



Elucidating the Genetic Architecture of Fiber Quality in Hemp (*Cannabis sativa* L.) Using a Genome-Wide Association Study

Jordi Petit¹, Elma M. J. Salentijn¹, Maria-João Paulo², Christel Denneboom¹, Eibertus N. van Loo¹ and Luisa M. Trindade^{1*}

¹ Wageningen UR Plant Breeding, Wageningen University & Research, Wageningen, Netherlands, ² Biometris, Wageningen University & Research, Wageningen, Netherlands

OPEN ACCESS

Edited by:

Gea Guerriero,
Luxembourg Institute of Science
and Technology, Luxembourg

Reviewed by:

Marc Behr,
Université libre de Bruxelles, Belgium
Jiasheng Wu,
Zhejiang Agriculture & Forestry
University, China
Eugenio López-Cortegano,
University of Edinburgh,
United Kingdom

*Correspondence:

Luisa M. Trindade
luisa.trindade@wur.nl

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 27 May 2020

Accepted: 25 August 2020

Published: 17 September 2020

Citation:

Petit J, Salentijn EMJ, Paulo M-J, Denneboom C, van Loo EN and Trindade LM (2020) Elucidating the Genetic Architecture of Fiber Quality in Hemp (*Cannabis sativa* L.) Using a Genome-Wide Association Study. *Front. Genet.* 11:566314. doi: 10.3389/fgene.2020.566314

Hemp (*Cannabis sativa* L.) is a bast-fiber crop with a great potential in the emerging bio-based economy. Yet, hemp breeding for fiber quality is restricted and that is mainly due to the limited knowledge of the genetic architecture of its fiber quality. A panel of 123 hemp accessions, with large phenotypic variability, was used to study the genetic basis of seven cell wall and bast fiber traits relevant to fiber quality. These traits showed large genetic variance components and high values of broad sense heritability 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 single nucleotide polymorphism (SNP) markers was used for a genome-wide association study (GWAS) approach to get insights into quantitative trait loci (QTLs) controlling fiber quality traits. In absence of a complete hemp genome sequence, identification of QTLs was based on the following characteristics: (i) association level to traits, (ii) fraction of explained trait variance, (iii) collinearity between QTLs, and (iv) detection across different environments. Using this approach, 16 QTLs were identified across locations for different fiber quality traits, including contents of glucose, glucuronic acid, mannose, xylose, lignin, and bast fiber content. Among them, six were found across the three environments. The genetic markers composing the QTLs that are common across locations are valuable tools to develop novel genotypes of hemp with improved fiber quality. Underneath the QTLs, 12 candidate genes were identified which are likely to be involved in the biosynthesis and modification of monosaccharides, polysaccharides, and lignin. These candidate genes were suggested to play an important role in determining fiber quality in hemp. This study provides new insights into the genetic architecture of fiber traits, identifies QTLs and candidate genes that form the basis for molecular breeding for high fiber quality hemp cultivars.

Keywords: GWAS, *Cannabis sativa*, hemp, fiber quality, cell wall composition, RAD-sequencing, QTL

INTRODUCTION

Hemp (*Cannabis sativa* L.) is a bast-fiber 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., 2008; 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 and more sustainable source of fibers relative to water and nutrient demanding crops, (e.g., cotton) and to non-renewable glass and fossil-based fibers (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 mostly a consequence of the strong decline in hemp production in the last century, when intensive breeding programs drove great improvements and amplified cultivation of major staple crops (Salentijn et al., 2015). Subsequently, hemp breeding was restricted and to date little is known about the genetic architecture that underlies hemp fiber quality.

Recently, large phenotypic variability in 28 traits related to fiber quality of hemp has been reported in a diverse panel of 123 hemp accessions native to different regions in the world (Petit et al., 2020). This study, conducted in trials in three European locations, reported that fiber 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 ($G \times E$), 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 $G \times E$ interactions to traits that are largely controlled by environmental components, with small contributions of genetic factors.

Fibers of high quality for different purposes and easily extractable from the stems are among the traits with the greatest appreciation by the hemp industry. More specifically, these include high bast fiber content, high content of cellulose, and low content of pectin in the cell walls, fine bast fibers and efficient decortication properties of the stems (Ranalli, 2004; Salentijn et al., 2015). The association of lignin content and fiber quality depends on their applications. Briefly, low lignin content in the bast fibers is associated to high quality of fiber for textile purposes. The lignin polymers hinder decortication and increases the stiffness of the fibers (Ranalli, 2004). On the other hand, the antioxidant properties of lignin and its adhesion functions, which are associated to increase the stability in the composites, boost the interest of lignin for innovative applications of the fibers (reviewed in Pickering et al., 2016). The contents of bast fiber, cellulose and lignin were reported to be largely determined by genetic components in the hemp accession panel from Petit et al. (2020), suggesting that the panel is a promising dataset to further study the genetic basis of these traits. In contrast, the variation of pectin, bast fiber fineness and decortication efficiency were largely controlled by environmental factors, which would hamper the study of their genetic components. For these low heritable traits,

Petit et al. (2020) suggested agronomic practices, as they showed adaptive behavior under certain environmental conditions.

Despite the large variability of hemp fiber quality and the large influence of genetics on important traits of fiber 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). Salentijn et al. (2019) reviewed the state of art of candidate gene studies for fiber 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). 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 quantitative trait loci (QTLs) in a genetic map. Despite the interest, to the best of our knowledge, no association studies for fiber 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 harboring large variability in fiber quality have hampered such studies on this crop (Sawler et al., 2015). As a result, no QTLs have been reported for hemp fiber quality. Furthermore, hemp breeding programs 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 fiber properties. This approach would allow the selection of promising individuals at early developmental stages, reducing the time and costs of breeding programs. Genetic association studies are thus of great value to upgrade breeding programs toward molecular approaches.

Insights in the genetic architecture of hemp fiber quality are essential to develop new breeding strategies, which seek for novel hemp cultivars with improved fiber properties. The objective of this study was to identify QTLs associated to fiber quality. The highly variable panel of 123 diploid hemp accessions ($2n = 20$) was used for this purpose (Petit et al., 2020). GWAS analyses were performed for fiber quality traits, with large genetic components. The study was performed in three test locations with contrasting environments.

MATERIALS AND METHODS

Plant Material

A panel of 123 hemp accessions was used in this study (Supplementary Table 1; Petit et al., 2020). The panel included mostly fiber 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, 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 1 m² in Italy and the Netherlands and of 1.5 m² in France.

Phenotyping and Data Collection

Phenotyping of cell wall traits was performed essentially as described in Petit et al. (2019) and in Petit et al. (2020). 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 fiber parameter (BCD%) were measured after harvesting the stems. Plants were cut from the base (10 cm above the ground) when the accumulative temperature degree days ($\Sigma^{\circ}\text{C}$, the accumulated Celsius degree day over a period at a base temperature of 1°C) were 1740.25 $\Sigma^{\circ}\text{C}$, 1421.1 $\Sigma^{\circ}\text{C}$, and 1843.3 $\Sigma^{\circ}\text{C}$ in CRA, FNPC, and VDS, respectively. These temperatures corresponded to the $\Sigma^{\circ}\text{C}$ 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 fiber 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. (2020).

DNA Extraction

Genomic DNA was isolated from young grinded hemp leaves (~20–400 mg, lyophilized material) using a cetyl trimethyl ammonium bromide (CTAB) method (Doyle and Doyle, 1987) with additional steps to remove proteins, polysaccharides, and RNA. First, the leaf material was treated with proteinase K in 500 μl TES buffer (100 mM TRIS pH = 8.0, 10 mM EDTA pH = 8.0, 2% SDS, 200 $\mu\text{g}/\text{ml}$ proteinase K) for 1 h at 60°C. Thereafter, the CTAB method was performed using a final concentration of 1% CTAB and 1.4 N NaCl. Polysaccharides were removed by incubation for 1 h on ice in 1.2 M NH₄Ac, centrifugation for 10 min at 13,000 rpm and discarding the pellet. The supernatant was then treated with RNaseA (100 μg RNaseA per 700 μl supernatant, 30 min at 37°C) and extracted with chloroform: isoamyl alcohol (IAA). The DNA was precipitated and dissolved in 550 μl 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 QubitTM 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 restriction site associated DNA sequencing (RAD-seq).

Restriction-Site-Associated DNA Sequencing (RAD-Seq)

The GWAS mapping panel was sequenced with RAD-seq to identify single nucleotide polymorphisms (SNPs) distributed over the genome to be used as molecular markers. High quality genomic DNA (2.5–5 μg at a concentration ≥ 25 ng/ μl) was digested using the restriction enzyme *EcoRI*. Then, RAD libraries with insert sizes of 300–550 bp, 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 \times 1 Gbp genomic data per sample. RAD library preparation and sequencing were performed by Beijing Genomics Institute (BGI, Hong Kong).

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 $Q \leq 12$, unknown bases > 3%, reads that lacks 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 or fraction of gap opens = 1; (e) max number or fraction of gap extensions = 50; (m) maximum entries in the queue = 100,000). 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 50.3–85.7%). Subsequently, SOAPsn was used to call SNPs in each sample (Li et al., 2009).

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 harbors DNA from 16 alleles, represented by mostly expected two different nucleotides but occasionally three and rarely four 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.

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 analyzed to study the level of population structure in the GWAS panel. Pairwise comparisons of coefficient of population differentiation (F_{ST}) 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} .

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 modeling, see section "Genome-Wide Association Study (GWAS) Analysis") 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 (~130–~600 kbp) 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.

In most tested genomic scaffolds, the r^2 decayed with the distance. Nevertheless, a baseline value of r^2 close to 0.1 was detected in all scaffolds independently of the distance between markers. This baseline value of r^2 (< 0.1) 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. Therefore, $r^2 \geq 0.1$ was used as threshold for collinearity of markers in the multiQTL models (see section "Genome-Wide Association Study (GWAS) Analysis"). 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 ~500 kbp 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.

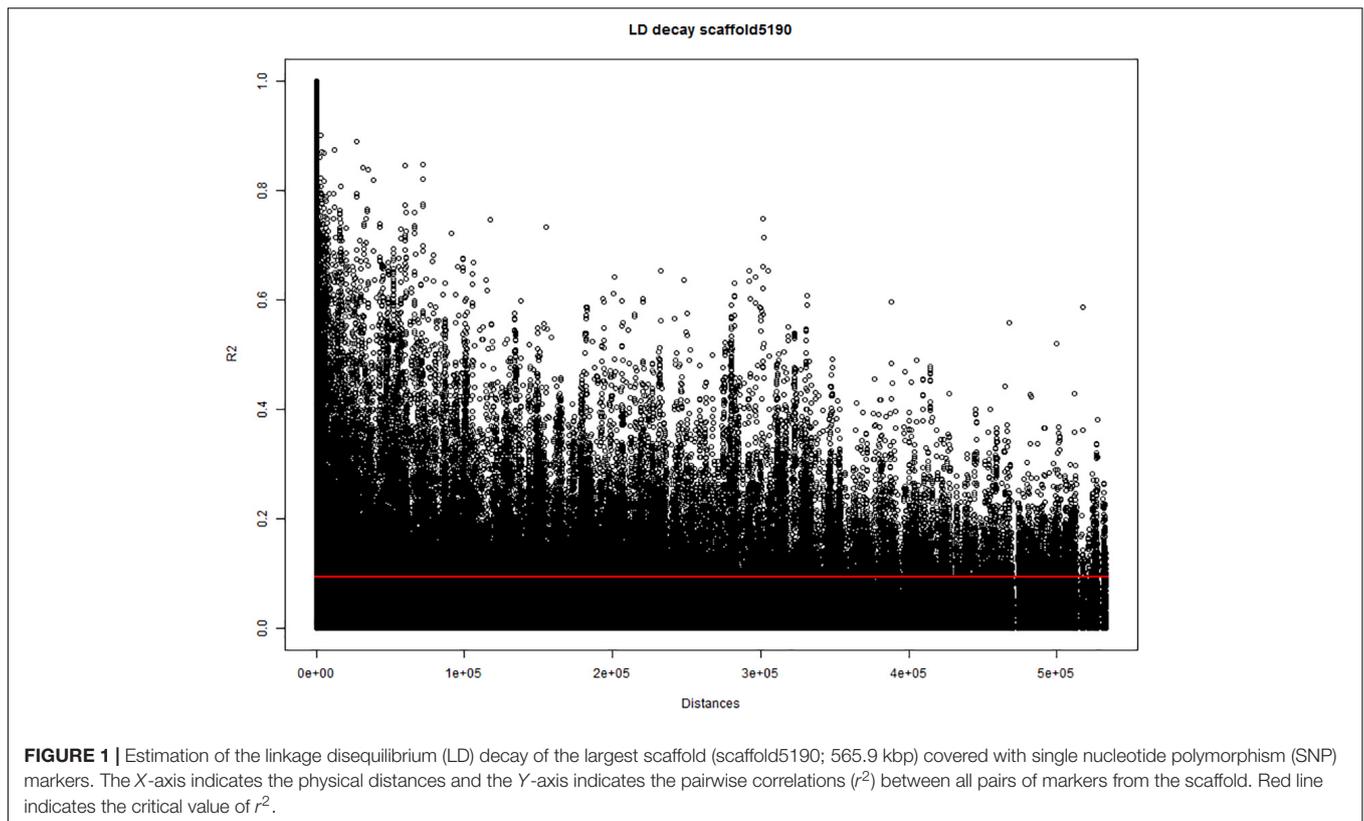
Genome-Wide Association Study (GWAS) Analysis

A Linear Mixed Model (LMM) was used to identify significant associations between SNP markers and fiber 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). Briefly, the effects of the SNP markers on the phenotypic variation were studied with the fixed effects of the model. The kinship was used to control for population structure effects and was set in the random effects of the model. The kinship was calculated using all selected markers (~600 k), following the equations to calculate the genomic relationship matrix developed by VanRaden (2008). LMM analyses were performed using restricted maximum likelihood (REML) algorithm and the kinship was calculated using R¹ version 3.4.3 statistical software. The linear mixed model equation for our association with kinship correction is expressed as:

$$\gamma = X\alpha + K\beta + e \quad (2)$$

Equation (2) is a standard linear mixed model in which γ represents the phenotype, X is the marker, α is the effect size of the marker, K is the kinship for population structure correction, β is the effect of the population structure and e is the residual effects. $X\alpha$ represents the fixed effects, and $K\beta$, and e the random effects. Wald statistic from the REML analysis was used to assess the significant level of the associations. To account for multiple testing and estimate the threshold for significant associations, 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 3,000 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

¹<http://www.r-project.org/>



were unlinked to the polymorphism controlling the variation of the traits. An independent analysis was performed separately for seven fiber 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 percentage of the total phenotypic variation explained by the genetic effects of the full MultiQTL model was calculated by the correlation (r^2) between the fitted trait values and the observed traits values. The explained variance is an approximation to the heritability of the trait.

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 2**).

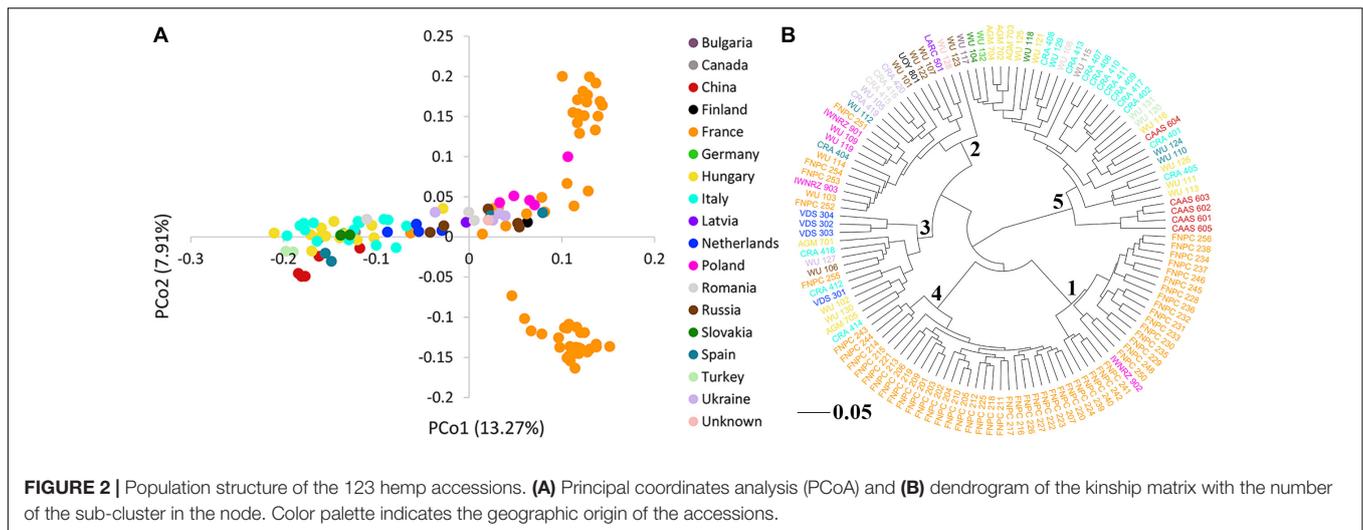
In order to find groups of significant markers, a principal component analysis (PCA) was performed. The PC1, PC2, and the $-\log_{10}P$ of the association between genotypes and phenotypes were plotted in 3D scatter plots to detect the clusters. The 3D scatter plots resembled the peak tops of the QTLs in Manhattan plot distributions.

To further characterize the genetics of fiber 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 harbor correlated ($r \geq 0.3$; $r^2 \geq 0.1$) representative QTL-markers from MultiQTL models of two or three locations.

Genome-wide association study (GWAS), cumulative distribution analyses, PCA and correlation analyses were performed in Genstat 19th edition software (VSN International, Hemel Hempstead, United Kingdom). The 3D scatter plots were performed in Excel version 14.0, using the macro Excel 3D Scatter Plot version 2.1 (Doka, 2013).

Transcriptome Annotation and Candidate Gene Identification

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.cabr.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; Conesa et al., 2005; Conesa and Gotz, 2008). The annotated transcriptomic data are deposited in the 4TU.ResearchData archive². Genomic scaffolds associated to the QTLs were analyzed 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.

The sequences of the candidate genes were blasted using Blast + (Camacho et al., 2009) to the transcriptome BioProject PRJNA435671 (Behr et al., 2019) to identify the corresponding *Arabidopsis* homologues.

RESULTS

Development of Markers for Genome-Wide Association Studies

Restriction site associated DNA sequencing generated 3,717.57 million clean reads (557.4 Gb) with an average of 29.7 million reads per sample (range 15.1–49.2 millions) with an average read length of 149.9 bp. 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 (843 Mb).

²<https://doi.org/10.4121/uuid:d17c382c-f292-4875-888b-47e6433d600f>

Low Level of Population Structure in the Hemp 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. The 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).

Effective Control of Population Structure in Highly Heritable Fiber 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 fiber content make these traits excellent candidates to study the genetic architecture of fiber quality of hemp (Supplementary Tables 3, 4). The high heritabilities of these seven traits allow the study of the additive control of fiber

TABLE 1 | Pairwise coefficients of population differentiation (F_{ST}) between the five clusters of the genome-wide association study (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	–

quality traits. In addition, their interactions between genotypes and environments ($G \times E$) allow to study the heritable variation in fiber quality sensitive to the environment (Petit et al., 2020).

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 fiber 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 for the seven fiber quality traits. These results show that the kinship matrix efficiently controlled the effects of population structure on the seven traits, which is key to identify true-positive associations. The phenotypic data of the seven traits can therefore be used to get insights into the genetic control of fiber quality.

Significant Associations Between Markers and Fiber Quality Traits in Hemp

The GWAS analyses resulted in the identification of over 2,500 markers, mapping to single loci on 1,515 different genomic scaffolds. These associations were found significant ($-\log_{10}P \geq 4.047$) for at least one trait and one location (**Supplementary File 1**). 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 14,465 bp) that showed 12 significant markers (**Supplementary File 1**). 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 $\sim 9,000$ bp, while the 11 significant markers from scaffold34707 clustered in a shorter sequence of ~ 600 bp. 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 significantly associated to several fiber quality traits. For example, many markers from scaffold4465 were significantly associated to contents of glucose, xylose, lignin (ADL, KL) and bast fiber content (**Supplementary File 1**). These results suggest a common genetic control between different traits.

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 eight 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 significant $G \times E$ interactions on the traits (**Supplementary Table 4**).

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 the QTLs in 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 fiber content, while only a single cluster of markers was identified for respectively contents of mannose and glucuronic acid.

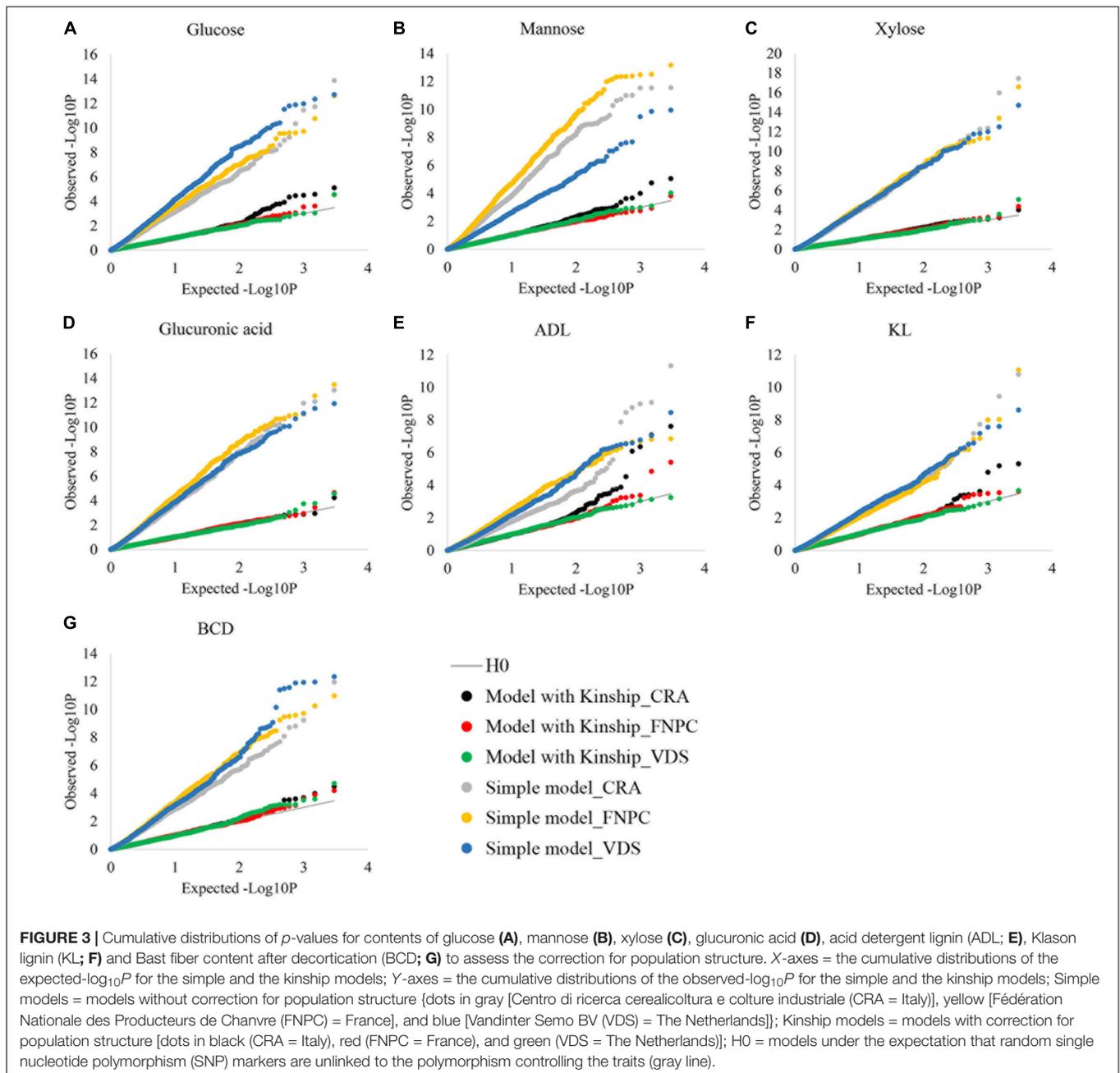
QTL Regions Across Locations for Fiber Quality

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 fiber 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 fiber content (QTL_{BCD1}); and two for content of KL (QTL_{KL1} and QTL_{KL2}).

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; **Tables 2, 3**). The common scaffolds between QTLs for several traits indicated co-localization of some QTLs in tightly close genomic regions.

Candidate Genes for Fiber 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 5**). 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 defense 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 showed high similarity to genes involved in cell wall biosynthesis. These candidate genes were identified in six QTLs across locations, as detailed in **Table 4**.



A *polygalacturonase* gene, with reported activity in pectin hydrolysis, and *2-phytyl-1,4- β -naphthoquinone methyltransferase*, with reported activity in flavonoid biosynthesis, were identified in QTL_{ADL1} for ADL content. Two genes involved in lignin and sugar metabolism 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 was mapped in the scaffold53823 where both QTL_{Glc1} and QTL_{KL1} are located. 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. Two genes involved in sugar metabolism (*glyceraldehyde-3-phosphate dehydrogenase (gapdh)* and *α -mannosidase*) and two genes indirectly involved in the lignin metabolism, [*chalcone synthase (chs)* and *stilbene synthase (sts)*], primarily involved in the precursor pathways of flavonoid and stilbene biosynthesis, 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 adenyltransferase (glgc)* involved

TABLE 2 | Quantitative trait loci (QTLs) for fiber 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
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	Fiber	CRA	3	41.8	scaffold56299_8302; scaffold13466_34599; scaffold35378_13209
BCD	Fiber	FNPC	3	61.08	scaffold26831_57993; scaffold25080_10706; scaffold62704_12566
BCD	Fiber	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 fiber 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; Explained variance, percentage of the total phenotypic variation explained by the genetic effects of the full MultiQTL model. The explained variance is an approximation of the heritability contribution of the full model. 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 single nucleotide polymorphism (SNP) within the scaffold (canSat3 assembly GAC_000230575.1).

in the metabolism of monosaccharides was identified in QTL_{Man1} for mannose content.

DISCUSSION

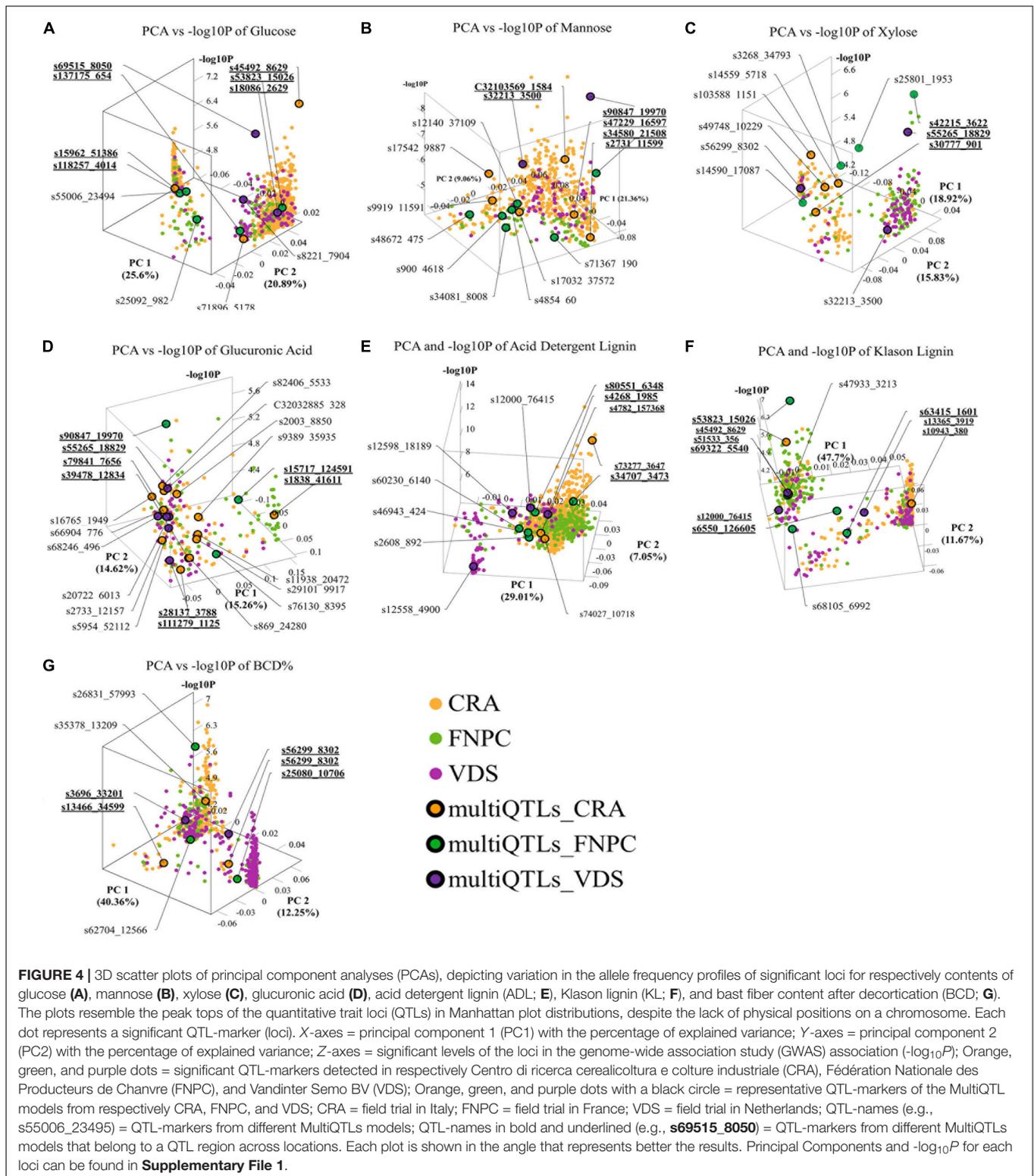
Genetic Architecture Underlying Hemp Fiber Quality

The study of the genetic architecture of hemp fiber quality is fundamental to develop molecular strategies to breed for new hemp cultivars with improved fiber properties. The genetics of hemp fiber quality have been poorly studied and genetic markers for these traits are lacking in 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 fiber crops, such as cotton (Liu et al., 2018) 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 limitation, 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 fiber 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 fiber 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 (Supplementary Table 4; Petit et al., 2020). Only



glucuronic acid was largely influenced by the environment (Supplementary Table 4). Petit et al. (2020) suggested that the content of this monosaccharide might have stronger general response to adapt to the environment, independently of the

heritable genetic control of the trait. This might be explained by functional redundancy between different cell wall components. The putative changes in plant fitness associated to the changes in glucuronic acid content across environments might be replaced

TABLE 3 | Identification of quantitative trait loci (QTL) regions across locations for seven fiber quality traits of hemp.

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	

QTLs common across locations were identified by correlation analyses ($r \geq 0.3$, $\sim r^2 \geq 0.1$). QTL across locations = name given to the QTL common across locations; Trait: ADL%dm, acid detergent lignin content; BCD, bast fiber 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, Netherlands.

by other cell wall components with similar functions, such as lignin (Sarkar et al., 2009). Despite the large environmental component of glucuronic acid, as its phenotypic variation is also explained by the genetic factors, it is also a good candidate for genetic association studies. Genetic studies on cell wall composition in other crops also showed large heritabilities for biomass quality related traits. In maize, heritabilities for biomass traits, such as contents of cellulose, hemicellulose and lignin were reported above 0.6 (Torres et al., 2015). 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 high heritabilities indicate that the genetic gains of fiber quality can be maximized in several crops.

In this report, sixteen QTLs across locations were found to be associated to fiber quality in hemp. For six of these QTLs, 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 lignin biosynthesis. The genes *phosphoenolpyruvate carboxylase* (*pepc*), *glyceraldehyde-3-phosphate dehydrogenase* (*gapdh*) and *glucose-1-phosphate adenyltransferase* (*glgc*), involved in sugar metabolism were identified underneath QTL_{Glc3}, QTL_{GlcA3}, and QTL_{Man1}, respectively. *Phosphoenolpyruvate carboxylase* (*pepc*) codifies for an enzyme that catalyzes the carboxylation of phosphoenolpyruvate to oxaloacetate to produce malate. The activity of this enzyme has been reported to play multiple roles, among them fiber elongation, carbon storage and energy production (Zieher, 2010). The accumulation of malate and sugars, is thought to play an important role in the fiber elongation, through osmotic regulation and charge balance. For instance, Li et al. (2010)

reported in cotton that in periods of rapid elongation phase, *pepc* genes were highly expressed, but weakly expressed at slow elongation periods. In addition, *pepc* is involved in carbon storage and energy production by converting phosphoenolpyruvate into oxaloacetate (Jeanneau et al., 2002). Phosphoenolpyruvate results from the glycolysis of glucose and consequently reduces the source of monosaccharides for other pathways, such as to biosynthesize cell wall polysaccharides. *Glyceraldehyde-3-phosphate dehydrogenase* (*gapdh*) is another important gene involved in glycolysis. This enzyme catalyzes the conversion of glyceraldehyde-3-phosphate to 1,3-bisphosphoglycerate with reduction of NAD⁺ to NADH. Downregulation of *gapdh* in *Arabidopsis* resulted in drastic changes in the sugar balance of the plant, arrested root development and dwarfism (Muñoz-Bertomeu et al., 2009). This study showed the importance of *gapdh* in plant primary metabolism. Thus, variation in *gapdh* expression might change the monosaccharide availability for different cellular biosynthetic processes. *Glucose-1-phosphate adenyltransferase* (*glgc*) is involved in the partitioning of α -D-glucose-1P for starch and cell wall polysaccharide biosynthesis. *Glucose-1-phosphate adenyltransferase* codes for an enzyme that catalyzes 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 substrate 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-apiiose (Wierzbicki et al., 2019).

TABLE 4 | Putative candidate genes in quantitative trait loci (QTLs) across locations associated to fiber quality traits.

QTL Across Locations	Trait	Scaffold	Candidate Gene (Transcript)	Description	Class of Metabolism	UniProt Code	References
QTL _{ADL1}	ADL%dm	scaffold4782	FN14269	<i>Polygalacturonase</i>	Polysaccharide	B9SPH9 and A0A1J3JMW1	Chan et al., 2010; Blande et al., 2017
			FN26643	<i>2-Phytyl-1,4-beta-naphthoquinone methyltransferase</i>	Lignin	W9RCM0 and A0A1R3I6R5	
QTL _{Glc1}	Glc%dm	scaffold53823	FN16068	<i>Cytochrome b₅</i>	Lignin	A0A1R3IKW6, W9R7N2 and W9R7N2	He et al., 2013
QTL _{Glc3}	Glc%dm	scaffold137175	FN29173	<i>Phosphoenolpyruvate carboxylase (pepc), housekeeping isozyme</i>	Monosaccharide	W9RZ79	He et al., 2013
QTL _{GlcA1}	GlcA%dm	scaffold55265	FN02726	<i>p-Coumaroyl shikimate 3-hydroxylase (c3h1)</i>	Lignin	A0A068FPP3	Kamdee et al., 2014
QTL _{GlcA3}	GlcA%dm	scaffold15717	FN00563	<i>Glyceraldehyde-3-phosphate dehydrogenase (gapdh)</i>	Monosaccharide	W9R8N5	He et al., 2013
			FN19953	<i>Chalcone synthase (chs)</i>	Lignin	A0A068PX04, Q1G6S4, Q94LW8 and R9WQY5	Lihova et al., 2006
			FN08380	<i>Stilbene synthase (sts)</i>	Lignin	Q6BAR5	Richter et al., 2005
			FN15157	<i>Alpha-mannosidase</i>	Polysaccharide	A0A087G6Y0 and A0A1S2XQU0	Parween et al., 2015
QTL _{KL1}	KL%dm	scaffold53823	FN16068	<i>Cytochrome b₅</i>	Lignin	A0A1R3IKW6 and W9R7N2	He et al., 2013
QTL _{KL2}	KL%dm	scaffold10943	FN05922	<i>Methyltransferase</i>	Lignin/Polysaccharide	W9S3D7	He et al., 2013
QTL _{KL3}	KL%dm	scaffold6550	FN11795	<i>Glucan endo-1,3-beta-glucosidase 3/F-box protein SKIP28</i>	Polysaccharide	W9SPC5, A0A1S2YDV5 and A0A1S3UV78	He et al., 2013; Kang et al., 2014; Parween et al., 2015
QTL _{Man1}	Man%dm	scaffold34580	FN16350	<i>Glucose-1-phosphate adenyltransferase (glgc)</i>	Monosaccharide	W9R328	He et al., 2013
QTL _{Xyl1}	Xyl%dm	scaffold55265	FN02726	<i>p-Coumaroyl shikimate 3-hydroxylase (c3h1)</i>	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 fiber 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 fiber 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. *Arabidopsis* homologues for the candidate genes can be found in Supplementary Table 6.

Alpha-mannosidase, *glucan endo-1,3-β-glucosidase 3*, and *polygalacturonase* are involved in the metabolism of glycoproteins and polysaccharides. *Alpha-mannosidase* codes for an enzyme involved in early N-glycan processing. The N-glycans are oligosaccharides (Glc₃Man₉GlcNAc₂) 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 signaling and interacting with the surrounding environment; and plant defense. Moreover, they are essential for 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 et al. (2014) suggested that N-glycosylation of glycoproteins plays a role in the correct targeting, assembly, or stability of cell wall biosynthetic or remodeling enzymes. 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 defense can modify cell wall composition [reviewed in Underwood (2012)]. *Glucan*

endo-1,3-β-glucosidase 3 codes for an enzyme involved in degradation of cell wall polysaccharides and plant defense against pathogens. In *Arabidopsis*, this enzyme has been implicated in pathogenesis when the plants were infected with tobamovirus (Zavaliev et al., 2013). The infection triggered the endogenous secretion of this enzyme in the cell wall, which degraded callose in 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 et al. (2013) 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 coding for enzymes involved in degradation of the cell wall, specifically homogalacturonan type of pectin. These enzymes function in a wide range of developmental processes, including cell separation and cell elongation to control the shape of the plant architecture (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 fibers (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 pgx1* mutants showed strong cell-elongation defects, alterations in pectin molecular mass and cell wall composition (Xiao et al., 2014).

Cytochrome b₅, p-coumaroyl shikimate 3-hydroxylase (c3h1), chalcone synthase (chs), stilbene synthase (sts), and 2-phytyl-1,4-β-naphthoquinone methyltransferase are involved directly or indirectly in lignin biosynthesis. A recent study in *Arabidopsis* described the product of *cytochrome b₅* 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 b₅* 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, 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 fiber quality without affecting plant fitness. Furthermore, anthocyanin biosynthesis and monolignol production share the phenylpropanoid pathway from phenylalanine to *p-coumaroyl CoA*. The latter is directly used in the flavonoid pathway through *chalcone synthase*, or used in the monolignol pathway through genes directly involved in the lignin metabolism (Behr et al., 2016). *Stilbene synthase* is another gene involved in the flavonoid and stilbene

biosynthesis. Overexpression of genes involved in the production of flavonoids might decrease the availability of metabolites for the monolignol pathway, which seems to be associated to differences in lignification of hemp fibers. These genes are known to be regulated at the level of transcription. MYB transcription factors (e.g., AtMYB75) are key regulators of the anthocyanin branch of the phenylpropanoid pathway to favor flavonoid biosynthesis at the expense of monolignol biosynthesis [Zhong and Ye, 2009 and Bhargava et al., 2010 (*Arabidopsis*); Shaipulah et al., 2016 (petunia)]. External treatment of chalcone in soybean was reported to inhibit lignin biosynthesis (Chen et al., 2011). In addition, Behr et al. (2016) reported differential expression of *chalcone synthase* gene between young and old hypocotyls of hemp. *Chalcone synthase* gene was more expressed in young hypocotyls to produce anthocyanins, acting as a photo-protectant when the hypocotyl emerges from the soil (Steyn et al., 2002; Wang et al., 2012). In older hypocotyls the expression of *chalcone synthase* gene is reduced to allow the biosynthesis of monolignol from *p-coumaroyl CoA*. 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). *2-phytyl-1,4-β-naphthoquinone methyltransferase* is an enzyme involved in the biosynthesis of phylloquinone. This gene plays an active role in the biosynthesis of aromatic amino acids, such as phenylalanine (reviewed in Tzin and Galili, 2010). Variation in *2-phytyl-1,4-β-naphthoquinone methyltransferase* might affect the lignin content similarly to *chalcone* and *stilbene genes*. They might affect the biosynthesis of monolignols and consequently the lignification of the fibers. These results are supported by the revisions of Boerjan et al. (2003) that reported profound effects on the biosynthesis of lignin due to availability of phenylalanine.

Moreover, the identification of the same genomic scaffolds in QTLs for different traits is an indication of co-localization of QTLs for fiber 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 et al. (2019) reported large correlations between cell wall components in hemp, such as a negative correlation between the contents of glucose and lignin ($r = -0.93$, which reflects the relationship between bast and woody hemp core, as glucose is the main component of bast and lignin of woody hemp core) and contents of xylose and glucuronic acid ($r^2 = 0.96$; Petit et al., 2020). 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 b₅*, 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. These results might be explained in the context of hemp stem development. The identified *cytochrome b₅* might be a reflection of the differential gene expression between young and old hypocotyls of hemp (Behr et al., 2016). Behr et al. (2016)

identified larger number of genes related to lignin metabolism in young than in old hypocotyls of hemp. In addition, Novaes et al. (2010) reviewed evidences that the negative correlation between lignin and cellulose (glucose) is partially mediated at the level of transcription. 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 are main components of 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 xylans are cross-linked (Pauly et al., 2013) and highly correlated (Petit et al., 2020). Allelic variation of *c3h1* can directly lead to phenotypic variation of lignin content, as reported in other crops (Ralph et al., 2006, 2012; Sykes et al., 2015). Thereafter, alteration of lignin content can affect the cross-linking to other polymers 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)].

The candidate genes identified in this study are of great significance for further studies, including reverse genetic approaches, such as candidate gene functional studies, but specially differential expression analysis from different stem tissues. Such studies would be very valuable for understanding the molecular mechanisms responsible for hemp fiber quality.

Importance of QTLs in Genetic Improvement of Hemp

The identification of QTLs across locations for hemp fiber quality will have relevant implications for the breeding programs 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) breeding schemes in hemp. The implementation of molecular markers in hemp breeding programs is expected to speed up the development of new hemp cultivars with improved fiber properties. The use of molecular markers has several advantages, including the early stage selection of promising plants. In addition, the use of markers circumvents the phenotyping of large number of samples for fiber quality, which usually involves large costs and time in breeding programs.

Furthermore, the identification of QTLs and candidate genes for fiber 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 fiber quality traits. Those studies would also provide more molecular breeding tools to accelerate the development of new hemp cultivars with improved fiber quality.

CONCLUSION

The present study provides valuable insights into the genetic and molecular architecture of fiber quality in hemp, which advocates for positive prospects to modernize breeding programs of hemp toward molecular approaches. The sixteen QTLs for fiber quality are the first candidates to include in marker assisted selection breeding schemes of hemp. The implementation of these QTLs in molecular breeding programs in hemp will accelerate the selection of interesting individuals and will bypass the phenotyping of large number of samples, leading to important cost reductions. In addition, the identification of QTLs and candidate genes for fiber quality enables the use of the hemp panel in further studies to extend the genetic architecture of other important traits in 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 sequence is available.

DATA AVAILABILITY STATEMENT

RADseq data, selected informative SNP markers and the annotated transcriptomic data of the hemp cultivar *C. sativa* 'Finola' are deposited in the 4TU.ResearchData archive (<https://doi.org/10.4121/12826832>, <https://doi.org/10.4121/uuid:d17c382c-f292-4875-888b-47e6433d600f>).

AUTHOR CONTRIBUTIONS

JP designed and performed the experiments, analyzed the data, and wrote the manuscript. M-JP and EL helped analyzing the data and revised the manuscript. ES helped designing and performing the experiments, helped analyzing the data, and revised the manuscript. CD helped designing and performing the experiments and revised the manuscript. LT coordinated and supervised this study, experimental strategy, and discussion of the outcomes and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

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 Program for research, technological developments, and demonstration under grant agreement number 311849.

ACKNOWLEDGMENTS

The authors are grateful to Richard Finkers and Francesco Pancaldi for their technical assistance in the annotation of the transcriptome.

REFERENCES

- Amaducci, S., Colauzzi, M., Zatta, A., and Venturi, G. (2008). Flowering dynamics in monoecious and dioecious hemp genotypes. *J. Ind. Hemp*. 13, 5–19. doi: 10.1080/15377880801898691
- 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, (Chichester: Wiley), 109–134.
- Babu, Y., and Bayer, M. (2014). Plant polygalacturonases involved in cell elongation and separation—the same but different? *Plants Basel* 3, 613–623. doi: 10.3390/plants3040613
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376. doi: 10.1371/journal.pone.0003376
- Behr, M., Legay, S., Žižková, E., Motyka, V., Dobrev, P. L., Hausman, J. F., et al. (2016). Studying secondary growth and bast fibre development: the hypocotyl peeks behind the wall. *Front. Plant Sci.* 7:1733. doi: 10.3389/fpls.2016.01733
- Behr, M., Lutts, S., Hausman, J.-F., Sergeant, K., Legay, S., and Guerreiro, G. (2019). *De novo* transcriptome assembly of textile hemp from datasets on hypocotyls and adult plants. *Data Brief* 27:104790. doi: 10.1016/j.dib.2019.104790
- Bhargava, A., Mansfield, S. D., Hall, H. C., Douglas, C. J., and Ellis, B. E. (2010). MYB75 functions in regulation of secondary cell wall formation in the arabidopsis inflorescence stem. *Plant Physiol.* 154, 1428–1438. doi: 10.1104/pp.110.162735
- 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 caerulea*. *Sci. Data* 4:160131. doi: 10.1038/sdata.2016.131
- Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. *Annu. Rev. Plant Biol.* 54, 519–546. doi: 10.1146/annurev.arplant.54.031902.134938
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinform.* 10:421. doi: 10.1186/1471-2105-10-421
- Chan, A. P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiui, D., et al. (2010). Draft genome sequence of the oilseed species *Ricinus communis*. *Nat. Biotechnol.* 28, 951–956. doi: 10.1038/nbt.1674
- Chen, W.-J., Yun, M.-S., Deng, F., and Yogo, Y. (2011). Chalcone suppresses lignin biosynthesis in illuminated soybean cells. *Weed Biol. Manag.* 11, 49–56. doi: 10.1111/j.1445-6664.2011.00404.x
- Conesa, A., and Gotz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008:619832. doi: 10.1155/2008/619832
- 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. doi: 10.1093/bioinformatics/bti610
- de Setta, N., Monteiro-Vitorello, C. B., Metcalfe, C. J., Cruz, G. M. Q., Del Bem, L. E., Vicentini, R., et al. (2014). Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics* 15:540. doi: 10.1186/1471-2164-15-540
- Doka, G. (2013). *Excel 3D Scatter Plot macros version 2.1*.
- Doyle, J. J., and Doyle, J. A. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Ebskamp, M. J. M. (2002). Engineering flax and hemp for an alternative to cotton. *Trends Biotechnol.* 20, 229–230. doi: 10.1016/s0167-7799(02)01953-4
- 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. doi: 10.1007/s10681-016-1641-2
- 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. doi: 10.1093/jxb/erm222
- Gou, M., Yang, X., Zhao, Y., Ran, X., Song, Y., and Liu, C.-J. (2019). Cytochrome b5 is an obligate electron shuttle protein for syringyl lignin biosynthesis in *Arabidopsis*. *Plant Cell* 31, 1344–1366. doi: 10.1105/tpc.18.00778
- 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, and R. C. Jennings, (Cambridge: Cambridge University Press), 95–106.
- Grassa, C. J., Wenger, J. P., Dabney, C., Poplawski, S. G., Motley, S. T., Michael, T. P., et al. (2018). A complete *Cannabis chromosome* assembly and adaptive admixture for elevated cannabidiol (CBD) content. *bioRxiv* [Preprint]. doi: 10.1101/458083
- He, N., Zhang, C., Qi, X., Zhao, S., Tao, Y., Yang, G., et al. (2013). Draft genome sequence of the mulberry tree *Morus notabilis*. *Nat. Commun.* 4:2445. doi: 10.1038/ncomms3445
- 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. *Front. Plant Sci.* 7:984. doi: 10.3389/fpls.2016.00984
- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. doi: 10.1038/ng.695
- Jeanneau, M., Vidal, J., Gousset-Dupont, A., Lebouteiller, B., Hodges, M., Gerentes, D., et al. (2002). Manipulating PEP levels in plants. *J. Exp. Bot.* 53, 1837–1845. doi: 10.1093/jxb/erf061
- 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 Biol. Technol.* 97, 68–76. doi: 10.1016/j.postharvbio.2014.06.004
- Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B. K., et al. (2014). Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat. Commun.* 5:5443. doi: 10.1038/ncomms6443
- Kruijer, W., Boer, M. P., Malosetti, M., Flood, P. J., Engel, B., Kooke, R., et al. (2015). Marker-based estimation of heritability in immortal populations. *Genetics* 199, 379–398. doi: 10.1534/genetics.114.167916
- Kumar, M., Campbell, L., and Turner, S. (2016). Secondary cell walls: biosynthesis and manipulation. *J. Exp. Bot.* 67, 515–531. doi: 10.1093/jxb/erv533
- Laverty, K. U., Stout, J. M., Sullivan, M. J., Shah, H., Gill, N., Holbrook, L., et al. (2019). A physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the THC/CBD acid synthase loci. *Genome Res.* 29, 146–156. doi: 10.1101/gr.242594.118
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, J., and Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 95, 221–227. doi: 10.1038/sj.hdy.6800717
- Li, R., Li, Y., Fang, X., Yang, H., Wang, J., Kristiansen, K., et al. (2009). SNP detection for massively parallel whole-genome resequencing. *Genome Res.* 19, 1124–1132. doi: 10.1101/gr.088013.108
- 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. doi: 10.1093/jxb/erp299
- Liebinger, E., Hüttner, S., Vavra, U., Fischl, R., Schoberer, J., Grass, J., et al. (2009). Class I alpha-mannosidases are required for N-glycan processing and root development in *Arabidopsis thaliana*. *Plant Cell* 21, 3850–3867. doi: 10.1105/tpc.109.072363
- Lihova, J., Shimizu, K. K., and Marhold, K. (2006). Allopolyploid origin of *Cardamine asarifolia* (Brassicaceae): incongruence between plastid and nuclear

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.566314/full#supplementary-material>

- ribosomal DNA sequences solved by a single-copy nuclear gene. *Mol. Phylogenet. Evol.* 39, 759–786. doi: 10.1016/j.ympev.2006.01.027
- Liu, R., Gong, J., Xiao, X., Zhang, Z., Li, J., Liu, A., et al. (2018). GWAS analysis and QTL identification of fiber quality traits and yield components in upland cotton using enriched high-density SNP markers. *Front. Plant Sci.* 9:1067. doi: 10.3389/fpls.2018.01067
- Lozovaya, V. V., Zabortina, 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.
- Malosetti, M., Van Eeuwijk, F. A., Boer, M. P., Casas, A. M., Elia, M., Moralejo, M., et al. (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. doi: 10.1007/s00122-011-1558-z
- Muñoz-Bertomeu, J., Cascales-Miñana, B., Mulet, J. M., Baroja-Fernández, E., Pozueta-Romero, J., Kuhn, J. M., et al. (2009). Plastidial Glyceraldehyde-3-Phosphate Dehydrogenase deficiency leads to altered root development and affects the sugar and amino acid balance in arabidopsis. *Plant Physiol.* 151, 541–558. doi: 10.1104/pp.109.143701
- Nguema-Ona, E., Vicré-Gibouin, M., Gotté, M., Plancot, B., Lerouge, P., Bardor, M., et al. (2014). Cell wall O-glycoproteins and N-glycoproteins: aspects of biosynthesis and function. *Front. Plant Sci.* 5:499. doi: 10.3389/fpls.2014.00499
- 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 Physiol.* 154, 555–561. doi: 10.1104/pp.110.161281
- Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290.
- Parween, S., Nawaz, K., Roy, R., Pole, A. K., Venkata Suresh, B., Misra, G., et al. (2015). An advanced draft genome assembly of a desi type chickpea (*Cicer arietinum* L.). *Sci. Rep.* 5:12806. doi: 10.1038/srep12806
- Pauly, M., Gille, S., Liu, L., Mansoori, N., De Souza, A., Schultink, A., et al. (2013). Hemicellulose biosynthesis. *Planta* 238, 627–642. doi: 10.1007/s00425-013-1921-1
- 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. *Front. Plant Sci.* 10:959. doi: 10.3389/fpls.2019.00959
- Petit, J., Salentijn, E. M. J., Paulo, M. J., Thourminot, C., Van Dinter, B. J., Magagnini, G., et al. (2020). Genetic variability of morphological, flowering and biomass quality traits in hemp (*Cannabis sativa* L.). *Front. Plant Sci.* 10:102. doi: 10.3389/fpls.2020.00102
- Pickering, K. L., Efendy, M. G., Aruan, and Le, T. M. (2016). A review of recent developments in natural fibre composites and their mechanical performance. *Compos. Part A Appl. Sci. Manuf.* 83, 98–112. doi: 10.1016/j.compositesa.2015.08.038
- Ralph, J., Akiyama, T., Coleman, H. D., and Mansfield, S. D. (2012). Effects on lignin structure of coumarate 3-hydroxylase downregulation in poplar. *Bioenerg. Res.* 5, 1009–1019. doi: 10.1007/s12155-012-9218-y
- Ralph, J., Akiyama, T., Kim, H., Lu, F., Schatz, P. F., Marita, J. M., et al. (2006). Effects of coumarate 3-hydroxylase down-regulation on lignin structure. *J. Biol. Chem.* 281, 8843–8853. doi: 10.1074/jbc.M511598200
- Ranalli, P. (2004). Current status and future scenarios of hemp breeding. *Euphytica* 140, 121–131. doi: 10.1007/s10681-004-4760-0
- 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*. *Physiol. Mol. Plant* 67, 248–260. doi: 10.1016/j.pmpp.2006.03.001
- 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.
- Salentijn, E. M. J., Petit, J., and Trindade, L. M. (2019). The complex interactions between flowering behavior and fiber quality in hemp. *Front. Plant Sci.* 10:614. doi: 10.3389/fpls.2019.00614
- Salentijn, E. M. J., Zhang, Q., Amaducci, S., Yang, M., and Trindade, L. M. (2015). New developments in fiber hemp (*Cannabis sativa* L.) breeding. *Ind. Crop. Prod.* 68, 32–41. doi: 10.1016/j.indcrop.2014.08.011
- 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. doi: 10.1093/jxb/erp245
- Sawler, J., Stout, J. M., Gardner, K. M., Hudson, D., Vidmar, J., Butler, L., et al. (2015). The genetic structure of marijuana and hemp. *PLoS One* 10:e0133292. doi: 10.1371/journal.pone.0133292
- Shaipulah, N. F., Muhlemann, J. K., Woodworth, B. D., Van Moerkercke, A., Verdonk, J. C., Ramirez, A. A., et al. (2016). CCoAOMT down-regulation activates anthocyanin biosynthesis in *Petunia*. *Plant Physiol.* 170, 717–731. doi: 10.1104/pp.15.01646
- Slavov, G. T., Nipper, R., Robson, P., Farrar, K., Allison, G. G., Bosch, M., et al. (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. doi: 10.1111/nph.12621
- Steyn, W. J., Wand, S. J. E., Holcroft, D. M., and Jacobs, G. (2002). Anthocyanins in vegetative tissues: a proposed unified function in photoprotection. *New Phytol.* 155, 349–361. doi: 10.1046/j.1469-8137.2002.00482.x
- Strasser, R. (2014). Biological significance of complex N-glycans in plants and their impact on plant physiology. *Front. Plant Sci.* 5:363. doi: 10.3389/fpls.2014.00363
- 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. *Ind. Crop. Prod.* 11, 107–118. doi: 10.1016/S0926-6690(99)00048-5
- Sykes, R. W., Gjersing, E. L., Foutz, K., Rottmann, W. H., Kuhn, S. A., Foster, C. E., et al. (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. doi: 10.1186/s13068-015-0316-x
- Torres, A. F., Noordam-Boot, C. M. M., Dolstra, O., Van Der Weijde, T., Combes, E., Dufour, P., et al. (2015). Cell wall diversity in forage maize: genetic complexity and bioenergy potential. *Bioenerg. Res.* 8, 187–202. doi: 10.1007/s12155-014-9507-8
- Tzin, V., and Galili, G. (2010). New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. *Mol. Plant* 3, 956–972. doi: 10.1093/mp/ssq048
- Underwood, W. (2012). The plant cell wall: a dynamic barrier against pathogen invasion. *Front. Plant Sci.* 3:85. doi: 10.3389/fpls.2012.00085
- van Bakel, H., Stout, J. M., Cote, A. G., Tallon, C. M., Sharpe, A. G., Hughes, T. R., et al. (2011). The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol.* 12:R102. doi: 10.1186/gb-2011-12-10-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 Sci.* 174, 205–220. doi: 10.1016/j.plantsci.2007.11.008
- 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. *Front. Plant Sci.* 7:2004. doi: 10.3389/fpls.2016.02004
- van der Werf, H. M. G., and Turunen, L. (2008). The environmental impacts of the production of hemp and flax textile yarn. *Ind. Crop. Prod.* 27, 1–10. doi: 10.1016/j.indcrop.2007.05.003
- 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. doi: 10.1186/1471-2229-12-130
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Wang, S., Gusovius, H.-J., Lühr, C., Musio, S., Uhrlaub, B., Amaducci, S., et al. (2018). Assessment system to characterise and compare different hemp varieties based on a developed lab-scaled decortication system. *Ind. Crop. Prod.* 117, 159–168. doi: 10.1016/j.indcrop.2018.02.083
- Wang, Y., Zhou, B., Sun, M., Li, Y., and Kawabata, S. (2012). UV-A light induces anthocyanin biosynthesis in a manner distinct from synergistic blue + UV-B light and UV-A/blue light responses in different parts of the hypocotyls in turnip seedlings. *Plant Cell Physiol.* 53, 1470–1480. doi: 10.1093/pcp/pcs088
- 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. *Front. Plant Sci.* 10:176. doi: 10.3389/fpls.2019.00176
- Wright, S. (1969). *Evolution and the Genetics of Populations, vol. II. The Theory of Gene Frequencies*. Chicago, IL: University of Chicago Press.
- 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. doi: 10.1105/tpc.114.123968

- Xie, D., Dai, Z., Yang, Z., Sun, J., Zhao, D., Yang, X., et al. (2018). Genome-wide association study identifying candidate genes influencing important agronomic traits of flax (*Linum usitatissimum* L.) using SLAF-seq. *Front. Plant Sci.* 8:2232. doi: 10.3389/fpls.2017.02232
- Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702
- Zavaliev, R., Levy, A., Gera, A., and Epel, B. L. (2013). Subcellular dynamics and role of arabisidopsis beta-1,3-glucanases in cell-to-cell movement of tobamoviruses. *Mol. Plant Microbe* 26, 1016–1030. doi: 10.1094/mpmi-03-13-0062-r
- Zhong, R., and Ye, Z. H. (2007). Regulation of cell wall biosynthesis. *Cur. Opin. Plant Biol.* 10, 564–572. doi: 10.1016/j.pbi.2007.09.001
- Zhong, R., and Ye, Z. H. (2009). Transcriptional regulation of lignin biosynthesis. *Plant Signal. Behav.* 4, 1028–1034. doi: 10.4161/psb.4.11.9875
- 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. doi: 10.1093/pcp/pcr185
- Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–824. doi: 10.1038/ng.2310
- Zieher, C. (2010). “Biochemistry of the fiber,” in *Physiology of Cotton*, eds J. M. Stewart, D. M. Oosterhuis, J. J. Heitholt, and J. R. Mauney, (Dordrecht: Springer), 361–378.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Petit, Salentijn, Paulo, Denneboom, van Loo and Trindade. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.