



Article Syntenic Cell Wall QTLs as Versatile Breeding Tools: Intraspecific Allelic Variability and Predictability of Biomass Quality Loci in Target Plant Species

Francesco Pancaldi ¹^(b), Eibertus N. van Loo ¹^(b), Sylwia Senio ¹, Mohamad Al Hassan ¹, Kasper van der Cruijsen ¹, Maria-João Paulo ², Oene Dolstra ¹, M. Eric Schranz ³ and Luisa M. Trindade ¹,*^(b)

- Plant Breeding, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
- ² Biometris, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
 ³ Biosystematics, Wageningen University & Research, Droevendaalsesteeg 1,
 - 6708 PB Wageningen, The Netherlands
- * Correspondence: luisa.trindade@wur.nl; Tel.: +31-317-481-538

Abstract: Syntenic cell wall QTLs (SQTLs) can identify genetic determinants of biomass traits in understudied species based on results from model crops. However, their effective use in plant breeding requires SQTLs to display intraspecific allelic variability and to predict causative loci in other populations/species than the ones used for SQTLs identification. In this study, genome assemblies from different accessions of Arabidopsis, rapeseed, tomato, rice, Brachypodium and maize were used to evaluate the intraspecific variability of SQTLs. In parallel, a genome-wide association study (GWAS) on cell wall quality traits was performed in miscanthus to verify the colocalization between GWAS loci and miscanthus SQTLs. Finally, an analogous approach was applied on a set of switchgrass cell wall QTLs retrieved from the literature. These analyses revealed large SQTLs intraspecific genetic variability, ranging from presence-absence gene variation to SNPs/INDELs and changes in coded proteins. Cell wall genes displaying gene dosage regulation, such as PAL and CAD, displayed presence-absence variation in Brachypodium and rapeseed, while protein INDELs were detected for the Brachypodium homologs of the rice brittle culm-like 8 locus, which may likely impact cell wall quality. Furthermore, SQTLs significantly colocalized with the miscanthus and switchgrass QTLs, with relevant cell wall genes being retained in colocalizing regions. Overall, SQTLs are useful tools to screen germplasm for relevant genes and alleles to improve biomass quality and can increase the efficiency of plant breeding in understudied biomass crops.

Keywords: QTL analysis; GWAS; biomass quality; cell wall; synteny; genomics; miscanthus; switchgrass

1. Introduction

The projection of known quantitative trait loci (QTLs) across species through genome synteny was recently shown to allow for the quick identification of conserved genomic regions underlying traits of interest in large panels of plants, including novel, underdomesticated, crops [1]. On the one hand, this is possible thanks to the established relationship between the occurrence of gene synteny (i.e., the conservation of gene presence and gene order across genomes) and the conservation of gene function across diverse living organisms [2–4]. On the other hand, the colocalization of syntenic regions with previously mapped QTLs ensures the relevance of the regions identified for the involvement in the trait of interest, since the occurrence of QTLs proves that genomic regions are causative of trait variability in particular species. The genomic regions identified across several species by combining synteny and QTL information can be referred to as syntenic quantitative trait loci (SQTLs) [1].



Citation: Pancaldi, F.; van Loo, E.N.; Senio, S.; Al Hassan, M.; van der Cruijsen, K.; Paulo, M.-J.; Dolstra, O.; Schranz, M.E.; Trindade, L.M. Syntenic Cell Wall QTLs as Versatile Breeding Tools: Intraspecific Allelic Variability and Predictability of Biomass Quality Loci in Target Plant Species. *Plants* **2023**, *12*, 779. https:// doi.org/10.3390/plants12040779

Academic Editor: Baohong Zhang

Received: 19 January 2023 Revised: 1 February 2023 Accepted: 2 February 2023 Published: 9 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). The concept of SQTLs was demonstrated for plant cell wall compositional traits [1]. These traits entail the relative amounts and the chemical–physical properties of the polysaccharides that constitute plant cell walls and underly feedstock quality of biomass crops—mainly cellulose, hemicelluloses, pectins, and lignin [5,6]. The generally high syntenic conservation of cell wall genes within previously mapped cell wall QTLs allowed for the identification of numerous SQTLs, highlighting the potential of this strategy for projecting important conserved loci underlying biomass quality across multiple plants [1]. Furthermore, cell wall SQTLs were shown to allow for a "fine-mapping" of the initial cell wall QTLs, by assessing the overlap of multiple initial QTLs on specific syntenic segments across multiple genomes [1]. Finally, cell wall SQTLs contain relevant cell wall genes, known from previous studies to affect cell wall quality in different species, and therefore likely representing (some of) the conserved causative genes of the initial cell wall QTLs [1].

What was just discussed demonstrates that SQTLs represent valuable tools during the pre-breeding steps of crop improvement. Specifically, their availability, combined with the dropping of genome sequencing costs, allows for the potential circumvention (of part of) the pre-breeding studies on trait genetics needed to start breeding programs in under-domesticated crops [1]. Given the wealth of genetic resources represented by underdomesticated plant species [7], the availability of tools to speed up their improvement is an extremely important asset for agriculture [8–10]. As an example, the improvement of biomass crops—which are all the plant species that can produce biomass to sustain biobased value chains [11,12]—could greatly benefit from this prospect. In fact, they include several under-domesticated species (see Mehmood, Ibrahim, Rashid, Nawaz, Ali, Hussain and Gull [12], and Pancaldi and Trindade [6] for a comprehensive list), whose breeding cycles are highly time-consuming [13], while several complex plant traits should be improved in these species to allow their cultivation on marginal lands to avoid competition with food production [6]. Nonetheless, the use of SQTLs in breeding contexts requires the availability of intraspecific allelic variability with a potential impact on traits of interest for the SQTL regions themselves, as genetic variability is the prerequisite for selection. Moreover, SQTLs must be able to predict the localization of causative loci of the traits for which they were mapped, also in other plant populations or species than the ones used for SQTLs identification.

The two latter aspects are the focus of this study. On the one hand, the intraspecific allelic genomic variability of cell wall SQTLs was assessed in six angiosperm species—*Arabidopsis thaliana, Solanum lycopersicum, Brassica napus, Zea mays, Oryza sativa,* and *Brachypodium distachyon*—for which multiple genomes representing diverse plant accessions were available. On the other hand, the SQTL value for predicting relevant loci in novel populations/species was tested by assessing the colocalization between cell wall SQTLs from *Miscanthus sinensis* and *Panicum virgatum* and genomic regions associated with cell wall variability identified through association mapping in miscanthus and switchgrass, respectively. Miscanthus and switchgrass are C4 biomass crops with high potential to fulfil diverse industrial applications, but at the same time still largely under-domesticated [5,13]. The study of the genetics underlying cell wall composition in these crops is thus pivotal to bringing them out of their current state [6,14]. The results obtained in this study demonstrate the validity of SQTLs for this purpose.

2. Results

2.1. General SQTL Alignment Patterns across Multiple Accessions of Six Angiosperm Species

To investigate the intraspecific genetic variability of SQTLs, the nucleotide sequences of 1184 cell wall SQTLs previously identified in six angiosperm species (Arabidopsis, rapeseed, tomato, maize, rice, and Brachypodium) [1] were aligned against 111 genome assemblies representing different plant accessions of the six species themselves (Figure 1). On average, alignments covered 89.3% of SQTLs initial bp length across all the genome assemblies tested (CV = 5.8%; Figure 2). Figure 2 shows that inter-specific syntenic regions spanned by SQTLs are, as expected, very well conserved overall, also at the intraspecific

level. Nevertheless, it also highlights that minor SQTL portions display presence/absence variation (PAV) patterns between reference and specific target assemblies. Specifically, from the boxplots of Figure 2, these patterns appear particularly relevant in the three Poaceae species and, to a minor extent, tomato, where several SQTLs have relatively large portions of their total length not represented in alignments. On the one hand, this might be due to technological reasons such as the sequencing methodology of the target assemblies (e.g., de novo vs. reference-based resequencing). On the other hand, it could also depend on local intraspecific genomic rearrangements involving SQTLs regions. This hypothesis was tested in maize, by using the annotation of the transposon regions from the reference maize genome assembly used for initial SQTL detection (B73 version 4.0). This way, it was found out that maize SQTL regions span a large number of transposons (74 transposons per SQTL on average), which may be involved in intraspecific structural genomic variability, leading to the alignment patterns found (Supplementary Figure S1).



Figure 1. Schematic representation of the workflow followed to analyze intraspecific allelic variability of SQTLs. SQTL nucleotide sequences have been extracted from reference genome assemblies of six species (**A**) and aligned against multiple genome assemblies representing diverse accessions of each species (**B**). Nucleotide polymorphisms and gene presence–absence variation were quantified and analyzed across accessions and SQTLs (**C**). Finally, the effect of genomic polymorphisms on gene coding sequences on protein sequences and structures was also assessed (**D**) and compared with known mutations responsible of relevant biomass phenotypes.

Irrespective of the source of alignment length variability, the fact that SQTLs sometimes do not entirely align to target genome assemblies might lead to PAV of SQTL genes across the assemblies tested. In this sense, a BLAST validation of the SQTL genes displaying PAV according to the outputs of the intraspecific SQTL alignments revealed that 4225 SQTL genes are missing in one or more target assemblies across all the species tested (3.3% of all the genes contained in SQTLs). The majority of these genes is absent in only one target assembly (2086 genes across all the species), while 475 genes across all the species tested are missing in >50% of the target assemblies assessed for a particular species (Supplementary Table S1). Finally, of all the genes displaying PAV patterns between reference and target assemblies, 178 are cell wall genes (4.2%; Supplementary Table S1). Among these, some appear relevant for cell wall quality variability. These include an endoglucanasecoding gene from maize homolog to arabidopsis KORRIGAN (ND_NP_001288520.1, which is absent in four maize target assemblies), five PAL genes from *Brachypodium distachyon* involved in lignin synthesis (BG_XP_003575404.1, BG_XP_003575365.1, BG_XP_003575403.1, BG_XP_003575240.2, and BG_XP_003575238.1, which are absent in 11, 4, 2, 1, and 1 assemblies, respectively), and two CAD genes operating the in muro monolignol polymerization from Brassica napus (BL_XP_013702441.1, BL_XP_013702446.1, both absent in one target assembly) (Supplementary Table S1). To conclude, in spite of the genes just mentioned, the overall relatively low level of intraspecific PAV agrees with both the close relatedness of intraspecific accessions and the high level of syntenic conservation of SQTL regions.



Figure 2. Boxplots representing the coverage of reference SQTL segments across the target accessions used for SQTL alignment in every species. Data points are percentages of SQTL segments contained in a target assembly (one data point per SQTL, per species). (**A**) *Arabidopsis thaliana;* (**B**) *Brassica napus;* (**C**) *Solanum lycopersicum;* (**D**) *Zea mays;* (**E**) *Brachypodium distachyon;* (**F**) *Oryza sativa.*

2.2. Large Nucleotide Variation within the Intra-Specific SQTL Aligning Regions

SQTL alignments were also used to quantify intraspecific nucleotide variation at SQTL regions consisting of single nucleotide polymorphisms (SNPs) and insertions/deletions (INDELs, <100 bp) between reference and target assemblies (Figure 1). The results revealed extensive intraspecific nucleotide variability of SQTL regions (Figure 3). Specifically, SQTLs displayed on average 2.2 SNPs/kbp and 0.3 INDELs/kbp across all the species and assemblies assessed. These figures correspond to an absolute mean of 2643 SNPs and 295 INDELs per SQTL, considering the average SQTL length across all species (1176 kbp).



Figure 3. Average number of SNPs/kbp (dark grey) and INDELs/kbp (light grey) over SQTLs and per species (**A**) and plant clade (Eudicots vs. Poaceae) (**B**).

A more detailed analysis revealed that the number of SNPs/kbp and INDELs/kbp varies substantially between species (ANOVA's p = 0.000; Figure 3A). Specifically, Arabidopsis and Brachypodium SQTLs displayed a particularly high number of SNPs/kbp between reference and target assemblies compared to the other species (LSD's p < 0.001; Figure 3A). Conversely, regarding INDELs/kbp, only arabidopsis SQTLs displayed a substantially higher number of INDELs compared to the average across all SQTLs from all species (LSD's p < 0.001; Figure 3A). Finally, both SNPs and INDELs occur with much higher frequency in dicot SQTLs (2.6 SNPs/kbp and 1.3 INDELs/kbp) compared to Poaceae SQTLs (0.8 SNPs/kbp and 0.2 INDELs/kbp; *t*-test's p < 0.001 for both; Figure 3B), in spite of the opposite trend observed for overall coverage of SQTL alignments (Section 2.1).

The occurrence of SNPs and INDELs on SQTLs between reference and target assemblies was also assessed with respect to SQTL portions specifically spanned by genes. In this regard, it was found out that the majority of nucleotide polymorphisms occurs within intergenic genomic areas, as only 20.3% of SQTL SNPs and 11.5% of SQTL INDELs was located on SQTL genes (Supplementary Table S2). Moreover, the majority of all polymorphisms occurring on gene regions is located on noncoding gene segments, as only 44.4% of the SNPs on gene regions and 36.7% of the INDELs on gene regions were specifically located on exons. Still, in absolute terms, the latter two percentages correspond to an average of 159 SNPs and 23 INDELs per SQTL located on gene exons across all target species and assemblies, which are considerable numbers for the potential trait effects that these polymorphisms may cause. Finally, when assessed on a single-gene level, our results showed that 60,207 SQTL genes displayed at least one polymorphism in one species and against one target assembly (47% of all QTL genes). Of these, 3156 are cell wall-related

6 of 25

genes (5.2% of total polymorphic genes). As average, each of these genes retains 16 SNPs and 3 INDELs across all species and all target assemblies, of which 7 SNPs and 1 INDEL are on exon regions.

2.3. Intraspecific SQTL Variability Leads to Changes in Cell Wall Protein Sequences with a Potential Functional Impact

The final step of the analysis of SQTL variability across the six species and related target genome assemblies consisted in assessing the effect of SNPs and INDELs on translated protein sequences. Overall, across all SQTLs, all species, and all target assemblies, 30,654 SQTL genes (23% of all SQTL genes) displayed polymorphisms leading to one or more protein sequence changes between reference and one or more target assemblies (Supplementary Dataset S1). Of these, 1861 are cell wall genes. On average, each of the SQTL genes displaying protein-impacting polymorphisms across target assemblies retained 5.1 SNPs and 0.2 INDELs with an effect on the translated protein sequences. Finally, 1126 SQTL genes of all the ones displaying protein-impacting polymorphisms (0.8% of all SQTL genes) have SNPs or INDELs between reference and target assemblies that lead to stop codons and (likely) truncated proteins. Of these, 421 are cell wall genes.

To assess the potential effect of the intraspecific protein sequence variability on cell wall quality traits, how such patterns impact functionally relevant and highly syntenically conserved candidate SQTL genes identified in our previous SQTL study was studied [1]. These genes include members of the COBRA (COB) and COBRA-like (COBL) gene families responsible of the Brittle-culm cell wall rice mutants [15]; important genes associated with lignin content variability in cell walls, as FERULATE 5-HYDROXYLASE (F5H), CINNAMOYL-CoA O-METHYLESTERASE (CCOAOMT), and PHENYLALANINE AMMONIA-LYASE (PAL) [16]; key hemicellulose- and cellulose-related genes for cell wall polysaccharides metabolism, as members of the CELLULOSE SYNTHASE (CESA), CEL-LULOSE SYNTHASE-LIKE (CSL), and IRREGULAR XYLEM (IRX) gene families [17,18]; important transcription factors for secondary cell wall development such as WRKY12, NAC SECONDARY WALL-THICKENING FACTOR1 (NST1), and C3H14 [19,20]. For all of these genes, protein sequences were aligned and annotated for functional domains and motifs, allowing for the detection of protein sequence variability across assemblies, as well as for the assessment of the effect of variability on protein polarity, hydrophobicity, and functional domain properties. Figures 4 and 5 and Supplementary Figures S2–S5 display the results of these analyses.

Among all the alignments computed, the ones of the three members of the rice BRIT-TLE CULM-LIKE loci OsBCL1, OsBCL8, and OsBCL9, and of their related syntenic homologs conserved through SQTLs [1], yielded particularly interesting results. In fact, all the proteins coded by these three rice genes displayed one or more amino acid changes with effect on the polarity, charge and/or hydrophobicity of the proteins themselves. Specifically, these changes took place in the rice cultivars "Carolina Gold Select" and "Azucena" compared to the reference rice "Nipponbare". In this regard, OsBCL9 displayed three amino acid changes within the COBRA domain of the protein, of which one involving a change in polarity (alanine to threonine at position 123) and leading to the removal of a β -sheet domain in the 2D protein configuration (Figure 4A). A similar pattern was observed for Os-BCL1 (Supplementary Figure S2). Finally, OsBCL8 displayed a cysteine to arginine change, leading to loss of amino acid charge within the protein C-terminus (position 653), right next to the GPI-anchoring-related ω -site (Figure 4B). In addition to the rice BCL loci, the other *BCL* genes from Brachypodium, sorghum, and maize syntenic to *OsBCL* genes and retained within SQTLs were also assessed for protein variability. Intraspecific alignments revealed that the Brachypodium homolog of the OsBCL8 gene (XP_003559754.1) displayed massive protein rearrangements in three of the 45 Brachypodium genome assemblies analyzed ("ABR3", "Bd2", and "Adi12"). These changes involve large deletions of the BdBCL8

protein sequence, including a consistent part of the *COBRA* domain in the line "ABR3", and the complete *COBRA* domain in the lines "Bd2" and "Adi12" (Figure 5).



Figure 4. Amino acid changes and their effects on protein structure for three *brittle culm*-like loci between the reference rice cv. "Nipponbare" and the two cultivars "Azucena" and "CarolinaGoldS-elect". (A) *OsBCL9*; (B) *OsBCL8*. In each figure, protein sequences are colored according to amino acid polarity, and protein domains and signal peptides are annotated. Amino acid changes indicated with * indicate a change in polarity, while sites annotated with ω indicate the predicted GPI-anchoring-related omega sites.



Figure 5. The syntenic conservation of the rice locus *OsBCL8* (HN_XP_015645132.1) across Poaceae through SQTLs (**A**), and the intraspecific allelic variability with an impact on protein sequence of the *Brachypodium distachyon* syntenic homolog of *OsBCL8* conserved through SQTLs (BG_XP_003559754.1), across 25 Brachypodium accessions compared to the reference genome (**B**). In Figure 5A, connections between genes indicate synteny through SQTLs, as described in [1]. In Figure 5B, protein sequences are colored according to amino acid polarity, and protein domains and signal peptides are annotated.

In addition to the particularly relevant examples found for the *BRITTLE CULM*-like genes, intraspecific amino acid changes and INDELs impacting protein properties were found in several of the other proteins coded by the candidate SQTL genes inspected (Supplementary Figures S3–S5). As for the rice brittle culm loci, protein changes were observed both within and outside protein functional domains or motifs. For example, a maize IRX9 protein involved in xylan biosynthesis (NP_001147664.1) displayed two regions with relatively large INDELs and different amino acid substitutions within the GT43 functional domain (Supplementary Figure S3). These changes take place in 12 of the 33 maize assemblies assessed and have effects on protein polarity and/or charge. Similarly, the enzyme coded by the Brachypodium secondary cell wall CESA7 gene (XP_003574029.1), also retained within SQTLs, displayed large deletions within the cellulose synthase catalytic domain in the accession "TR9" compared to the reference Brachypodium genome (Supplementary Figure S4). Moreover, amino acid substitutions within the protein functional domain with an effect on protein polarity were also observed. Overall, these patterns were linked to major changes in the secondary structure of the CESA7 protein, including the deletion of seven α -helix and one β -sheet domains (Supplementary Figure S4). Finally, multiple INDELs and amino acid substitutions with impact on protein chemical properties were also detected across the arabidopsis accessions for the AtC3H14 gene (a transcription factor regulating secondary cell wall thickening), across different Brachypodium accessions for the *BdWRKY12* gene (a major transcription factor regulating cell wall biosynthesis

in plant vessels), and between the rice reference genome and three rice cultivars (Koshihikari, Kitaake, and Azucena) for the *OsNAC43* gene (also a major transcription factor for secondary cell wall biosynthesis in different plant tissues) (Supplementary Figure S5).

2.4. SQTLs Are Valid Tools to Predict Important Cell Wall Genomic Loci in Miscanthus sinensis and Panicum virgatum

In addition to evaluating the level and relevance of the intraspecific genetic variability of SQTLs in diverse species, another aim of this study was to assess the validity of SQTLs for predicting genomic loci associated with biomass quality traits in novel plant populations and species. This aspect was studied by assessing the degree of colocalization between the SQTLs previously detected in *Miscanthus sinensis* and *Panicum virgatum* [1] and cell wall related QTLs mapped by genome-wide association analysis (GWAS) in a *Miscanthus sinensis* collection and QTLs mapped by other researchers in an F1 progeny of a biparental cross of two switchgrass lines diverging for cell wall quality traits [21]. These populations represent respectively a different intraspecific population than the one from which QTLs used for SQTLs mapping came from (miscanthus), and a separate species than the ones from which initial QTLs have been selected at the moment of SQTLs detection (thus, a hypothetical novel, under-domesticated, species; switchgrass).

The GWAS conducted on the miscanthus population identified a total of 91 QTLs associated to eight traits related to cell wall content and composition (Figure 6 and Supplementary Table S3). These 91 QTLs cover 6.8% of the miscanthus genome and include a total of 148 cell wall genes. First, the general degree of overlap between the GWAS QTL regions and the 254 SQTLs previously detected in miscanthus (which cover 32.7% of the miscanthus genome; Supplementary Table S4) was assessed. This analysis revealed that 67 SQTLs (26.4% of all the miscanthus SQTLs) colocalized (for parts of their regions) with the 91 GWAS QTLs. Moreover, it was observed that 35 of the 91 GWAS QTLs (38%) colocalized for >50% of their bp length with genomic regions where miscanthus SQTLs are also present (Figure 6). To test if these figures highlight significant colocalization of the QTL loci identified by GWAS with the miscanthus SQTLs, a permutation analysis was performed by constructing 100 sets of 91 random genomic regions mirroring the bp size distribution of the GWAS QTLs (Supplementary Dataset S2). For each set, the proportion of random QTLs colocalizing with SQTLs was computed, and a binomial test was performed to assess if this proportion was significantly lower than the one observed for the real GWAS QTLs. The results showed that, as an average across the 100 random QTLs sets analyzed, 17 QTLs of the 91 included in each set (19%) colocalized for >50% of their length with SQTLs. This figure is significantly lower compared to what observed in real QTLs (binomial test significant in 91 of the 100 tests performed at $\alpha = 0.01$), highlighting that SQTLs colocalize significantly with the GWAS QTLs.

An analogous procedure to the one just described for miscanthus was performed in switchgrass to test the colocalization between the 56 cell wall-related QTLs identified by Ali, Serba, Walker, Jenkins, Schmutz, Bhamidimarri, and Saha [21] (Supplementary Table S5), and the 254 SQTLs previously detected in switchgrass and conserved across Poaceae (Supplementary Table S6). In switchgrass, it was shown that 53 SQTLs (20.8% of all the switchgrass SQTLs) (partially) colocalize with the 56 cell wall QTLs, while 33 QTLs (59%) colocalize for >50% of their length with SQTLs (Figure 7). Permutation analysis demonstrated that the colocalization of QTLs and SQTLs is significant, as random QTL sets produced significantly lower proportions of colocalizing QTLs (binomial test significant in 99% of iterations at $\alpha = 0.01$) (Supplementary Dataset S3).



Figure 6. The results of the miscanthus GWAS and of the colocalization between miscanthus SQTLs and QTLs plotted onto the miscanthus genome. From outside to inside, the first strip shows the LOD scores of the markers from the GWAS, with the markers included in the 91 QTLs colored in blue; the second strip displays the genomic ranges of the 91 QTLs from the GWAS, with the QTLs colocalizing with SQTLs for >50% of their length colored in red; the third strip displays the positions of miscanthus SQTLs; the fourth strip highlights the most relevant candidate genes identified from the GWAS analysis.

Since colocalization between SQTLs and QTLs was demonstrated for both miscanthus and switchgrass, as a final step, we analyzed which cell wall genes are retained in colocalizing regions between SQTLs and QTLs, as well as the level of syntenic conservation of those genes in angiosperms. In this regard, miscanthus results showed that several important cell wall genes were retained within the colocalizing regions between SQTLs and GWAS QTLs. These include, among others, different central cell wall transcription factors, as one of the two *MsBLH6* (GQ_01G471400 and GQ_02G130600), the *MsWRKY12* (GQ_12G168300), the miscanthus homolog of arabidopsis *VND4* (GQ_12G158800), and the *MsMYB103* (GQ_07G174800) (Figure 6 and Table 1). In addition, they comprise important lignin genes, such as a miscanthus homolog of the maize *BRITTLE-STALK 2* locus (*COBL* gene; GQ_12G165800) (Figure 6 and Table 1). Finally, synteny analysis showed that each of these genes is syntenic to other 84 genes on average (range 26–190). Most of the synteny is, as expected, toward Poaceae (Figure 8A,B). However, some syntenic connections involve also relatively distant eudicot species, such as rapeseed or tomato, which have copies of their *WRKY12*, *VND4*, and *CCOAOMT* genes syntenic to miscanthus.



Figure 7. The genomic map of the cell wall QTLs collected for switchgrass from the work of [21] and of the switchgrass SQTLs. From the outside, the first band represents the position of the 56 switchgrass QTLs collected from [21], with the QTLs colocalizing with SQTLs colored in red. The second band represents the positions of the switchgrass SQTLs. The third band displays relevant cell wall candidate genes found in the colocalizing regions between SQTLs and QTLs.

Gene_ID	Chrom	Conserved in SQTL	Cell Wall Function	Notes	References
GQ_01G229400	Chr01	MSQTL_122	CSL	Homolog of AtCSLA9, involved in the synthesis of mannans.	[22]
GQ_01G468200	Chr01	MSQTL_122	CSL	Homolog of AtCSLA9, involved in the synthesis of mannans.	[22]
GQ_01G471400	Chr01	MSQTL_5	BLH6	Transcription factor promoting secondary cell wall synthesis in grasses	[20,23]
GQ_01G471500	Chr01	MSQTL_5	BGAL	Beta-galactosidase involved in the degradation of several polysaccharides and cell wall remodeling	[24]
GQ_01G474600	Chr01	MSQTL_122	FRA3	Homolog of AtFRA3, which coordinates actin organization during cellulose deposition	[25]
GQ_01G478800	Chr01	MSQTL_212	4CL	Important gene of the lignin synthesis pathway	[16]
GQ_03G236400	Chr03	MSQTL_216	Endoglucanase/KOR	Important genes for cellulose and plant cell wall metabolism	[26,27]
GQ_03G240500	Chr03	MSQTL_248	Endoglucanase/KOR	Important genes for cellulose and plant cell wall metabolism	[26,27]
GQ_07G169800	Chr07	MSQTL_180	CCOAOMT	Central gene for lignin synthesis, affecting both lignin amount and monolignols ratio	[16]
GQ_07G169900	Chr07	MSQTL_180	LAC	Gene involved in the in muro lignin deposition	[16]
GQ_07G174800	Chr07	MSQTL_180	MYB103	Regulates F5H expression and S-lignin deposition in arabidopsis	[28]
GQ_07G435600	Chr07	MSQTL_245	ERF39	Binds to promoters of CESA1/3/6, synthesizing primary-like cell walls	[29]

Table 1. List of cell wall genes found within the miscanthus QTLs from the GWAS and conserved through SQTLs. The IDs of SQTL are reported as in Pancaldi, Vlegels, Rijken, van Loo, and Trindade [1].

Gene_ID	Chrom	Conserved in SQTL	Cell Wall Function	Notes	References
GQ_07G435700	Chr07	MSQTL_245	ERF39	Binds to promoters of CESA1/3/6, synthesizing primary-like cell walls	[29]
GQ_07G477100	Chr07	MSQTL_445	BXL	Inhibits xylan synthesis	[30]
GQ_12G086100	Chr12	MSQTL_118	XND1/WND1A	Major transcription factor in regulating secondary cell wall synthesis	[31]
GQ_12G091100	Chr12	MSQTL_118	Endoglucanase/KOR	Important genes for cellulose and plant cell wall metabolism	[26,27]
GQ_12G150500	Chr12	MSQTL_38	PAL	First step of the lignin pathway	[16]
GQ_12G153400	Chr12	MSQTL_38	UGT72E3	Gene influencing the kinetics of lignin deposition	[32]
GQ_12G158800	Chr12	MSQTL_2	VND4	Master regulator of secondary cell walls	[20,31]
GQ_12G165800	Chr12	MSQTL_91	COBL/COBRA	Homolog of brittle stalk2 locus of maize	[33]
GQ_12G168300	Chr12	MSQTL_2	WRKY12	Involved in regulating secondary cell wall and flowering in <i>Miscanthus</i> <i>lutarioriparius</i>	[20,34]
GQ_18G102600	Chr18	MSQTL_178	CSL	Gene involved in hemicellulose biosynthesis	[17]
GQ_18G114300	Chr18	MSQTL_185	MYB20/MYB43	MYB20, MYB43, and MYB85 regulate secondary cell wall formation	[20,31,35]

Table 1. Cont.

For switchgrass, the most relevant cell wall genes found in colocalizing regions between SQTLs and QTLs from Ali, Serba, Walker, Jenkins, Schmutz, Bhamidimarri, and Saha [21] are summarized in Figure 7 and Table 2. Among others, these include a homolog of the arabidopsis *FRAGILE FIBRE 1* locus (*FRA1*; IV_6NG323500) and a homolog of the arabidopsis *KOR* gene (IV_1NG408000). Moreover, an *IRX* gene involved in xylan synthesis (IV_5NG144100), and two *MYB42/MYB85* genes important for lignin biosynthesis (IV_2NG449200 and IV_6NG352900). All these genes are retained in QTLs from Ali, Serba, Walker, Jenkins, Schmutz, Bhamidimarri, and Saha [21] that are associated to traits for which the genes themselves appear highly functionally relevant (Table 2). Moreover, synteny analysis of these genes showed that, as for miscanthus, they are highly syntenic, even if synteny is again mostly restricted to Poaceae (Figure 8C,D).



Figure 8. Analysis of the syntenic relationships of the most relevant genes from miscanthus (**A**,**B**) and switchgrass (**C**,**D**) that are conserved through SQTLs and colocalizing with the QTLs described in Section 2.4. (**A**) Absolute number of syntenic genes and of different species to which each miscanthus gene is syntenic. (**B**) Average number of syntenic relationships with species belonging to different angiosperm clades across the genes in panel A. (**C**) Absolute number of syntenic relationships with species and of different species to which each switchgrass gene is syntenic. (**D**) Average number of syntenic relationships with species belonging to different angiosperm clades across the genes is syntenic. (**D**) Average number of syntenic relationships with species belonging to different angiosperm clades across the genes in panel C.

Table 2. List of cell wall genes from the switchgrass QTLs mapped by Ali, Serba, Walker, Jenkins, Schmutz, Bhamidimarri, and Saha [21] and conserved through SQTLs. QTL traits refer to the traits associated to the QTLs from Ali, Serba, Walker, Jenkins, Schmutz, Bhamidimarri, and Saha [21] where the genes were found as retained. The SQTL IDs are reported as in Pancaldi, Vlegels, Rijken, van Loo, and Trindade [1].

Gene_ID	Chrom	QTL Trait(s) [21]	Conserved in SQTL	Cell Wall Function
IV_1KG461400	Chr_01K	Glucose; total cell wall sugar	MSQTL_195;	4CL
IV_6NG354300	Chr_06N	Total cell wall sugar	MSQTL_80;	BGAL
IV_3NG135100	Chr_03N	Glucose	MSQTL_25;	BGLU46
IV_9KG034800	Chr_09K	Xylose	MSQTL_102;	BGLU46
IV_1NG408000	Chr_01N	Total cell wall sugar	MSQTL_185;	Endoglucanase/KOR

Gene_ID	Chrom	QTL Trait(s) [21]	Conserved in SQTL	Cell Wall Function
IV_6NG323500	Chr_06N	Glucose; total cell wall sugar	MSQTL_45;	FRA1
IV_1NG306100	Chr_01N	Total cell wall sugar	MSQTL_18;	GALT/HPGT
IV_2NG448300	Chr_02N	Klason lignin	MSQTL_54;	GAUT1
IV_2KG434900	Chr_02K	Glucose	MSQTL_230;	GH17
IV_5NG130900	Chr_05N	Klason lignin	MSQTL_123;	GLCAT
IV_5NG144100	Chr_05N	Glucose; xylose	MSQTL_130;	IRX10/10L
IV_2KG435600	Chr_02K	Glucose	MSQTL_230;	LAC
IV_5KG613400	Chr_05K	Glucose	MSQTL_350;	LAC
IV_1NG407200	Chr_01N	Total cell wall sugar	MSQTL_185;	MED/REF
IV_1NG407100	Chr_01N	Total cell wall sugar	MSQTL_185;	MYB20/MYB43
IV_6NG354000	Chr_06N	Total cell wall sugar	MSQTL_80;	MYB4/MYB6/MYB7/MYB21/MYB32
IV_2NG449200	Chr_02N	Klason lignin	MSQTL_248;	MYB42/MYB85
IV_6NG352900	Chr_06N	Total cell wall sugar	MSQTL_80;	MYB42/MYB85
IV_6NG354100	Chr_06N	Total cell wall sugar	MSQTL_80;	PRX
IV_9KG292900	Chr_09N	Klason lignin	MSQTL_122;	RHM
IV_1KG460000	Chr_01K	Glucose; total cell wall sugar	MSQTL_164;	XET/XTH

Table 2. Cont.

2.5. SQTLs as Tools to Circumvent Limitations of Genetic Mapping Approaches

In the final step of this research, it was evaluated whether SQTLs, given the incorporation of information from multiple genomic loci previously shown to determine variability in a trait of interest (in this case cell wall quality), can be used to overcome limitations of genetic mapping approaches. Specifically, a common issue encountered in association mapping-and especially GWAS-is that if patterns of population structure overlap with patterns of phenotypic variability across accessions for the trait(s) evaluated, population structure correction can hide (part of) the relevant loci that govern the trait(s) [36]. The GWAS conducted in *Miscanthus sinensis* was therefore used to study if SQTLs can, at least partly, overcome this limitation. First, the analysis of population structure computed to perform the "standard" GWAS described in Section 2.4 was co-analyzed with the phenotypic data on cell wall traits to assess covariation between population structure and phenotypic traits across the miscanthus accessions of the GWAS panel (Supplementary Figures S6 and S7). Interestingly, this covariation was found for multiple traits tested, including total cell wall (NDF; dry matter percentage), total ADF (dry matter percentage), cellulose (dry matter percentage), cellulose (NDF percentage), hemicellulose (dry matter percentage), and hemicellulose (NDF percentage; Supplementary Figures S6 and S7). A GWAS without population structure correction was therefore run, and colocalization between SQTLs and the LOD peaks found by this GWAS and not already included among the 91 QTLs from the "standard" GWAS with population structure correction (Section 2.4) was assessed. These peaks would normally be discarded as false positive associations due to population structure. However, their colocalization with SQTLs indicates that in other species, QTLs for similar traits were found on these exact genomic regions, highlighting their potential relevance. As an extra control, the SNP markers included within the LOD peaks colocalizing with SQTLs and not included among the 91 QTLs of the "standard" GWAS were extracted and tested with ANOVA for significant differences in the major allele frequency (MAF) between the different population structure groups depicted in Supplementary Figure S6. Upon detection of significant MAF differences between covarying population structure and phenotypic groups, the peak regions were declared as "extra QTLs" found by colocalization with SQTLs (Figure 9).



Figure 9. The steps followed for the detection of the "extra" GWAS QTLs. (**A**) Significant variation for one trait across the four main population structure groups of the GWAS panel. (**B**) An LOD peak from the GWAS without population structure correction for the same trait as in panel A. This peak was not included in the "standard" GWAS and colocalizes with a known SQTL region (yellow bars). (**C**) Assessment of the variation in the major allele frequency across the four population structure groups such as in panel A for all the markers included within the LOD peak of panel B. MAF differs significantly across population structure groups (asterisks denote significant differences at $\alpha = 0.05$). (**D**) Identification of the "extra QTL".

In total, 17 "extra QTLs" were detected (Supplementary Table S7). Their analysis demonstrated that they contain a total of 13 cell wall genes. Among these, the most relevant ones appeared to be a miscanthus homolog of the arabidopsis *KOR* gene (GQ_05G142200) that was found within a QTL associated to cellulose dry matter content. Moreover, an *MsCSL* gene (GQ_19G142500), a homolog of the Arabidopsis and maize *CslD5*, was also found within four colocalizing "extra QTLs" associated with total cell wall dry matter content, ADF dry matter content, cellulose dry matter content, and cellulose NDF percentage. Finally, two peroxidases (*PRX*; GQ_06G157300, GQ_19G142300), and a *XYLOGLUCAN ENDO-TRANS-GLYCOSYLASE/HYDROLASE* (*XTH*; GQ_19G142300).

3. Discussion

3.1. SQTLs as Reservoirs of Allelic Variation with a Potential Use for Biomass Improvement

In a previous publication, it was shown that SQTLs are useful breeding tools to project known genetic information on the architecture of traits of interest—specifically cell wall quality traits—from model species to understudied crops [1]. This research aimed at assessing whether SQTLs can potentially guide breeding activities by spanning genomic regions displaying allelic variability for target traits and by pinpointing relevant loci associated with biomass quality variability in novel plant populations and crops. Regarding the first point, this study clearly showed that SQTL regions are reservoirs of intraspecific allelic variability. A minor part of this variability entails PAV of SQTL genes. Gene PAV is commonly found in intraspecific comparative genomic studies, such as pangenome analyses, and is a known source of trait variability [37,38]. In this sense, the SQTL cell wall genes displaying intraspecific PAV may potentially impact biomass traits in target accessions in several ways. For example, in the case that a SQTL gene is part of a multigene family, its PAV could affect overall gene copy number and lead to differential gene dosage across accessions [39,40]. Alternatively, in more extreme cases, PAV may lead to a loss-of-function, with a likely large impact on plant traits [40]. In this study, the SQTL genes showing PAV between reference and target assemblies included PAL and CAD genes, which displayed PAV among some of the Brachypodium distachyon and Brassica napus accessions assessed. PAL and CAD are multigene families, and gene dosage is thought to be important for their functionality [41–43]. Moreover, in the close relative of *Brassica napus*, Brassica rapa, intraspecific copy number and PAV of PRX genes, which together with CAD determine the efficiency of lignin production at the final steps of the lignin pathway [44], was suggested to affect lignin metabolism and related morphological traits [45]. Therefore, intraspecific PAV of SQTL genes might be relevant for determining variability in cell wall quality traits, and genomic analysis of SQTL regions in (novel) crop panels might quickly highlight promising target accessions for inclusion in breeding programs based on the assessment of gene PAV.

In addition to PAV, nucleotide polymorphisms were also found in high numbers in the intraspecific analyses of SQTL regions. Specifically, multiple genes that were previously defined as relevant SQTLs candidates [1] display intraspecific nucleotide polymorphisms that impact, sometimes considerably, the sequences of their proteins. In this sense, the examples reported for the rice BCL8 locus and its Brachypodium syntenic homolog identified through SQTLs are particularly relevant. In fact, the protein modifications reported for these genes sometimes break the integrity of the COBRA domain of the protein or affect the membrane-anchoring domains at the C-terminus. These mutations are highly similar to the ones found in the well characterized *OsBC1* and *OsBCL1* mutants, which are close homologs of the OsBCL8 gene [46,47]. Both these mutants display alterations of the plant mechanical strength and of cell wall compositional properties as effect of the COBL gene mutations [15,46,47]. Thus, the mutations found in this study might potentially lead to similar effects. Interestingly, the rice cv. "Azucena"—which is one of the rice cultivars displaying a change in protein polarity in the C-terminus of the BCL8 protein—displays differences in the relative content of cell wall components, including cellulose, compared to the reference of this study, cv. "Nipponbare" [48]. Nevertheless, future research is needed to determine a relationship between these mutations and cell wall composition. The same goes for all the other mutations discussed in Section 2.3.

To conclude, irrespective of the functional relevance of the specific patterns of genomic variability observed for the SQTL regions in the intraspecific comparisons performed, it is still highly relevant that SQTL genes were proven to show such patterns. This demonstrates that the high conservation of gene presence and order across diverse genomes does not preclude the existence of intraspecific allelic variability for those highly conserved regions. Therefore, synteny and SQTLs can be effectively used to mine interesting alleles by genomically analyzing SQTL regions in target species and searching for specific mutations previously identified as particularly relevant in model crops. If this step would be

performed during the screening of useful material for breeding programs, it would lead to faster and more effective breeding activities in novel (under-domesticated) biomass species.

3.2. SQTLs as Valid Tools toward the Improvement of Breeding Activities in Novel Crops

In the second part of this study, it was tested if SQTLs represent effective tools to predict relevant loci associated to biomass quality in novel mapping panels or crop species. The results showed that two sets of QTLs from *Miscanthus sinensis* and *Panicum virgatum*, respectively, colocalized significantly with SQTL regions of these two species as compared to random genomic loci. This result demonstrates that, in a breeding context involving novel species, SQTLs can be valuable tools to pinpoint relevant target loci that are likely responsible of variability in traits of interest. Moreover, by using the annotations of SQTL genes from the multiple species that are represented within SQTLs—as performed here for cell wall genes—it is possible to filter candidate genes based on functional and literature information. Finally, syntenic conservation of target loci through SQTLs can quickly display the functional conservation of interesting target genes through multiple (model) species for which studies might be available to better evaluate the functional relevance of cell wall genes within QTLs and SQTLs.

By applying the approach just described, in addition to the colocalization between SQTLs and miscanthus and switchgrass QTLs, several genes involved in these colocalizations were identified that represent valuable candidates for determining variability in biomass composition in these species. Specifically, some of the genes found from the miscanthus GWAS and conserved through SQTLs appear of particular interest. For example, the transcription factor *BLH6* is known to deeply affect the properties of plant cell walls, not only within different species [23,49], but also between different species clades, such as Poaceae vs. eudicots [20]. Therefore, the finding of this gene within GWAS QTLs colocalizing with SQTLs makes it a very good candidate for modulating the relative ratio of different cell wall polysaccharides and lignin in miscanthus. In addition to *BLH6*, genes such as WRKY12, VND4, and MYB103 are all central cell wall genes, demonstrating the modulation of cell wall regulation and composition across a range of species [19,20,28,50–53]. All these genes are thus important candidates for the cell wall quality variability observed in the miscanthus panel used for the GWAS described in this manuscript. Ideally, this information might be therefore used to map relevant mutations at these genes and screen plant material for those mutations. Remarkably, the intraspecific genomic comparisons performed in this research revealed that nucleotide variability leading to amino acid substitutions and protein INDELs were observed among the Brachypodium accessions. These two approaches—SQTL-guided association mapping and genomic screening/prediction of mutations at target candidate genes—could be combined in novel species to enable quicker (pre-)breeding for biomass compositional traits.

In the last part of this study, whether SQTLs can be used to make the population structure correction more efficient in GWA studies was also tested. While it needs to be clearly stated that the accounting of population structure in GWAS is pivotal to enabling accurate mapping, the results obtained here indicate that the inspection of colocalization between SQTLs and significant SNPs normally discarded as false-positives after population structure correction could represent a useful complement of a "standard" GWAS pipeline. Specifically, as they are defined, SQTLs represent genomic regions that are syntenically conserved between species and in multiple model species were shown to be associated to QTLs for a trait of interest [1]. Therefore, they could help in retrieving some of the "missing heritability" that can get blurred below the threshold of false discovery [36], in the case of the SQTL areas significantly covarying with a target trait in other (related) crops. However, the "eye" of the researcher is crucial in order to be able to evaluate the relevance of potential candidate genes found within the "extra GWAS QTLs" pinpointed by using SQTLs. In this sense, in our miscanthus panel, some of the candidate genes retained within the 17 detected "extra QTLs" of Section 2.5 appear highly relevant in the context of cell wall variability. Therefore, these genes might be considered, depending on the needs, in breeding settings

to improve miscanthus biomass quality, together with all the other ones discussed in the previous paragraph(s).

4. Materials and Methods

4.1. Intraspecific SQTL Alignment

Multiple chromosome-level genome assemblies representing different accessions of *Arabidopsis thaliana, Brassica napus, Solanum lycopersicon, Brachypodium distachyon, Oryza sativa,* and *Zea mays* were collected from online databases after the literature search (Supplementary Table S8). These six species were chosen because of the availability of multiple good-quality chromosome-level genome assemblies from either pan-genomic studies [37,54,55] or genomic databases, as well as their relevance for plant research. Moreover, they include both eudicots and grasses for which SQTLs were available from our previous study [1]. The reference assembly of each of the six species above which SQTLs were initially detected was used to extract the reference nucleotide sequences of cell wall SQTLs. These sequences were then aligned against the accessions collected by using the NUCmer package of the MUMmer software [56,57]. NUCmer was run with the following parameters: –minmatch 100 and –mincluster 200. The NUCmer show-snps command was also called (default parameters) to detect SNPs and INDELs between reference SQTL sequences and aligning regions on target accessions.

4.2. Analysis of Intraspecific SQTL Alignments

Custom R scripts were developed to process the NUCmer outputs and extract different information. First, the aligning regions and coordinates of SQTLs in every target accession; second, SQTL coverage for each alignment; third, the SQTL (cell wall) genes included in each alignment; finally, the SNPs and INDELs found along alignments. To extract these data, the custom R scripts made also use of a previously developed list of cell wall genes from 169 angiosperm genomes [1], as well as of the IRanges and GenomicRanges R packages [58]. Moreover, to quantify gene PAV in SQTLs alignments, the data on missing genes in alignment produced by NUCmer were validated by a BLAST search [59] of the genes identified as missing in specific alignments against the assemblies involved in the alignments themselves (*E*-value = 1×10^{-3}). Finally, all the statistical analyses reported in this manuscript were performed using R or SPSS v27.0 (IBM Corp., Armonk, NY, USA).

4.3. Analysis of Cell Wall Protein Sequence Changes

The assessment of the effects of SQTLs genomic variability on protein sequences was performed by using a custom R script were the coordinates of the exons of every SQTL gene included in alignments and concurring to code the main gene transcripts were used to retrieve gene CDS and translated proteins in the target genome assemblies. On the retrieved proteins, sequence changes between reference and target assemblies were assessed with ClustalW [60]. Moreover, HMMsearch [61] was used to annotate protein functional domains available on the PFAM database [62]. Furthermore, protein signal peptides, including N- and C-terminus and related functional signals were annotated by integrating the predictions provided by SignalP v6.0 [63], DeepTMHMM [64], NetGPI [65], and DeepLoc v2.0 [66]. Finally, changes in protein structure and properties due to sequence changes were assessed by using NetSurfP v3.0 [67].

4.4. Genome-Wide Association Study on Miscanthus sinensis

The GWAS on *Miscanthus sinensis* was performed by using a miscanthus collection established in 2013 in Wageningen (The Netherlands) and composed of 94 accessions originated from various international gene banks and breeding programs. Genotypes were planted in square-like plots with 16 clonal replicas, and the four central plants of each plot were harvested every spring for five years, starting in 2017.

Phenotyping was performed for eight cell wall quality traits, including NDF cell wall as percentage of dry matter, ADF cell wall as percentage of dry matter, ADL lignin as percentage of dry matter, lignin as percentage of NDF, cellulose as percentage of dry matter, cellulose as percentage of NDF, hemicellulose as percentage of dry matter, and hemicellulose as percentage of NDF (Supplementary Table S9). Phenotyping was performed by first chopping the harvested miscanthus stems into pieces of 4 cm, and by drying (60 °C, 48 h) and weighing stem pieces to determine dry matter content. The dried stems were then milled, and a subset of samples was used for training a near-infrared spectrometry (NIRS) model for cell wall composition, by performing NDF, ADF, and ADL biochemical analyses following the ANKOM Technology protocols (ANKOM Technology Corporation). The NIRS model was in turn used to phenotype the cell wall traits on the milled feedstock of all the genotypes of the collection.

The genomic DNA of all the accessions was isolated from random young leaves from the four central plants of all the plots in the collection, following a CTAB-based protocol [68]. Extracted DNA from every sample was digested by using the restriction enzyme EcoR1, ligated to unique adapters, pooled, purified, amplified, and finally sequenced using an Illumina HiSeq X10/4000 system. Sequencing was performed by BGI (Shenzen, Guangdong, China), and generated 371.52 Gb of cleaned data. Reads were aligned to the *Miscanthus sinensis* reference genome [69], resulting in the identification of ~7.0 million SNPs. SNPs were filtered for only biallelic SNPs displaying 100% call rate across the 94 miscanthus accessions and a minor allele frequency > 20%. Moreover, following chromosome-wide LD analysis with the LD.decay function from R package Sommer [70] (Supplementary Figure S6), SNPs were further filtered to keep only one marker for genomic bins corresponding to one third of the average chromosomal LD distance. This way, a set of 57,891 markers relatively evenly distributed over the 19 miscanthus chromosomes was obtained.

The filtered SNPs were used to estimate population structure by using van Raden kinship [71] and principal coordinate analysis (PCoA; ape R package—Paradis and Schliep [72]) (Supplementary Figure S7). A dendrogram of the kinship matrix was also produced (ape R package—Paradis and Schliep [72]). Moreover, patterns of population structure among accessions were compared with the ones inferred from a principal component analysis (PCA) on the phenotypic data (Supplementary Figure S8), to assess covariation between population and phenotypic accession clusters (Supplementary Figure S9).

Genome-wide associations between the filtered SNPs and the eight cell wall traits above were performed by using a linear mixed model (LMM) incorporating SNP data and the kinship matrix, as implemented in the statgenGWAS R package [73]. FDR correction (1%) was used to account for multiple testing [74], while QQ-plots of observed vs. expected p-values of associations were computed to assess the effectiveness of population structure correction (Supplementary Figure S10). GWAS analyses were performed separately for each trait. Chromosome-scale LD windows were used to define significant regions around the significant markers found, within which candidate genes were looked for, by using the set of angiosperm cell wall genes developed in our previous SQTL study [1] and the Arabidopsis- and rice-based annotations of the miscanthus genes from Phytozome. Moreover, a separate GWAS was also performed without incorporating the kinship matrix in the model, to perform the study described in Section 2.5.

4.5. Colocalization of SQTLs and Cell Wall Loci Mapped on M. sinensis

Colocalization between SQTLs and the 91 QTLs found by the miscanthus GWAS was performed by developing 100 sets of 91 random QTL regions from the miscanthus genome mirroring the size distribution of the QTLs from GWAS results (custom R script). The proportion of QTLs colocalizing for >50% of their bp length with SQTLs was then calculated for every set, and binomial tests were performed to assess if random QTLs co-localized with SQTLs significantly less than the QTLs from the GWAS (custom R script).

In addition to calculating the statistical significance of the colocalization between miscanthus SQTLs and QTLs, the cell wall genes in colocalizing regions were identified by using the set of angiosperm cell wall genes developed in our previous SQTL study [1]. Moreover, angiosperm-wide syntenic conservation of those genes was analyzed by retriev-

ing their syntenic homologs from the synteny network developed by Pancaldi, et al. [75] and Pancaldi, Vlegels, Rijken, van Loo, and Trindade [1].

4.6. Retrieval of the P. virgatum Cell Wall QTLs and Analysis of Their Colocalization with SQTLs

A total of 56 QTLs related to cell wall traits in switchgrass were retrieved from the results of Ali, Serba, Walker, Jenkins, Schmutz, Bhamidimarri, and Saha [21] (Supplementary Table S5). Colocalization between these QTLs and the switchgrass SQTLs was analyzed with an analogous procedure to miscanthus. Specifically, 100 sets of 56 random QTL regions from the switchgrass genome mirroring the size distribution of the 56 QTLs from Ali, Serba, Walker, Jenkins, Schmutz, Bhamidimarri, and Saha [21] were computed. The proportion of QTLs colocalizing for >50% of their bp length with SQTLs was then calculated for every set, and binomial tests were performed to assess presence and significance of a decrement in such proportion (custom R script).

As performed in miscanthus, the cell wall genes in colocalizing regions were identified by using the set of angiosperm cell wall genes developed in our previous SQTL study [1]. Moreover, angiosperm-wide syntenic conservation of those genes was analyzed by retrieving their syntenic homologs from the synteny network developed by Pancaldi, van Loo, Schranz, and Trindade [75] and Pancaldi, Vlegels, Rijken, van Loo, and Trindade [1].

5. Conclusions

This study aimed at assessing the validity of SQTLs to assist breeding of (novel) biomass crops, by (i) analyzing the intraspecific level of SQTLs allelic genomic variability in panels of available good-quality plant genomes, (ii) assessing the SQTLs predictive value of relevant biomass-quality-related loci in miscanthus and switchgrass, and (iii) suggesting possible side-uses of SQTLs to complement standard approaches of genetic mapping. The results showed that SQTLs are valuable tools to improve breeding activities in novel crops, where they can be applied in different ways to either screen plant material for genes or alleles of interest, or to pinpoint relevant loci based on the information from model species in novel accession panels or crops. Moreover, the research performed to achieve the goals above allowed the finding of relevant alleles at candidate SQTL loci, as well as of important candidate cell wall genes for miscanthus biomass improvement. Future research could investigate the phenotypic relevance hypothesized for the intraspecific patterns of SQTLs variability revealed in this study, by phenotyping cell wall composition in the plant lines showing promising variability. Moreover, the miscanthus loci highlighted in this study could be targeted by genetic modification to study their specific relevance for miscanthus biomass quality. Finally, SQTLs might start to be included in breeding programs of underdomesticated biomass crops as proposed in this research, to test their usefulness in novel, important, breeding settings.

Supplementary Materials: The following supporting information can be downloaded at https: //www.mdpi.com/article/10.3390/plants12040779/s1, Figure S1: Relative distribution of maize SQTLs (red regions) and maize transposons (as annotated by MaizeGDBonto the B73 genome v4.0) along the chromosomes of the maize B73 genome (v4.0); Figure S2: Amino acid changes and their effects on the protein structure of OsBCL1between the reference rice cv. "Nipponbare" and the two cultivars "Azucena" and "CarolinaGoldSelect". Coloring of amino acids reflects amino acid polarity, and protein domains and signal peptides are annotated. Amino acid changes indicated with * indicate a change in polarity, while sites annotated with ω indicate the predicted GPI-anchoring-related omega sites; Figure S3: Amino acid substitutions and INDELs of ZmIRX9between the reference maize B73 genome and 12 maize accessions. Coloring of amino acids reflects amino acid polarity, and the GT43 protein domain is annotated. Amino acid changes indicated with * indicate a change in polarity/charge in some of the accessions compared to the reference genome; Figure S4: Amino acid substitutions and INDELs of BdCESA7between the reference Brachypodium genome and the BdTR9 assembly. (A) Multiple protein sequence alignment displaying substitutions and INDELs. Coloring of amino acids reflects amino acid polarity, and the Cellulose synthase protein domain is annotated. (B) Predicted 2D protein structure of the reference BdCESA7 protein. (C) Predicted

2D protein structure of the BdCESA7 protein from line BdTR9; Figure S5: Amino acid substitutions and INDELs of AtC3H14, BdWRKY12, and OsNAC43between the reference genome assemblies and multiple target genome assemblies displaying polymorphisms for the proteins coded by these genes. Coloring of amino acids reflects amino acid polarity, and the functional protein domains are annotated. (A) AtC3H14; (B) BdWRKY12; (C) OsNAC43; Figure S6: Linkage disequilibrium decay plots of the first four Miscanthus sinensis chromosomes; Figure S7: PCoAof the kinship matrix of all the Miscanthus GWAS accessions, representing the population structure within the Miscanthus collection. Four main population structure groups emerge (different colors). The definition of the groups was based also on hierarchical clustering of the accessions based on SNP data; Figure S8: PCA of Miscanthus GWAS accessions based on phenotypic data. Accession points are colored based on the four identified population structure groups; Figure S9: Phenotypic variability per trait across the four population structure groups. Colors reflect the population structure groups defined as in Supplementary Figure S7; Figure S10: QQ-plots of expected vs. observed p-values of SNP associations after correction for population structure; Table S1: Missing SQTL genes across all the target assemblies analysed; Table S2: All the data on SNPs and INDELs found on SQTL genes; Table S3: The QTL intervals identified through the GWAS in miscanthus; Table S4: The Miscanthus SQTLs intervals; Table S5: The Switchgrass cell wall QTLs identified by Ali et al., 2020; Table S6: The Switchgrass SQTLs intervals; Table S7: The 17 "extra QTLs" mapped in the GWAS run without population structure correction; Table S8: The genome assemblies used in the comparative genomics study; Table S9: Full phenotypic data used as input for the Miscanthus GWAS; Supplementary Datasets S1–S3 are available at: https://data.4tu.nl/ (accessed on 1 February 2023), under doi: 10.4121/21896757.

Author Contributions: Conceptualization, F.P. and L.M.T.; Methodology, F.P., L.M.T., E.N.v.L., M.E.S., M.A.H., K.v.d.C., M.-J.P. and O.D.; Formal analysis, F.P., S.S., M.A.H. and K.v.d.C.; Writing—original draft, F.P.; Writing—review and editing, F.P., E.N.v.L., S.S., M.A.H., K.v.d.C., M.-J.P., O.D., M.E.S. and L.M.T.; Supervision, E.N.v.L., M.E.S. and L.M.T.; Project administration, L.M.T. and F.P.; Funding acquisition, L.M.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research is part of a project that received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 727698.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Pancaldi, F.; Vlegels, D.; Rijken, H.; van Loo, E.N.; Trindade, L.M. Detection and Analysis of Syntenic Quantitative Trait Loci Controlling Cell Wall Quality in Angiosperms. *Front. Plant Sci.* 2022, 13, 855093. [CrossRef] [PubMed]
- Dewey, C.N. Positional orthology: Putting genomic evolutionary relationships into context. *Briefings Bioinform.* 2011, 12, 401–412. [CrossRef] [PubMed]
- Zhao, T.; Schranz, M.E. Network approaches for plant phylogenomic synteny analysis. *Curr. Opin. Plant Biol.* 2017, 36, 129–134. [CrossRef] [PubMed]
- Kerstens, M.H.; Schranz, M.E.; Bouwmeester, K. Phylogenomic analysis of the APETALA2 transcription factor subfamily across angiosperms reveals both deep conservation and lineage-specific patterns. *Plant J.* 2020, 103, 1516–1524. [CrossRef]
- van der Weijde, T.; Alvim Kamei, C.L.; Torres, A.F.; Evermerris, W.; Edolstra, O.; Visser, R.G.F.; Trindade, L.M. The potential of C4 grasses for cellulosic biofuel production. *Front. Plant Sci.* 2013, 4, 107. [CrossRef]
- Pancaldi, F.; Trindade, L.M. Marginal Lands to Grow Novel Bio-Based Crops: A Plant Breeding Perspective. Front. Plant Sci. 2020, 11, 227. [CrossRef]
- 7. Tadele, Z. Orphan crops: Their importance and the urgency of improvement. *Planta* 2019, 250, 677–694. [CrossRef]
- Salentijn, E.M.; Pereira, A.; Angenent, G.C.; van der Linden, C.G.; Krens, F.; Smulders, M.J.; Vosman, B. Plant translational genomics: From model species to crops. *Mol. Breed.* 2006, 20, 1–13. [CrossRef]
- 9. Armstead, I.; Huang, L.; Ravagnani, A.; Robson, P.; Ougham, H. Bioinformatics in the orphan crops. *Briefings Bioinform*. 2009, 10, 645–653. [CrossRef]
- 10. Kamei, C.L.A.; Severing, E.I.; Dechesne, A.; Furrer, H.; Dolstra, O.; Trindade, L.M. Orphan Crops Browser: A bridge between model and orphan crops. *Mol. Breed.* **2016**, *36*, 9. [CrossRef]
- 11. Trindade, L.; Dolstra, O.; van Loo, E.R.; Visser, R. Plant Breeding and Its Role in A Biobased Economy. In *The Biobased Economy*; Routledge: London, UK, 2010; pp. 90–105.
- 12. Mehmood, M.A.; Ibrahim, M.; Rashid, U.; Nawaz, M.; Ali, S.; Hussain, A.; Gull, M. Biomass production for bioenergy using marginal lands. *Sustain. Prod. Consum.* **2016**, *9*, 3–21. [CrossRef]

- Clifton-Brown, J.; Harfouche, A.; Casler, M.D.; Dylan Jones, H.; Macalpine, W.J.; Murphy-Bokern, D.; Smart, L.B.; Adler, A.; Ashman, C.; Awty-Carroll, D. Breeding progress and preparedness for mass-scale deployment of perennial lignocellulosic biomass crops switchgrass, miscanthus, willow and poplar. *GCB Bioenergy* 2019, *11*, 118–151. [CrossRef]
- 14. van der Cruijsen, K.; Al Hassan, M.; van Erven, G.; Dolstra, O.; Trindade, L.M. Breeding Targets to Improve Biomass Quality in Miscanthus. *Molecules* **2021**, *26*, 254. [CrossRef]
- 15. Zhang, B.; Zhou, Y. Rice brittleness mutants: A way to open the 'Black Box'of monocot cell wall biosynthesis free access. J. Integr. Plant Biol. 2011, 53, 136–142. [CrossRef]
- 16. Zhong, R.; Cui, D.; Ye, Z.H. Secondary cell wall biosynthesis. New Phytol. 2019, 221, 1703–1723. [CrossRef]
- Little, A.; Schwerdt, J.G.; Shirley, N.J.; Khor, S.F.; Neumann, K.; O'Donovan, L.A.; Lahnstein, J.; Collins, H.M.; Henderson, M.; Fincher, G.B.; et al. Revised Phylogeny of the *Cellulose Synthase* Gene Superfamily: Insights into Cell Wall Evolution. *Plant Physiol.* 2018, 177, 1124–1141. [CrossRef]
- Brown, D.; Wightman, R.; Zhang, Z.; Gomez, L.D.; Atanassov, I.; Bukowski, J.P.; Tryfona, T.; McQueen-Mason, S.J.; Dupree, P.; Turner, S. Arabidopsis genes IRREGULAR XYLEM (IRX15) and IRX15L encode DUF579-containing proteins that are essential for normal xylan deposition in the secondary cell wall. *Plant J.* 2011, *66*, 401–413. [CrossRef]
- Zhong, R.; Lee, C.; Zhou, J.; McCarthy, R.L.; Ye, Z.-H. A Battery of Transcription Factors Involved in the Regulation of Secondary Cell Wall Biosynthesis in *Arabidopsis. Plant Cell* 2008, 20, 2763–2782. [CrossRef]
- Rao, X.; Dixon, R.A. Current Models for Transcriptional Regulation of Secondary Cell Wall Biosynthesis in Grasses. *Front. Plant Sci.* 2018, *9*, 399. [CrossRef]
- Ali, S.; Serba, D.D.; Walker, D.; Jenkins, J.; Schmutz, J.; Bhamidimarri, S.; Saha, M.C. Genome-wide quantitative trait loci detection for biofuel traits in switchgrass (*Panicum virgatum* L.). *GCB Bioenergy* 2020, 12, 923–940. [CrossRef]
- 22. Liepman, A.H.; Wilkerson, C.G.; Keegstra, K. Expression of cellulose synthase-like (*Csl*) genes in insect cells reveals that *CslA* family members encode mannan synthases. *Proc. Natl. Acad. Sci. USA* 2005, *102*, 2221–2226. [CrossRef] [PubMed]
- Hirano, K.; Kondo, M.; Aya, K.; Miyao, A.; Sato, Y.; Antonio, B.A.; Namiki, N.; Nagamura, Y.; Matsuoka, M. Identification of Transcription Factors Involved in Rice Secondary Cell Wall Formation. *Plant Cell Physiol.* 2013, 54, 1791–1802. [CrossRef] [PubMed]
- Chandrasekar, B.; van der Hoorn, R.A. Beta galactosidases in Arabidopsis and tomato—A mini review. *Biochem. Soc. Trans.* 2016, 44, 150–158. [CrossRef] [PubMed]
- 25. Zhong, R.; Burk, D.H.; Morrison, W.H., III; Ye, Z.-H. FRAGILE FIBER3, an Arabidopsis gene encoding a type II inositol polyphosphate 5-phosphatase, is required for secondary wall synthesis and actin organization in fiber cells. *Plant Cell* **2004**, *16*, 3242–3259. [CrossRef] [PubMed]
- Zuo, J.; Niu, Q.-W.; Nishizawa, N.; Wu, Y.; Kost, B.; Chua, N.-H. KORRIGAN, an Arabidopsis endo-1, 4-β-glucanase, localizes to the cell plate by polarized targeting and is essential for cytokinesis. *Plant Cell* 2000, *12*, 1137–1152. [PubMed]
- Glass, M.; Barkwill, S.; Unda, F.; Mansfield, S.D. Endo-β-1,4-glucanases impact plant cell wall development by influencing cellulose crystallization. *J. Integr. Plant Biol.* 2015, 57, 396–410. [CrossRef]
- Öhman, D.; Demedts, B.; Kumar, M.; Gerber, L.; Gorzsás, A.; Goeminne, G.; Hedenström, M.; Ellis, B.; Boerjan, W.; Sundberg, B. MYB 103 is required for FERULATE-5-HYDROXYLASE expression and syringyl lignin biosynthesis in A rabidopsis stems. *Plant J.* 2013, 73, 63–76. [CrossRef]
- Saelim, L.; Akiyoshi, N.; Tan, T.T.; Ihara, A.; Yamaguchi, M.; Hirano, K.; Matsuoka, M.; Demura, T.; Ohtani, M. Arabidopsis Group IIId ERF proteins positively regulate primary cell wall-type CESA genes. J. Plant Res. 2018, 132, 117–129. [CrossRef]
- Goujon, T.; Minic, Z.; El Amrani, A.; Lerouxel, O.; Aletti, E.; Lapierre, C.; Joseleau, J.P.; Jouanin, L. AtBXL1, a novel higher plant (*Arabidopsis thaliana*) putative beta-xylosidase gene, is involved in secondary cell wall metabolism and plant development. *Plant J.* 2003, 33, 677–690. [CrossRef]
- Zhong, R.; Ye, Z.-H. Secondary Cell Walls: Biosynthesis, Patterned Deposition and Transcriptional Regulation. *Plant Cell Physiol.* 2014, 56, 195–214. [CrossRef]
- Baldacci-Cresp, F.; Le Roy, J.; Huss, B.; Lion, C.; Créach, A.; Spriet, C.; Duponchel, L.; Biot, C.; Baucher, M.; Hawkins, S.; et al. UDP-GLYCOSYLTRANSFERASE 72E3 Plays a Role in Lignification of Secondary Cell Walls in Arabidopsis. Int. J. Mol. Sci. 2020, 21, 6094. [CrossRef]
- Sindhu, A.; Langewisch, T.; Olek, A.; Multani, D.S.; McCann, M.C.; Vermerris, W.; Carpita, N.C.; Johal, G. Maize *Brittle stalk2* Encodes a COBRA-Like Protein Expressed in Early Organ Development But Required for Tissue Flexibility at Maturity. *Plant Physiol.* 2007, 145, 1444–1459. [CrossRef]
- 34. Yu, Y.; Hu, R.; Wang, H.; Cao, Y.; He, G.; Fu, C.; Zhou, G. MIWRKY12, a novel Miscanthus transcription factor, participates in pith secondary cell wall formation and promotes flowering. *Plant Sci.* **2013**, *212*, 1–9. [CrossRef]
- Geng, P.; Zhang, S.; Liu, J.; Zhao, C.; Wu, J.; Cao, Y.; Fu, C.; Han, X.; He, H.; Zhao, Q. MYB20, MYB42, MYB43, and MYB85 Regulate Phenylalanine and Lignin Biosynthesis during Secondary Cell Wall Formation. *Plant Physiol.* 2019, 182, 1272–1283. [CrossRef]
- 36. Brachi, B.; Morris, G.P.; Borevitz, J.O. Genome-wide association studies in plants: The missing heritability is in the field. *Genome Biol.* **2011**, *12*, 232. [CrossRef]

- Gordon, S.P.; Contreras-Moreira, B.; Woods, D.P.; Des Marais, D.L.; Burgess, D.; Shu, S.; Stritt, C.; Roulin, A.C.; Schackwitz, W.; Tyler, L.; et al. Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure. *Nat. Commun.* 2017, *8*, 1–13. [CrossRef]
- 38. Song, J.-M.; Guan, Z.; Hu, J.; Guo, C.; Yang, Z.; Wang, S.; Liu, D.; Wang, B.; Lu, S.; Zhou, R.; et al. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants* **2020**, *6*, 34–45. [CrossRef]
- Kondrashov, F.A. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. R. Soc. B Boil. Sci.* 2012, 279, 5048–5057. [CrossRef]
- Gabur, I.; Chawla, H.S.; Snowdon, R.J.; Parkin, I.A. Connecting genome structural variation with complex traits in crop plants. *Theor. Appl. Genet.* 2018, 132, 733–750. [CrossRef]
- 41. Wagner, A.; Donaldson, L.; Ralph, J. Lignification and lignin manipulations in conifers. In *Advances in Botanical Research*; Elsevier: Amsterdam, The Netherlands, 2012; Volume 61, pp. 37–76.
- 42. Preisner, M.; Kulma, A.; Zebrowski, J.; Dymińska, L.; Hanuza, J.; Arendt, M.; Starzycki, M.; Szopa, J. Manipulating cinnamyl alcohol dehydrogenase (CAD) expression in flax affects fibre composition and properties. *BMC Plant Biol.* 2014, 14, 50. [CrossRef]
- Lu, J.; Shi, Y.; Li, W.; Chen, S.; Wang, Y.; He, X.; Yin, X. RcPAL, a key gene in lignin biosynthesis in *Ricinus communis* L. BMC Plant Biol. 2019, 19, 181. [CrossRef] [PubMed]
- Liu, Q.; Luo, L.; Zheng, L. Lignins: Biosynthesis and Biological Functions in Plants. Int. J. Mol. Sci. 2018, 19, 335. [CrossRef] [PubMed]
- Lin, K.; Zhang, N.; Severing, E.I.; Nijveen, H.; Cheng, F.; Visser, R.G.; Wang, X.; de Ridder, D.; Bonnema, G. Beyond genomic variation-comparison and functional annotation of three Brassica rapagenomes: A turnip, a rapid cycling and a Chinese cabbage. BMC Genom. 2014, 15, 1–17. [CrossRef] [PubMed]
- 46. Li, Y.; Qian, Q.; Zhou, Y.; Yan, M.; Sun, L.; Zhang, M.; Fu, Z.; Wang, Y.; Han, B.; Pang, X. *BRITTLE CULM1*, Which Encodes a COBRA-Like Protein, Affects the Mechanical Properties of Rice Plants. *Plant Cell* **2003**, *15*, 2020–2031. [CrossRef]
- Dai, X.; You, C.; Wang, L.; Chen, G.; Zhang, Q.; Wu, C. Molecular characterization, expression pattern, and function analysis of the OsBC1L family in rice. *Plant Mol. Biol.* 2009, 71, 469–481. [CrossRef]
- Jahn, C.E.; Mckay, J.K.; Mauleon, R.; Stephens, J.; McNally, K.L.; Bush, D.R.; Leung, H.; Leach, J.E. Genetic Variation in Biomass Traits among 20 Diverse Rice Varieties. *Plant Physiol.* 2010, 155, 157–168. [CrossRef]
- 49. Liu, Y.; You, S.; Taylor-Teeples, M.; Li, W.L.; Schuetz, M.; Brady, S.M.; Douglas, C.J. BEL1-LIKE HOMEODOMAIN6 and KNOTTED ARABIDOPSIS THALIANA7 interact and regulate secondary cell wall formation via repression of *REVOLUTA*. *Plant Cell* **2014**, *26*, 4843–4861. [CrossRef]
- Wang, H.; Avci, U.; Nakashima, J.; Hahn, M.G.; Chen, F.; Dixon, R.A. Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. *Proc. Natl. Acad. Sci. USA* 2010, 107, 22338–22343. [CrossRef]
- 51. Zhou, J.; Zhong, R.; Ye, Z.-H. Arabidopsis NAC domain proteins, VND1 to VND5, are transcriptional regulators of secondary wall biosynthesis in vessels. *PLoS ONE* **2014**, *9*, e105726. [CrossRef]
- 52. Yang, L.; Zhao, X.; Yang, F.; Fan, D.; Jiang, Y.; Luo, K. PtrWRKY19, a novel WRKY transcription factor, contributes to the regulation of pith secondary wall formation in Populus trichocarpa. *Sci. Rep.* **2016**, *6*, 18643. [CrossRef]
- Wu, L.; Zhang, M.; Zhang, R.; Yu, H.; Wang, H.; Li, J.; Wang, Y.; Hu, Z.; Wang, Y.; Luo, Z. Down-regulation of OsMYB103L distinctively alters beta-1, 4-glucan polymerization and cellulose microfibers assembly for enhanced biomass enzymatic saccharification in rice. *Biotechnol. Biofuels* 2021, 14, 1–15. [CrossRef]
- Hurgobin, B.; Golicz, A.A.; Bayer, P.E.; Chan, C.K.K.; Tirnaz, S.; Dolatabadian, A.; Schiessl, S.V.; Samans, B.; Montenegro, J.D.; Parkin, I.A. Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid Brassica napus. *Plant Biotechnol. J.* 2017, 16, 1265–1274. [CrossRef]
- 55. Jiao, W.-B.; Schneeberger, K. Chromosome-level assemblies of multiple Arabidopsis genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nat. Commun.* **2020**, *11*, 1–10. [CrossRef]
- 56. Delcher, A.L.; Phillippy, A.; Carlton, J.; Salzberg, S.L. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* 2002, *30*, 2478–2483. [CrossRef]
- Kurtz, S.; Phillippy, A.; Delcher, A.L.; Smoot, M.; Shumway, M.; Antonescu, C.; Salzberg, S.L. Versatile and open software for comparing large genomes. *Genome Biol.* 2004, 5, R12. [CrossRef]
- 58. Lawrence, M.; Huber, W.; Pages, H.; Aboyoun, P.; Carlson, M.; Gentleman, R.; Morgan, M.T.; Carey, V.J. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **2013**, *9*, e1003118. [CrossRef]
- Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. J. Mol. Biol. 1990, 215, 403–410. [CrossRef]
- 60. Chenna, R.; Sugawara, H.; Koike, T.; Lopez, R.; Gibson, T.J.; Higgins, D.G.; Thompson, J.D. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **2003**, *31*, 3497–3500. [CrossRef]
- 61. Eddy, S.R. Accelerated profile HMM searches. PLoS Comput. Biol. 2011, 7, e1002195. [CrossRef]
- 62. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [CrossRef]

- Teufel, F.; Almagro Armenteros, J.J.; Johansen, A.R.; Gíslason, M.H.; Pihl, S.I.; Tsirigos, K.D.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* 2022, 40, 1023–1025. [CrossRef] [PubMed]
- 64. Hallgren, J.; Tsirigos, K.D.; Pedersen, M.D.; Armenteros, J.J.A.; Marcatili, P.; Nielsen, H.; Krogh, A.; Winther, O. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv* **2022**. [CrossRef]
- 65. Gíslason, M.H.; Nielsen, H.; Armenteros, J.J.A.; Johansen, A.R. Prediction of GPI-anchored proteins with pointer neural networks. *Curr. Res. Biotechnol.* **2021**, *3*, 6–13. [CrossRef]
- 66. Thumuluri, V.; Almagro Armenteros, J.J.; Johansen, A.R.; Nielsen, H.; Winther, O. DeepLoc 2.0: Multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res.* **2022**, *50*, W228–W234. [CrossRef] [PubMed]
- Høie, M.H.; Kiehl, E.N.; Petersen, B.; Nielsen, M.; Winther, O.; Nielsen, H.; Hallgren, J.; Marcatili, P. NetSurfP-3.0: Accurate and fast prediction of protein structural features by protein language models and deep learning. *Nucleic Acids Res.* 2022, 50, W510–W515. [CrossRef]
- 68. Tai, T.H.; Tanksley, S.D. A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue. *Plant Mol. Biol. Rep.* **1990**, *8*, 297–303. [CrossRef]
- 69. Mitros, T.; Session, A.M.; James, B.T.; Wu, G.A.; Belaffif, M.B.; Clark, L.V.; Shu, S.; Dong, H.; Barling, A.; Holmes, J.R. Genome biology of the paleotetraploid perennial biomass crop Miscanthus. *Nat. Commun.* **2020**, *11*, 1–11. [CrossRef]
- 70. Covarrubias-Pazaran, G. Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS ONE* **2016**, *11*, e0156744. [CrossRef]
- 71. VanRaden, P.M. Efficient methods to compute genomic predictions. J. Dairy Sci. 2008, 91, 4414–4423. [CrossRef]
- 72. Paradis, E.; Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **2018**, *35*, 526–528. [CrossRef]
- 73. van Rossum, B.-J.; Kruijer, W.; van Eeuwijk, F.; Boer, M.; Malosetti, M.; Bustos-Korts, D.; Millet, E.; Paulo, J.; Verouden, M.; Wehrens, R. Package 'statgenGWAS'. *R Package, version* 1.0.7, 2020.
- 74. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300. [CrossRef]
- 75. Pancaldi, F.; van Loo, E.N.; Schranz, M.E.; Trindade, L.M. Genomic Architecture and Evolution of the *Cellulose synthase* Gene Superfamily as Revealed by Phylogenomic Analysis. *Front. Plant Sci.* **2022**, *13*, 870818. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.