as substitutes for plastics and mineral fibres (Peças et al., 2018).

Bast fibre crops, in particular hemp (*Cannabis sativa* L.), are excellent candidates for the production of sustainable natural fibres. This is essentially because hemp has a large fibre yield (Struik et al., 2000;Amaducci et al., 2015;Piotrowski and Carus, 2019) and is an environmental-friendly crop (**Table 1**, **General Introduction**) (Ebskamp, 2002;Pfister et al., 2011). Regardless of the benefits of hemp as a source of natural fibres, the crop has faced a worldwide progressive decline in the last century. This is the result of a gradual replacement of hemp fibres for textile applications by cotton and synthetic fibres (Allegret et al., 2013). As a consequence, hemp remains a poorly developed crop and that is mainly due to the limited knowledge of the molecular mechanisms underlying fibre quality and the limited molecular breeding of the crop (Salentijn et al., 2015). Consequently, hemp breeding programmes mostly remain in conventional breeding schemes, awaiting for cutting-edge molecular plant breeding tools. These tools are desired to improve hemp as a raw material for the production of biomaterials, such as bio-composites. For these applications the most relevant breeding targets include fibres with high bast fibre content, high cellulose content, large crystalline structures and low lignin content (Ranalli, 2004;Salentijn et al., 2015). To boost the development of hemp breeding programmes, we must first get insights into the extent of genetic variability in hemp fibre quality and understand how we can exploit this variation.

In this thesis a range of agronomical, biochemical, molecular and genetic experiments were performed to identify quantitative trait loci (QTLs) associated to fibre quality traits to upgrade hemp breeding programmes. In the following sections, the insights generated in the previous chapters are comprehensively combined, discussed and contextualized in the light of genetic improvement of hemp fibre quality.

2. Cutting-edge knowledge on the fibre quality of hemp in the context of genetic improvement

2.1 Advanced high-throughput phenotyping of the cell wall composition of hemp

Rapid and cost-efficient cell wall phenotyping is necessary to screen large numbers of hemp accessions. Near-infrared reflectance spectroscopy (NIRS) is a high-throughput technology widely used to develop prediction models for biomass quality traits (contents of cellulose, hemicellulose and lignin) in fibre crops, such as maize (Barrière et al., 2008;Lorenzana et al., 2010;Torres et al., 2015b), miscanthus (van der Weijde et al., 2017), sorghum (Vermerris et al., 2007) or switchgrass (Vogel et al., 2011;DeMartini et al., 2013). NIRS prediction models are based on biochemical and spectral data from several samples. High quality biochemical data from many samples are needed to develop accurate prediction models. This requires the use of high-throughput methods for the extraction and biochemical analysis of hemp, which were not available at the start of this thesis. High-throughput cell wall extraction methods are available for grasses (Goering and van Soest, 1970), such as maize (Torres et al., 2013) and miscanthus (van der Weijde et al., 2016). However, these methods are not suitable to extract or analyse all cell wall components of hemp, in particular the pectic fraction. This is because these methods are based on aqueous detergents (Pettolino et al., 2012) and some pectins are hydrophilic, such as rhamnogalacturonan type I pectins (Huffman and Caballero, 2003). These pectins are removed from the cell wall fraction with the aqueous detergents (Pettolino et al., 2012). Therefore, the first analyses we performed in this thesis comprised the optimization of biochemical methods to generate accurate biochemical cell wall data from hemp.

The protocol for cell wall extraction developed in this thesis allows to obtain enough material in a single extraction, that is sufficient for the complete characterization of the cell wall. This improvement is an advantage compared to previous biochemical protocols for dicot plants because they extract lower amounts of cell wall (Foster et al., 2010;Pettolino et al., 2012). Using the previous methods, several rounds of extractions were needed to produce enough cell wall for a complete characterisation. Therefore, the improvements of our method allow to reduce the time of the complete cell wall study. The new method also allows to study the contents of monosaccharides derived from hydrophilic pectin (**Figures 2** and **3**, **Chapter 2**), and that is because it is based on alcohol solvents. Furthermore, the optimizations of the protocols for cell wall hydrolyses increase the repeatability of cell wall composition analyses (**Figure 3** and **Tables 3-4**, **Chapter 2**) and allow to analyse 60 hemp samples per run. These improvements increase the throughput of the biochemical analysis (**Table 2**, **Chapter 2**). Therefore, these new biochemical methods provide the tools needed to accurately characterize the cell wall composition from several hemp samples.

Furthermore, the biochemical methods developed in **Chapter 2** were combined with the NIRS technology in **Chapter 3,** to improve the high-throughput phenotyping of hemp cell walls. NIRS prediction models were developed for eleven cell wall traits, including cell wall content, contents of different monosaccharides and lignin content. The complete cell wall extraction and the large repeatability of the biochemical data allowed an accurate analysis of cell wall composition and contributed to the development of high quality prediction models (**Tables 4** and **5**, **Chapter 3**). These methods enabled the phenotyping of a large hemp panel and consequently the identification of QTLs for hemp fibre quality. In addition, these prediction models will contribute to reduce the phenotyping costs and time in future breeding programmes of hemp.

2.2 Genetic variability in fibre quality of hemp

The identification of QTLs for hemp fibre quality requires the study of phenotypic variability from large hemp panels. Yet, variability studies of hemp cell wall composition have only been performed in panels with a small number of accession. Toonen and co-workers used four hemp accessions ('Chamaeleon', 'Felina 34', 'Kompolti hybrid TC' and 'Green Sibling') to study cell wall content, cellulose content and lignin content of hemp (Toonen et al., 2004). Likewise, flowering time studies in hemp have been conducted among only six hemp accessions ('Carmagnola', 'Felina 34', 'Fibranova', 'Futura 75', 'Futura 77' and 'Tiborszallasi') (Amaducci et al., 2008b). Therefore, the extent of phenotypic variability in fibre quality, flowering time and sex determination of hemp in large hemp panels was still largely unknown at the beginning of this thesis.

To characterize the variability in fibre quality of hemp, we first assembled a hemp accession panel of 123 accessions from diverse geographic origin (16 different origins); different cultivation purpose (fibre and oil accessions) and different population type (breeding material, landraces and wild races) (**Table 1**, **Chapter 3**). The analysis of the hemp panel revealed that there was a large phenotypic variability in 30 traits relevant to fibre quality, including cell wall composition, agronomic traits, flowering time and sex determination traits (**Table 3**, **Chapter 3**). The variability reported in the present study was comparable to the large variability reported in other (bast) fibre crops, such as flax (You et al., 2017), with similar applications to hemp (contents of cellulose (54-65%) and lignin (8-10%), bast fibre content (34-46%) and flowering time (45-71 days) (You et al., 2017)). In addition, the variability of the present hemp panel was larger than previous studies in fibre quality of hemp with smaller sets of accessions (Toonen et al., 2004).

The European THC restriction policies on cultivation of hemp in open fields limited the choice of hemp accessions that could be used in the trials. This is because only accessions with a THC content below 0.3% were included in the study. Also, it is difficult to find, and to get access to fibre hemp resources that are expected to vary and to multiply enough seeds from each accessions for three trial locations. With hemp being an outcrossing species,

the seed batches for the trial were harvested from small inbred families, one per accession. Notwithstanding these difficulties, a hemp panel of 123 accessions was assembled. The results showed that the panel was large enough to include an extensive range of genotypic and phenotypic variation. In addition, the panel showed low levels of population structure (**Table 1** and **Figure 2**, **Chapter 4**). These results indicated that the panel is a good set for genetic association studies to identify QTLs for fibre quality. Moreover, the hemp accessions with outstanding fibre properties identified in this study can already be introduced in breeding programmes, as they fulfil the THC policy requirements for open field cultivation.

2.3 Relevant targets to maximize the genetic gains of hemp fibre quality

To maximise the genetic gains of fibre quality, breeding targets must be heritable, being largely determined by genetic components (Davis, 2008). Prior to this thesis, little was known about the heritability of important traits relevant to fibre quality of hemp. The extensive variability in hemp fibre quality characterized herein, allowed to assess the genetic component of the 30 fibre quality traits. Large heritability values were detected in several fibre quality traits of hemp (**Figure 1** and **Table 9, Chapter 3**). These heritabilities were comparable to the heritability values for biomass traits (contents of cellulose, hemicellulose and lignin) reported in other fibre crops (e.g. eucalyptus (Raymond and Schimleck, 2002;Schimleck et al., 2004;Poke et al., 2006), maize (Torres et al., 2015a), miscanthus (Slavov et al., 2014;van der Weijde et al., 2017), poplar (Klasnja et al., 2003;Davis, 2008) and switchgrass (McLaughlin et al., 2006;Boe and K. Lee, 2007)). In addition, heritabilities of hemp flowering time traits were also comparable to previous results in other crops (e.g. almond (Sánchez-Pérez et al., 2014), apricot (Campoy et al., 2011), adabidopsis (Sasaki et al., 2015), cotton (Kushanov et al., 2017), flax (Soto-Cerda et al., 2014;You et al., 2017) and rice (Takahashi et al., 2001;Huang et al., 2010)). Large heritability indicates a stable ranking of phenotypes across different environments, which suggests predictable and repeatable phenotypes. Therefore, our study revealed that breeding for contents of cellulose, mannan, xylan, lignin and bast fibre have great potential to increase the genetic gains of hemp fibre quality. In addition, the large heritability values in flowering traits indicate that hemp genotypes with specific flowering times and more stable sex expression can be achieved through breeding programmes.

2.4 Influence of the environment component on the fibre quality of hemp

Changes in environmental factors, such as photoperiod and temperature regimes are known to trigger the beginning of flowering in hemp. In addition, flowering is associated with a shift in carbon partitioning. These physiological changes affect the stem morphology and the cell wall composition of hemp (Struik et al., 2000;van der Werf and Turunen, 2008;Amaducci et al., 2012;Hall et al., 2012). Thus, hemp fibre quality is strongly influenced

by the environment (Faux et al., 2013;Hall et al., 2014b;Amaducci et al., 2015;Sawler et al., 2015). In general, hemp harvest for fibre occurs around flowering time and late to medium late flowering cultivars are used for fibre production because of the higher biomass yield (Amaducci et al., 2015).

Despite the large heritabilities of many fibre quality traits, environmental factors that influence fibre quality will have implications for breeding targets. Nonetheless, the effect of the environment in the fibre quality traits was largely unknown at the start of this thesis.

2.4.1 Adaptive behaviour of hemp fibre quality in different environments

The environmental differences across the three locations significantly affected the phenotypic variation of the 30 fibre quality traits of hemp (**Figure 2**, **Chapter 3**). The environment affected most accessions in a similar manner. These results are in agreement with the strong influence of the environment on fibre quality and flowering traits in hemp (Faux et al., 2013;Hall et al., 2014b;Amaducci et al., 2015;Sawler et al., 2015). This component of the phenotypic variation can be understood as an adaptive behaviour of fibre quality traits under specific environmental factors, which is independent of the heritable fraction of the phenotypic variation. Therefore, this component can have special implications for fibre quality traits that are poorly controlled by the genetic component, (e.g. pectin related traits).

Our results do suggest that the environment plays an important role in the content of pectin from the stem. The content of pectin in most accessions grown in The Netherlands was lower than that in most accessions cultivated in France. Hemp plants were cultivated under different environments differing in many factors, including precipitation (**Table 2**, **Chapter 3**). The amount of water available might contribute to these phenotypic differences across locations. It is known that the pectin content can be modified as a response to water availability (reviewed in Le Gall et al. (2015)). The water holding capacity function of pectin (Voragen et al., 2009) protects the plant when exposed to water stress. Piro, Leucci and co-workers reported in wheat changes in pectin content in response to drought (Piro et al., 2003;Leucci et al., 2008). Similar findings were reported in resurrection plants (Moore et al., 2008b;Gribaa et al., 2013;Moore et al., 2014). It was proposed that homogalacturonan type pectin and arabinan side-chain substitutions of rhamnogalacturonan type I pectin (RG-I) maintain the flexibility of the plant cell walls under water deficit stress, in *Myrothamnus flabillifolia*. Pectin protects the cell walls from shrinking. In this way, pectin preserves the functionality of the stem tissues (Moore et al., 2008a).

Yet, many environmental factors differed between locations in our study and further research is needed to identify specific environmental factor(s) responsible for the phenotypic differences. Detailed characterization of the role of individual environmental factors in fibre quality will help to design agricultural practices to emulate the environmental

factors responsible to specific phenotypes in hemp. These agricultural practices could be implemented in the cultivation of hemp to improve the fibre quality associated to traits with low heritable components.

2.4.2 Heritable variation sensitive to environmental factors

Phenotypic variation can also be explained by the interactions between the genotypes and the environments (*G*×E). *G*×E interactions evaluate the fraction of the phenotypic variation that is heritable and dependent on environmental factors. *G*×E interactions allow to study the sensitivity of individual accessions under specific environments. This fraction of the phenotypic variation is generally badly appreciated by plant breeders, as selection for traits with large *G*×E interactions might result in biased selection decisions, owed to the unknown effects in untested environments (van der Weijde et al., 2017). **Chapter 3** of this thesis is the first study in hemp describing significant *G*×E interactions in all 30 fibre quality traits. Similar *G*×E interactions were reported in biomass quality traits in many crops (e.g. alfalfa (Sheaffer et al., 1998), maize (Dolstra et al., 1992;Cox et al., 1994;Argillier et al., 1997;Barrière et al., 2008;Torres et al., 2015a), miscanthus (van der Weijde et al., 2017) and switchgrass (Hopkins et al., 1995)). In addition, Sasaki an co-workers also reported significant *G*×E interactions in flowering time in arabidopsis (Sasaki et al., 2015). Despite these significant interactions, *G*×E interactions were small in important fibre quality traits (contents of cellulose, hemicellulose and lignin, and flowering time) in hemp and the other plant species. Therefore, in terms of breeding for these traits in hemp, the extent of the *G*×E interactions is not likely to interfere in selection decisions.

Nonetheless, our study reported that several other traits, such as pectin related traits, most agronomic traits and two fibre traits, decortication efficiency and fibre fineness, showed large *G*×E interactions in combination with small genetic components (**Figure 2**, **Chapter 3**). The use of these traits in breeding programmes may lead to unexpected phenotypes. This is because the specific environmental factors and alleles responsible for these sensitivities of certain hemp accessions are currently still unknown.

In depth characterization of the genetic basis and the environmental factors responsible for the *G*×E interactions in relevant traits will offer novel approaches to breed for certain fibre qualities for specific environments, through local adaptation (El-Soda et al., 2014).

2.5 Breeding for fibre quality requires a better understanding of the interactions between different traits

The interactions between traits relevant to fibre quality of hemp are important to maximise the fibre production of excellent quality. Nonetheless, these interactions are largely unknown in hemp. This knowledge is particularly important because the breeding for some fibre quality traits can lead to trade-offs in other traits. In addition, understanding the

interactions between traits can play an essential role in the improvement of low heritable traits that cannot be improved through breeding programmes.

Results from **Chapter 2** showed that thick stems had larger fraction of secondary bast fibres in the bast area compared to thin stems. In addition, thin stems had higher proportion of bast area relatively to the whole stem compared to thick stems (**Figure 4 - 6**, **Chapter 2**). In contrast to primary bast fibres, the secondary bast fibres are poorly appreciated by the textile and bio-composite industries (Müssig and Martens, 2003;Chernova and Gorshkova, 2007;Amaducci et al., 2015). This is because secondary bast fibres have larger amount of lignin compared to primary bast fibres (Liu et al., 2015). Previous studies described that growing hemp at high plant densities is good for fibre production, since it produces hemp plants with thin stems and yields more bast fibre (Struik et al., 2000;Westerhuis et al., 2009;Hall et al., 2014a;Amaducci et al., 2015;Westerhuis, 2016). Our results support this observation in terms of ratio bast/woody fibres and additionally suggest that thin stems produce bast fibres of better quality than thick stems, as thin stems contain a lower fraction of secondary bast fibres. Therefore, the production of fibres of better quality can be improved by modifying the stem diameter. This morphological goal can be reached through agricultural practices, as the low heritability of stem diameter (**Table 9**, **Chapter 3**) would hamper the modification of stem thickness through breeding programmes.

In **Chapter 3**, the positive correlation between the content of galacturonic acid from the stem (the main component of pectin (Willats et al., 2001)) and the content of woody fibres left on the bast fibres after the decortication (χ, shives) indicates that high pectin content hinders the decortication (**Figure 4**, **Chapter 3**). Pectin composing the middle lamella of the procambium (**Figure 1**, **General Introduction**) is involved in the decortication efficiency (Müssig and Martens, 2003;Müssig et al., 2008;Amaducci et al., 2015;Musio et al., 2018). This is essentially because of the adhesion function of pectin between bast and woody fibres (reviewed in Willats et al. (2001), Voragen et al. (2009) and Xiao and Anderson (2013)). Pectin creates crosslinks between other molecules of pectin (Willats et al., 2001) and lignin (Kang et al., 2019) to maintain the structure of the stem. Therefore, high pectin content is likely to create more crosslinks in the middle lamella, hampering the decortication of the stems. Consequently, our results support the negative effect of pectin in fibre quality (Ranalli, 2004;Müssig et al., 2008;Salentijn et al., 2015).

Results from **Chapter 2**, suggested that secondary bast fibres are composed by rhamnogalacturonan type I pectin (RG-I) with higher degree of substitutions. This was supported by the higher contents of arabinan and galactan (monosaccharides of the RG-I substitutions (Willats et al., 2001)) in stems with large proportion of secondary than primary bast fibres (**Figure 5**, **Chapter 2**). Secondary bast fibres connect with the woody fibres through the vascular cambium (**Figure 1**, **General introduction**) (Chernova et al., 2018). Thereby, the results from **Chapter 2** and **Chapter 3** suggest that the larger amount

of secondary bast fibres hampers the decortication efficiency and pectin play here a major role.

Moreover, the size of the bast area showed a positive correlation with the content of pectin in the bast fibres (**Figure 5, Chapter 2**). Pectin is also found in the bast fibre, filling the space between fibre bundles and inside the fibre bundles between the elementary fibres (**Figure 1**, **General Introduction**). This fraction of pectin is responsible to support the structure of the bast fibres and the fibre bundles (Willats et al., 2001;Xiao and Anderson, 2013). In addition, it influences the strength and elasticity of the fibres (Willats et al., 2006). Fuentes and co-workers reported that stems with a higher content of middle lamella in the bast fibre tissue were associated to good mechanical properties, due to large fibre strength (Fuentes et al., 2017). An obvious next step would be to investigate the relationship between morphology of the stem, pectin content and mechanical properties. Understanding the relationship between these three factors is of paramount importance to better understand the role of pectin content in fibre quality. The presence of high pectin levels in the vascular cambium has negative implications for fibre quality. Yet, pectin content in the bast fibre bundles might strengthen the microfibrils of cellulose (Willats et al., 2001). In this way, pectin might improve fibre quality (Fuentes et al., 2017).

Finally, in **Chapter 3**, monoecious accessions were found to have higher contents of bast fibre and cellulose, and a lower content of lignin compared to dioecious (**Figure 4**, **Chapter 3**). The fibres produced by monoecious hemp have been reported to be more homogeneous in composition, a characteristic which is associated with better quality of fibres (Bócsa and Karus, 1998;Mandolino and Carboni, 2004). Dioecious hemp plants generally have a longer vegetative growth period and start to flower later than monoecious genotypes (Amaducci et al., 2008b;Amaducci et al., 2008c). As a consequence, dioecious plants generate a higher stem yield, though not always a higher bast fibre yield (van der Werf et al., 1994;Höppner and Menge-Hartmann, 2007;Amaducci et al., 2008b;Amaducci et al., 2008c;Amaducci et al., 2015). In addition, the sexual dimorphism between dioecious males and dioecious females is responsible to the phenotypic heterogeneity (Mandolino and Carboni, 2004;Faux et al., 2016). Our results confirm that monoecious hemp genotypes have better fibre quality than dioecious hemp genotypes. These interactions reiterate the importance of sex determination and the selection of cultivars to produce fibres of high quality in hemp (Amaducci et al., 2015).

2.6 Identification of QTLs for hemp fibre quality

Up to date, hemp breeding programmes are mostly based on conventional breeding schemes. These practices commonly consist on the use of natural variation, the generation of varietal parents through mass selection and cross breeding, and the development of cultivars through inbreeding and hybrid breeding (Salentijn et al., 2015). Conventional

breeding requires the phenotyping of many plants in several generations, increasing the costs and time span of breeding programmes.

Marker-assisted selection (MAS), by contrast, reduces the need to phenotype large number of samples and assist the selection decisions. This enables to collect material from only selected genotypes. Nonetheless, MAS schemes have not yet been implemented in fibre hemp breeding programmes (Salentijn et al., 2015).

The identification of quantitative trait loci (QTLs) can be performed through QTL mapping on a (bi)parental population or a panel of diverse accessions. To the latter is called genome-wide association studies (GWAS) and it is the approach we follow in this thesis. GWAS makes use of the natural genetic variation from a large panel of accessions (Huang and Han, 2014). This offers a fine resolution, almost to the base pair resolution. In contrast, a (bi)parental population approach can be limited to the genetic diversity present into the parents of the segregating population, despite its high statistical power for detecting QTLs. Therefore, when interested in the study of the whole genetic architecture of a trait including the number of QTLs and their effects, GWAS approach appears to be an advantage (Cavanagh et al., 2008;Korte and Farlow, 2013;Wallace et al., 2014). Moreover, GWAS has the advantage to use ancient recombination events and it does not need to create several filial generations such as in (bi)parental population approaches, which is labour and cost intensive. This reduces the time span to identify QTLs associated to fibre quality.

However, the lack of genetic tools and the fragmented genome sequence of hemp did not provide a good basis for GWAS in hemp (Sawler et al., 2015) at the start of this thesis. To overcome these hindrances, in **Chapter 4**, we first developed an alternative methodology to identify independent QTLs with the help of a GWAS approach but without the help of a complete genome sequence. This method was based on the identification of non-random associations between markers that were found to be significant for a trait in a GWAS, after correcting for putative population structure effects. This correction ensured that the associations between markers were not based on genetic drift or selection but most likely based on physical proximity in tight genomic regions. Association between markers also reflected that these correlated markers explained a similar part of the phenotypic variance in the GWAS population. The marker that represented better this variance was assigned as the marker representing a QTL. This method enabled genome-wide association analyses for hemp, and it can be extended to genetic studies in other orphan species, for which no complete and assembled genome is available (Graham, 2013).

In the present study, we identified 24 QTLs across locations for fibre quality relevant traits of hemp, including cell wall composition, bast fibre content, flowering time and sex determination (**Table 3**, **Chapter 4** and **Table 3**, **Chapter 6**). These QTLs will have positive implications for hemp breeding programmes, that aim to develop cultivars with large fibre quality. QTLs for cell wall composition and bast fibre content will directly improve traits

related the fibres. In addition, QTLs for flowering time traits will help to select later and shorter flowering hemp. A later flowering phenotype is associated with slowing down the physiological processes that affect fibre quality negatively, such as lignification of the stems (Amaducci et al., 2015;Liu et al., 2015). Moreover, QTLs for sex determination will promote monoecy and consequently will decrease the heterogeneity of fibre quality and seed yield. The use in breeding programmes of the QTLs identified across two locations will develop genotypes that perform well under certain environments but not necessarily under other environments. Furthermore, when breeding for QTLs identified across the three locations, the resulting genotypes will perform well for specific traits under more environments. Therefore, the two or three markers from each QTL across locations would be the first candidates to include in MAS schemes of hemp. This shift is expected to accelerate the development of commercial hemp cultivars with improved fibre quality, suitable for bio-based applications, such as bio-composites.

The candidate genes identified in the QTLs across locations provide insights into the molecular mechanisms affecting cell wall composition, flowering time and sex determination in hemp (**Table 4**, **Chapter 4** and **Table 4**, **Chapter 6**). This is the first study reporting molecular reasoning for sex determination of hemp, which represents a significant step towards understanding the sex determination system of cannabis. We proposed a system for monoecious hemp similar to the M locus from spinach (**Chapter 5** and **6**) (West and Golenberg, 2018), where the QTLs detected are male-promoting and female-supressing factors. This system can be explained by an alteration of phytohormones, gibberellic acid through the candidate gene *gibberellic acid insensitive* (DELLA protein) and auxins through the candidate genes *auxins response factors* (ARF proteins) (**Figure 4**, **Chapter 6**). DELLA protein regulates the expression of masculinizing *B-class* genes (West and Golenberg, 2018), while ARF proteins regulate the female flower development (Hardtke and Berleth, 1998;Goetz et al., 2006;Liu et al., 2014;Li et al., 2016). To sum up, the candidate genes that were identified near the QTLs are of great significance to further validate the QTLs and the biology behind, through functional analysis, such as reverse genetic approaches.

3. Future studies on fibre quality of hemp

3.1 Further development of prediction models

Large variation in the contents of crystalline polysaccharides was described among six hemp accessions in **Chapter 2**. However, the extent of phenotypic variability, heritability and environmental adaptability of crystalline structures in larger hemp panels (e.g. the hemp GWAS panel described in **Chapter 3**) are still unknown. This knowledge is desired because the percentage of crystalline polysaccharides (cellulose and mannan) in the bast

fibres of hemp is associated to improved mechanical properties, such as fibre strength (Thygesen et al., 2006).

Ozaki and co-workers reported that the content of crystalline structures is predictable with the NIRS technology (Ozaki et al., 2017). Therefore, the combination of the biochemical protocols for crystalline structures developed in **Chapter 2** and the NIRS technology is expected to develop prediction models for these important traits. These models will provide the tools to further study the crystalline structures of hemp and provide deeper insights into the genetic regulation of crystalline structures.

In addition, few studies have reported the presence of RG-I pectin in the gelatinous layer (G-layer, S3 layer) of the hemp bast fibre (Gorshkova et al., 2010;Behr et al., 2016;Chernova et al., 2018). RG-I pectin in the G-layer has been well studied in flax (Gorshkova et al., 2015;Guedes et al., 2017) and it has been associated to mechanical properties of flax fibres (Arnould et al., 2017). Yet, this information is limited in hemp (Blake et al., 2008). The crystalline structure protocol developed in **Chapter 2**, also allowed to study the contents of monosaccharides derived from RG-I pectin in the crystalline fraction of the cell wall. Chernova and co-workers described that RG-I pectins are entrapped between the microfibrils of cellulose and mannan (Chernova et al., 2018). Therefore, the RG-I pectins detected in our protocols were assumed to belong to the secondary cell wall or the G-layer because they are the cell wall layers with larger crystalline polysaccharides (Crônier et al., 2005). The use of these biochemical protocols, along with the NIRS technology is also expected to develop prediction models to study the RG-I pectin from the crystalline fractions in many hemp accessions. These models will be especially important to get deeper insights into the role of RG-I in the mechanical properties of hemp fibre and thus the implication of RG-I for fibre quality.

3.2 Optimizing fibre quality at the tissue level

The majority of fibre quality variability studies in fibre crops, including bast fibre crops such as flax (You et al., 2017), focused the evaluation of the composition on the entire stem samples. However, it is known that stems of bast fibre crops are organized in different fibres, bast and woody fibres, as depicted in **Figure 1**, **General Introduction**, with different suborganizations (**Figure 6**, **Chapter 2**) and cell wall composition (**Figure 2**, **Chapter 2**) (Crônier et al., 2005). For example, bast fibre of hemp has larger amount of cellulose and pectin than that of the woody fibres. Moreover, woody fibre and secondary bast fibres are characterized by a higher proportion of lignin content than primary bast fibres (Crônier et al., 2005;Liu et al., 2015).

Results from **Chapter 2** and **Chapter 3** suggest a specific relationship between the tissue, from which certain cell wall components were extracted, and specific fibre quality

properties. This is the case of pectin polysaccharides. Therefore, variability studies of cell wall composition at the tissue level are likely to be key developments to advance fibre quality of hemp. These studies may help to answer several questions, such as, the putative role of RG-I pectin in the crystallization of cellulose and in the mechanical properties of the fibres (Thygesen et al., 2006;Gorshkova et al., 2010;Behr et al., 2016;Chernova et al., 2018). Moreover, such studies may help to minimize cell wall components without interfering with plant fitness. For instance, reducing pectin from the vascular cambium to the level that does not hamper the decortication of the stems, but not reducing pectin to the level that is crucial for the plants. However, the separation of stem into different tissues is not straight-forward, nor high-throughput to allow variability studies where several samples are needed. For this reason, cell wall composition studies of the bast fibres and their relationship with fibre quality have been reported only in a limited number of accessions. Among them, **Chapter 2** is an example. A promising approach that would allow these type of studies are the development of prediction models of cell wall composition proposed in section 3.1. Moreover, the combination of the stem morphological methods for hemp developed in **Chapter 2** with immunohistochemistry (Blake et al., 2008;Behr et al., 2016) would allow the study of certain cell wall components from specific stem tissues, without the need of separating stem samples into their different tissues. This approach combined with computational image-analysis would speed-up the quantification of specific cell wall components (Roeder et al., 2012).

4. Concluding remarks

The new insights reported in this thesis expand the knowledge of hemp fibre quality for and beyond breeding approaches. In this thesis, we highlight the importance of genetic components but, also the paramount influence of the environment in the fibre quality of hemp. Our study advocates for breeding programmes using heritable phenotypic variation less sensitive to the environment, to generate an extent of fibre qualities. Furthermore, the production of sustainable and excellent fibres should take advantage of the adaptive behaviour of hemp fibre quality, through specific agricultural practices. The combination of agricultural practices and elite genotypes, developed via breeding efforts, will boost the production of excellent fibres qualities.

To conclude, bio-based crops, such as hemp, can and should be optimized to replace non-renewable and unsustainable materials and with them reduce GHG emissions. This is very important because every action to reduce GHG footprint is needed, in a time when the breaking down of our climate should not be disregarded. In this respect, plant breeding is an essential player in the transition towards sustainable products.

References

Ahmad, R., Tehsin, Z., Malik, S.T., Asad, S.A., Shahzad, M., Bilal, M., Shah, M.M., and Khan, S.A. (2016). Phytoremediation Potential of Hemp (Cannabis sativa L.): Identification and Characterization of Heavy Metals Responsive Genes. *CLEAN – Soil, Air, Water* 44, 195-201.

Ainsworth, C. (2000). Boys and Girls Come Out to Play: The Molecular Biology of Dioecious Plants. *Annals of Botany* 86, 211-221.

Allegret, S., Arnaud, L., and Bouloc, P. (2013). *Hemp : industrial production and uses*. Wallingford, Oxfordshire ; UK: Cabi.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-3402.

Amaducci, S., Colauzzi, M., Bellocchi, G., Cosentino, S.L., Pahkala, K., Stomph, T.J., Westerhuis, W., Zatta, A., and Venturi, G. (2012). Evaluation of a phenological model for strategic decisions for hemp (Cannabis Sativa L.) biomass production across European sites. *Industrial Crops and Products* 37, 100-110.

Amaducci, S., Colauzzi, M., Bellocchi, G., and Venturi, G. (2008a). Modelling post-emergent hemp phenology (Cannabis sativa L.): Theory and evaluation. *European Journal of Agronomy* 28, 90-102.

Amaducci, S., Colauzzi, M., Zatta, A., and Venturi, G. (2008b). Flowering Dynamics in Monoecious and Dioecious Hemp Genotypes. *Journal of Industrial Hemp* 13, 5-19.

Amaducci, S., and Gusovius, H.J. (2010). "Hemp-Cultivation, Extraction and Processing," in *Industrial Applications of Natural Fibres: Structure, Properties and Technical Applications*, ed. J. Müssig.), 109-134.

Amaducci, S., Scordia, D., Liu, F.H., Zhang, Q., Guo, H., Testa, G., and Cosentino, S.L. (2015). Key cultivation techniques for hemp in Europe and China. *Industrial Crops and Products* 68, 2-16.

Amaducci, S., Zatta, A., Pelatti, F., and 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 Research* 107, 161-169.

Amaducci, S., Zatta, A., Raffanini, M., and Venturi, G. (2008d). Characterisation of hemp (Cannabis sativa L.) roots under different growing conditions. *Plant and Soil* 313, 227-235.

Andre, C.M., Hausman, J.-F., and Guerriero, G. (2016). Cannabis sativa: The Plant of the Thousand and One Molecules. *Frontiers in Plant Science* 7.

Anterola, A.M., and Lewis, N.G. (2002). Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. *Phytochemistry* 61, 221-294.

Argillier, O., Barrière, Y., Traineau, R., Emile, J.C., and Hebert, Y. (1997). Genotype × environment interactions for digestibility traits in silage maize estimated from in vivo measurements with standard sheep. *Plant Breeding* 116, 423-427.

Arnould, O., Siniscalco, D., Bourmaud, A., Le Duigou, A., and Baley, C. (2017). Better insight into the nano-mechanical properties of flax fibre cell walls. *Industrial Crops and Products* 97, 224-228.

Atal, C.K. (1961). Effect of gibberellin on the fibers of hemp. *Economic Botany* 15, 133-139.

Babu, Y., and Bayer, M. (2014). Plant Polygalacturonases Involved in Cell Elongation and Separation-The Same but Different? *Plants (Basel, Switzerland)* 3, 613-623.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., and Johnson, E.A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLOS ONE* 3, e3376.

Baker, D., Pryce, G., Giovannoni, G., and Thompson, A.J. (2003). The therapeutic potential of cannabis. *Lancet Neurol* 2, 291-298.

Barrière, Y., Thomas, J., and Denoue, D. (2008). QTL mapping for lignin content, lignin monomeric composition, p-hydroxycinnamate content, and cell wall digestibility in the maize recombinant inbred line progeny F838×F286. *Plant Science* 175, 585-595.

Barth, M., and Carus, M. (2015). "Carbon Footprint and Sustainability of Different Natural Fibres for Biocomposites and Insulation Material: Study providing data for the automotive and insulation industry.". Report nova-Institut GmbH).

Behr, M., Legay, S., Žižková, E., Motyka, V., Dobrev, P.I., Hausman, J.F., Lutts, S., and Guerriero, G. (2016). Studying secondary growth and bast fiber development: The hemp hypocotyl peeks behind the wall. *Frontiers in Plant Science* 7.

Behr, M., Lutts, S., Hausman, J.-F., and Guerriero, G. (2018a). Jasmonic acid to boost secondary growth in hemp hypocotyl. *Planta* 248, 1029-1036.

Behr, M., Sergeant, K., Leclercq, C.C., Planchon, S., Guignard, C., Lenouvel, A., Renaut, J., Hausman, J.-F., Lutts, S., and Guerriero, G. (2018b). Insights into the molecular regulation of monolignol-derived product biosynthesis in the growing hemp hypocotyl. *BMC Plant Biology* 18, 1.

Bell, E., and Mullet, J.E. (1993). Characterization of an Arabidopsis Lipoxygenase Gene Responsive to Methyl Jasmonate and Wounding. *Plant Physiology* 103, 1133-1137.

Berenji, J., Sikora, V., G., F., and Beherec, O. (2013). "Genetics and selection of hemp", in: *Hemp : industrial production and uses*. (Wallingford, Oxfordshire ; UK: Cabi).

Bernardo, R. (2016). Bandwagons I, too, have known. *Theoretical and Applied Genetics* 129, 2323-2332.

Bielecka, M., Kaminski, F., Adams, I., Poulson, H., Sloan, R., Li, Y., Larson, T.R., Winzer, T., and Graham, I.A. (2014). Targeted mutation of Δ12 and Δ15 desaturase genes in hemp produce major alterations in seed fatty acid composition including a high oleic hemp oil. *Plant Biotechnology Journal* 12, 613-623.

Biemelt, S., Tschiersch, H., and Sonnewald, U. (2004). Impact of Altered Gibberellin Metabolism on Biomass Accumulation, Lignin Biosynthesis, and Photosynthesis in Transgenic Tobacco Plants. *Plant Physiology* 135, 254-265.

Blake, A.W., Marcus, S.E., Copeland, J.E., Blackburn, R.S., and Knox, J.P. (2008). In situ analysis of cell wall polymers associated with phloem fibre cells in stems of hemp, Cannabis sativa L. *Planta* 228, 1-13.

Blande, D., Halimaa, P., Tervahauta, A.I., Aarts, M.G.M., and Kärenlampi, S.O. (2017). De novo transcriptome assemblies of four accessions of the metal hyperaccumulator plant Noccaea caerulescens. *Scientific Data* 4, 160131.

Bócsa, I., and Karus, M. (1998). *The cultivation of hemp: botany, varieties, cultivation and harvesting*. Sebastopol: Hemptech.

Boe, A., and K. Lee, D. (2007). Genetic Variation for Biomass Production in Prairie Cordgrass and Switchgrass. *Crop Science - CROP SCI* 47.

Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin Biosynthesis. *Annual Review of Plant Biology* 54, 519-546.

Bonatti, P.M., Ferrari, C., Focher, B., Grippo, C., Torri, G., and Cosentino, C. (2004). Histochemical and supramolecular studies in determining quality of hemp fibres for textile applications. *Euphytica* 140, 55-64.

Borthwick, H.A., and Scully, N.J. (1954). Photoperiodic Responses of Hemp. *Botanical Gazette* 116, 14-29.

Bouché, F., Detry, N., and Perilleux, C. (2015). Heat can erase epigenetic marks of vernalization in Arabidopsis. *Plant Signal Behav* 10, e990799.

Bouché, F., Lobet, G., Tocquin, P., and Perilleux, C. (2016). FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana. *Nucleic Acids Res* 44, D1167-1171.

Bourmaud, A., Morvan, C., Bouali, A., Placet, V., Perré, P., and Baley, C. (2013). Relationships between micro-fibrillar angle, mechanical properties and biochemical composition of flax fibers. *Industrial Crops and Products* 44, 343-351.

Brown, D.M., Zeef, L.a.H., Ellis, J., Goodacre, R., and Turner, S.R. (2005). Identification of Novel Genes in Arabidopsis Involved in Secondary Cell Wall Formation Using Expression Profiling and Reverse Genetics. *The Plant Cell* 17, 2281-2295.

Cai, Y., Chen, L., Liu, X., Guo, C., Sun, S., Wu, C., Jiang, B., Han, T., and Hou, W. (2018). CRISPR/Cas9-mediated targeted mutagenesis of GmFT2a delays flowering time in soya bean. *Plant Biotechnology Journal* 16, 176-185.

Callaway, J.C. (2004). Hempseed as a nutritional resource: An overview. *Euphytica* 140, 65-72.

Campoy, J.A., Ruiz, D., Egea, J., Rees, D.J.G., Celton, J.M., and Martínez-Gómez, P. (2011). Inheritance of Flowering Time in Apricot (Prunus armeniaca L.) and Analysis of Linked Quantitative Trait Loci (QTLs) using Simple Sequence Repeat (SSR) Markers. *Plant Molecular Biology Reporter* 29, 404-410.

Carpita, N.C. (1996). Structure and biogenesis of the cell walls of grasses. *Annu Rev Plant Physiol Plant Mol Biol* 47, 445-476.

Carpita, N.C., and Gibeaut, D.M. (1993). Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J* 3, 1-30.

Carus, M., and Dammer, L. (2018). The Circular Bioeconomy—Concepts, Opportunities, and Limitations. *Industrial Biotechnology* 14, 83-91.

Cavanagh, C., Morell, M., Mackay, I., and Powell, W. (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology* 11, 215-221.

Chailakhyan, M.K., and Khryanin, V.N. (1978). The influence of growth regulators absorbed by the root on sex expression in hemp plants. *Planta* 138, 181-184.

Chan, A.P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., Melake-Berhan, A., Jones, K.M., Redman, J., Chen, G., Cahoon, E.B., Gedil, M., Stanke, M., Haas, B.J., Wortman, J.R., Fraser-Liggett, C.M., Ravel, J., and Rabinowicz, P.D. (2010). Draft genome sequence of the oilseed species Ricinus communis. *Nat Biotechnol* 28, 951-956.

Chaohua, C., Gonggu, Z., Lining, Z., Chunsheng, G., Qing, T., Jianhua, C., Xinbo, G., Dingxiang, P., and Jianguang, S. (2016). A rapid shoot regeneration protocol from the cotyledons of hemp (Cannabis sativa L.). *Industrial Crops and Products* 83, 61-65.

Chen, H. (2014). *Biotechnology of lignocellulose: Theory and practice.*

Chen, Q., Payyavula, R.S., Chen, L., Zhang, J., Zhang, C., and Turgeon, R. (2018). FLOWERING LOCUS T mRNA is synthesized in specialized companion cells in Arabidopsis and Maryland Mammoth tobacco leaf veins. *Proceedings of the National Academy of Sciences* 115, 2830-2835.

Chen, W.-J., Yun, M.-S., Deng, F., and Yogo, Y. (2011). Chalcone suppresses lignin biosynthesis in illuminated soybean cells. *Weed Biology and Management* 11, 49-56.

Cheng, J.Z., Zhou, Y.P., Lv, T.X., Xie, C.P., and Tian, C.E. (2017). Research progress on the autonomous flowering time pathway in Arabidopsis. *Physiol Mol Biol Plants* 23, 477-485.

Chernova, T.E., and Gorshkova, T.A. (2007). Biogenesis of plant fibers. *Russian Journal of Developmental Biology* 38, 221-232.

Chernova, T.E., Mikshina, P.V., Salnikov, V.V., Ibragimova, N.N., Sautkina, O.V., and Gorshkova, T.A. (2018). Development of distinct cell wall layers both in primary and secondary phloem fibers of hemp (Cannabis sativa L.). *Industrial Crops and Products* 117, 97-109.

Cho, L.-H., Pasriga, R., Yoon, J., Jeon, J.-S., and An, G. (2018). Roles of Sugars in Controlling Flowering Time. *Journal of Plant Biology* 61, 121-130.

Christie, J.M., Blackwood, L., Petersen, J., and Sullivan, S. (2014). Plant Flavoprotein Photoreceptors. *Plant and Cell Physiology* 56, 401-413.

Cober, E.R., and Morrison, M.J. (2010). Regulation of seed yield and agronomic characters by photoperiod sensitivity and growth habit genes in soybean. *Theoretical and Applied Genetics* 120, 1005-1012.

Conesa, A., and Gotz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics* 2008, 619832.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674-3676.

Copley, T.R., Duceppe, M.-O., and O'donoughue, L.S. (2018). Identification of novel loci associated with maturity and yield traits in early maturity soybean plant introduction lines. *BMC Genomics* 19, 167.

Cox, W.J., Cherney, J.H., Pardee, W.D., and Cherney, D.J.R. (1994). Forage Quality and Harvest Index of Corn Hybrids Under Different Growing Conditions. *Agronomy Journal* 86, 277-282.

Crônier, D., Monties, B., and Chabbert, B. (2005). Structure and chemical composition of bast fibers isolated from developing hemp stem. *Journal of Agricultural and Food Chemistry* 53, 8279-8289.

Dam, J. (2008). Natural fibres and the environment : environmental benefits of natural fibre production and use.

Dam, J., and Bos, H. (2004). *The environmental impact of fibre crops in industrial applications.*

Das, A., Chaudhury, S., Kalita, M.C., and Mondal, T.K. (2015). In silico identification, characterization and expression analysis of miRNAs in Cannabis sativa L. *Plant Gene* 2, 17-24.

Davin, N., Edger, P.P., Hefer, C.A., Mizrachi, E., Schuetz, M., Smets, E., Myburg, A.A., Douglas, C.J., Schranz, M.E., and Lens, F. (2016). Functional network analysis of genes differentially expressed during xylogenesis in soc1ful woody Arabidopsis plants. *The Plant Journal* 86, 376-390.

Davis, J.M. (2008). "Genetic Improvement of Poplar (Populus spp.) as a bioenergy crop," in *Genetic Improvement of Bioenergy Crops*, ed. W. Vermerris.), 1-449.

De Bruijn, P.B., Jeppsson, K.-H., Sandin, K., and Nilsson, C. (2009). Mechanical properties of lime–hemp concrete containing shives and fibres. *Biosystems Engineering* 103, 474-479.

De Meijer, E.P.M., and Van Der Werf, H.M.G. (1994). Evaluation of current methods to estimate pulp yield of hemp. *Industrial Crops and Products* 2, 111-120.

De Setta, N., Monteiro-Vitorello, C.B., Metcalfe, C.J., Cruz, G.M.Q., Del Bem, L.E., Vicentini, R., Nogueira, F.T.S., Campos, R.A., Nunes, S.L., Turrini, P.C.G., Vieira, A.P., Ochoa Cruz, E.A., Corrêa, T.C.S., Hotta, C.T., De Mello Varani, A., Vautrin, S., Da Trindade, A.S., De Mendonça Vilela, M., Lembke, C.G., Sato, P.M., De Andrade, R.F., Nishiyama, M.Y., Cardoso-Silva, C.B., Scortecci, K.C., Garcia, A.a.F., Carneiro, M.S., Kim, C., Paterson, A.H., Bergès, H., D'hont, A., De Souza, A.P., Souza, G.M., Vincentz, M., Kitajima, J.P., and Van Sluys, M.-A. (2014). Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics* 15, 540.

Demartini, J.D., Pattathil, S., Miller, J.S., Li, H., Hahn, M.G., and Wyman, C.E. (2013). Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. *Energy & Environmental Science* 6, 898-909.

Deyholos, M.K., and Potter, S. (2014). Engineering bast fiber feedstocks for use in composite materials. *Biocatalysis and Agricultural Biotechnology* 3, 53-57.

Dill, A., Jung, H.-S., and Sun, T.-P. (2001). The DELLA motif is essential for gibberellin-induced degradation of RGA. *Proceedings of the National Academy of Sciences* 98, 14162-14167.

Ding, Z., Millar, A.J., Davis, A.M., and Davis, S.J. (2007). TIME FOR COFFEE Encodes a Nuclear Regulator in the Arabidopsis thaliana Circadian Clock. *The Plant Cell* 19, 1522-1536.

Do, C.T., Pollet, B., Thevenin, J., Sibout, R., Denoue, D., Barriere, Y., Lapierre, C., and Jouanin, L. (2007). Both caffeoyl Coenzyme A 3-O-methyltransferase 1 and caffeic acid O-methyltransferase 1 are involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in Arabidopsis. *Planta* 226, 1117-1129.

Doi, K., Izawa, T., Fuse, T., Yamanouchi, U., Kubo, T., Shimatani, Z., Yano, M., and Yoshimura, A. (2004). Ehd1, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1. *Genes & development* 18, 926-936.

Doka, G. (2013). Excel 3D Scatter Plot macros version 2.1 http://www.doka.ch/Excel3Dscatterplot.htm.

Dolstra, O., Medema, J.H., and De Jong, A.W. (1992). Genetic improvement of cell-wall digestibility in forage maize (Zea mays L.). I. Performance of inbred lines and related hybrids. *Euphytica* 65, 187-194.

Doyle, J.J., and Doyle, J.A. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19, 11-15.

Durand, E., Bouchet, S., Bertin, P., Ressayre, A., Jamin, P., Charcosset, A., Dillmann, C., and Tenaillon, M.I. (2012). Flowering time in maize: linkage and epistasis at a major effect locus. *Genetics* 190, 1547-1562.

Ebskamp, M.J.M. (2002). Engineering flax and hemp for an alternative to cotton. *Trends in Biotechnology* 20, 229-230.

Ehrenreich, I.M., Hanzawa, Y., Chou, L., Roe, J.L., Kover, P.X., and Purugganan, M.D. (2009). Candidate Gene Association Mapping of Arabidopsis Flowering Time. *Genetics* 183, 325-335.

El-Soda, M., Malosetti, M., Zwaan, B.J., Koornneef, M., and Aarts, M.G. (2014). Genotypexenvironment interaction QTL mapping in plants: lessons from Arabidopsis. *Trends Plant Sci* 19, 390-398.

Elfordy, S., Lucas, F., Tancret, F., Scudeller, Y., and Goudet, L. (2008). Mechanical and thermal properties of lime and hemp concrete ("hempcrete") manufactured by a projection process. *Construction and Building Materials* 22, 2116-2123.

Falcioni, R., Moriwaki, T., De Oliveira, D.M., Andreotti, G.C., De Souza, L.A., Dos Santos, W.D., Bonato, C.M., and Antunes, W.C. (2018). Increased Gibberellins and Light Levels Promotes Cell Wall Thickness and Enhance Lignin Deposition in Xylem Fibers. *Frontiers in Plant Science* 9.

Fao. 2003. Fertilizer Use by Crop in Uzbekistan Available: http://www.fao.org/3/Y4711E/y4711e00.htm#Contents.

Fao. 2012. Biofuel co-products as livestock feed - Opportunities and challenges. Available: http://www.fao.org/3/i3009e/i3009e00.htm.

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

Faux, A.M., Draye, X., Flamand, M.C., Occre, A., and Bertin, P. (2016). Identification of QTLs for sex expression in dioecious and monoecious hemp (Cannabis sativa L.). *Euphytica* 209, 357-376.

Faux, A.M., Draye, X., Lambert, R., D'andrimont, R., Raulier, P., and Bertin, P. (2013). The relationship of stem and seed yields to flowering phenology and sex expression in monoecious hemp (Cannabis sativa L.). *European Journal of Agronomy* 47, 11-22.

Foster, C.E., Martin, T.M., and Pauly, M. (2010). Comprehensive compositional analysis of plant cell walls (lignocellulosic biomass) part II: carbohydrates. *Journal of visualized experiments : JoVE*.

Fournier-Level, A., Korte, A., Cooper, M.D., Nordborg, M., Schmitt, J., and Wilczek, A.M. (2011). A Map of Local Adaptation in Arabidopsis thaliana. *Science* 334, 86-89.

Freeman, D.C., Harper, K.T., and Charnov, E.L. (1980). Sex change in plants: Old and new observations and new hypotheses. *Oecologia* 47, 222-232.

Frollini, E., Leao, A., H. C. Mattoso, L., Rowell, R., S. Han, J., and S. Rowell, J. (2000). *Characterization and Factors Effecting Fiber Properties*.

Fry, S.C. (2003). "Cell walls and fibers," in *Encyclopedia of Applied Plant Sciences,* ed. B.G.M. Brian Thomas, Denis J. Murphy. Academic Press).

Fuentes, C.A., Willekens, P., Petit, J., Thouminot, C., Müssig, J., Trindade, L.M., and Van Vuure, A.W. (2017). Effect of the middle lamella biochemical composition on the non-linear behaviour of technical fibres of hemp under tensile loading using strain mapping. *Composites Part A: Applied Science and Manufacturing* 101, 529-542.

Galoch, E. (2015). *The hormonal control of sex differentiation in dioecious plants of hemp (Cannabis sativd). The influence of plant growth regulators on sex expression in male and female plants*.

Galvao, V.C., and Fankhauser, C. (2015). Sensing the light environment in plants: photoreceptors and early signaling steps. *Curr Opin Neurobiol* 34, 46-53.

Gibson, L.J. (2012). The hierarchical structure and mechanics of plant materials. *J R Soc Interface* 9, 2749-2766.

Gitonga, V.W., Koning-Boucoiran, C.F., Verlinden, K., Dolstra, O., Visser, R.G., Maliepaard, C., and Krens, F.A. (2014). Genetic variation, heritability and genotype by environment interaction of morphological traits in a tetraploid rose population. *BMC Genet* 15, 146.

Goering, H.K., and Van Soest, P.J. (1970). "Forage fiber analyses (apparatus, reagents, procedures, and some applications)". (Washington, D.C. :: Agricultural Research Service, U.S. Dept. of Agriculture).

Goetz, M., Vivian-Smith, A., Johnson, S.D., and Koltunow, A.M. (2006). AUXIN RESPONSE FACTOR8 is a negative regulator of fruit initiation in Arabidopsis. *Plant Cell* 18, 1873-1886.

Gonzalez-Carranza, Z.H., Elliott, K.A., and Roberts, J.A. (2007). Expression of polygalacturonases and evidence to support their role during cell separation processes in Arabidopsis thaliana. *J Exp Bot* 58, 3719-3730.

Gorshkova, T., Brutch, N., Chabbert, B., Deyholos, M., Hayashi, T., Lev-Yadun, S., Mellerowicz, E.J., Morvan, C., Neutelings, G., and Pilate, G. (2012). Plant Fiber Formation: State of the Art, Recent and Expected Progress, and Open Questions. *Critical Reviews in Plant Sciences* 31, 201-228.

Gorshkova, T., Mokshina, N., Chernova, T., Ibragimova, N., Salnikov, V., Mikshina, P., Tryfona, T., Banasiak, A., Immerzeel, P., Dupree, P., and Mellerowicz, E.J. (2015). Aspen Tension Wood Fibers Contain beta-(1-->4)-Galactans and Acidic Arabinogalactans Retained by Cellulose Microfibrils in Gelatinous Walls. *Plant Physiol* 169, 2048-2063.

Gorshkova, T.A., Gurjanov, O.P., Mikshina, P.V., Ibragimova, N.N., Mokshina, N.E., Salnikov, V.V., Ageeva, M.V., Amenitskii, S.I., Chernova, T.E., and Chemikosova, S.B. (2010). Specific type of secondary cell wall formed by plant fibers. *Russian Journal of Plant Physiology* 57, 328-341.

Gou, J.-Y., Felippes, F.F., Liu, C.-J., Weigel, D., and Wang, J.-W. (2011). Negative Regulation of Anthocyanin Biosynthesis in Arabidopsis by a miR156-Targeted SPL Transcription Factor. *The Plant Cell* 23, 1512-1522.

Gou, M., Yang, X., Zhao, Y., Ran, X., Song, Y., and Liu, C.-J. (2019). Cytochrome b_5 Is an Obligate Electron Shuttle Protein for Syringyl Lignin Biosynthesis in Arabidopsis. *The Plant Cell* 31, 1344-1366.

Graham, I. (2013). "Using molecular breeding to improve orphan crops for emerging economies," in *Successful Agricultural Innovation in Emerging Economies: New Genetic Technologies for Global Food Production,* eds. D.J. Bennett & R.C. Jennings. (Cambridge: Cambridge University Press), 95-106.

Grant, S., Houben, A., Vyskot, B., Siroky, J., Pan, W.-H., Macas, J., and Saedler, H. (1994). Genetics of sex determination in flowering plants. *Developmental Genetics* 15**,** 214-230.

Grassa, C.J., Wenger, J.P., Dabney, C., Poplawski, S.G., Motley, S.T., Michael, T.P., Schwartz, C.J., and Weiblen, G.D. (2018). A complete Cannabis chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. *bioRxiv***,** 458083.

Gribaa, A., Dardelle, F., Lehner, A., Rihouey, C., Burel, C., Ferchichi, A., Driouich, A., and Mollet, J.C. (2013). Effect of water deficit on the cell wall of the date palm (Phoenix dactylifera 'Deglet nour', Arecales) fruit during development. *Plant Cell Environ* 36**,** 1056-1070.

Griffiths, J., Murase, K., Rieu, I., Zentella, R., Zhang, Z.-L., Powers, S.J., Gong, F., Phillips, A.L., Hedden, P., Sun, T.-P., and Thomas, S.G. (2006). Genetic Characterization and Functional Analysis of the GID1 Gibberellin Receptors in Arabidopsis. *The Plant Cell* 18**,** 3399-3414.

Grohe, B. (2004). Heat conductivities of insulation mats based on water glass bonded non-textile hemp or flax fibres. *Holz als Roh- und Werkstoff* 62**,** 352-357.

Guedes, F.T.P., Laurans, F., Quemener, B., Assor, C., Laine-Prade, V., Boizot, N., Vigouroux, J., Lesage-Descauses, M.C., Leple, J.C., Dejardin, A., and Pilate, G. (2017). Non-cellulosic polysaccharide distribution during G-layer formation in poplar tension wood fibers: abundance of rhamnogalacturonan I and arabinogalactan proteins but no evidence of xyloglucan. *Planta* 246**,** 857-878.

Guerriero, G., Behr, M., Legay, S., Mangeot-Peter, L., Zorzan, S., Ghoniem, M., and Hausman, J.-F. (2017). Transcriptomic profiling of hemp bast fibres at different developmental stages. *Scientific Reports* 7**,** 4961.

Guo, H., Yang, H., Mockler, T.C., and Lin, C. (1998). Regulation of flowering time by Arabidopsis photoreceptors. *Science* 279**,** 1360-1363.

Gupta, O.P., Karkute, S.G., Banerjee, S., Meena, N.L., and Dahuja, A. (2017). Contemporary Understanding of miRNA-Based Regulation of Secondary Metabolites Biosynthesis in Plants. *Frontiers in Plant Science* 8.

Hall, J., Bhattarai, S.P., and Midmore, D.J. (2012). Review of Flowering Control in Industrial Hemp. *Journal of Natural Fibers* 9**,** 23-36.

Hall, J., Bhattarai, S.P., and Midmore, D.J. (2014a). Effect of industrial hemp (*Cannabis sativa* L.) planting density on weed suppression, crop growth, physiological responses, and fibre yield in the subtropics. *Renewable Bioresources* 2.

Hall, J., Bhattarai, S.P., and Midmore, D.J. (2014b). The Effects of Photoperiod on Phenological Development and Yields of Industrial Hemp. *Journal of Natural Fibers* 11**,** 87-106.

Hardtke, C.S., and Berleth, T. (1998). The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. *Embo j* 17**,** 1405-1411.

Hasan, M.A., Hussain, M.H., Chowdhury, A.S., Dhar, S.B., Abedin, M., and Fima, I.N. (2016). Computational identification of potential microRNAs and their targets from expressed sequence tags of marijuana (Cannabis sativa). *Meta Gene* 10**,** 45-55.

Hayama, R., Yokoi, S., Tamaki, S., Yano, M., and Shimamoto, K. (2003). Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature* 422**,** 719-722.

He, N., Zhang, C., Qi, X., Zhao, S., Tao, Y., Yang, G., Lee, T.-H., Wang, X., Cai, Q., Li, D., Lu, M., Liao, S., Luo, G., He, R., Tan, X., Xu, Y., Li, T., Zhao, A., Jia, L., Fu, Q., Zeng, Q., Gao, C., Ma, B., Liang, J., Wang, X., Shang, J., Song, P., Wu, H., Fan, L., Wang, Q., Shuai, Q., Zhu, J., Wei, C., Zhu-Salzman, K., Jin, D., Wang, J., Liu, T., Yu, M., Tang, C., Wang, Z., Dai, F., Chen, J., Liu, Y., Zhao, S., Lin, T., Zhang, S., Wang, J., Wang, J., Yang, H., Yang, G., Wang, J., Paterson, A.H., Xia, Q., Ji, D., and Xiang, Z. (2013). Draft genome sequence of the mulberry tree Morus notabilis. *Nature Communications* 4**,** 2445.

He, Y., Michaels, S.D., and Amasino, R.M. (2003). Regulation of Flowering Time by Histone Acetylation in Arabidopsis. *Science* 302**,** 1751-1754. doi: 10.1126/science.1091109.

Hedden, P., and Phillips, A.L. (2000). Gibberellin metabolism: new insights revealed by the genes. *Trends Plant Sci* 5**,** 523-530.

Heikrujam, M., Sharma, K., Prasad, M., and Agrawal, V. (2014). *Review on different mechanisms of sex determination and sex- linked molecular markers in dioecious crops- A current update*.

Hennink, S. (1994). Optimisation of breeding for agronomic traits in fibre hemp (Cannabis sativa L.) by study of parent-offspring relationships. *Euphytica* 78, 69-76.

Henry, B., Laitala, K., and Klepp, I.G. (2019). Microfibres from apparel and home textiles: Prospects for including microplastics in environmental sustainability assessment. *Sci Total Environ* 652, 483-494.

Heslop-Harrison, J. (1956). Auxin and Sexuality in Cannabis sativa. *Physiologia Plantarum* 9, 588-597.

Hill, C.B., and Li, C. (2016). Genetic Architecture of Flowering Phenology in Cereals and Opportunities for Crop Improvement. *Frontiers in Plant Science* 7.

Hille, F., Richter, H., Wong, S.P., Bratovic, M., Ressel, S., and Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. *Cell* 172, 1239-1259.

Holbery, J., and Houston, D. (2006). Natural-fiber-reinforced polymer composites in automotive applications. *JOM* 58, 80-86.

Hopkins, A.A., Vogel, K.P., Moore, K.J., Johnson, K.D., and Carlson, I.T. (1995). Genotypic Variability and Genotype × Environment Interactions among Switchgrass Accessions from the Midwestern USA. *Crop Science* 35, 565-571.

Höppner, F., and Menge-Hartmann, U. (2007). Yield and quality of fibre and oil of fourteen hemp cultivars in Northern Germany at two harvest dates. *Landbauforschung Volkenrode* 57, 219-232.

Houston, K., Tucker, M.R., Chowdhury, J., Shirley, N., and Little, A. (2016). The Plant Cell Wall: A Complex and Dynamic Structure As Revealed by the Responses of Genes under Stress Conditions. *Frontiers in Plant Science* 7.

Hu, Y.-X., Tao, Y.-B., and Xu, Z.-F. (2017). Overexpression of Jatropha Gibberellin 2-oxidase 6 (JcGA2ox6) Induces Dwarfism and Smaller Leaves, Flowers and Fruits in Arabidopsis and Jatropha. *Frontiers in Plant Science* 8.

Huang, X., and Han, B. (2014). Natural Variations and Genome-Wide Association Studies in Crop Plants. *Annual Review of Plant Biology* 65, 531-551.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, K., Huang, T., Zhou, T., Jing, Y., Li, W., Lin, Z., Buckler, E.S., Qian, Q., Zhang, Q.F., Li, J., and Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 42, 961-967.

Huffman, F.G., and Caballero, B. (2003). "URONIC ACIDS," in *Encyclopedia of Food Sciences and Nutrition (Second Edition)*. (Oxford: Academic Press), 5890-5896.

Hughes, M. (2012). Defects in natural fibres: Their origin, characteristics and implications for natural fibre-reinforced composites. *Journal of Materials Science* 47, 599-609.

Hyun, Y., Richter, R., Vincent, C., Martinez-Gallegos, R., Porri, A., and Coupland, G. (2016). Multi-layered Regulation of SPL15 and Cooperation with SOC1 Integrate Endogenous Flowering Pathways at the Arabidopsis Shoot Meristem. *Dev Cell* 37, 254-266.

Iiyama, K., Lam, T.B.T., and Stone, B.A. (1994). Covalent Cross-Links in the Cell Wall. *Plant physiology* 104, 315-320.

Immink, R.G.H., Posé, D., Ferrario, S., Ott, F., Kaufmann, K., Valentim, F.L., De Folter, S., Van Der Wal, F., Van Dijk, A.D.J., Schmid, M., and Angenent, G.C. (2012). Characterization of SOC1's Central Role in Flowering by the Identification of Its Upstream and Downstream Regulators. *Plant Physiology* 160, 433-449.

Jankauskiene, Z., Butkute, B., Gruzdeviene, E., Cesevičiene, J., and Fernando, A.L. (2015). Chemical composition and physical properties of dew- and water-retted hemp fibers. *Industrial Crops and Products* 75, 206-211.

Jeanneau, M., Vidal, J., Gousset-Dupont, A., Lebouteiller, B., Hodges, M., Gerentes, D., and Perez, P. (2002). Manipulating PEPC levels in plants. *J Exp Bot* 53, 1837-1845.

Jenkins, G.I. (2014). The UV-B Photoreceptor UVR8: From Structure to Physiology. *The Plant Cell* 26, 21-37.

Jiang, D., Yang, W., He, Y., and Amasino, R.M. (2007). Arabidopsi Relatives of the Human Lysine-Specific Demethylase1 Repress the Expression of *FWA* and *FLOWERING LOCUS C* and Thus Promote the Floral Transition. *The Plant Cell* 19, 2975-2987. doi: 10.1105/tpc.107.052373.

Johnson, E., Bradley, M., Harberd, N.P., and Whitelam, G.C. (1994). Photoresponses of Light-Grown phyA Mutants of Arabidopsis (Phytochrome A Is Required for the Perception of Daylength Extensions). *Plant Physiol* 105, 141-149.

Jones, D., and Brischke, C. (2017). "3 - Nonwood bio-based materials," in *Performance of Bio-based Building Materials*. Woodhead Publishing), 97-186.

Jung, C., Pillen, K., Staiger, D., Coupland, G., and Von Korff, M. (2017). Editorial: Recent Advances in Flowering Time Control. *Frontiers in Plant Science* 7.

Kamdee, C., Imsabai, W., Kirk, R., Allan, A.C., Ferguson, I.B., and Ketsa, S. (2014). Regulation of lignin biosynthesis in fruit pericarp hardening of mangosteen (Garcinia mangostana L.) after impact. *Postharvest Biology and Technology* 97, 68-76.

Kang, X., Kirui, A., Dickwella Widanage, M.C., Mentink-Vigier, F., Cosgrove, D.J., and Wang, T. (2019). Lignin-polysaccharide interactions in plant secondary cell walls revealed by solid-state NMR. *Nature Communications* 10, 347.

Kang, Y.J., Kim, S.K., Kim, M.Y., Lestari, P., Kim, K.H., Ha, B.K., Jun, T.H., Hwang, W.J., Lee, T., Lee, J., Shim, S., Yoon, M.Y., Jang, Y.E., Han, K.S., Taeprayoon, P., Yoon, N., Somta, P., Tanya, P., Kim, K.S., Gwag, J.G., Moon, J.K., Lee, Y.H., Park, B.S., Bombarely, A., Doyle, J.J., Jackson, S.A., Schafleitner, R., Srinives, P., Varshney, R.K., and Lee, S.H. (2014). Genome sequence of mungbean and insights into evolution within Vigna species. *Nat Commun* 5, 5443.

Karus, M., and Vogt, D. (2004). European hemp industry: Cultivation, processing and product lines. *Euphytica* 140, 7-12.

Keller, A., Leupin, M., Mediavilla, V., and Wintermantel, E. (2001). Influence of the growth stage of industrial hemp on chemical and physical properties of the fibres. *Industrial Crops and Products* 13, 35-48.

Kelliher, T., Starr, D., Su, X., Tang, G., Chen, Z., Carter, J., Wittich, P.E., Dong, S., Green, J., Burch, E., Mccuiston, J., Gu, W., Sun, Y., Strebe, T., Roberts, J., Bate, N.J., and Que, Q. (2019). One-step genome editing of elite crop germplasm during haploid induction. *Nature Biotechnology* 37, 287-292.

Klasnja, B., Kopitovic, S., and Orlovic, S. (2003). Variability of some wood properties of eastern cottonwood (Populus deltoides Bartr.) clones. *Wood Science and Technology* 37, 331-337.

Klein, J., Saedler, H., and Huijser, P. (1996). A new family of DNA binding proteins includes putative transcriptional regulators of theAntirrhinum majus floral meristem identity geneSQUAMOSA. *Molecular and General Genetics MGG* 250, 7-16.

Klocko, A.L., Ma, C., Robertson, S., Esfandiari, E., Nilsson, O., and Strauss, S.H. (2016). FT overexpression induces precocious flowering and normal reproductive development in Eucalyptus. *Plant Biotechnology Journal* 14, 808-819.

Kojima, S., Takahashi, Y., Kobayashi, Y., Monna, L., Sasaki, T., Araki, T., and Yano, M. (2002). Hd3a, a Rice Ortholog of the Arabidopsis FT Gene, Promotes Transition to Flowering Downstream of Hd1 under Short-Day Conditions. *Plant and Cell Physiology* 43, 1096-1105.

Kolmos, E., Nowak, M., Werner, M., Fischer, K., Schwarz, G., Mathews, S., Schoof, H., Nagy, F., Bujnicki, J.M., and Davis, S.J. (2009). Integrating ELF4 into the circadian system through combined structural and functional studies. *HFSP Journal* 3, 350-366.

Komiya, R., Ikegami, A., Tamaki, S., Yokoi, S., and Shimamoto, K. (2008). *Hd3a* and *RFT1* are essential for flowering in rice. *Development* 135, 767-774.

Komiya, R., Yokoi, S., and Shimamoto, K. (2009). A gene network for long-day flowering activates *RFT1* encoding a mobile flowering signal in rice. *Development* 136, 3443-3450.

Kong, F., Liu, B., Xia, Z., Sato, S., Kim, B.M., Watanabe, S., Yamada, T., Tabata, S., Kanazawa, A., Harada, K., and Abe, J. (2010). Two Coordinately Regulated Homologs of *FLOWERING LOCUS T* Are Involved in the Control of Photoperiodic Flowering in Soybean. *Plant Physiology* 154, 1220-1231.

Kong, S.-G., and Okajima, K. (2016). Diverse photoreceptors and light responses in plants. *Journal of Plant Research* 129, 111-114.

Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant methods* 9, 29-29.

Kruijer, W., Boer, M.P., Malosetti, M., Flood, P.J., Engel, B., Kooke, R., Keurentjes, J.J., and Van Eeuwijk, F.A. (2015). Marker-based estimation of heritability in immortal populations. *Genetics* 199, 379-398.

Kumar, M., Campbell, L., and Turner, S. (2016). Secondary cell walls: biosynthesis and manipulation. *J Exp Bot* 67, 515-531.

Kushanov, F.N., Buriev, Z.T., Shermatov, S.E., Turaev, O.S., Norov, T.M., Pepper, A.E., Saha, S., Ulloa, M., Yu, J.Z., Jenkins, J.N., Abdukarimov, A., and Abdurakhmonov, I.Y. (2017). QTL mapping for flowering-time and photoperiod insensitivity of cotton Gossypium darwinii Watt. *PloS one* 12, e0186240-e0186240.

Langewisch, T., Lenis, J., Jiang, G.-L., Wang, D., Pantalone, V., and Bilyeu, K. (2017). The development and use of a molecular model for soybean maturity groups. *BMC Plant Biology* 17, 91.

Laverty, K.U., Stout, J.M., Sullivan, M.J., Shah, H., Gill, N., Holbrook, L., Deikus, G., Sebra, R., Hughes, T.R., Page, J.E., and Van Bakel, H. (2019). A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci. *Genome Res* 29, 146-156.

Le Gall, H., Philippe, F., Domon, J.M., Gillet, F., Pelloux, J., and Rayon, C. (2015). Cell Wall Metabolism in Response to Abiotic Stress. *Plants (Basel)* 4, 112-166.

Lee, J., and Lee, I. (2010). Regulation and function of SOC1, a flowering pathway integrator. *J Exp Bot* 61, 2247-2254.

Leivar, P., and Monte, E. (2014). PIFs: Systems Integrators in Plant Development. *The Plant Cell* 26, 56-78.

Leizer, C., Ribnicky, D., Poulev, A., Dushenkov, S., and Raskin, I. (2000). The Composition of Hemp Seed Oil and Its Potential as an Important Source of Nutrition. *Journal of Nutraceuticals, Functional & Medical Foods* 2, 35-53.

Leucci, M., Lenucci, M., Piro, G., and Dalessandro, G. (2008). Water stress and cell wall polysaccharides in the apical root zone of wheat cultivars varying in drought tolerance. *Journal of plant physiology* 165, 1168-1180.

Levy, Y.Y., Mesnage, S., Mylne, J.S., Gendall, A.R., and Dean, C. (2002). Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. *Science* 297, 243-246.

Lewis, N.G. (1999). A 20th century roller coaster ride: a short account of lignification. *Current Opinion in Plant Biology* 2, 153-162.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.

Li, J., and Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* 95, 221-227.

Li, R., Li, Y., Fang, X., Yang, H., Wang, J., Kristiansen, K., and Wang, J. (2009). SNP detection for massively parallel whole-genome resequencing. *Genome Res* 19, 1124-1132.

Li, S.-B., Xie, Z.-Z., Hu, C.-G., and Zhang, J.-Z. (2016). A Review of Auxin Response Factors (ARFs) in Plants. *Frontiers in plant science* 7, 47-47.

Li, X., Wang, S., Du, G., Wu, Z., and Meng, Y. (2013). Variation in physical and mechanical properties of hemp stalk fibers along height of stem. *Industrial Crops and Products* 42, 344-348.

Li, X.R., Wang, L., and Ruan, Y.L. (2010). Developmental and molecular physiological evidence for the role of phosphoenolpyruvate carboxylase in rapid cotton fibre elongation. *J Exp Bot* 61, 287-295.

Liebminger, E., Hüttner, S., Vavra, U., Fischl, R., Schoberer, J., Grass, J., Blaukopf, C., Seifert, G.J., Altmann, F., Mach, L., and Strasser, R. (2009). Class I alpha-mannosidases are required for N-glycan processing and root development in Arabidopsis thaliana. *The Plant cell* 21, 3850-3867.

Lifschitz, E., Ayre, B.G., and Eshed, Y. (2014). Florigen and anti-florigen – a systemic mechanism for coordinating growth and termination in flowering plants. *Frontiers in Plant Science* 5.

Lihova, J., Shimizu, K.K., and Marhold, K. (2006). Allopolyploid origin of Cardamine asarifolia (Brassicaceae): incongruence between plastid and nuclear ribosomal DNA sequences solved by a single-copy nuclear gene. *Mol Phylogenet Evol* 39, 759-786.

Lima, M.D.F., Eloy, N.B., Siqueira, J.a.B.D., Inzé, D., Hemerly, A.S., and Ferreira, P.C.G. (2017). Molecular mechanisms of biomass increase in plants. *Biotechnology Research and Innovation* 1, 14-25.

Lin, C. (2000). Photoreceptors and Regulation of Flowering Time. *Plant Physiology* 123, 39-50.

Lin, C., Ahmad, M., Gordon, D., and Cashmore, A.R. (1995). Expression of an Arabidopsis cryptochrome gene in transgenic tobacco results in hypersensitivity to blue, UV-A, and green light. *Proc Natl Acad Sci U S A* 92, 8423-8427.

Linger, P., Müssig, J., Fischer, H., and Kobert, J. (2002). Industrial hemp (Cannabis sativa L.) growing on heavy metal contaminated soil: fibre quality and phytoremediation potential. *Industrial Crops and Products* 16, 33-42.

Lisson, S.N., Mendham, N.J., and Carberry, P.S. (2000a). Development of a hemp (*Cannabis sativa* L.) simulation model 2.The flowering response of two hemp cultivars to photoperiod. *Australian Journal of Experimental Agriculture* 40, 413-417.

Lisson, S.N., Mendham, N.J., and Carberry, P.S. (2000b). Development of a hemp (*Cannabis sativa* L.) simulation model 4. Model description and validation. *Australian Journal of Experimental Agriculture* 40, 425-432.

Liu, B., Kanazawa, A., Matsumura, H., Takahashi, R., Harada, K., and Abe, J. (2008). Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. *Genetics* 180, 995-1007.

Liu, F., Quesada, V., Crevillén, P., Bäurle, I., Swiezewski, S., and Dean, C. (2007). The Arabidopsis RNA-Binding Protein FCA Requires a Lysine-Specific Demethylase 1 Homolog to Downregulate FLC. *Molecular Cell* 28, 398-407. doi: 10.1016/j.molcel.2007.10.018.

Liu, H., Roeder, K., and Wasserman, L. 2010. Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models. *arXiv e-prints* [Online]. [Accessed June 01, 2010].

Liu, L., Li, B., Xiang, Y., Zhang, R., Yu, J., and Fang, B. (2018a). Effect of Growth Period and Sampling Section on the Chemical Composition and Microstructure of Raw Hemp Fibers. *BioResources* 13, 1961-1976.

Liu, M., Fernando, D., Daniel, G., Madsen, B., Meyer, A.S., Ale, M.T., and Thygesen, A. (2015). Effect of harvest time and field retting duration on the chemical composition, morphology and mechanical properties of hemp fibers. *Industrial Crops and Products* 69, 29-39.

Liu, N., Wu, S., Van Houten, J., Wang, Y., Ding, B., Fei, Z., Clarke, T.H., Reed, J.W., and Van Der Knaap, E. (2014). Down-regulation of AUXIN RESPONSE FACTORS 6 and 8 by microRNA 167 leads to floral development defects and female sterility in tomato. *J Exp Bot* 65, 2507-2520.

Liu, Q., Harberd, N.P., and Fu, X. (2016). SQUAMOSA Promoter Binding Protein-like Transcription Factors: Targets for Improving Cereal Grain Yield. *Mol Plant* 9, 765-767.

Liu, R., Gong, J., Xiao, X., Zhang, Z., Li, J., Liu, A., Lu, Q., Shang, H., Shi, Y., Ge, Q., Iqbal, M.S., Deng, X., Li, S., Pan, J., Duan, L., Zhang, Q., Jiang, X., Zou, X., Hafeez, A., Chen, Q., Geng, H., Gong, W., and Yuan, Y. (2018b). GWAS Analysis and QTL Identification of Fiber Quality Traits and Yield Components in Upland Cotton Using Enriched High-Density SNP Markers. *Frontiers in Plant Science* 9.

Liu, X., Li, Y., Han, B., Zhang, Q., Zhou, K., Zhang, X., and Hashemi, M. (2012). Yield response of continuous soybean to one-season crop disturbance in a previous continuous soybean field in Northeast China. *Field Crops Research* 138, 52-56.

Liu, Z., Wang, H., and Hui, L. (2018c). "Pulping and Papermaking of Non-Wood Fibers.").

Lorenzana, R., Lewis, M., Jung, H.-J., and Bernardo, R. (2010). Quantitative Trait Loci and Trait Correlations for Maize Stover Cell Wall Composition and Glucose Release for Cellulosic Ethanol. *Crop Science* 50.

Lozovaya, V.V., Zabotina, O.A., and Widholm, J.M. (1996). Synthesis and Turnover of Cell-Wall Polysaccharides and Starch in Photosynthetic Soybean Suspension Cultures. *Plant Physiol* 111, 921-929.

Lu, S., Zhao, X., Hu, Y., Liu, S., Nan, H., Li, X., Fang, C., Cao, D., Shi, X., Kong, L., Su, T., Zhang, F., Li, S., Wang, Z., Yuan, X., Cober, E.R., Weller, J.L., Liu, B., Hou, X., Tian, Z., and Kong, F. (2017). Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield. *Nature Genetics* 49, 773.

Lynd, L.R. (1996). Overviwe and evaluation of fuel ethanol from cellulosic biomass: Technology, Economics, the Environment, and Policy. *Annual Review of Energy and the Environment* 21, 403-465.

Malosetti, M., Van Eeuwijk, F.A., Boer, M.P., Casas, A.M., Elia, M., Moralejo, M., Bhat, P.R., Ramsay, L., and Molina-Cano, J.L. (2011). Gene and QTL detection in a three-way barley cross under selection by a mixed model with kinship information using SNPs. *Theor Appl Genet* 122, 1605-1616.

Manabe, Y., Nafisi, M., Verhertbruggen, Y., Orfila, C., Gille, S., Rautengarten, C., Cherk, C., Marcus, S.E., Somerville, S., Pauly, M., Knox, J.P., Sakuragi, Y., and Scheller, H.V. (2011). Loss-of-Function Mutation of *REDUCED WALL ACETYLATION2* in Arabidopsis Leads to Reduced Cell Wall Acetylation and Increased Resistance to *Botrytis cinerea*. *Plant Physiology* 155, 1068-1078.

Mandolino, G., and Carboni, A. (2004). Potential of marker-assisted selection in hemp genetic improvement. *Euphytica* 140, 107-120.

Mandolino, G., Carboni, A., Bagatta, M., Moliterni, V.M.C., and 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., and Ranalli, P. (1999). Identification of DNA markers linked to the male sex in dioecious hemp (Cannabis sativa L.). *Theoretical and Applied Genetics* 98, 86-92.

Mandolino, G., and Ranalli, P. (2002). The Applications of Molecular Markers in Genetics and Breeding of Hemp. *Journal of Industrial Hemp* 7, 7-23.

Mankowska, G., Grabowska, L., and P., B. (2006). "Description of selected genotypes of hemp (Cannabis sativa L.) stored in the INF gene bank," in *Renewable Resources and Plant Biotechnology,* ed. G.E.Z.a.F.P. Ryszard Kozlowski. Nova Science Publishers), 173-181.

Marrot, L., Lefeuvre, A., Pontoire, B., Bourmaud, A., and Baley, C. (2013). Analysis of the hemp fiber mechanical properties and their scattering (Fedora 17). *Industrial Crops and Products* 51, 317-327.

Matsoukas, I.G. (2014). Interplay between sugar and hormone signaling pathways modulate floral signal transduction. *Frontiers in Genetics* 5.

Matsunaga, S., and Kawano, S. (2001). Sex Determination by Sex Chromosomes in Dioecious Plants. *Plant Biology* 3, 481-488.

Mazian, B. (2018). Influence of field retting duration on the biochemical, microstructural, thermal and mechanical properties of hemp fibres harvested at the beginning of flowering. *Industrial crops and products* v. 116, pp. 170-181-2018 v.2116.

Mccallum, C.M., Comai, L., Greene, E.A., and Henikoff, S. (2000). Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiol* 123, 439-442.

Mcdougall, G.J., Morrison, I.M., Stewart, D., Weyers, J.D.B., and Hillman, J.R. (1993). Plant fibres: Botany, chemistry and processing for industrial use. *Journal of the Science of Food and Agriculture* 62, 1-20.

Mcgarry, R.C., and Ayre, B.G. (2012). Manipulating plant architecture with members of the CETS gene family. *Plant Sci* 188-189, 71-81.

Mclaughlin, S.B., Kiniry, J.R., Taliaferro, C.M., and De La Torre Ugarte, D. (2006). "Projecting Yield and Utilization Potential of Switchgrass as an Energy Crop," in *Advances in Agronomy*. Academic Press), 267-297.

Mediavilla, V., Leupin, M., and Keller, A. (2001). Influence of the growth stage of industrial hemp on the yield formation in relation to certain fibre quality traits. *Industrial Crops and Products* 13, 49-56.

Meijer, E.P.M.D. (1994). *Diversity in cannabis.* De Meijer.

Meijer, E.P.M.D., and Keizer, L.C.P. (1996). Patterns of diversity in Cannabis. *Genetic Resources and Crop Evolution* 43, 41-52.

Meijer, E.P.M.D., Van Der Kamp, H.J., and Van Eeuwijk, F.A. (1992). Characterisation of Cannabis accessions with regard to cannabinoid content in relation to other plant characters. *Euphytica* 62, 187-200.

Mellerowicz, E.J., and Gorshkova, T.A. (2012). Tensional stress generation in gelatinous fibres: a review and possible mechanism based on cell-wall structure and composition. *Journal of Experimental Botany* 63, 551-565.

Melzer, S., Lens, F., Gennen, J., Vanneste, S., Rohde, A., and Beeckman, T. (2008). Flowering-time genes modulate meristem determinacy and growth form in Arabidopsis thaliana. *Nature Genetics* 40, 1489-1492.

Michaels, S.D., and Amasino, R.M. (1999). *FLOWERING LOCUS C* Encodes a Novel MADS Domain Protein That Acts as a Repressor of Flowering. *The Plant Cell* 11, 949-956.

Millane, R.P., and Hendrixson, T.L. (1994). Crystal structures of mannan and glucomannans. *Carbohydrate Polymers* 25, 245-251.

Ming, R., Bendahmane, A., and Renner, S.S. (2011). Sex Chromosomes in Land Plants. *Annual Review of Plant Biology* 62, 485-514.

Mishra, P., and Panigrahi, K.C. (2015). GIGANTEA – an emerging story. *Frontiers in Plant Science* 6.

Moinuddin, S.G., Jourdes, M., Laskar, D.D., Ki, C., Cardenas, C.L., Kim, K.W., Zhang, D., Davin, L.B., and Lewis, N.G. (2010). Insights into lignin primary structure and deconstruction from Arabidopsis thaliana COMT (caffeic acid O-methyl transferase) mutant Atomt1. *Org Biomol Chem* 8, 3928-3946.

Moliterni, V.M.C., Cattivelli, L., Ranalli, P., and Mandolino, G. (2004). The sexual differentiation of Cannabis sativa L.: A morphological and molecular study. *Euphytica* 140, 95-106.

Monna, L., Lin, X., Kojima, S., Sasaki, T., and Yano, M. (2002). Genetic dissection of a genomic region for a quantitative trait locus, Hd3, into two loci, Hd3a and Hd3b, controlling heading date in rice. *Theor Appl Genet* 104, 772-778.

Montalvo, J.G., and Faught, S.E. (1999). Headspace Resistance Standards for the Shirley Developments Ltd. Micromat Tester. *Textile Research Journal* 69, 269-277.

Moon, J., Suh, S.-S., Lee, H., Choi, K.-R., Hong, C.B., Paek, N.-C., Kim, S.-G., and Lee, I. (2003). The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis. *The Plant Journal* 35**,** 613-623.

Moore, J.P., Farrant, J.M., and Driouich, A. (2008a). A role for pectin-associated arabinans in maintaining the flexibility of the plant cell wall during water deficit stress. *Plant Signal Behav* 3**,** 102-104.

Moore, J.P., Nguema-Ona, E., Fangel, J.U., Willats, W.G., Hugo, A., and Vivier, M.A. (2014). Profiling the main cell wall polysaccharides of grapevine leaves using high-throughput and fractionation methods. *Carbohydr Polym* 99**,** 190-198.

Moore, J.P., Vicre-Gibouin, M., Farrant, J.M., and Driouich, A. (2008b). Adaptations of higher plant cell walls to water loss: drought vs desiccation. *Physiol Plant* 134**,** 237-245.

Morrison, I.M. (1988). Hydrolysis of plant cell walls with trifluoroacetic acid. *Phytochemistry* 27**,** 1097-1100.

Mouradov, A., Cremer, F., and Coupland, G. (2002). Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell* 14 Suppl**,** S111-130.

Murtaza, G., Ghafoor, A., and Qadir, M. (2006). Irrigation and soil management strategies for using saline-sodic water in a cotton–wheat rotation. *Agricultural Water Management* 81**,** 98-114.

Musio, S., Müssig, J., and Amaducci, S. (2018). Optimizing Hemp Fiber Production for High Performance Composite Applications. *Frontiers in Plant Science* 9.

Müssig, J. (2001). *Untersuchung der Eignung heimischer Pflanzenfasern für die Herstellung von naturfaserverstärkten Duroplasten - vom Anbau zum Verbundwerkstoff.* Düsseldorf, DE: VDI Verlag GmbH (ISBN 3-18-363005-2).

Müssig, J., and Amaducci, S. (2018). Scanner based image analysis to characterise the influence of agronomic factors on hemp (Cannabis sativa L.) fibre width. *Industrial Crops and Products* 113**,** 28-37.

Müssig, J., Fischer, H., Graupner, N., and Drieling, A. (2010). "Testing Methods for Measuring Physical and Mechanical Fibre Properties (Plant and Animal Fibres)," in *Industrial Applications of Natural Fibres,* ed. J. Müssig.).

Müssig, J., and Martens, R. (2003). Quality Aspects in Hemp Fibre Production—Influence of Cultivation, Harvesting and Retting. *Journal of Industrial Hemp* 8**,** 11-32.

Müssig, J., Zatta, A., and Venturi, G. (2008). *An Innovative Production System for Hemp Fibre for Textile Destinations: From Laboratory Results to Industrial Validation*.

Mutasa-Göttgens, E., and Hedden, P. (2009). Gibberellin as a factor in floral regulatory networks. *Journal of Experimental Botany* 60**,** 1979-1989.

Nakajima, M., Shimada, A., Takashi, Y., Kim, Y.-C., Park, S.-H., Ueguchi-Tanaka, M., Suzuki, H., Katoh, E., Iuchi, S., Kobayashi, M., Maeda, T., Matsuoka, M., and Yamaguchi, I. (2006). Identification and characterization of Arabidopsis gibberellin receptors. *The Plant Journal* 46**,** 880-889.

Nguema-Ona, E., Vicré-Gibouin, M., Gotté, M., Plancot, B., Lerouge, P., Bardor, M., and Driouich, A. (2014). Cell wall O-glycoproteins and N-glycoproteins: aspects of biosynthesis and function. *Frontiers in plant science* 5**,** 499-499.

Notaguchi, M., Abe, M., Kimura, T., Daimon, Y., Kobayashi, T., Yamaguchi, A., Tomita, Y., Dohi, K., Mori, M., and Araki, T. (2008). Long-distance, graft-transmissible action of Arabidopsis FLOWERING LOCUS T protein to promote flowering. *Plant Cell Physiol* 49**,** 1645-1658.

Novaes, E., Kirst, M., Chiang, V., Winter-Sederoff, H., and Sederoff, R. (2010). Lignin and biomass: a negative correlation for wood formation and lignin content in trees. *Plant physiology* 154**,** 555-561.

Nykter, M. (2006). *Microbial quality of hemp (Cannabis sativa L.) and flax (Linum usitatissimum L.) from plants to thermal insulation.* University of Helsinki.

O'brien, T.P., Feder, N., and Mccully, M.E. (1964). Polychromatic staining of plant cell walls by toluidine blue O. *Protoplasma* 59**,** 368-373.

Onofri, C., and Mandolino, G. (2017). "Genomics and Molecular Markers in Cannabis sativa L."), 319-342.

Ozaki, Y., Genkawa, T., Futami, Y., Lindon, J.C., Tranter, G.E., and Koppenaal, D.W. (2017). "Near-Infrared Spectroscopy," in *Encyclopedia of Spectroscopy and Spectrometry (Third Edition)*. (Oxford: Academic Press), 40-49.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20, 289-290.

Park, S., Baker, J.O., Himmel, M.E., Parilla, P.A., and Johnson, D.K. (2010). Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. *Biotechnology for Biofuels* 3, 10.

Parween, S., Nawaz, K., Roy, R., Pole, A.K., Venkata Suresh, B., Misra, G., Jain, M., Yadav, G., Parida, S.K., Tyagi, A.K., Bhatia, S., and Chattopadhyay, D. (2015). An advanced draft genome assembly of a desi type chickpea (Cicer arietinum L.). *Sci Rep* 5, 12806.

Pauly, M., Gille, S., Liu, L., Mansoori, N., De Souza, A., Schultink, A., and Xiong, G. (2013). Hemicellulose biosynthesis. *Planta* 238, 627-642.

Peças, P., Carvalho, H., Salman, H., and Leite, M. (2018). Natural Fibre Composites and Their Applications: A Review. *Journal of Composites Science* 2, 66.

Peil, A., Flachowsky, H., Schumann, E., and Weber, W.E. (2003). Sex-linked AFLP markers indicate a pseudoautosomal region in hemp (Cannabis sativa L.). *Theoretical and Applied Genetics* 107, 102-109.

Peng, J., Carol, P., Richards, D.E., King, K.E., Cowling, R.J., Murphy, G.P., and Harberd, N.P. (1997). The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin responses. *Genes & development* 11, 3194-3205.

Petit, J., Gulisano, A., Dechesne, A., and Trindade, L.M. (2019a). Phenotypic Variation of Cell Wall Composition and Stem Morphology in Hemp (Cannabis sativa L.): Optimization of Methods. *Frontiers in Plant Science* 10.

Petit, J., Salentijn, E.M.J., Paulo, M.J., Denneboom, C., Van Loo, E.N., and Trindade, L.M. (2019b). Elucidating the genetic architecture of fibre quality in hemp (Cannabis sativa L.) using a Genome-Wide Association Study. *In preparation*.

Petit, J., Salentijn, E.M.J., Paulo, M.J., Thourminot, C., Van Dinter, B.J., Magagnini, G., Gusovius, H.J., Tang, K., Amaducci, S., Wang, S., Uhrlaub, B., Mussig, J., and Trindade, L.M. (2019c). Genetic variability of morphological, flowering and biomass quality traits in hemp (Cannabis sativa L.). *Frontiers in Plant Science (Accepted)*.

Pettolino, F.A., Walsh, C., Fincher, G.B., and Bacic, A. (2012). Determining the polysaccharide composition of plant cell walls. *Nat Protoc* 7, 1590-1607.

Pfister, S., Bayer, P., Koehler, A., and Hellweg, S. (2011). Environmental Impacts of Water Use in Global Crop Production: Hotspots and Trade-Offs with Land Use. *Environmental Science & Technology* 45, 5761-5768.

Phillips, A.L., Ward, D.A., Uknes, S., Appleford, N., Lange, T., Huttly, A.K., Gaskin, P., Graebe, J.E., and Hedden, P. (1995). Isolation and Expression of Three Gibberellin 20-Oxidase cDNA Clones from Arabidopsis. *Plant Physiology* 108, 1049-1057.

Piattoni, C.V., Ferrero, D.M.L., Dellaferrera, I., Vegetti, A., and Iglesias, A.Á. (2017). Cytosolic Glyceraldehyde-3-Phosphate Dehydrogenase Is Phosphorylated during Seed Development. *Frontiers in Plant Science* 8.

Piotrowski, S., and Carus, M. (2010). "Natural fibres in technical applications: market and trends.," in *Industrial Applications of Natural Fibres: Structure, Properties and Technical Applications*, ed. J. Müssig. (Ltd, Chichester, UK: John Wiley & Sons).

Piotrowski, S., and Carus, M. (2019). *Ecological benefits of hemp and flax cultivation and products.*

Piro, G., Leucci, M., Waldron, K., and Dalessandro, G. (2003). Exposure to water stress causes changes in the biosynthesis of cell wall polysaccharides in roots of wheat cultivars varying in drought tolerance. *Plant Science* 165, 559-569.

Placet, V. (2009). Characterization of the thermo-mechanical behaviour of Hemp fibres intended for the manufacturing of high performance composites. *Composites Part A: Applied Science and Manufacturing* 40, 1111-1118.

Placet, V., Cissé, O., and Lamine Boubakar, M. (2014). Nonlinear tensile behaviour of elementary hemp fibres. Part I: Investigation of the possible origins using repeated progressive loading with in situ microscopic observations. *Composites Part A: Applied Science and Manufacturing* 56, 319-327.

Poke, F.S., Potts, B.M., Vaillancourt, R.E., and Raymond, C.A. (2006). Genetic parameters for lignin, extractives and decay in Eucalyptus globulus. *Ann. For. Sci.* 63, 813-821.

Ponnu, J., Wahl, V., and Schmid, M. (2011). Trehalose-6-Phosphate: Connecting Plant Metabolism and Development. *Frontiers in Plant Science* 2.

Popper, Z.A., Michel, G., Herve, C., Domozych, D.S., Willats, W.G., Tuohy, M.G., Kloareg, B., and Stengel, D.B. (2011). Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu Rev Plant Biol* 62, 567-590.

Preston, J., and Hileman, L. (2013). Functional Evolution in the Plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) Gene Family. *Frontiers in Plant Science* 4.

Pu, Y., Hu, F., Huang, F., Davison, B.H., and Ragauskas, A.J. (2013). Assessing the molecular structure basis for biomass recalcitrance during dilute acid and hydrothermal pretreatments. *Biotechnology for Biofuels* 6, 15.

Ralph, J., Akiyama, T., Coleman, H.D., and Mansfield, S.D. (2012). Effects on Lignin Structure of Coumarate 3-Hydroxylase Downregulation in Poplar. *Bioenergy research* 5, 1009-1019.

Ralph, J., Akiyama, T., Kim, H., Lu, F., Schatz, P.F., Marita, J.M., Ralph, S.A., Reddy, M.S., Chen, F., and Dixon, R.A. (2006). Effects of coumarate 3-hydroxylase down-regulation on lignin structure. *J Biol Chem* 281, 8843-8853.

Ram, H.Y.M., and Jaiswal, V.S. (1972). Induction of male flowers on female plants of Cannabis sativa by gibberellins and its inhibition by abscisic acid. *Planta* 105, 263-266.

Ramanna, M.S. (1976). Are there heteromorphic sex chromosomes in spinach (Spinacia oleracea L.)? *Euphytica* 25, 277-284.

Ranalli, P. (2004). Current status and future scenarios of hemp breeding. *Euphytica* 140, 121-131.

Ranocha, P., Denance, N., Vanholme, R., Freydier, A., Martinez, Y., Hoffmann, L., Kohler, L., Pouzet, C., Renou, J.P., Sundberg, B., Boerjan, W., and Goffner, D. (2010). Walls are thin 1 (WAT1), an Arabidopsis homolog of Medicago truncatula NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. *Plant J* 63, 469-483.

Raymond, C.A., and Schimleck, L.R. (2002). Development of near infrared reflectance analysis calibrations for estimating genetic parameters for cellulose content in Eucalyptus globulus. *Canadian Journal of Forest Research* 32, 170-176.

Razumova, O.V., Alexandrov, O.S., Divashuk, M.G., Sukhorada, T.I., and Karlov, G.I. (2016). Molecular cytogenetic analysis of monoecious hemp (Cannabis sativa L.) cultivars reveals its karyotype variations and sex chromosomes constitution. *Protoplasma* 253, 895-901.

Reed, J.W., Nagatani, A., Elich, T.D., Fagan, M., and Chory, J. (1994). Phytochrome A and Phytochrome B Have Overlapping but Distinct Functions in Arabidopsis Development. *Plant Physiol* 104, 1139-1149.

Renaud, E.N., Lammerts Van Bueren, E.T., Myers, J.R., Paulo, M.J., Van Eeuwijk, F.A., Zhu, N., and Juvik, J.A. (2014). Variation in broccoli cultivar phytochemical content under organic and conventional management systems: implications in breeding for nutrition. *PLoS One* 9, e95683.

Richter, H., Pezet, R., Viret, O., and Gindro, K. (2005). Characterization of 3 new partial stilbene synthase genes out of over 20 expressed in Vitis vinifera during the interaction with Plasmopara viticola. *Physiological and Molecular Plant Pathology* 67, 248-260.

Rieu, I., Eriksson, S., Powers, S.J., Gong, F., Griffiths, J., Woolley, L., Benlloch, R., Nilsson, O., Thomas, S.G., Hedden, P., and Phillips, A.L. (2008a). Genetic analysis reveals that C19-GA 2-oxidation is a major gibberellin inactivation pathway in Arabidopsis. *Plant Cell* 20, 2420-2436.

Rieu, I., Ruiz-Rivero, O., Fernandez-Garcia, N., Griffiths, J., Powers, S.J., Gong, F., Linhartova, T., Eriksson, S., Nilsson, O., Thomas, S.G., Phillips, A.L., and Hedden, P. (2008b). The gibberellin biosynthetic genes AtGA20ox1 and AtGA20ox2 act, partially redundantly, to promote growth and development throughout the Arabidopsis life cycle. *Plant J* 53, 488-504.

Roberts, K., Grief, C., Hills, G.J., and Shaw, P.J. (1985). Cell wall glycoproteins: structure and function. *J Cell Sci Suppl* 2, 105-127.

Roeder, A.H.K., Cunha, A., Burl, M.C., and Meyerowitz, E.M. (2012). A computational image analysis glossary for biologists. *Development* 139, 3071-3080.

Ruelens, P., De Maagd, R.A., Proost, S., Theißen, G., Geuten, K., and Kaufmann, K. (2013). FLOWERING LOCUS C in monocots and the tandem origin of angiosperm-specific MADS-box genes. *Nature Communications* 4, 2280.

Ryu, C.H., Lee, S., Cho, L.H., Kim, S.L., Lee, Y.S., Choi, S.C., Jeong, H.J., Yi, J., Park, S.J., Han, C.D., and An, G. (2009). OsMADS50 and OsMADS56 function antagonistically in regulating long day (LD)-dependent flowering in rice. *Plant Cell Environ* 32**,** 1412-1427.

Salentijn, E.M.J., Pereira, A., Angenent, G.C., Van Der Linden, C.G., Krens, F., Smulders, M.J.M., and Vosman, B. (2007). Plant translational genomics: from model species to crops. *Molecular Breeding* 20**,** 1-13.

Salentijn, E.M.J., Petit, J., and Trindade, L.M. (2019). The Complex Interactions Between Flowering Behavior and Fiber Quality in Hemp. *Frontiers in Plant Science* 10.

Salentijn, E.M.J., Zhang, Q., Amaducci, S., Yang, M., and Trindade, L.M. (2015). New developments in fiber hemp (Cannabis sativa L.) breeding. *Industrial Crops and Products* 68**,** 32-41.

Samanfar, B., Molnar, S.J., Charette, M., Schoenrock, A., Dehne, F., Golshani, A., Belzile, F., and Cober, E.R. (2017). Mapping and identification of a potential candidate gene for a novel maturity locus, E10, in soybean. *Theor Appl Genet* 130**,** 377-390.

Sánchez-Pérez, R., Del Cueto, J., Dicenta, F., and Martínez-Gómez, P. (2014). Recent advancements to study flowering time in almond and other Prunus species. *Frontiers in plant science* 5**,** 334-334.

Sánchez-Retuerta, C., Suaréz-López, P., and Henriques, R. (2018). Under a New Light: Regulation of Light-Dependent Pathways by Non-coding RNAs. *Frontiers in Plant Science* 9.

Sankari, H.S. (2000). Comparison of bast fibre yield and mechanical fibre properties of hemp (Cannabis sativa L.) cultivars. *Industrial Crops and Products* 11**,** 73-84.

Sapino, S., Carlotti, M.E., Peira, E., and Gallarate, M. (2005). Hemp-seed and olive oils: their stability against oxidation and use in O/W emulsions. *J Cosmet Sci* 56**,** 227-251.

Sarkar, P., Bosneaga, E., and Auer, M. (2009). Plant cell walls throughout evolution: towards a molecular understanding of their design principles. *J Exp Bot* 60**,** 3615-3635.

Sasaki, E., Zhang, P., Atwell, S., Meng, D., and Nordborg, M. (2015). "Missing" G x E Variation Controls Flowering Time in Arabidopsis thaliana. *PLoS genetics* 11**,** e1005597-e1005597.

Sawler, J., Stout, J.M., Gardner, K.M., Hudson, D., Vidmar, J., Butler, L., Page, J.E., and Myles, S. (2015). The Genetic Structure of Marijuana and Hemp. *PLoS One* 10**,** e0133292.

Schimleck, L.R., Kube, P.D., and Raymond, C.A. (2004). Genetic improvement of kraft pulp yield in Eucalyptus nitens using cellulose content determined by near infrared spectroscopy. *Canadian Journal of Forest Research* 34**,** 2363-2370.

Schindele, P., Wolter, F., and Puchta, H. (2018). Transforming plant biology and breeding with CRISPR/Cas9, Cas12 and Cas13. *FEBS Lett* 592**,** 1954-1967.

Schneider, K., Kienow, L., Schmelzer, E., Colby, T., Bartsch, M., Miersch, O., Wasternack, C., Kombrink, E., and Stuible, H.P. (2005). A new type of peroxisomal acyl-coenzyme A synthetase from Arabidopsis thaliana has the catalytic capacity to activate biosynthetic precursors of jasmonic acid. *J Biol Chem* 280**,** 13962-13972.

Schönberg, C., Oksanen, T., Suurnäkki, A., Kettunen, H., and Buchert, J. (2001). "The Importance of Xylan for the Strength Properties of Spruce Kraft Pulp Fibres", in: *Holzforschung*.).

Shabir, G., Aslam, K., Khan, A.R., Shahid, M., Manzoor, H., Noreen, S., Khan, M.A., Baber, M., Sabar, M., Shah, S.M., and Arif, M. (2017). Rice molecular markers and genetic mapping: Current status and prospects. *Journal of Integrative Agriculture* 16**,** 1879-1891.

Sheaffer, C.C., Cash, D., Ehlke, N.J., Henning, J.C., Jewett, J.G., Johnson, K.D., Peterson, M.A., Smith, M., Hansen, J.L., and Viands, D.R. (1998). Entry × Environment Interactions for Alfalfa Forage Quality. *Agronomy Journal* 90**,** 774-780.

Shen, Y., Xiang, Y., Xu, E., Ge, X., and Li, Z. (2018). Major Co-localized QTL for Plant Height, Branch Initiation Height, Stem Diameter, and Flowering Time in an Alien Introgression Derived Brassica napus DH Population. *Frontiers in Plant Science* 9.

Shenk, J.S., and Westerhaus, M.O. (1991). Populations Structuring of Near Infrared Spectra and Modified Partial Least Squares Regression. *Crop Science* 31**,** 1548-1555.

Shim, J.S., and Imaizumi, T. (2015). Circadian clock and photoperiodic response in Arabidopsis: from seasonal flowering to redox homeostasis. *Biochemistry* 54**,** 157-170.

Slavov, G.T., Nipper, R., Robson, P., Farrar, K., Allison, G.G., Bosch, M., Clifton-Brown, J.C., Donnison, I.S., and Jensen, E. (2014). Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass Miscanthus sinensis. *New Phytol* 201**,** 1227-1239.

Sluiter, A., Hames, B., Ruiz, R., Scarlata, C., Sluiter, J., Templeton, D., Crocker, D. (2012). "Determination of structural carbohydrates and lignin in biomass.," in *Laboratory Analytical Procedure*. (Golden, CO, USA: National Renewable Energy Laboratory).

Small, E. (2015). Evolution and Classification of Cannabis sativa (Marijuana, Hemp) in Relation to Human Utilization. *The Botanical Review* 81**,** 189-294.

Smith, H. (2000). Phytochromes and light signal perception by plants—an emerging synthesis. *Nature* 407**,** 585-591.

Soldatova, N.A., and Khryanin, V.N. (2010). The effects of heavy metal salts on the phytohormonal status and sex expression in marijuana. *Russian Journal of Plant Physiology* 57**,** 96-100.

Sorieul, M., Dickson, A., Hill, S.J., and Pearson, H. (2016). Plant Fibre: Molecular Structure and Biomechanical Properties, of a Complex Living Material, Influencing Its Deconstruction towards a Biobased Composite. *Materials (Basel, Switzerland)* 9**,** 618.

Soto-Cerda, B.J., Duguid, S., Booker, H., Rowland, G., Diederichsen, A., and Cloutier, S. (2014). Genomic regions underlying agronomic traits in linseed (Linum usitatissimum L.) as revealed by association mapping. *Journal of integrative plant biology* 56**,** 75-87.

Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A., Holmberg, A., Amini, B., Bhalerao, R., Larsson, M., Villarroel, R., Van Montagu, M., Sandberg, G., Olsson, O., Teeri, T.T., Boerjan, W., Gustafsson, P., Uhlen, M., Sundberg, B., and Lundeberg, J. (1998). Gene discovery in the wood-forming tissues of poplar: analysis of 5, 692 expressed sequence tags. *Proc Natl Acad Sci U S A* 95**,** 13330-13335.

Strasser, R. (2014). Biological significance of complex N-glycans in plants and their impact on plant physiology. *Frontiers in plant science* 5**,** 363-363.

Struik, P.C., Amaducci, S., Bullard, M.J., Stutterheim, N.C., Venturi, G., and Cromack, H.T.H. (2000). Agronomy of fibre hemp (Cannabis sativa L.) in Europe. *Industrial Crops and Products* 11**,** 107-118.

Suaria, G., Avio, C.G., Mineo, A., Lattin, G.L., Magaldi, M.G., Belmonte, G., Moore, C.J., Regoli, F., and Aliani, S. (2016). The Mediterranean Plastic Soup: synthetic polymers in Mediterranean surface waters. *Scientific Reports* 6**,** 37551.

Sykes, R.W., Gjersing, E.L., Foutz, K., Rottmann, W.H., Kuhn, S.A., Foster, C.E., Ziebell, A., Turner, G.B., Decker, S.R., Hinchee, M.A., and Davis, M.F. (2015). Down-regulation of p-coumaroyl quinate/shikimate 3'-hydroxylase (C3'H) and cinnamate 4-hydroxylase (C4H) genes in the lignin biosynthetic pathway of Eucalyptus urophylla x E. grandis leads to improved sugar release. *Biotechnol Biofuels* 8**,** 128.

Takahashi, Y., Shomura, A., Sasaki, T., and Yano, M. (2001). *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the α subunit of protein kinase CK2. *Proceedings of the National Academy of Sciences* 98**,** 7922-7927.

Tang, K., Fracasso, A., Struik, P.C., Yin, X., and Amaducci, S. (2018). Water- and Nitrogen-Use Efficiencies of Hemp (Cannabis sativa L.) Based on Whole-Canopy Measurements and Modeling. *Frontiers in Plant Science* 9.

Tang, K., Struik, P.C., Yin, X., Thouminot, C., Bjelková, M., Stramkale, V., and Amaducci, S. (2016). Comparing hemp (Cannabis sativa L.) cultivars for dual-purpose production under contrasting environments. *Industrial Crops and Products* 87**,** 33-44.

Techen, N., Chandra, S., Lata, H., Elsohly, M.A., and Khan, I.A. (2010). Genetic identification of female Cannabis sativa plants at early developmental stage. *Planta Med* 76**,** 1938-1939.

Thomas, S.G., Phillips, A.L., and Hedden, P. (1999). Molecular cloning and functional expression of gibberellin 2- oxidases, multifunctional enzymes involved in gibberellin deactivation. *Proceedings of the National Academy of Sciences of the United States of America* 96**,** 4698-4703.

Thygesen, A., Daniel, G., Lilholt, H., and Thomsen, A.B. (2006). Hemp Fiber Microstructure and Use of Fungal Defibration to Obtain Fibers for Composite Materials. *Journal of Natural Fibers* 2**,** 19-37.

Toonen, M.a.J., Maliepaard, C., Reijmers, T.H., Van Der Voet, H., Mastebroek, H.D., Van Den Broeck, H.C., Ebskamp, M.J.M., Kessler, W., and Kessler, R.W. (2004). Predicting the chemical composition of fibre and core fraction of hemp (Cannabis sativa L.). *Euphytica* 140**,** 39-45.

Torres, A.F., Noordam-Boot, C.M.M., Dolstra, O., Van Der Weijde, T., Combes, E., Dufour, P., Vlaswinkel, L., Visser, R.G.F., and Trindade, L.M. (2015a). Cell Wall Diversity in Forage Maize: Genetic Complexity and Bioenergy Potential. *BioEnergy Research* 8, 187-202.

Torres, A.F., Van Der Weijde, T., Dolstra, O., Visser, R.G.F., and Trindade, L.M. (2013). Effect of Maize Biomass Composition on the Optimization of Dilute-Acid Pretreatments and Enzymatic Saccharification. *BioEnergy Research* 6, 1038-1051.

Torres, A.F., Visser, R.G.F., and Trindade, L.M. (2015b). Bioethanol from maize cell walls: genes, molecular tools, and breeding prospects. *GCB Bioenergy* 7, 591-607.

Ueguchi-Tanaka, M., Ashikari, M., Nakajima, M., Itoh, H., Katoh, E., Kobayashi, M., Chow, T.-Y., Hsing, Y.-I.C., Kitano, H., Yamaguchi, I., and Matsuoka, M. (2005). GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. *Nature* 437, 693-698.

Underwood, W. (2012). The plant cell wall: a dynamic barrier against pathogen invasion. *Frontiers in plant science* 3, 85-85.

Updegraff, D.M. (1969). Semimicro determination of cellulose inbiological materials. *Analytical Biochemistry* 32, 420-424.

Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A., and Coupland, G. (2004). Photoreceptor Regulation of CONSTANS Protein in Photoperiodic Flowering. *Science* 303, 1003-1006.

Van Bakel, H., Stout, J.M., Cote, A.G., Tallon, C.M., Sharpe, A.G., Hughes, T.R., and Page, J.E. (2011). The draft genome and transcriptome of Cannabis sativa. *Genome Biology* 12, R102.

Van Den Broeck, H.C., Maliepaard, C., Ebskamp, M.J.M., Toonen, M.a.J., and Koops, A.J. (2008). Differential expression of genes involved in C1 metabolism and lignin biosynthesis in wooden core and bast tissues of fibre hemp (Cannabis sativa L.). *Plant Science* 174, 205-220.

Van Den Oever, M.J.A., Bas, N., Van Soest, L.J.M., Melis, C., and Van Dam, J.E.G. (2003). Improved method for fibre content and quality analysis and their application to flax genetic diversity investigations. *Industrial Crops and Products* 18, 231-243.

Van Der Weijde, T. (2016). *Targets and tools for optimizating lignocellulosic biomass quality of miscanthus.* Wageningen University.

Van Der Weijde, T., Dolstra, O., Visser, R.G.F., and Trindade, L.M. (2017). Stability of Cell Wall Composition and Saccharification Efficiency in Miscanthus across Diverse Environments. *Frontiers in Plant Science* 7.

Van Der Weijde, T., Torres, A.F., Dolstra, O., Dechesne, A., Visser, R.G.F., and Trindade, L.M. (2016). Impact of Different Lignin Fractions on Saccharification Efficiency in Diverse Species of the Bioenergy Crop Miscanthus. *BioEnergy Research* 9, 146-156.

Van Der Werf, H.M.G. (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, H.M.G., Harsveld Van Der Veen, J.E., Bouma, A.T.M., and Ten Cate, M. (1994). Quality of hemp (*Cannabis sativa* L.) stems as a raw material for paper. *Industrial Crops and Products* 2, 219-227.

Van Der Werf, H.M.G., M, M.E.W.J., and Haverkort, A.J. (1996). The potential of hemp (*Cannabis sativa* L.) for sustainable fibre production: a crop physiological appraisal. *Annals of Applied Biology* 129, 109-123.

Van Der Werf, H.M.G., and Turunen, L. (2008). The environmental impacts of the production of hemp and flax textile yarn. *Industrial Crops and Products* 27, 1-10.

Vander Mijnsbrugge, K., Meyermans, H., Van Montagu, M., Bauw, G., and Boerjan, W. (2000). Wood formation in poplar: identification, characterization, and seasonal variation of xylem proteins. *Planta* 210, 589-598.

Vannozzi, A., Dry, I.B., Fasoli, M., Zenoni, S., and Lucchin, M. (2012). Genome-wide analysis of the grapevine stilbene synthase multigenic family: genomic organization and expression profiles upon biotic and abiotic stresses. *BMC Plant Biol* 12, 130.

Vanous, A., Gardner, C., Blanco, M., Martin-Schwarze, A., Lipka, A.E., Flint-Garcia, S., Bohn, M., Edwards, J., and Lübberstedt, T. (2018). Association Mapping of Flowering and Height Traits in Germplasm Enhancement of Maize Doubled Haploid (GEM-DH) Lines. *The Plant Genome* 11.

Vanraden, P.M. (2008). Efficient methods to compute genomic predictions. *J Dairy Sci* 91, 4414-4423.

Vermerris, W., Saballos, A., Ejeta, G., Mosier, N.S., Ladisch, M.R., and Carpita, N.C. (2007). Molecular Breeding to Enhance Ethanol Production from Corn and Sorghum Stover. *Crop Science* 47**,** S-142-S-153.

Via, S., and Lande, R. (1985). Genotype-Environment Interaction and the Evolution of Phenotypic Plasticity. *Evolution* 39**,** 505-522.

Vogel, K.P., Dien, B.S., Jung, H.G., Casler, M.D., Masterson, S.D., and Mitchell, R.B. (2011). Quantifying Actual and Theoretical Ethanol Yields for Switchgrass Strains Using NIRS Analyses. *BioEnergy Research* 4**,** 96-110.

Volkmann, D., and Baluška, F. (2006). Gravity: one of the driving forces for evolution. *Protoplasma* 229**,** 143-148.

Voragen, A.G.J., Coenen, G.-J., Verhoef, R.P., and Schols, H.A. (2009). Pectin, a versatile polysaccharide present in plant cell walls. *Structural Chemistry* 20**,** 263.

Wallace, J.G., Larsson, S.J., and Buckler, E.S. (2014). Entering the second century of maize quantitative genetics. *Heredity (Edinb)* 112**,** 30-38.

Wambua, P., Ivens, J., and Verpoest, I. (2003). Natural fibres: can they replace glass in fibre reinforced plastics? *Composites Science and Technology* 63**,** 1259-1264.

Wang, B., Zhu, L., Zhao, B., Zhao, Y., Xie, Y., Zheng, Z., Li, Y., Sun, J., and Wang, H. (2019). Development of a Haploid-Inducer Mediated Genome Editing System for Accelerating Maize Breeding. *Molecular Plant* 12**,** 597-602.

Wang, S., Gusovius, H.-J., Lühr, C., Musio, S., Uhrlaub, B., Amaducci, S., and Müssig, J. (2018). Assessment system to characterise and compare different hemp varieties based on a developed lab-scaled decortication system. *Industrial Crops and Products* 117**,** 159-168.

Wang, S., Yin, Y., Ma, Q., Tang, X., Hao, D., and Xu, Y. (2012). Genome-scale identification of cell-wall related genes in Arabidopsis based on co-expression network analysis. *BMC plant biology* 12**,** 138-138.

Watanabe, S., Hideshima, R., Xia, Z., Tsubokura, Y., Sato, S., Nakamoto, Y., Yamanaka, N., Takahashi, R., Ishimoto, M., Anai, T., Tabata, S., and Harada, K. (2009). Map-Based Cloning of the Gene Associated With the Soybean Maturity Locus E3. *Genetics* 182**,** 1251-1262.

Watanabe, S., Xia, Z., Hideshima, R., Tsubokura, Y., Sato, S., Yamanaka, N., Takahashi, R., Anai, T., Tabata, S., Kitamura, K., and Harada, K. (2011). A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. *Genetics* 188**,** 395-407.

Wenden, B., Kozma-Bognár, L., Edwards, K.D., Hall, A.J.W., Locke, J.C.W., and Millar, A.J. (2011). Light inputs shape the Arabidopsis circadian system. *The Plant Journal* 66**,** 480-491.

Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A., and Coupland, G. (2006). CONSTANS and the CCAAT Box Binding Complex Share a Functionally Important Domain and Interact to Regulate Flowering of Arabidopsis. *The Plant Cell* 18**,** 2971-2984.

West, N.W., and Golenberg, E.M. (2018). Gender-specific expression of GIBBERELLIC ACID INSENSITIVE is critical for unisexual organ initiation in dioecious Spinacia oleracea. *New Phytol* 217**,** 1322-1334.

Westergaard, M., and Demerec, M. (1958). "The Mechanism of Sex Determination in Dioecious Flowering Plants," in *Advances in Genetics*. Academic Press), 217-281.

Westerhuis, W. (2016). *Hemp for textiles: plant size matters*. Includes bibliographic references. - With summaries in English and Dutch, Wageningen University.

Westerhuis, W., Amaducci, S., Struik, P.C., Zatta, A., Van Dam, J.E.G., and Stomph, T.J. (2009). Sowing density and harvest time affect fibre content in hemp (Cannabis sativa) through their effects on stem weight. *Annals of Applied Biology* 155**,** 225-244.

Wierzbicki, M.P., Maloney, V., Mizrachi, E., and Myburg, A.A. (2019). Xylan in the Middle: Understanding Xylan Biosynthesis and Its Metabolic Dependencies Toward Improving Wood Fiber for Industrial Processing. *Frontiers in Plant Science* 10.

Wigge, P.A. (2011). FT, a mobile developmental signal in plants. *Curr Biol* 21**,** R374-378.

Willats, W.G.T., Knox, J.P., and Mikkelsen, J.D. (2006). Pectin: new insights into an old polymer are starting to gel. *Trends in Food Science & Technology* 17**,** 97-104.

Willats, W.G.T., Mccartney, L., Mackie, W., and Knox, J.P. (2001). Pectin: cell biology and prospects for functional analysis. *Plant Molecular Biology* 47**,** 9-27.

Wong, A.C.S., Hecht, V.F.G., Picard, K., Diwadkar, P., Laurie, R.E., Wen, J., Mysore, K., Macknight, R.C., and Weller, J.L. (2014). Isolation and functional analysis of CONSTANS-LIKE genes suggests that a central role for CONSTANS in flowering time control is not evolutionarily conserved in Medicago truncatula. *Frontiers in Plant Science* 5.

Wright, S. (1969). *Evolution and the Genetics of Populations, vol. II. The Theory of Gene Frequencies.* . University of Chicago Press, Chicago.

Wu, G., Park, M.Y., Conway, S.R., Wang, J.W., Weigel, D., and Poethig, R.S. (2009). The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis. *Cell* 138, 750-759.

Xia, Z., Watanabe, S., Yamada, T., Tsubokura, Y., Nakashima, H., Zhai, H., Anai, T., Sato, S., Yamazaki, T., Lu, S., Wu, H., Tabata, S., and Harada, K. (2012). Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. *Proc Natl Acad Sci U S A* 109, E2155-2164.

Xiao, C., and Anderson, C.T. (2013). Roles of pectin in biomass yield and processing for biofuels. *Front Plant Sci* 4, 67.

Xiao, C., Somerville, C., and Anderson, C.T. (2014). POLYGALACTURONASE INVOLVED IN EXPANSION1 functions in cell elongation and flower development in Arabidopsis. *Plant Cell* 26, 1018-1035.

Xie, D., Dai, Z., Yang, Z., Sun, J., Zhao, D., Yang, X., Zhang, L., Tang, Q., and Su, J. (2018). Genome-Wide Association Study Identifying Candidate Genes Influencing Important Agronomic Traits of Flax (Linum usitatissimum L.) Using SLAF-seq. *Frontiers in plant science* 8, 2232-2232.

Xu, X., Paik, I., Zhu, L., and Huq, E. (2015). Illuminating Progress in Phytochrome-Mediated Light Signaling Pathways. *Trends in Plant Science* 20, 641-650.

Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., Zhou, H., Yu, S., Xu, C., Li, X., and Zhang, Q. (2008). Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. *Nature Genetics* 40, 761-767.

Yamamoto, K., Oda, Y., Haseda, A., Fujito, S., Mikami, T., and Onodera, Y. (2014). Molecular evidence that the genes for dioecism and monoecism in Spinacia oleracea L. are located at different loci in a chromosomal region. *Heredity (Edinb)* 112, 317-324.

Yang, G., Tang, H., Nie, Y., and Zhang, X. (2011). Responses of cotton growth, yield, and biomass to nitrogen split application ratio. *European Journal of Agronomy* 35, 164-170.

Yano, M., Kojima, S., Takahashi, Y., Lin, H., and Sasaki, T. (2001). Genetic Control of Flowering Time in Rice, a Short-Day Plant. *Plant Physiology* 127, 1425-1429.

You, F.M., Jia, G., Xiao, J., Duguid, S.D., Rashid, K.Y., Booker, H.M., and Cloutier, S. (2017). Genetic Variability of 27 Traits in a Core Collection of Flax (Linum usitatissimum L.). *Frontiers in Plant Science* 8.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., Mcmullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., and Buckler, E.S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38, 203-208.

Zabaniotou, A., and Kamaterou, P. (2019). Food waste valorization advocating Circular Bioeconomy - A critical review of potentialities and perspectives of spent coffee grounds biorefinery. *Journal of Cleaner Production* 211, 1553-1566.

Zavaliev, R., Levy, A., Gera, A., and Epel, B.L. (2013). Subcellular dynamics and role of Arabidopsis beta-1,3-glucanases in cell-to-cell movement of tobamoviruses. *Mol Plant Microbe Interact* 26, 1016-1030.

Zeng, X., Liu, H., Du, H., Wang, S., Yang, W., Chi, Y., Wang, J., Huang, F., and Yu, D. (2018). Soybean MADS-box gene GmAGL1 promotes flowering via the photoperiod pathway. *BMC Genomics* 19, 51.

Zhai, H., Lü, S., Wu, H., Zhang, Y., Zhang, X., Yang, J., Wang, Y., Yang, G., Qiu, H., Cui, T., and Xia, Z. (2015). Diurnal Expression Pattern, Allelic Variation, and Association Analysis Reveal Functional Features of the E1 Gene in Control of Photoperiodic Flowering in Soybean. *PLOS ONE* 10, e0135909.

Zhang, J., Zhou, X., Yan, W., Zhang, Z., Lu, L., Han, Z., Zhao, H., Liu, H., Song, P., Hu, Y., Shen, G., He, Q., Guo, S., Gao, G., Wang, G., and Xing, Y. (2015). Combinations of the Ghd7, Ghd8 and Hd1 genes largely define the ecogeographical adaptation and yield potential of cultivated rice. *New Phytologist* 208, 1056-1066.

Zhao, C., Takeshima, R., Zhu, J., Xu, M., Sato, M., Watanabe, S., Kanazawa, A., Liu, B., Kong, F., Yamada, T., and Abe, J. (2016). A recessive allele for delayed flowering at the soybean maturity locus E9 is a leaky allele of FT2a, a FLOWERING LOCUS T ortholog. *BMC Plant Biology* 16**,** 20.

Zhao, H., Kwak, J.H., Wang, Y., Franz, J.A., White, J.M., and Holladay, J.E. (2006). Effects of Crystallinity on Dilute Acid Hydrolysis of Cellulose by Cellulose Ball-Milling Study. *Energy & Fuels* 20**,** 807-811.

Zhao, Q., and Dixon, R.A. (2011). Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends Plant Sci* 16**,** 227-233.

Zhao, Q., Nakashima, J., Chen, F., Yin, Y., Fu, C., Yun, J., Shao, H., Wang, X., Wang, Z.-Y., and Dixon, R.A. (2013). *LACCASE* Is Necessary and Nonredundant with *PEROXIDASE* for Lignin Polymerization during Vascular Development in Arabidopsis. *The Plant Cell* 25**,** 3976-3987.

Zhao, T., Liu, H., Roeder, K., Lafferty, J., and Wasserman, L. (2012). The huge Package for High-dimensional Undirected Graph Estimation in R. *Journal of machine learning research : JMLR* 13, 1059-1062.

Zhong, R., and Ye, Z.H. (2007). Regulation of cell wall biosynthesis. *Curr Opin Plant Biol* 10, 564-572.

Zhong, R., and Ye, Z.H. (2012). MYB46 and MYB83 bind to the SMRE sites and directly activate a suite of transcription factors and secondary wall biosynthetic genes. *Plant Cell Physiol* 53, 368-380.

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44, 821-824.

Zhu, M., and Zhao, S. (2007). Candidate gene identification approach: progress and challenges. *International journal of biological sciences* 3, 420-427.

Zieher, C. (2010). "Biochemistry of the Fiber," in *Physiology of Cotton,* eds. J.M. Stewart, D.M. Oosterhuis, J.J. Heitholt & J.R. Mauney. (Dordrecht: Springer Netherlands), 361-378.

Summary

Hemp (*Cannabis sativa* L.), a high yielding bast fibre crop, is an environmentally-friendly crop that produces biomass of excellent quality for multiple purposes. The bio-composite applications of hemp fibres have received a lot of attention in the last decades by governments and industry. These bio-composites have the capacity to replace non-renewable and unsustainable resources (e.g. synthetic and mineral fibres). However, the molecular knowledge of fibre quality from hemp is limited. Consequently, breeding programmes to increase the yield and quality of hemp fibre have mostly focused on conventional methods, due to limited tools available for the development of marker-assisted selection (MAS) programmes for hemp fibre quality. To address these challenges, it is envisioned the development of molecular markers associated to fibre quality, which are expected to accelerate the development of new hemp cultivars with high quality fibres.

The biochemical composition of the cell walls is a key factor defining the quality of the fibre in hemp, as the fibres are fundamentally cell walls. Nevertheless, prior to this thesis there were no high-throughput methods available that enabled the chemical characterization of the cell wall in hemp. Consequently, the analysis of many accessions is expensive and time consuming, hampering the breeding for high fibre quality hemp cultivars. In **Chapter 2** of this thesis, we developed and optimized the throughput of five biochemical methods including cell wall extraction, biochemical composition of cell wall polysaccharides, quantification of lignin, quantifications of crystalline polysaccharides and morphology of the stems. The methods presented here revealed to be highly repeatable. In addition, the cell wall extraction was optimized to extract enough material for the complete characterization of the cell wall of hemp in a single run, while reducing the time of the entire analysis. Yet, wet biochemistry phenotyping methods are still costly and laborious. To overcome these limitations, high-throughput prediction models based on near-infrared spectroscopy (NIRS) were developed in **Chapter 3**. A range of prediction models for cell wall components were developed to phenotype large numbers of hemp accessions in a rapid and cost-efficient way.

Hemp fibre quality is a quantitative trait and little is known about the phenotypic variability, the heritability behaviour and the effect of the environment on fibre quality of hemp. The aim of **Chapter 3** is to gain deeper insights into the variability of fibre quality within a hemp panel of 123 accessions. The plants were cultivated in three locations with contrasting environments, corresponding to climates of northern, central and southern Europe. The accessions were phenotyped for 30 traits relevant to fibre quality of hemp. The study showed extensive phenotypic variability in all fibre quality traits in the hemp panel. In addition, the characterization of the phenotypic variation and heritability of the traits lead to the identification of fibre quality traits, that potentially lead to high genetic gains when used in breeding programmes. Relevant traits to maximize the genetic gains of hemp

fibre quality are most cell wall components (contents of monosaccharides derived from cellulose and hemicellulose; and lignin content), bast fibre content and flowering traits. Furthermore, all traits analysed showed significant *G*×E interactions, defined as heritable phenotypic fraction sensitive to environmental factors. Yet, these interactions were small in targets for hemp breeding with large heritabilities and they are not expected to interfere in breeding decisions. Moreover, all fibre quality traits showed an adaptive behaviour to environmental factors non associated to genetic components. Highly adaptive traits were contents of pectin-related monosaccharides, most agronomic traits and several fibre traits (fineness and decortication efficiency). Therefore, these results indicate that the influence of the environment on fibre quality should be considered while designing a breeding programme for the fibre quality of hemp.

The development of updated breeding programmes for hemp fibre quality requires the identification of quantitative trait loci (QTLs) for the various traits that are relevant to the quality of bast fibre in hemp. To this end, the hemp accession panel was genotyped using Restriction site Associated DNA Sequencing (RAD-seq). A large set (>600.000) of genome-wide SNP markers was selected to further investigate the genetics of fibre quality. The genotypes and the phenotypes were used to perform Genome-Wide Association Studies (GWAS) to get insights into the genetic architecture of highly heritable cell wall components and bast fibre content. Yet, the incomplete sequenced genome of hemp hindered the mapping of the genetic components that were identified via GWAS. To go beyond these bounds, we first developed a methodology to identify independent QTLs (**Chapter 4**). The method used the collinearity between allelic frequencies of markers significantly associated to the traits. The analyses identified 16 QTLs for fibre quality traits across two or three locations with different environments. These QTLs are excellent genetic tools for the development of novel fibre cultivars of hemp that perform well in different environments. Furthermore, putative candidate genes for the biosynthesis and modification of monosaccharides, polysaccharides and lignin were detected in direct vicinity of the QTLs. This study is the first GWAS in hemp, and we have reported the first QTLs associated to fibre quality of hemp. Thus, this study is a significant step towards the development of a marker-assisted selection (MAS) for hemp fibre quality.

Flowering time strongly underlies a shift in the carbon partitioning of hemp from producing biomass to produce flowers and seeds. This shift influences the quality of bast fibre. In addition, sex determination is known to influence fibre quality as hemp is characterized by protandry and male plants die after flowering, while female plants remain alive until seed maturity. Meanwhile, monoecious plants are characterized by a uniform flowering and fibre production. It is well-known that genetic factors play a role in the outcome of flowering time and sex determination but the major molecular mechanisms controlling these traits are not characterized in hemp. To gain deeper insights into this knowledge

gap, we reviewed the state-of-art of flowering time, sex determination and fibre quality in short-day plants and dioecious crops, respectively, and compare them with hemp (**Chapter 5**). In this review we proposed possible mechanisms for the genetic control of these complex traits in hemp.

In **Chapter 6**, we attempt to close the gap in molecular knowledge on flowering time and sex determination, using a GWAS approach. The study revealed six QTLs across two or three locations for flowering and two for sex traits. In addition, putative candidate genes for these traits were identified in the vicinity of the QTLs. Candidate genes involved in the photoperiod, vernalization and endogenous flowering pathways were identified in the QTLs for flowering traits. Moreover, candidate genes involved in the balance of phytohormones, such as *gibberellic acid insensitive* gene (DELLA proteins) and *auxin response factors* genes (ARFs) were identified in a QTL for sex determination. Sex QTLs were associated to the stability of sex determination in monoecious plants and the candidate genes support this association. This chapter elucidates relevant molecular mechanisms of flowering and sex traits in hemp, providing additional tools to further develop MAS programmes for hemp.

Finally, the knowledge generated within the framework of this thesis was integrated in the light of breeding for hemp to increase the production of high quality and sustainable hemp fibres. We provided high-throughput methods to phenotype biomass quality traits, specifically adapted to hemp. We provided deeper insights into the variability of fibre quality of hemp and its relation to the environment. Complementary practices alternative to breeding for poorly heritable traits were discussed. In addition, we identified QTLs and molecular pathways of fibre quality relevant traits. Overall, the collection of the new insights from this thesis can be used to improve hemp fibre quality and expand the use of hemp as a feasible alternative to non-renewable (e.g. mineral and fossil) and unsustainable materials (e.g. cotton).

Resum

El cànem (*Cannabis sativa* L.), un cultiu de fibra de líber molt rendible, és un cultiu ecològic que produeix biomassa d'excel·lent qualitat per a múltiples propòsits. Les aplicacions de bio-compostos de fibres de cànem han rebut molta atenció en les darreres dècades per part dels governs i de la indústria. Ja que aquests bio-compostos tenen la capitat de substituir recursos no renovables i insostenibles, com ara fibres sintètiques i minerals. Tanmateix, el coneixement molecular de la qualitat de la fibra de cànem és limitat. En conseqüència, els programes de millora genètica per augmentar-ne el rendiment i la qualitat s'han centrat majoritàriament en mètodes convencionals, degut a la manca d'eines moleculars disponibles per al desenvolupament de programes de selecció assistida per marcadors (SAM) associats a la qualitat de la fibra. El desenvolupament d'aquests marcadors s'espera que pugui accelerar la millora de noves varietats de cànem d'alta qualitat.

La composició bioquímica de les parets cel·lulars és un factor clau per definir la qualitat de la fibra del cànem, ja que les fibres són fonamentalment parets cel·lulars. No obstant això, a l'inici d'aquesta tesi no hi havia mètodes d'alt rendiment disponibles que permetessin la caracterització bioquímica de la paret cel·lular del cànem. Per tant, l'anàlisi de moltes varietats era costosa i requeria molt de temps, dificultant així la millora genètica de la fibra de cànem. Al **capítol 2** d'aquesta tesi vam desenvolupar i optimitzar el rendiment de cinc mètodes bioquímics incloent l'extracció de la paret cel·lular, composició bioquímica de polisacàrids de paret cel·lular, quantificació de lignina, quantificacions de polisacàrids cristal·lins i morfologia de les tiges. Els mètodes presentats aquí van revelar ser molt repetibles. A més, el mètode d'extracció de paret cel·lular es va optimitzar per produir suficient material per a la caracterització completa de la paret cel·lular del cànem en una sola extracció, alhora que es va reduir el temps de l'anàlisi complet. No obstant això, els mètodes de fenotipat bioquímics encara són costosos i laboriosos. Per superar aquest obstacle, al **capítol 3** vam desenvolupar models de predicció d'alt rendiment basats en espectroscòpia d'infrarojos propers (NIRS).

La qualitat de la fibra de cànem és un caràcter complex i es coneix poc sobre la seva diversitat fenotípica, el seu comportament hereditari i l'efecte del medi ambient sobre la qualitat de la seva fibra. L'objectiu del **capítol 3** és aprofundir en aquests temes en un panell de 123 varietat de cànem. Les plantes es van conrear en tres localitats amb ambients contrastants, corresponents a climes del nord, centre i sud d'Europa. Les varietats es van fenotipar per 30 caràcters rellevants per la qualitat de la fibra. L'estudi va mostrar una àmplia diversitat fenotípica en tots els caràcters. A més, aquest estudi va permetre la identificació de caràcters que poden comportar grans guanys genètics en programes de millora de la qualitat de la fibra de cànem. Aquests caràcters són la majoria dels components de la paret cel·lular (contingut de monosacàrids que composen cel·lulosa i hemicel·lulosa i contingut de lignina), contingut de fibra de líber i caràcters de floració. A més, tots els caràcters analitzats van mostrar interaccions significatives entre genotips i medi ambient (interaccions *GxE*),

definits com la fracció fenotípica heretable i sensible als factors mediambientals. Tot i això, aquestes interaccions eren petites en caràcters amb components genètics alts i no es preveu que interfereixin en les decisions de millora. A més, tots els caràcters de qualitat de la fibra van mostrar un comportament adaptatiu als factors mediambientals no associats als components genètics. Els caràcters altament adaptatius van ser el contingut de monosacàrids relacionats amb la pectina, la majoria de caràcters agronòmics i diversos de fibra (finor i eficiència de la decorticació). Per tant, aquests resultats indiquen que cal tenir en compte la influència del medi ambient en la qualitat de la fibra quan es dissenya un programa de millora per aquests caràcters en el cànem.

La modernització dels programes de millora de la qualitat de la fibra de cànem requereix la identificació de loci de caràcters quantitatius (QTLs) associats a caràcters importants per la qualitat. Amb aquest objectiu, el panell de cànem es va genotipar utilitzant la seqüenciació d'ADN associada a dianes de restricció (RAD-Seq). Un gran grup (>600.000) de marcadors moleculars (SNPs), que cobreix tot el genoma, es van identificar amb la seqüenciació. Els genotips i fenotips es van utilitzar per realitzar estudis d'associació de tot el genoma (GWAS) per investigar l'arquitectura genètica dels components de la paret cel·lular i del contingut de fibra de líber altament heretables. Una sèrie de marcadors van ser associats significativament amb aquests caràcters. No obstant això, la seqüència incompleta del genoma del cànem va dificultar el posicionament d'aquests marcadors en el genoma. Per superar aquestes limitacions, vam desenvolupar una metodologia per identificar QTLs independents (**Capítol 4**). Aquest mètode utilitza la col·linealitat entre les freqüències al·lèliques dels marcadors associades significativament als caràcters. Les anàlisis va identificar 16 QTLs per caràcters de qualitat de fibra en dues o tres localitzacions amb condicions mediambientals diferents. Aquests QTLs són excel·lents eines genètiques per al desenvolupament de noves varietats de fibra de cànem que tinguin un bon rendiment en diferents ambients. A més, es van detectar gens candidats per a la biosíntesi i modificació de monosacàrids, polisacàrids i lignina a les seqüències contigües dels QTLs. Aquest estudi és el primer GWAS realitzat en el cànem i hem identificat els primers QTLs associats a la qualitat de la fibra d'aquest cultiu. Aquest capítol és un pas significatiu cap al desenvolupament d'una selecció assistida per marcadors per a la qualitat de la fibra de cànem.

En el cànem, el temps de floració determina el canvi de compartimentació del carboni de la producció de biomassa a la producció de flors i llavors. Aquest canvi influeix en la qualitat de la fibra de líber. A més, se sap que la determinació del sexe influeix en la qualitat de la fibra ja que el cànem dioic (flors masculines i femenines en plantes diferents) es caracteritza per protàndria (les plantes masculines floreixen abans que les plantes femenines). Les plantes masculines moren després de la floració mentre que les plantes femenines romanen vives fins a la maduresa de les llavors. Mentrestant, les plantes de cànem monoiques (flors masculines i femenines en la mateix planta) es caracteritzen per una floració i producció de

fibra uniformes. És ben sabut que els factors genètics tenen un paper clau en el temps de floració i la determinació sexual del cànem, però els mecanismes moleculars que controlen aquests caràcters romanen desconeguts. Per obtenir més informació sobre aquest tema, vam revisar el coneixement disponible sobre el temps de floració, la determinació sexual i la qualitat de la fibra en altres cultius de dia curt i dioics i les vam comparar amb el cànem (**Capítol 5**). En aquesta revisió vam proposar possibles mecanismes per al control genètic d'aquests caràcters complexos en el cànem.

Al **Capítol 6**, intentem completar el buit en el coneixement molecular sobre el temps de floració i la determinació sexual mitjançant estudis d'associació GWAS. Aquests estudis van revelar sis QTLs en dues o tres ubicacions per a la floració i dos per caràcters sexuals. A més, vam identificar gens candidats per a aquests caràcters en les regions contigües dels QTLs. Gens implicats en el fotoperíode, la vernalització i les vies florals endògenes es van identificar en els QTLs per als caràcters de floració. D'altra banda, es van identificar gens implicats en l'equilibri de les fitohormones, com les proteïnes DELLA (àcid giberélic) i els *factors de resposta de les auxines* (auxines) en un QTL per a la determinació sexual. QTLs identificats per a caràcters sexuals estaven associats a l'estabilitat de la determinació del sexe en plantes monoiques i els gens candidats donen suport a aquesta associació. Aquest capítol explica importants mecanismes moleculars per la floració i la determinació sexual del cànem. A més, proporciona eines addicionals per al desenvolupament de programes de millora (SAM) per al cànem.

Finalment, el coneixement generat en el marc d'aquesta tesi va ser integrat per reforçar la millora genètica del cànem per tal d'incrementar la producció de fibres de cànem d'alta qualitat i sostenibles. Vam proporcionar mètodes d'alt rendiment per fenotipar caràcters de qualitat de biomassa, adaptats específicament al cànem. Vam proporcionar una visió més profunda sobre la diversitat de qualitat de fibra del cànem i la seva relació amb el medi ambient. Vam debatre pràctiques complementàries i alternatives a la millora genètica per a caràcters poc heretables. A més, vam identificar QTLs i rutes moleculars per a caràcters rellevant per a la qualitat de la fibra. En conjunt, el nou coneixement generat en aquesta tesi pot ser utilitzat per millorar la qualitat de la fibra del cànem i ampliar l'ús del cànem com a alternativa factible a materials no renovables (per exemple, d'origen mineral i fòssil) i materials insostenibles (per exemple, el cotó).

Resumen

El cáñamo (Cannabis sativa L.), un cultivo de fibra de líber muy rentable, es un cultivo ecológico que produce biomasa de excelente calidad para múltiples propósitos. Las aplicaciones de bio-compuestos de fibras de cáñamo han recibido mucha atención en las últimas décadas por parte de los gobiernos y de la industria. Estos bio-compuestos, tienen la capacidad de sustituir recursos no renovables e insostenibles como fibras sintéticas y minerales. Sin embargo, el conocimiento molecular sobre la calidad de la fibra de cáñamo es limitado. En consecuencia, los programas de mejora genética para aumentar el rendimiento y la calidad se han centrado mayoritariamente en métodos convencionales, debido a la falta de herramientas moleculares disponibles para el desarrollo de programas de selección asistida por marcadores (SAM) asociados a la calidad de la fibra. Se espera que el desarrollo de estos marcadores pueda acelerar la mejora de nuevas variedades de cáñamo de alta calidad.

La composición bioquímica de las paredes celulares es un factor clave para definir la calidad de la fibra del cáñamo, ya que las fibras son fundamentalmente paredes celulares. Sin embargo, al inicio de esta tesis, no existían métodos de alto rendimiento que permitieran la caracterización bioquímica de la pared celular del cáñamo. Por lo tanto, el análisis de muchas variedades era costoso y requería mucho tiempo, dificultando así la mejora genética del cáñamo enfocada hacia la mejora de la calidad de la fibra. En el capítulo 2 de esta tesis desarrollamos y optimizamos el rendimiento de cinco métodos bioquímicos, incluyendo la extracción de la pared celular, composición bioquímica de polisacáridos de pared celular, cuantificación de lignina, cuantificaciones de polisacáridos cristalinos y morfología de los tallos. Los métodos presentados aquí revelaron ser muy reproducibles. Además, el método de extracción de pared celular se optimizó para producir suficiente material para la caracterización completa de la pared celular del cáñamo en una sola extracción, al tiempo que se redujo el tiempo del análisis completo. Sin embargo, los métodos bioquímicos de fenotipado todavía son costosos y laboriosos. Para superar este obstáculo, en el capítulo 3 desarrollamos modelos de predicción de alto rendimiento basados en espectroscopía de infrarrojos cercanos (NIRS).

La calidad de la fibra de cáñamo es un rasgo complejo y poco se conoce sobre su diversidad fenotípica, su comportamiento hereditario, y el efecto del medio ambiente sobre la calidad de su fibra. El objetivo del capítulo 3 es profundizar en la diversidad de la calidad de la fibra de un panel de 123 variedades de cáñamo. Las plantas se cultivaron en tres localidades con ambientes contrastantes, correspondientes a climas del norte, centro y sur de Europa. Las variedades se fenotiparon por 30 rasgos relevantes para la calidad de la fibra. El estudio mostró una amplia diversidad fenotípica en todos los rasgos. Además, la caracterización de la variación fenotípica y la herencia de los rasgos permitieron la identificación de rasgos que pueden conllevar grandes beneficios genéticos en programas de mejora genética. Estos

rasgos son la mayoría de los componentes de la pared celular (contenido de monosacáridos que componen celulosa y hemicelulosa, y contenido de lignina), contenido de fibra de líber y rasgos de floración. Además, todos los rasgos analizados mostraron interacciones significativas entre genotipos y medio ambiente (interacciones GxE), definidos como la fracción fenotípica heredable y sensible a los factores medioambientales. Sin embargo, estas interacciones eran pequeñas en rasgos con altos componentes genéticos y no se prevé que interfieran en las decisiones de mejora. Además, todos los rasgos de calidad de la fibra mostraron un comportamiento adaptativo a los factores medioambientales no asociados a los componentes genéticos. Los rasgos altamente adaptativos fueron el contenido de monosacáridos relacionados con la pectina, la mayoría de rasgos agronómicos y varios rasgos de fibra (finura y eficiencia de la decorticación). Por lo tanto, estos resultados indican que hay que tener en cuenta la influencia del medio ambiente en la calidad de la fibra cuando se diseña un programa de mejora para estos rasgos en el cáñamo.

La modernización de los programas de mejora para la calidad de la fibra de cáñamo requiere la identificación de loci de rasgos cuantitativos (QTLs) asociados a rasgos importantes para la calidad. Con este objetivo, el panel de cáñamo se genotipó utilizando la secuenciación de ADN asociada a dianas de restricción (RAD-Seq). Un gran grupo (> 600.000) de marcadores moleculares (SNPs), que cubre todo el genoma, se identificaron con la secuenciación. Los genotipos y fenotipos se utilizaron para realizar estudios de asociación de todo el genoma (GWAS) para investigar la arquitectura genética de los componentes de la pared celular y del contenido de fibra de líber, altamente heredables. Una serie de marcadores fueron asociados significativamente con estos rasgos. Sin embargo, la secuencia incompleta del genoma del cáñamo dificultó el posicionamiento de estos marcadores en el genoma. Para superar estas limitaciones, desarrollamos una metodología para identificar QTLs independientes (Capítulo 4). Este método utiliza la colinearidad entre las frecuencias alélicas de los marcadores asociadas significativamente a los rasgos. Los análisis identificaron 16 QTLs para rasgos de calidad de fibra en dos o tres lugares con condiciones medioambientales diferentes. Estos QTLs son excelentes herramientas genéticas para el desarrollo de nuevas variedades de fibra de cáñamo que tengan un buen rendimiento en diferentes ambientes. Además, se detectaron genes candidatos para la biosíntesis y modificación de monosacáridos, polisacáridos y lignina en las secuencias contiguas de los QTLs. Este estudio es el primer GWAS realizado en cáñamo y hemos identificado los primeros QTLs asociados a la calidad de la fibra de este cultivo. Este estudio es un paso significativo hacia el desarrollo de una selección asistida por marcadores para la calidad de la fibra de cáñamo.

En el cáñamo, el tiempo de floración determina el cambio de compartimentación del carbono de la producción de biomasa a la producción de flores y semillas. Este cambio influye en la calidad de la fibra de líber. Además, se sabe que la determinación del sexo influye en la calidad de la fibra, ya que el cáñamo dioico (flores masculinas y femeninas en

plantas diferentes) se caracteriza por protandria (las plantas machos florecen antes que las plantas hembras). Las plantas masculinas mueren tras la floración, mientras que las plantas hembras permanecen vivas hasta la madurez de las semillas. Mientras tanto, las plantas de cáñamo monoicas (flores masculinas y femeninas en la misma planta) se caracterizan por una floración y producción de fibra uniformes. Es bien sabido que los factores genéticos tienen un papel clave en el tiempo de floración y la determinación sexual del cáñamo, pero los mecanismos moleculares que controlan estos rasgos siguen siendo desconocidos. Para obtener más información sobre este tema, revisamos el conocimiento disponible sobre el tiempo de floración, la determinación sexual y la calidad de la fibra en otros cultivos de día corto y dioicos y las comparamos con el cáñamo (Capítulo 5). En esta revisión propusimos posibles mecanismos para el control genético de estos rasgos complejos en el cáñamo.

En el Capítulo 6, intentamos completar el vacío en el conocimiento molecular sobre el tiempo de floración y la determinación sexual mediante estudios de asociación GWAS. El estudio reveló seis QTLs en dos o tres ubicaciones para la floración y dos por rasgos sexuales. Además, identificamos genes candidatos para estos rasgos en las regiones contiguas de los QTLs. Genes implicados en el fotoperiodo, la vernalización y las vías florales endógenas se identificaron en los QTLs para los rasgos de floración. Por otra parte, se identificaron genes implicados en el equilibrio de las fitohormonas, como las proteínas DELLA (ácido Giberélico) y los factores de respuesta de las auxinas (auxinas) en un QTL para la determinación sexual. QTLs identificados para rasgos sexuales estaban asociados a la estabilidad de la determinación del sexo en plantas monoicas y los genes candidatos apoyan esta asociación. Este capítulo explica importantes mecanismos moleculares para la floración y la determinación sexual del cáñamo. Además, proporciona herramientas adicionales para el desarrollo de programas de mejora (SAM) para el cáñamo.

Finalmente, el conocimiento generado en el marco de esta tesis fue integrado para reforzar la mejora genética del cáñamo para incrementar la producción de fibras de cáñamo de alta calidad y sostenibles. Proporcionamos métodos de alto rendimiento para fenotipar rasgos de calidad de biomasa, adaptados específicamente al cáñamo. Proporcionamos una visión más profunda sobre la diversidad de calidad de fibra del cáñamo y su relación con el medio ambiente. Debatimos prácticas complementarias y alternativas a la mejora genética para rasgos poco heredables. Además, identificamos QTLs y rutas moleculares para rasgos relevantes para la calidad de la fibra. En conjunto, el nuevo conocimiento generado en esta tesis puede ser utilizado para mejorar la calidad de la fibra del cáñamo y ampliar el uso del cáñamo como alternativa factible a materiales no renovables (por ejemplo, de origen mineral y fósil) y materiales no sostenibles (por ejemplo, el algodón).

Acknowledgements

After having written all what I know about hemp, I could not think about another way to finish this five years of challenges, excitement, experiences, emotions, frustrations, sometimes panic, but specially professional and personal growth. It's time to find the right words of probably the most read section of a PhD thesis. I am aware of the responsibility to express my gratitude to all the countless people involved in this work or in my life during this unforgettable journey. It's not an easy task given that the list is long, but I hope not to forget anyone.

First and foremost I would like to thank my promotor and supervisor, **Luisa Trindade**. I cannot thank you enough for giving me the opportunity to do this research project. I still remember our first interview meeting, where after a discussion about biochemistry, genetics, bioinformatics and bio-based economy, you proposed me to work on the genetics of hemp and my innocent mind had to ask you: "Sorry Luisa but, what exactly is hemp? Isn't that… marijuana? Is that legal?". Dear Luisa, thank you immensely for giving me the chance to discover what science is and how to become an independent scientist. Thank you for stimulating me to grow. Thank you also for all the time of proofreading and correcting my manuscripts in your busy life, even during your business trips. Also, thank you for having always the door opened for me, not only for professional issues but also for personal ones and for your willingness to make time to discuss whenever needed. I appreciate your positive encouragement, your guidance and your kindness throughout this challenging journey. Muito obrigado!

Second of all, I would like to express my gratitude undoubtedly to **Elma Salentijn**, my co-promotor and daily supervisor. I highly appreciated your endless willingness to help me anytime I needed it, whether I was stuck in writing, in analyzing the data or just trying to understand our huge Multihemp database. Dear Elma, thank you very much for being always there, cheering me up in every single moment. I also enjoyed and learnt a lot with you and Luisa during our travels to attend conferences and project meetings. Bedankt voor alles!

Furthermore, I am truly grateful to **Maria João Paulo** and **Robert van Loo,** my statistical mentors. I have to admit that at the beginning I had troubles to follow you, all these models, hypotheses to prove, etc. You inspired me, stimulated my curiosity and taught me a lot. You made me enjoy statistics and especially to understand them! Thank you for our countless discussions, and not only about statistics, I also loved our discussions about politics with you. Your constant curiosity about Catalunya made me feel not that far away from home.

Of course the content of this thesis cannot be attributed just to me or to my supervisors. A big contribution to the work presented here has been the work of five **MSC students**, who performed experiments under my supervision. **Wen Fang**, **Dominik Ewald**, **Wei Xiong**, **Li Gu**, and **Agata Gulisano** I hope you enjoyed and learned working with hemp and me, as much as I did with you.

I would like to give my special thanks to all members of the **Biobased Economy Group**. I learned a lot from all of you. Thank you also for all the fun moments we have lived together in the BBE activities, Tuesdays lunch meetings and dinners! I have to admit that basically in about six years in the Netherlands, only with you I was having Dutch typical lunch (but with my tinny olive oil bottle, of course). I would also like to thank all the people from the other groups of **Plant Breeding** who helped me in one way or another one throughout these years. Becoming a Doctor means to become an independent scientist, but without collaboration, help and support from our colleagues, this wouldn't be possible. My sincere appreciation is extended to **Annemarie Dechesne**. You have been my biochemical mentor. Thank you for your support, for teaching me, for trusting me and for providing me self-confidence to work independently in the biochemical lab. I also want to thank **Christel Denneboom** for all the help in the molecular lab. I know that the extraction of DNA from the hemp panel became sort of a nightmare for you. Without your help, the sequencing and the GWAS wouldn't have been possible. **Marcel Visser**, thank you for all your help and discussions in the biochemical lab. At the beginning, I truly believed you belonged to the BBE group. For me, you were part of it. **Heleen Furrer**, thank you for fighting constantly against me to book the scales. You always won! I highly appreciate your willingness to help and your endless smile. Let me tell you a secret, I also didn't like the radio station.

Bezhad, my Iranian (now also Dutch) bro! You have been next to my desk and in the lab from the very beginning and supporting me until the very end. At the beginning you were the exact colleague I needed to start learning about cell walls and biochemistry from scratch. You had no problem to help every single time I said: Bezhaaaaaad, I have another question… I don't understand this… I don't know how to do that… You never refused to help me. Professionally talking, thank you for the countless and deep discussions we have had and I hope to keep on having them. Personally talking, there is not enough ink to describe my gratitude to you. You have been a key person throughout my stage in the Netherlands. I know our friendship will be a lasting one. **Agata,** my partner-in-crime, a thousand words wouldn't be enough to thank you. Remember, you can get married with however you want (hopefully with Valerio, please), but you are married with me in science forever, Petit et al. (2019a)!!! I found something special on you, since the very beginning in our informal interview in Droef95 with a beer. You always have claimed to be my favorite student, then you became a great colleague but you have always been an essential friend. I already miss you. **Francesco**, I have the feeling that we have always been working together. Hours and hours of discussions, you coming to my office showing off about some results or figures in R, or me coming to you about some new hypotheses or ideas. I know you are and will be an excellent scientist and I'll be proud to say: "he is my friend". But! Remember, I am a better swimmer than you! *Also good mentioning that because of us, Kiripiri went bankrupt. **Vivi**, durante todos estos años has sido la persona que ha sabido calmarme cuando hacía falta pero permitir mi locura cuando era necesario. Tus aganas de mantenernos juntos haciendo cosas fuera del trabajo

han sido clave para crear una amistad de verdad. Siempre te he considerado part del BBE group. Mil gracias por todo y espero algún día poder visitar Ecuador y las Galápagos con vosotros. A special thanks also goes to **Xuan Xu**, my "older sistaaaar". I still don't know how did you survive writing you thesis with a young hyperactive PhD student in the same office. You left quite some time ago already, but I still keep you in my mind as a nice colleague but especially as a good friend. You're always more than welcome to Catalunya or wherever I will be. **Andrés**, I think I haven't met a smarter person than you in all senses. Our discussions about genetics, about sustainability, basically about everything have always been really inspiring. I think you also have enjoyed a lot having a friend that, besides a colleague, was also a hippie soul living surrounded by chickens. You once told me: "You are doing a great job. It is taking a bit longer than expected, but you are enjoying your life in between." I couldn't agree more. **Michiel Klaassen**, I enjoyed a lot going to California with you, Andrés, Luisa and Xuan. I especially enjoyed our days in San Francisco. I think that the conference trip was nice to know each other a bit better. I wish you all he best in your professional and personal life. **Kasper**, you are the new me! A bit older and with less hair but you remind me when I came for the first time in the biochemical lab. You are a cool person, keep the standards high! **Mohamed**, we didn't spend much time together, but your jokes and sense of humor made me always draw a smile on my face during my hard times.

Thanks also to my (former)officemates **Agata Gulisano**, **Aurélie Jouanin**, **Bezhad Rashidi**, **Eshan Motazedi**, **Izan Sharuil**, **Kasper van der Cruijsen**, **Kim Magnée**, **Mathilde Daniau**, **Mohamed al Hassan**, **Tim van der Weijde**, **Viviana Jaramillo**, **Ying Liu** and **Xuan Xu**, I hope you enjoyed me as a colleague and officemate as much as I did with you! Sorry if sometimes I have annoyed you. My hyperactivity, stress and bipolarity were guilty. I would also like to thank my fellow (former)PhD colleagues I had the chance to meet throughout all these years. The list is long to mention everyone by name, so apologize me if you came looking for your name and it doesn't appear listed. I would particularly like to mention **Alejandro Therese**, **Anne Giesbergs**, **Carolina Aguilera**, **Charlotte Prodhomme**, **Daniel Monino**, **Eleni Koseoglou**, **Rafael Chan**, **Jarst van Belle**, **Mas Muniroh**, **Marine Paupiere**, **Manos Domazakis**, **Michela Appiano**, **Miguel Santillán**, **Myluska Caro**, **Pauline van Haperen**, **Peter Bourke** and **Peter Dinh**, thank you all for making my time in the department so enjoyable!

Marcela, mi mentora, mi inspiradora, mi impulsora. Tú pusiste la semilla de hacer un Doctorado en mi cerebro. Lo creías tan fuertemente que al final conseguí uno! Nunca podré agradecerte suficiente todo lo que hiciste por mí. Des del principio fuiste como una hermana, me has cuidado siempre, en lo bueno y en lo malo, incluso cuando mi familia vino a verme, nos hiciste un tour por la Uni. Y aún ahora siguen preguntándome por ti, Rafa y la pequeña Elena. Recuerda que siempre tendréis una casa viva donde viva. Te quiero mucho. Pura vida!

A special thanks is also deserved to the people of **Unifarm**, particularly **Wim van der Slikke**, **Gerrit Huisman** and **Herman Meurs**. The biochemical, molecular and genetic analyses wouldn't have been possible without your constant help during the sample preparation. Ontzettend bedankt voor alles!

Of course, these acknowledgements would be uncomplete without the special mention of the secretariat from Plant Breeding. **Nicole Trefflich**, **Letty Dijker**, **Danielle van der Wee** and **Janneke van Deursen**, my life in the Netherlands has been easier thanks of you. Thank you for your constant willingness to answer every single big or insignificant question from this little annoying Catalan. Thank you for being such a friendly and valuable team!

To my fellow members of the EPS PhD council: **Amalia**, **Francesca**, **Jarst**, **Jesse**, **Kiki**, **Kim**, **Malaika**, **Manos**, **Martina**, **Sandeep**, **Sara**, **Shanice**, **Peter**, **Tieme** and **Tom**, working together was a nice experience and our dinner meetings (sometimes chaotic when held in Orion) were pretty funny. I think we did great things together, specially organizing the Get2gethers! I also want to thank the **committee members** of my thesis, mostly known as The Opponents, for accepting to read this thesis and taking the time to evaluate it.

To all the people in **Multihemp** that I had the chance to work with, thank you for the great cooperation, the discussions and the fruitful project meetings. A special thanks to **Stefano Amaducci**, **Jörg Müssig**, **Jörg-Hans Gusovius**, **Kailei Tang**, **Giampaulo Grassi**, **Claire Thouminot**, **Carlos Fuentes**, **Aart W. van Vuure**, **Simon Mcqueen**, **Alexandra Lanot**, **Stephan Piotrowski** and **Michael Carus**. I feel privileged to have been part of this panel of hemp experts!

I would also like to thank all European taxpayers. Without the taxes paid by every European, public education wouldn´t be possible and many people, including me, couldn´t have gone to the University. In addition, it is thanks to these taxes that this research was financially supported.

As I said before, nobody gets through a PhD without the support from the colleagues, but equally important is all the people who contribute to keep safe the mental health of the candidates. I was lucky to find a lovely community named **Droevendaal** to live in, where people are always willing to share everything they have: experiences, energy, knowledge, time, food, love, …, life. Thank you my dear community for these beautiful six years abroad. You made these years pass in the blink of an eye and I kept on saying every single day: "Another day in the paradise" because I believed it! I truly believe that I have lived probably one of the best periods of my life with you. Among this people, I find the dynamic family I chose in the last years, never the same but always the one I needed: **Adèle**, **Adrià**, **Adrián**, **Agata**, **Ainhoa**, **Alberto**, **Alexandre**, **Angel**, **Antonia**, **Berta**, **Clotti**, **Carmen Díez**, **Carmen Saez**, **Charlotte**, **Cristina**, **David**, **Dimitri**, **Eliot**, **Elvira**, **Emilie**, **Emilio**, **Fiona**, **Franco**, **George**, **Giorgos Valsamakis**, **Giulio**, **Greta**, **Hestia**, **Íñigo**, **Isabel**, **Jan**, **Jokin**, **Juan**, **Judit**,

Júlia, **Kasia**, **Laura's**, **Leire**, **Lili**, **Lindy**, **Lucía**, **Maaike**, **Mara**, **Marco**, **Marghe**, **María Arredondo**, **Maria Aurell**, **Maria Borràs**, **Mariana**, **Marko**, **Marta Roca**, **Martí**, **Matika**, **Maura**, **Max**, **Milos**, **Natalia**, **Nicolò**, **Óscar**, **Pau's**, **Paula**, **Philip**, **Popi**, **Ramon**, **Raquel**, **Rob**, **Rodri, Romanos**, **Sanne**, **Serena** (and **Hakuna**), **Shan**, **Silvi**, **Silvia Quarta**, **Sofia**, **Taco**, **Teo**, **Thomas**, **Tobia**, **Tim**, **Valerio**, **Vanesa**, **Vera**, **Vero**, **Vicky**, **Víctor**, **Victòria**, **Xabi** (Partner-in-crime in ¨gusanería¨), **Yannis** and the people I'm forgetting, thank you for accepting me in your lives! I would also like to give my special thanks to **Ibex**, the climbing association I belonged to, especially to its board and to my climbing instructors for introducing me to this beautiful sport: **Inge**, **Ingi Alofs**, **Jesse**, **Marion**, **Nick**, **Robin**, **Stan** and **Tinko.** Likewise, I feel the need to thanks all the staff from **De Bongerd swimming pool** because without swimming and climbing I doubt I would have achieved this. Sport has been my therapy.

Ramon, ens vam conéixer a una conferència a Orion fa uns anys, després vam anar a viure junts i des d'aquell moment, cada vegada que baixo a Catalunya una veueta em diu: "ves a veure en Ramon". Crec que mai ningú m'ha conegut ni m'ha entés tant bé com ho has fet tu, ni jo mateix. Quan m'estava estressant, per art de màgia apareixies a la oficina i em calmaves. Hi ha persones que van i venen, tu vas venir i t'hi vas quedar. Els dos sabem que és dur, però si volem, podem. I si jo puc, tu també! T'estimo moltíssim.

Adrià, tu sempre has dit que "Sharing is caring, but it is also having less. Definitivament si és amb tu preferixo tenir molt menys que tenir-ho tot. Merci per tot, per ser un cabró a vegades (moltes) peró també per aguantar-me sempre, per estimar-me tant, per fer-me sentir viu i sobretot per ensenyar-me que a la vida hem de fer feliços i alliberar-nos, i a qui no li agradi? Doncs, dos pedres! Definitivament ets com el germà amb qui sempre discutir a veure qui la té més gran. T'estimo moltíssim.

Família SaliPetiSoldeMom, ens va unir la nostra bogeria a la carrera i des de llavors, no hem parat. La nostra "empresa" algún dia existirà! Sempre heu sigut clau, gràcies per tot, us estimo!

Abert, **Edu**, **Martí**, **Miquel** i **Xavi**, han passat sis anys des que vaig marxar però mai m'he sentit com si visqués lluny de vosaltres. Ens coneixem des de que tinc memòria. M'ho heu aguantat tot, m'heu respectat i acceptat sempre, en tots els sentits. En algún moment tenia por de ser rebutjat però vosaltres, sempre heu estat allà. Tot el que he aconseguit, en part, també és vostre. Sou els meus "trolls" herois! Estimar-vos és poc! Infinites gràcies per existir en la meva vida.

Pot semblar estrany donar les gràcies a un poble peró ho considero necessari. Et vaig deixar quan tenia 18 anys, ara deu anys més tard encara et trobo a faltar. Et porto i et portaré sempre allà on vagi i amb tu, els records, la **família** (única), la colla (**LGDL**, us estimo moltíssim!), el Cau. Jo sóc d'un poble, on la gent no veu les coses passar, sinó que fa que les coses passin. Jo sóc de **Tàrrega**!

Finalment i no menys important, vull donar les gràcies als que sempre ho han fet tot possible. **Papa**, **mama** i **Aïda**, possiblement sou els que heu patit més amb mi. No sóc gens fàcil, ho sé. Tot i que a vegades puc ser esquerp, us estimo moltíssim i això no ho hauria aconseguit sense vosaltres, ni el vostre suport. Gràcies família per cuidar-me, ajudar-me i permetre-me fer sempre el que he volgut per ser una mica més feliç. Gràcies Papa per animar-me a continuar quan més ho necessitava. Aquesta tesi també és vostra.

I look back in time and I only see good people. Gràcies per tot!

Tàrrega, January 2020

About the author

Jordi Petit Pedró was born on the 4th of July, 1991, in Tàrrega (Catalunya), a city in the North-East of Spain. After completing high school studies he moved to Barcelona to study a Bachelor in Genetics at the Autonomous University of Barcelona, where he got the degree in 2013. In the last year of his Bachelor studies, he focused on plant genetics.

As his interest in plant genetics increased, he started a MSc studies in Molecular Biotechnology at the University of Barcelona, focusing on plant genetics and biotechnology. During his master, he entered the Erasmus Programme and was able to do his Master thesis at Wageningen University and Research with Dr Arnaud Bovy and Dr Marcela Víquez-Zamora. His Master thesis was about QTL mapping for quality in tomato, which he did at the department of Plant Breeding. He got his MSc degree in 2014.

Soon after, he started his PhD research at the same department on the field of Plant breeding for Bio-based Economy under the supervision of Prof. Dr Luisa Trindade and Dr Elma Salentijn. His focus was the study of fibre quality variability in hemp to get insides into its genetic architecture. The results of his PhD project are described in this thesis.

List of publications

Petit, J., Gulisano, A., Dechesne, A., and Trindade, L.M. (2019). Phenotypic Variation of Cell Wall Composition and Stem Morphology in Hemp (*Cannabis sativa* L.): Optimization of Methods. *Frontiers in Plant Science* 10. doi: 10.3389/fpls.2019.00959

Salentijn, E.M.J., **Petit, J.**, and Trindade, L.M. (2019). The Complex Interactions Between Flowering Behavior and Fiber Quality in Hemp. *Frontiers in Plant Science* 10. doi: 10.3389/fpls.2019.00614

Petit, J., Salentijn, E.M.J., Paulo, M.J., Thourminot, C., Van Dinter, B.J., Magagnini, G., Gusovius, H.J., Tang, K., Amaducci, S., Wang, S., Uhrlaub, B., Mussig, J., and Trindade, L.M. (2019). Genetic variability of morphological, flowering and biomass quality traits in hemp (*Cannabis sativa* L.). *Frontiers in Plant Science (Accepted)*.

Petit, J., Salentijn, E.M.J., Paulo, M.J., Denneboom, C., Van Loo, E.N., and Trindade, L.M. (2019). Elucidating the genetic architecture of fibre quality in hemp (*Cannabis sativa* L.) using a Genome-Wide Association Study. *In preparation*.

Petit, J., Salentijn, E.M.J., Paulo, M.J., Denneboom, C., Van Loo, E.N., and Trindade, L.M. (2019b). Genetic architecture of flowering time and sex determination in hemp (*Cannabis sativa* L.): a Genome-Wide Association Study. *In preparation*.

Fuentes, C.A., Willekens, **P., Petit**, J., Thouminot, C., Müssig, J., Trindade, L.M., and Van Vuure, A.W. (2017). Effect of the middle lamella biochemical composition on the non-linear behaviour of technical fibres of hemp under tensile loading using strain mapping. *Composites Part A: Applied Science and Manufacturing* 101**,** 529-542. doi: 10.1016/j.compositesa.2017.07.017

Education Statement of the Graduate School Experimental Plant Sciences

Issued to: Jordi Petit Pedró
Date: 21 February 2020
Group: Laboratory of Plant Breeding
University: Wageningen University & Research

1) Start-up phase

	date
▶ **First presentation of your project**	
Genome-wide association mapping for hemp breeding	12 May 2015
▶ **Writing or rewriting a project proposal**	
Genome-wide association mapping for hemp breeding	02 Mar 2015
▶ **Writing a review or book chapter**	
▶ **MSc courses**	
▶ **Laboratory use of isotopes**	
Subtotal Start-up phase	*7.5 credits**

2) Scientific Exposure

	date
▶ **EPS PhD student days**	
EPS PhD student days 'Get2Gether 2016', Soest, NL	28-29 Jan 2016
EPS PhD student days 'Get2Gether 2017', Soest, NL	9-10 Feb 2017
EPS PhD student days 'Get2Gether 2018', Soest, NL	15-16 Feb 2018
▶ **EPS theme symposia**	
EPS Theme 3 Symposium: 'Metabolism and Adaptation', Wageningen, NL	10 Feb 2015
EPS Theme 2 Symposium: 'Interaction between Plants and Biotic Agents', Leiden, NL	22 Jan 2016
EPS Theme 3 Symposium: 'Metabolism and Adaptation', Amsterdam, NL	23 Feb 2016
▶ **NWO Lunteren days and other National Platforms**	
NWO-ALW meeting 'Experimental Plant Sciences' Lunteren, NL	13-14 Apr 2015
NWO-ALW meeting 'Experimental Plant Sciences' Lunteren, NL	11-12 Apr 2016
NWO-ALW meeting 'Experimental Plant Sciences' Lunteren, NL	10-11 Apr 2017
NWO-ALW meeting 'Experimental Plant Sciences' Lunteren, NL	9-10 Apr 2018
▶ **Seminars (series), workshops and symposia**	
Workshop: Wageningen UR Software Carpentry, Wageningen, NL	9 Fer 2015
Symposium: Plant Breeding Research Day, Wageningen, NL	29 Sep 2015
Seminar: Affymetrix, Genomics in plant and animal breeding and conservation: Strategies for success in plants, birds, and livestock, Wageningen, NL	4 Nov 2015
Seminar: Wilfred Vermerris, "An introduction to sorghum breeding", Wageningen, NL	16 Nov 2016
Seminar: Wilfred Vermerris, "Genetic and genomic resources for sorghum", Wageningen, NL	17 Nov 2016
Symposium: WURomics 2016, Wageningen, NL	15 Dec 2016
Seminar: Dr. Diana Santelia, "Rewiring starch metabolism for plant environmental adaptation", Wageningen, NL	1 Nov 2018
Symposium: Genotype to Phenotype modelling of Plant Adaptation, Wageningen, NL	16 Nov 2018
Workshop: Breeding for sustainability: Perspectives from a company and academic plant breeder, Wageningen, NL	20 Nov 2018
Seminar: Dr. Andew J. Simkin, "Metabolic engineering to enhance photosynthesis and increase crop yield", Wageningen, NL	21 Mar 2019
Seminar: Dr. Monika Doblin, "Designing walls for a sustainable future", Wageningen, NL	5 Jul 2019
Workshop: Breeding for diversity, Opportunities and challenges, Wageningen, NL	Oct 30 2019
▶ **Seminar plus**	

▶ **International symposia and congresses**	
MultiHemp Meeting 2 - York, UK	26-28 Jan 2015
MultiHemp Meeting 3 - Paris, France	18-20 Jan 2016
MultiHemp Meeting 4 - Milan, Italy	19-20 Jan 2017
12th International Conference of the European Industrial Hemp, Cologne, Germany	20-21 May 2015
EPSR PhD Retreat, Barcelona, Spain	20-23 June 2016
VIII. Cell wall research conference, Asilomar CA, USA	18-22 June 2018
▶ **Presentations**	
Talks:	
MultiHemp Meeting 3, Paris, France	18-20 Jan 2016
MultiHemp Meeting 4, Milan, Italy	18-20 Jan 2017
EPS Lunteren 2018, Lunteren, NL	9-10 Apr 2018
Plant Breeding and Biotechnology Symposium, Wageningen, NL	11-13 Jun 2019
Posters:	20-23 Jun 2016
EPSR PhD Retreat, Barcelona, Spain	20-23 Jun 2016
EPS Lunteren 2017, Lunteren, NL	10-11 Apr 2017
VIII. Cell wall research conference, Asilomar CA, USA (Speed talk and poster)	18-22 Jun 2018
▶ **IAB interview**	
▶ **Excursions**	
Company visit Koppert Biological Systems (EPS PhD council company visit)	26 Oct 2018
Subtotal Scientific Exposure	*20.2 credits**

3) In-Depth Studies	*date*
▶ **EPS courses or other PhD courses**	
EPS: Genome Assembly, Wageningen, NL	28-29 Apr 2015
PE&RC SENSE: Introduction to R for Statistical Analysis, Wageningen, NL	12-13 May 2016
PE&RC SENSE: Linear Models, Wageningen, NL	7-9 Jun 2017
PE&RC SENSE: Mixed Linear Models, Wageningen, NL	29-30 Jun 2017
PE&RC SENSE: Generalized Linear Models, Wageningen, NL	26-27 Jun 2017
SLU – WUR: Plant Breeding and Biotechnology, Wageningen, NL	11-13 Jun 2019
▶ **Journal club**	
▶ **Individual research training**	
Subtotal In-Depth Studies	*4.9 credits**

4) Personal development	*date*
▶ **Skill training courses**	
EPS: PhD introduction course, Wageningen, NL	20 Jan 2015
In'to languages: Scientific Writing, Wageningen, NL	Sep & Nov 2016
In'to languages: Presentation with impact, Wageningen, NL	Nov & Dec 2016
▶ **Organisation of PhD students day, course or conference**	
EPS Get2gether 2017 (Sponsorship)	Jan 2017
EPS Get2gether 2018 (Sponsorship)	Feb 2017
▶ **Membership of Board, Committee or PhD council**	
EPS PhD Council member	Mar 2016 - Mar 2018
Subtotal Personal Development	*5.8 credits**

TOTAL NUMBER OF CREDIT POINTS*	**38.4**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits.

**A credit represents a normative study load of 28 hours of study.*

The presented research has received funding from the European Union consortium MultiHemp (Multipurpose hemp for industrial bioproducts and biomass) under grant agreement number 311849. This consortium operated under the European Union's Seventh Programme for research, technological developments and demonstration (FP7/2012-2017).

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover photo: Agata Gulisano
Backcover photo: Jordi Petit Pedró
Cover & book layout: Iliana Boshoven-Gkini | AgileColor.com
Printed by: GVO drukkers & vormgevers, Ede (NL) | gvo.nl
This thesis is printed in FSC® certified paper.