Received: 7 July 2022 | Revised: 27 September 2022 | Accepted: 29 September 2022

DOI: 10.1111/gcbb.13006

GCB-BIOENERGY BIOPRODUCTS FOR A SUSTAINABLE BIOECONOMY Open Access

WILEY

RESEARCH ARTICLE

Investigating the potential of Andean lupin as a lignocellulosic feedstock for Europe: First genome-wide association study on *Lupinus mutabilis* biomass quality

Agata Gulisano[1] | **Annemarie Dechesne**[1] | **Maria-João Paulo**[2] | **Luisa M. Trindade**[1]

[1]Wageningen University and Research Plant Breeding, Wageningen University, Wageningen, The Netherlands

[2]Wageningen University and Research Biometris, Wageningen Research, Wageningen, The Netherlands

Correspondence
Luisa M. Trindade, Wageningen University and Research Plant Breeding, Wageningen University, Wageningen, The Netherlands.
Email: luisa.trindade@wur.nl

Funding information
European Union's Horizon 2020, Grant/Award Number: 720726

Abstract

The development of multipurpose crops will drive the transformation of agricultural "waste" into added-value products, helping to meet biomass demands without competing with food production or increasing environmental pressure. *Lupinus mutabilis*, has been proposed not only as a valid source of protein and oil for Europe but also as a possible source of lignocellulosic feedstock for the biorefinery industry. In this study, the quality of *L. mutabilis* lignocellulosic biomass and its genetic architecture are investigated for the first time, using a panel of 223 accessions planted across three locations in two different European cropping conditions. Biomass quality was evaluated based on the estimation of neutral detergent fiber, cellulose, hemicellulose and acid detergent lignin fractions, and on the basis of the monosaccharide composition of cell wall polysaccharides. The broad variation in yield and composition of biomass encountered in the panel confirms the potential of *L. mutabilis* as lignocellulosic feedstock and points out the value of this panel as a breeding tool for the improvement of biomass quality. A genome-wide association study was conducted to identify single-nucleotide polymorphisms (SNPs) associated with biomass quality, both across locations and per specific location. Scanning of 16,781 SNPs across the whole genome identified 46 unique quantitative trait loci for biomass quality, 4 of which were detected as common either among traits or GWAS models. For each of the traits analyzed, between 3 and 10 SNPs were detected, explaining 2.7%–15.9% of the phenotypic variation. Underlying these loci, 28 genes were proposed as candidate genes for biomass quality. Important genes involved in cellulose and sucrose synthesis (*CESA4, SPP1,WRKY33, GONST2*), monolignol biosynthesis (*SKIP31, WAT1, CCR-SNL6*) and pectin degradation (*RAV1, PE*) were identified and will require validation to confirm their value for application in *L. mutabilis* breeding.

KEYWORDS

agricultural residues, biomass quality, CESA4, GWAS, *Lupinus mutabilis*, molecular markers, SNP

1 | INTRODUCTION

The need to lower the dependency on non-renewable energy sources has become a global priority. Replacing petrochemical feedstock with sustainable biomass production for bioenergy and other biobased products represents a key step in this process. Bioenergy cropping as it has been done in the past years has led to monocultures of major crops exclusively for biofuel production (e.g. maize or sugarcane), attracting skepticism as energy cropping becomes a direct competitor with food production. However, the use of other biomass sources such as agricultural residues, discarded as "lignocellulosic waste" after harvest, can represent a more sustainable alternative. Due to their high availability, agricultural residues represent an important renewable source of cellulose and fermentable sugars. Moreover, these resources have low energetic and economical costs, are recyclable, have a low carbon footprint, and can contribute to simplified waste disposal and added-value production for farmers (Garcia et al., 2016; Pennells et al., 2021).

Different sources of lignocellulose vary in chemical composition, cell wall structure, and degree of recalcitrance to deconstruction, which in turn make biomass more or less suitable for the production of different biobased products. In general, lignocellulosic biomass consists of 35%–55% cellulose, 25%–40% hemicellulose, and 15%–25% lignin (Kumar et al., 2009), which are the main fractions of plant cell walls in monocots. Biomass quality is defined by the relative content of molecules of interest and by the recalcitrance of plant cell wall to deconstruction, which is key in the extraction process of target molecules (Pancaldi & Trindade, 2020). Modern breeding technologies can be used to optimize crops to suit the industrial needs and achieve efficient exploitation of agricultural residues, for example, by selecting for favorable alterations in cell wall properties.

In the past decades, important developments in research have paved the way for the breeding of plants for biofuel and other biobased products. The advancement in molecular marker technology and the genetic analysis of natural variation has made it possible to identify genomic regions that are associated with improved biomass quality and to deploy these regions through breeding. Several studies have reported the identification of quantitative trait loci (QTL) linked to biomass quality for major feedstock crops such as maize (Li, Wang, et al., 2016; López-Malvar et al., 2021; Méchin et al., 2001; Torres et al., 2013) or miscanthus (Gifford et al., 2015; van der Weijde et al., 2017) but also for agricultural residues such as rice and wheat straw (Malik et al., 2019; Nguyen et al., 2020). These QTL reflects mainly natural variation in lignin, cellulose, and hemicellulose content and have aided the selection of genotypes with improved digestibility and/or content and functionality of specific cell wall components. Furthermore, the increasing use of genome-wide association studies (GWAS) as an approach to dissecting the genetic architecture of complex traits, has also contributed to pinpointing relevant genes for the improvement of biomass quality. This approach, in combination with the unravelling of cell wall structure and synthesis in model species, has resulted in a list of homolog genes that could be mined or manipulated to improve biomass quality through pathway engineering. Gene families involved in the synthesis of cellulose, hemicellulose, and lignin along with genes affecting cellulose crystallinity, hemicellulose substitutions and monolignol composition in lignin polymers have been targeted as they are the most relevant for biomass quality (Pancaldi & Trindade, 2020).

The development of tailored multipurpose crops for both food and non-food applications, can enable the profitable processing of agricultural "waste" into added-value products. In doing so, multipurpose crops can become crucial in meeting biomass demands without competing for food production or increasing environmental pressure (particularly on soil and land). Examples of comprehensive use of lignocellulosic residues are the use of tomato leaves, stems and juice by-products for the realization of antibacterial fiber-based packaging (Tiekstra, 2014), the preparation of composites based on pea hulls and starch (Chen et al., 2009) or the use of rice and wheat straw for the production of biofuels and other biobased chemicals (Garcia et al., 2016; Schnitzer et al., 2014; Tian et al., 2018).

Within the frame of Horizon 2020, the project LIBBIO aimed at investigating the possibility of growing *Lupinus mutabilis* as a multipurpose crop on European marginal land and the use of the non-edible parts as a source of lignocellulosic feedstock for the biorefinery industry. *L. mutabilis* is originally an Andean legume, but its introduction to Europe could have many advantages as it represents a potentially superior alternative to the current sources of protein and oil. Its high-quality grains are characterized by a protein and oil content similar to that of soybean and its ability to fix nitrogen and to adapt to low input farming in temperate conditions make this crop suitable for cultivation in marginal lands, also at higher latitudes (Gulisano et al., 2019). Given the similar characteristics of Andean lupin and soy, there is evidence that *L. mutabilis* protein isolates can be as much valuable to the food and chemical industry as soy isolates (Carvajal-Larenas et al., 2016). An advantage relatively to soy is that *L. mutabilis* can fit into sustainable cropping rotation schemes, requiring less water and pesticides, and adapting more easily to poor soils and higher latitudes (Gulisano et al., 2019). In rotation with cereal crops, the ability to actively enrich the soil through nitrogen fixation

and phosphates mobilization can reduce the need for nitrogenous fertilizers, provide valuable disease breaks and boost cereal yields (Gladstones, 1970). The recent evaluation of over 200 accessions of *L. mutabilis* across Europe has revealed a high biomass production in Central-North european conditions, particularly for indeterminate types of Andean lupin (Gulisano et al., 2022). Uses of Andean lupin as green-fodder or silage have been investigated also in combination with other crops such as pea (Mikić et al., 2013) or maize (Austria, LIBBIO), but they are discouraged due to the presence of bitter alkaloids. Clearly, debittered seeds are more profitable for food uses or oil extraction; therefore, a separate utilization of agricultural residues as a biomass source for the biorefinery industry is desirable. To our knowledge, the use of lupin agricultural residues for the biorefinery industry has been investigated so far only in two minor species, *Lupinus nootkatensis* and *Lupinus rotundiflorus*. Acceptable yields of fermentable sugars have been reported in both cases, confirming the potential of lupin for biogas and bio-methane production (Kamm et al., 2006; Radillo et al., 2011).

The aim of this work is to analyze the biomass quality of *L. mutabilis* and to identify QTL and candidate genes associated with biomass quality. The natural variation present in a panel of 223 *L. mutabilis* accessions was exploited to characterize the variation in biomass yield and composition across two different cropping conditions in Europe.

2 | MATERIALS AND METHODS

2.1 | Plant materials

A panel of 223 *L. mutabilis* accessions was used in this study. The panel included mostly landraces, varieties, and wild material from the center of origin of the species, the Andean region, provided by the Instituto Nacional de Investigaciones Agropecuarias of Quito (INIAP, Ecuador). Additionally, some breeding material was provided by European breeding institutes, namely the Instituto Superior de Agronomia (ISA) and the Julius Kühn-Institut (JKI). Overall, the panel included material originating from Ecuador (95 accessions), Peru' (65), Bolivia (15), Chile (1), Belarus (7) and of unknown origin (18) and lines resulting from the ISA (18) and JKI (4) breeding programs. The panel was tested in three field trials: in Portugal as a winter crop between November and May in 2019, in The Netherlands as a summer crop between April and October in 2019 and 2020. Each location had a randomized complete block design with three replicates per accession. The experimental units were plots including 20 plants (4 rows with 5 plants per row), at a distance of 30 × 30 cm. Phenotyping of morphological and phenological parameters was conducted on the six central plants of each plot, to gather data on biomass yield (fresh weight), dry matter (DM) content and yield of agricultural residues (total dry biomass yield—dry seed yield). Detailed information about field trial design and data collection are described in detail in Gulisano et al. (2022).

2.2 | Biomass quality analysis

Mature stem tissues were collected per single plot, by harvesting the central part of the main stem (~20 cm) of the six plants phenotyped per plot. For each accession, three replicates were obtained per location. Stem tissues were successively dried at 50°C until constant weight and ground to 1 mm. A total of 1941 samples were processed and scanned using a Foss DS2500 near-infrared spectrometer (FOSS, Co. LLC) to obtain representative Near-InfraRed spectra for each of the samples under analysis. High-throughput phenotyping of biomass quality was achieved by estimation of cell wall traits using multivariate prediction models based on Near-InfraRed Spectroscopy (NIRS), after a calibration curve for each trait was developed. A subset of 110 samples was selected for calibration based on the variation of the NIRS spectra, using WinISI 4.9 statistical software (FOSS), and then biochemically analyzed to develop predictive models.

Sample selection is based on Mahalanobis distance (H) values, which measure the distances in spectral data from the mean population and from each other, to select the most diverse grouping of spectra. For the biochemical analysis, the van Soest acid detergent fiber method was used to assess fiber composition by determination of neutral detergent fiber (NDF), acid detergent fiber (ADF), and acid detergent lignin (ADL) in dry stem tissues (Van Soest et al., 1991). Furthermore, biochemical analyses of monosaccharide composition of cell wall polysaccharides were performed to precisely assess the content of glucose and of major pectic components, not quantified by the van Soest method. Specifically, content of glucose, galacturonic acid, rhamnose, arabinose, and galactose were measured as % of the alcohol-insoluble residue (AIR), which is a good estimate of the cell wall content, according to the protocol developed by Petit et al. (2019). Calibration equations for each of the trait biochemically measured were fitted using partial least squares (PLS) and modified partial least square (MPLS) regression methods implemented in WinISI software. The coefficient of determination of cross validations (RSQ) ranged between 0.68 and 0.95 and the ratio performance deviation (RPD = standard deviation/standard error of cross-validation) between 1.5 and 2 indicating a good predictive capacity for all traits.

Monosaccharide components were estimated and directly predicted as % of AIR (% AIR) while predicted detergent fiber fractions were used to calculate the concentrations (in g/kg dm) of cell wall (NDF), cellulose (CEL, equals ADF-ADL), hemicellulosic polysaccharides (HEM, equals NDF—ADF), and acid detergent lignin (ADL) in stem DM.

2.3 | Phenotypic data analysis

Descriptive statistics for each trait were calculated using basic functions in R software. An ANOVA model was performed on phenotypic traits means to determine significant effects of genotype (G), environment (E), and genotype-by-environment interaction ($G \times E$). The analysis of the data was performed both per individual field trials and on all environments, using adjusted means across trials (multi-trial model). In addition, a random-effects model was used to calculate variance components and estimates of broad sense heritability on the multi-trial model. Broad sense heritability was calculated across the three environments as in Renaud et al. (2014):

$$H^2 = V_G / \left(V_G + \frac{V_{GE}}{nE} + \frac{V_\varepsilon}{nE \times \text{nBlock}} \right)$$

where V_G, V_{GE}, and V_ε represent, respectively, the estimated genetic, $G \times E$, and error variance components, while nE represents the number of environments and nBlock the number of blocks in each environment. Phenotypic (Pearson's R) correlations between traits were calculated by generating a correlation matrix using the packages *Himsc* and *corrplot*.

2.4 | Genotyping

2.4.1 | DNA extraction

Genomic DNA was isolated from young grinded *L. mutabilis* leaves (~20–400 mg, freeze dried material) using acetyl trimethyl ammonium bromide (CTAB) method (Doyle & Doyle, 1987) with additional steps to remove proteins, polysaccharides, and RNA (as described in Petit et al. (2020)). DNA samples were controlled for quality and DNA concentration on agarose gel and by Qubit™ Fluorometric quantitation to provide high-quality genomic DNA for massive sequencing. Due to a certain degree of outcrossing behavior in *L. mutabilis*, each accession used in the GWAS panel might have a certain degree of genetic heterogeneity, despite being phenotypically homogeneous. Therefore, to cover all allelic variation within accessions, the genomic DNA of 10 individual plants per accession was isolated and pooled, resulting in 223 samples for genotyping by restriction site associated DNA sequencing (RAD-seq).

2.4.2 | RAD sequencing

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. 1 µg of high-quality genomic DNA (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. The 223 samples were paired end sequenced on the HiSeq 4000 or Xten System (Illumina platform) to provide 1 Gbp genomic data per sample. RAD library preparation and sequencing were performed by Beijing Genomics Institute (BGI).

2.4.3 | 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 $Q \leq 10$, reads with unknown bases >10%, reads that contain the sequence of the adaptors and reads that lacks a part of the multiplexing barcode and could not be identified. The clean sequence reads of each sample were mapped to the *Lupinus angustifolius* 'Tanjil' (LupAngTanjil_v1.0 refSeq GCF_001865875.1) genome reference (Hane et al., 2017) using Burrows–Wheeler Alignment Tool based on BWA-MEM algorithm. Picard-tools (v2.7.1) was used to sort the Sequence Alignment Map (SAM) files by coordinate and convert them to Binary Alignment Map (BAM) files. The average mapping rate was 76.9%, and the properly paired average 63.7%. Subsequently, BCFtools (v1.9) was used to call SNPs in each sample based on genotype likelihoods (specific BCFtools mpileup parameters: [d] max-depth = 10,000). In total, 3,747,406 putative SNPs were identified.

2.4.4 | Single-nucleotide polymorphism marker selection

Samples for each accession consisted of pools of 10 diploid plants to cover all allelic variation present in each accession. Each sample, therefore, harbors DNA from 20 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 allele frequencies 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 223 *L. mutabilis* accessions. Markers with a minor allele frequency below 5% and with a major allele frequency above 95% in the mapping panel were removed. Only biallelic markers were selected, having a frequency sum of the two major alleles equal or above 95% and a sequencing depth >40×. After removal of SNPs not yet assigned to any chromosome, a total a set of 16,781 SNPs was selected for the genetic analysis. Each SNP was scored as the proportion of the major allele in the pooled sample of plants from the same accession. Quality SNP marker selection was performed in Python using a custom script. Finally, missing values were imputed using the average frequency of the concerned marker to minimize unwanted shifting of the data.

2.4.5 | Estimation of linkage disequilibrium

Linkage Disequilibrium between SNPs genome-wide across was estimated on *L. mutabilis* GWAS panel following the approach of Vos et al. (2017). Short-range LD decay was based on the 90th percentile of Pearson allele frequencies correlation (r^2). The LD decay graph was drawn by fitting a smooth spline of r^2 over physical distances, measured as pairwise differences between marker position (Kbp), using the RQSMOOTH procedure in Genstat. Critical values of r^2 as evidence of linkage, were set based on a fixed value of 0.1 (Nordborg et al., 2002; Remington et al., 2001; Robbins et al., 2011). The intersection between the smooth spline line and the baselines ($r^2 = 0.1$) was used to define the distance at which LD decays.

2.4.6 | Genome-wide association study and analysis of population structure

A single trait GWAS was performed for each cell wall trait individually using the efficient mixed-model association approach (EMMA algorithm; Kang et al., 2008) which was implemented in StatgenGWAS package v1.0.5 (van Rossum et al., 2020). Based on the observation of the generated scatterplot of observed LOD scores versus expected LOD scores (Turner, 2018), the Van Raden kinship (van Rossum et al., 2020) was selected as best correction in combination with the addition of the origin of accessions as a covariate. Finally, SNPs with a minor allele frequency below 0.02 (112 SNPs) were also excluded from the GWAS analysis. To characterize the genetics of *L. mutabilis* biomass quality across different environments and growing conditions, GWAS was performed across location (combining the three trials) and per location (PT, NL-Sc, and NL-Wi) to identify also markers more specific for certain environments. The phenotypic variation associated with each SNP was estimated using the function propSnpVar implemented in the StatgenGWAS package (van Rossum et al., 2020). and is estimated as $\widehat{\beta}^2 \mathrm{var(SNP)} / \mathrm{var}(\mu_y)$, where $\widehat{\beta}$ is the snp effect, var(SNP) is the variance for the SNP and $\mathrm{var}(\mu_y)$ is the variance of the adjusted phenotypic means. To account for multiple testing and estimate the threshold for significant associations, two different corrections were used: a more conservative Bonferroni correction based on the number of independent markers at 5% significance (Li & Ji, 2005), leading to a threshold of 5.52 for $-\log_{10}(p)$, and one less stringent correction with a cutoff of $p = 1 \times 10^{-4}$ for the detection of relevant associations.

2.4.7 | Candidate gene identification

All candidate genes were selected based on the information contained in the NCBI *L. angustifolius* Annotation Release 100 for the genome assembly GCF_001865875.1 of LupAngTanjil_v1.0 (https://ftp.ncbi.nlm.nih.gov/genomes/all/annotation_releases/3871/100/). The overlap of associated SNPs with annotated coding sequences in *L. angustifolius*, was determined using the SNPs' positions available in the BED file. Based on the LD analysis, several genes in or near (within 80 Kbp up- and downstream of the lead SNP) the associated SNP were selected as candidate genes. Special attention was given to genes with predicted functions related to the biosynthesis and modification of the cell wall.

3 | RESULTS

3.1 | Biomass yields in different European cropping conditions

In this work, we assess the potential of *L. mutabilis* as biomass crop by investigating natural variation in biomass yield and quality in a panel of 223 accessions, across two different cropping conditions for Europe. Our results showed a high impact of cropping conditions on plant development and, as a consequence, on biomass yield (Figure 1). In winter Mediterranean conditions, on

average, 12.9 t/ha of fresh aboveground biomass were harvested with an average DM content of 56%, leading to a production of 4.9 t/ha of agricultural residues (dry weight). In summer-North European cropping conditions, 55.86 t/ha of fresh biomass were harvested in NL-Sc and 84.18 in NL-Wi, leading to, respectively, 14.7 and 20.9 t/ha of agricultural residues. The DM content of biomass at harvest was stable in North European trials (28.7% and 26.8%, respectively, in NL-Sc and NL-Wi), demonstrating that plants were harvested at the same developmental stage in both years (ripening of seeds, BBCH 85–88). Differences in biomass yield across the two trials in The Netherlands can be explained by the extreme heatwave hitting The Netherlands during summer 2019 and negatively affecting plant development. Conversely, the differences in biomass yield and DM content between Mediterranean and North-European cropping of *L. mutabilis* reflect the different development and growing stage of plants at harvest in these two conditions under study. In winter Mediterranean cropping, *L. mutabilis* ceased vegetative development soon after flowering as it encountered drought, reaching complete maturity at harvest with small and very dry plants. In summer North European conditions, a generally higher water availability prompted vegetative development by favoring a constant overlap between vegetative and reproductive phases until harvest, when plants had a bigger canopy and were still holding a higher water content. Nevertheless, the presence of a high range of variation in biomass and agricultural residues yield

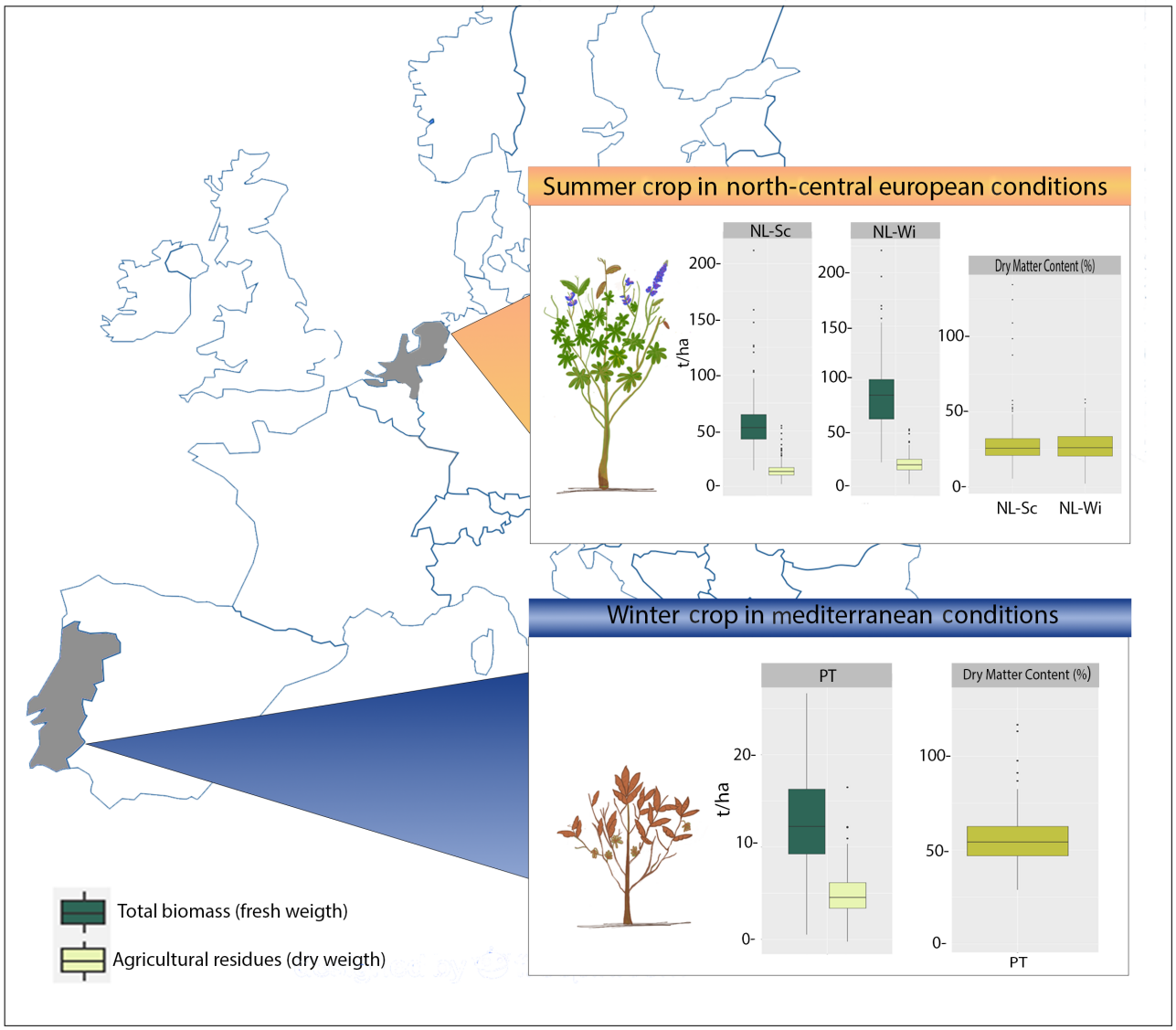

FIGURE 1 Potential yield of *Lupinus mutabilis* as a biomass crop under two different cropping conditions in Europe. Fresh weight of total biomass (t/ha), dry weight of agricultural residues after seed harvest (i.e., stems and leaves, t/ha) and dry matter content (as % of total fresh weight) are reported for 223 *L. mutabilis* accessions growing in Portugal as winter crop and in the Netherlands as summer crop.

within the panel under study, suggests that the selection of accessions for biomass production in Mediterranean and North-Central European environments has economic potential and can already provide 10–30 t/ha of agricultural residues for the biobased industry.

3.2 | Variation and heritability of biomass quality traits in GWAS panel

This study represents the first evaluation of *L. mutabilis* biomass quality and shades light on the suitability of this crop as a source of biomass for the biobased industry. The presence of a significant range of variation for both the fiber fractions and the monosaccharide components of the cell wall measured across accessions (see Figure 2 and mean and ranges reported in Table 1) candidates this panel as a good reference for GWAS studies. Overall, cellulose and glucose represented respectively the higher fiber fraction (511.5 ± 21.4 g/kg dm) and monosaccharide component (54% ± 6.7% AIR) of the cell wall, while hemicellulose and lignin were present in a similar ratio. We also report a high concentration of galacturonic acid (3.8% AIR), constituting together with rhamnose the backbone of pectin in the cell wall. The remaining components, arabinose and galactose, are present in lower concentrations and can contribute to the formation of both pectic and hemicellulosic components of the cell wall. Based on the phenotypic evaluation of the three different trials, the analysis of variance revealed a significant effect of genotype, environments and genotype by environment interaction ($G \times E$) on each of the traits measured ($p < 0.01$). Only for HEM $G \times E$ interaction effect was not significant. All the traits possessed a moderate to high heritability (ranging from 0.45 to 0.79) suggesting that variations of cell wall components are highly affected by genetic factors (see Table 1). When looking at mean values across environments (Figure 1), differences in fiber fractions content (NDF, CEL, HEM, and ADL) across environments appeared not significant, while differences in monosaccharides compositions were significant ($p < 0.001$) (depicted in Figure 2). The different level of accuracy in the biochemical assessment of fiber fractions and monosaccharide content might have contributed to this result, as shown also by the higher degree of variance associated to residuals for these traits (41.2%–81.8%, Table 1). Nevertheless, our findings suggest that even when the yield of the different biomass fiber fractions would remain stable across locations, there are differences in the monosaccharide composition of these fractions. Furthermore, the observed differences between biomass quality traits in NL-Sc and NL-Wi depicted in Figure 2 suggest that even in similar locations meteorological extremes characterizing two different years can have a significant effect on biomass quality. Only galactose and arabinose appear as consistently higher in the cell wall of biomass growing in North-European cropping conditions.

3.3 | Relationships between biomass quality traits

A correlation analysis was performed between the nine traits analyzed, including both fiber fractions and monosaccharide composition. To avoid relationships associated only to specific environments, correlation coefficients were calculated from the averaged phenotypic values across the three environments. As expected, positive correlations were detected between the fiber fractions of NDF and CEL ($r = 0.79$; $p < 0.001$) and NDF and HEM ($r = 0.44$; $p < 0.001$), as the NDF fraction comprises both hemicellulose and cellulose fractions. The positive correlation between ADL and HEM ($r = 0.28$; $p < 0.001$), components closely associated in the cell wall, and the negative correlation between ADL and CEL ($r = -0.38$; $p < 0.001$) are also in line with the known structure of plant cell wall. Furthermore, we detect a positive correlation not only between glucose and cellulose ($r = 0.73$; $p < 0.001$) but also between galactose and glucose ($r = 0.81$: $p < 0.001$) and consequently between galactose and cellulose ($r = 0.68$; $p < 0.001$). Instead, galacturonic acid, rhamnose, and arabinose are highly correlated with each other ($r = 0.8$; $p < 0.001$) and share similar correlation patterns with the other traits, including negative correlations with NDF, CEL, HEM, glucose, and galactose and positive correlations with ADL. All the correlations between traits are depicted in Figure 3.

3.4 | Genotyping of the GWAS panel

RAD-seq generated 2053375 million clean reads (~308 Gb) with on average of 9 million reads per sample and a read length of 150 bp. Following SNP selection, a total of 187,468 SNPs markers were identified in the collection of 223 accessions of *L. mutabilis*. Out of these, 16,781 SNPs had a coverage >40× and were physically mapped across the 20 chromosomes of *L. angustifolius*. These SNPs showed an homogenous distribution (Figure S1) and were retained for all further analysis. An average of 839 SNPs were identified per chromosome, ranging from 479 SNPs on chromosome 14 to 1477 SNPs on chromosome 11. The average marker density was approximately one marker every 28 Kbp across the 609-Mbp *L. angustifolius* genome (Hane et al., 2017).

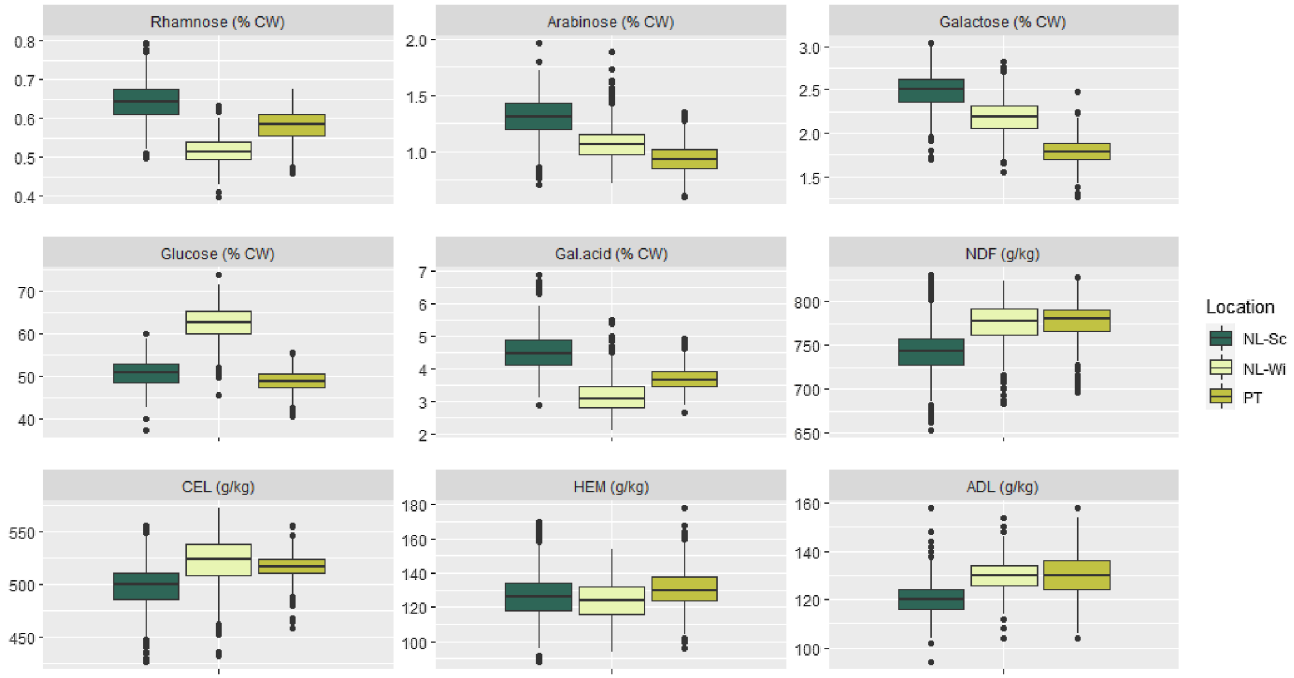

FIGURE 2 Predicted biomass quality of *Lupinus mutabilis*. Differences in monosaccharide composition were all significant among the three trials ($p < 0.001$), while differences in fiber fractions (NDF, CEL, HEM, ADL) were not significant.

TABLE 1 Summary of statics for 12 biomass quality traits for 223 diverse *Lupinus mutabilis* accessions (data averaged over the three trials). Mean and range represent the phenotypic variation span assessed for each trait; the distribution of total variance is shown as the percentage of variance components contributing to phenotypic variation; finally broad sense heritability (H^2) values indicate the extent to which a trait is genetically determined.

			Variance components (%)					
	Mean	Range	Genotype	Environment	E: Block	$G \times E$	Residual	H^2
Rhamnose	0.58	0.40–0.79	6.5	69.0	3.8	2.3	18.5	0.70
Arabinose	1.11	0.60–1.97	5.7	54.4	9.6	3.7	26.6	0.57
Galactose	2.16	1.27–3.04	5.5	77.3	0.8	2.0	14.3	0.71
Glucose	53.77	37.48–73.69	1.8	82.6	0.8	1.3	13.5	0.48
Gal. acid	3.83	2.14–6.86	11.6	64.8	3.1	3.9	16.6	0.79
NDF g/kg	765.11	654–830	7.5	33.7	6.4	5.7	46.7	0.51
CEL g/kg	511.47	426–572	17.9	27.2	5.7	5.5	43.7	0.73
HEM g/kg	127.08	88–178	7.4	5.5	5.3	0.0	81.8	0.45
ADL g/kg	126.55	94–158	12.5	33.5	6.6	6.2	41.2	0.65

3.5 | Linkage disequilibrium and population structure analyses

Linkage disequilibrium was estimated between all SNP markers over 223 accessions of *L. mutabilis*. Based on the intersection between the smooth spline line of the 90th percentile and the baseline for $r^2 = 0.1$, LD was estimated to decay around 80 Kbp of distance (Figure S2). A similar decay (LD = 77.45 Kbp) was recently estimated in a set of domesticated *L. angustifolius* accessions (Mousavi-Derazmahalleh et al., 2018). The level of diversity and stratification present in the collection under study were also examined before performing the GWAS. Controlling for population structure is a standard procedure for GWAS but it becomes particularly relevant in this case, as our collection comprises genotypes from many different sources including both wild material and breeding lines. The Van Raden kinship analysis and the principal coordinate analysis of the distance matrix inferred from this kinship, revealed the presence of two

subgroups in the collection (Figure 4). This grouping confirms a division of the accessions based on their origin and type (wild material vs. breeding material). One group includes most of the wild material from Ecuador, Peru, and Bolivia (see Figure 4, red circle), while the seconds group consist mainly of breeding lines and elite material from Ecuador and Peru (see Figure 4, blue circle). Even though PCoA1 and PCoA2 explained only 17.63% of the total genetic variance, this grouping corroborates previous findings. A similar separation pattern within this GWAS panel was already inferred from phenotypic observations (Gulisano et al., 2022) and is now confirmed in a genomic relationship matrix based on genotype information for each individual marker (VanRaden, 2008).

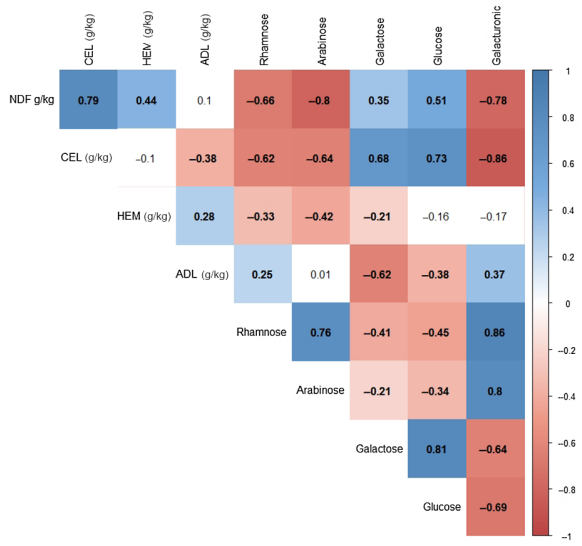

FIGURE 3 Phenotypic correlation between the nine biomass quality traits analyzed. Significant correlations were detected at a significance level of p value <0.001, and blank cells represent not-significant correlations.

3.6 | GWAS of biomass quality in mature *L. mutabilis* stems

To dissect the genetic basis of natural variation in the cell wall components under study, a GWAS analysis was performed by fitting a linear mixed model with population structure and origin of the accessions. The analysis was firstly performed across-locations and then separately for each of the locations, to identify both associated loci stable across locations and location-specific loci for the assessed traits. The quantile–quantile (QQ) plots of each trait confirmed that the associations were best controlled for population stratification by fitting both the Van Raden kinship matrix and the origin of accessions as covariate to the model. Manhattan plots and QQ plots generated by the GWAS analysis both across-location and per single locations are reported in Figure S3. Overall, 51 unique SNPs were identified in 46 QTL. At a significance level of $p < 1 \times 10^{-4}$ 17 SNPs were detected across-locations, of which nine SNPs were identified for fiber fraction components and eight SNPs for individual monosaccharide contents. For all the traits analyzed, the phenotypic variance explained by each allele ranged from 2.7% to 11.8%. In details, four SNPs were found in three loci significantly associated to NDF (corresponding to cellulose, hemicellulose and lignin), of which two were located on chromosome 8 and two on chromosome 9. For cellulose (CEL) one significant SNP was detected on chromosome 14, while for hemicellulose (HEM) three SNPs were detected on chromosome 2, 3 and 11, respectively. One SNP was detected in association with ADL on chromosome 12. Regarding the monosaccharide content of cell wall polysaccharides, one SNP was associated to the content of arabinose (chr 20), two SNPs to galactose (chr 18), two to glucose (chr 6), and three to galacturonic acid (chr 8 and 20). No significant

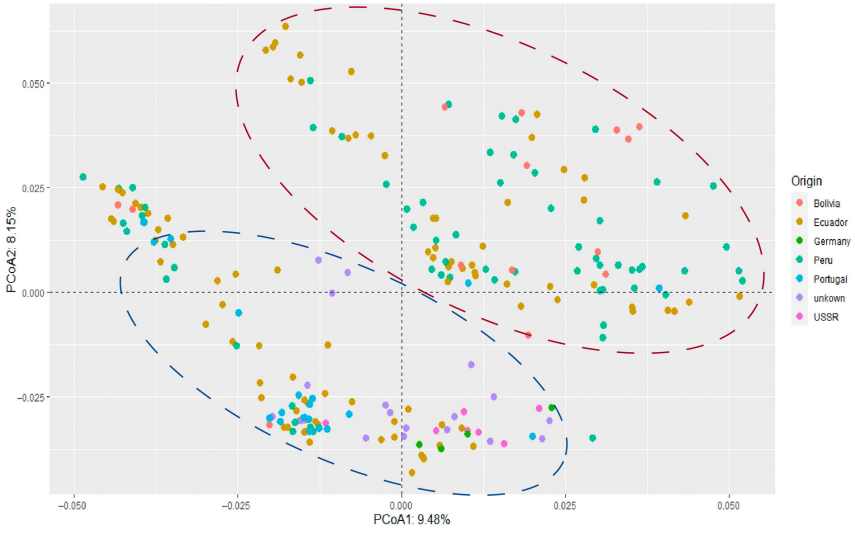

FIGURE 4 Principal coordinate analysis (PcoA) based on kinship matrix visualizes the estimated genetic structure of *Lupinus. mutabilis* collection under study and highlights separate grouping of wild material (red circle) and breeding lines (blue circle).

association with SNPs was detected for rhamnose across locations. In the location-specific GWAS, 41 SNPs were found among all traits and locations. Of these, 20 SNPs were associated to fiber fractions and 21 to monosaccharide components. The number of markers detected and the explained variance of the model for the same trait generally differed across locations. In detail, 18 SNPs were detected in PT, 19 in NL-Sc, and 4 in NL-Wi. The low number of association reported for NL-Wi might be explained by the strong effect of extreme climatic events affecting this location and the phenotypic variation of the data collected. Out of all detected associations, four SNPs were detected as common markers either among traits or among the different GWAS models performed (i.e., across-locations and location-specific). SNP M7867 (chr 8, pos 23103579)—significant also at the higher Bonferroni threshold—was detected in association with NDF and galacturonic acid, while SNP M16431 (chr 20, pos 11887436) was associated with arabinose and galacturonic acid both across-location and in the specific location PT19. Similarly, SNP M7413 (chr 8) was associated with both arabinose and galacturonic acid and M10510 (chr 12) with NDF and galacturonic acid in NL-Sc, while M7865 (chr 8) was associated to galacturonic acid both across-locations and specifically in NL-Sc trial. Noticeably, among all traits and locations, we identify set of consecutive SNPs for related traits on different chromosomes, generating "hot spots" for hemicellulose and pectin related monosaccharides (such as arabinose, galactose, rhamnose and galacturonic acid) on chromosome 8 and 20 and for glucose on chromosome 6 (see Figure 5).

3.7 | Candidate genes for biomass quality in *L. mutabilis*

The reference genome assembly GCF_001865875.1 of LupAngTanjil_v1.0 on the NCBI genome database (https://www.ncbi.nlm.nih.gov/data-hub/gene/table/taxon/3871/) was used to annotate the genes that were near the significant SNPs. After excluding repeated loci due to co-localizing of SNPs for different traits and detection of consecutive SNPs leading to the same locus, a total of 68 genes were proposed as candidate genes across-locations for biomass quality in *L. mutabilis*. Among those candidate genes, 17 encoded uncharacterized proteins, and the proteins encoded by the remaining genes include, transcription factors (ERF, bHLH, WRKY, bZIP-type transcription factor families), enzymes involved in protein and lipid metabolism, stress response, photoperiod and flowering pathways, cell wall biosynthesis and other biological processes. Among the 68 genes, 10 genes were identified as associated to biomass quality across the three locations, in details to QTL for NDF, hemicellulose, cellulose, ADL and content of glucose, galactose, arabinose, and galacturonic acid.

In association with NDF, on chromosome 8, we find a gene encoding the transcription factor *RAV1* (LOC109353869), a negative regulator of plant growth and development, and an *EXTENSIN-2-LIKE* gene responsible for the strengthening of primary cell wall. Instead, on chromosome 11, a gene encoding for WAT-1 related protein At2g37460-like (LOC109359103) and involved in the regulation of secondary cell wall (SCW) thickness was identified in association with hemicellulose. Linked

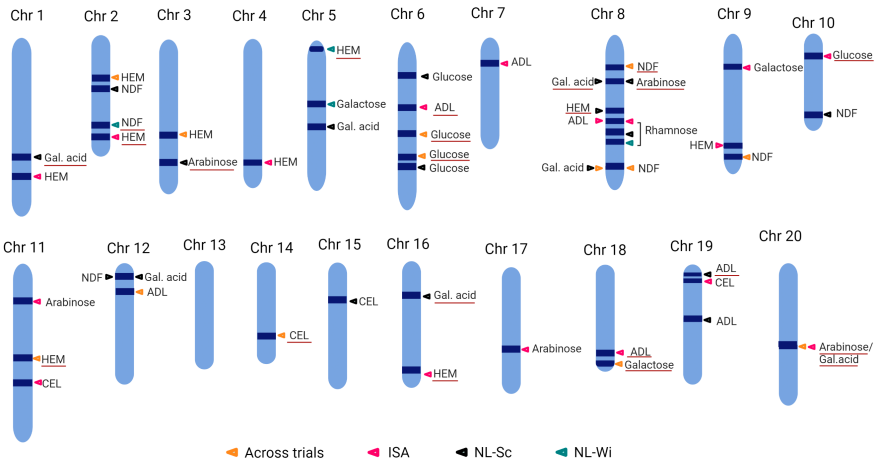

FIGURE 5 Schematic representation of the quantitative trait loci (QTL) found in association with *Lupinus mutabilis* biomass quality traits and mapped on *Lupinus angustifolius* genome. Colors of the arrows indicate the field trial locations where the respective single-nucleotide polymorphisms (SNPs) were detected. The dark blue bars on the chromosomes indicate the confidence intervals for the QTL based on the linkage disequilibrium decay (±80 Kbp flanking regions on either side of the associated SNP). The traits underlined in red are those for which relevant candidate genes with a role in cell wall composition were identified (Table 2).

to the content of cellulose, we report the finding of an *EXPANSIN-A8-LIKE* gene (chr 14) causing loosening and extension of plant cell wall in Arabidopsis and rice, while in association with ADL we find *WRKY33* transcription factor (chr 12). *WRKY33* is a negative regulator of *CESA8*, a cellulose synthase terminal complex involved in the formation of SCW, which play a role in cell wall remodeling and stress response, in particular to drought (Wang et al., 2013). Another transcription factor of the family WRKY, *WRKY15*, described as a negative regulator of cell wall lignification, was identified association with glucose content (chr 6). In association with glucose on chromosome 6 we also report the finding of *GONST2*, an GDP-mannose transporter associated with changes in cellulose content (Jing et al., 2018). Lastly, two genes were associated to changes in galactose content on chromosome 18: *KINUA*, encoding a kinesin like protein KIN-UA functioning as a microtubule protein and influencing the structure and orientation of cellulose microfibrils, and a gene encoding a LRR receptor-like serine/threonine-protein kinase probably involved in defense-related cell wall modifications/reinforcements. On chromosome 20, we identify a *sucrose phosphatase* involved in sucrose biosynthesis in relation with variation in galacturonic acid and arabinose content in the cell wall.

The location-specific GWAS studies revealed the presence of new genes whose expression might be specific for certain environmental conditions. In PT, seven genes associated to ADL, HEM and glucose content were identified. Lignin-related genes *phosphatidylinositol 4-phosphate 5-kinase 6-like and F-box protein SKIP31* were identified, *respectively*, on chromosome 7 and 18. Furthermore, a *WAT1-related protein At4g19185-like* (chr 18) regulating SCW thickness in stem fibers and an *expansin-A7* (chr 6) were also identified in association with ADL. Two genes on chromosome 2 were associated with hemicellulose, encoding a *pectinesterase-like* and a *protein STRUBBELIG-RECEPTOR FAMILY 7-like* involved respectively in pectin degradation and primary cell wall biosynthesis. Finally, a gene encoding a calmodulin-binding protein with a role in biomass formation, *protein IQ-DOMAIN 32-like*, was identified on chromosome 10 in a positive association with the content of glucose. In NL-Sc 11 genes were identified, associated with ADL, HEM, and content of arabinose and galacturonic acid. A CESA gene encoding a *cellulose synthase A catalytic subunit 4 [UDP-forming]* was identified for ADL on chromosome 19 together with a *Peroxidase 7-like* involved in lignin biosynthesis and degradation and a *pyruvate kinase isozyme A* involved in carbohydrate degradation. Notably, on chromosome 8 we identify a variety of genes associated with hemicellulose and its components such as arabinose and galacturonic acid. The SNP M7413 (chr 8, pos 7680541), co-localizing for arabinose and galacturonic acid, is contained in the gene LOC109354452 which encodes a *cysteine-rich, receptor-like protein kinase 10* associated in Arabidopsis with dwarf genotypes and collapsed xylem vessels. In the same region we also identify genes encoding an *extensin* and a *beta-glucosidase BoGH3B-like* involved in xyloglucan degradation. In addition, two more genes were associated with galacturonic acid: the *expansin A4* already identified across location on chromosome 16 and a *filament-like plant protein 3* on chromosome 1, with a role in SCW deposition. Finally, two more genes associated with Arabinose were found on chromosome 3, encoding namely an *inositol-3-phosphate synthase-like* involved in the synthesis of cell wall pectin and a *transcription factor RLM1-like* activating the transcription of genes involved in cell wall biosynthesis and maintenance.

In the last location, NL-Wi, the lowest number of associations was found and only three candidate genes were identified. On chromosome 2, linked to NDF, two genes encoded cell wall associated enzymes acting in the modification of cell wall via demethylesterification and de-esterification of cell wall pectin (a *pectinesterase-like* and a *pectinesterase 3-like*). Finally, a *cinnamoyl-CoA reductase-like SNL6* probably involved in lignin biosynthesis was identified on chromosome 5 for hemicellulose.

4 | DISCUSSION

The identification of QTL and putative candidate genes for *L. mutabilis* biomass quality has relevant implications for the breeding of this crop, when aiming at the utilization of its agriculture residues as biomass source for the biorefinery industry. To our knowledge, this is the first genetic study of *L. mutabilis* both including a substantial number of accessions and assessing biomass quality in this species. This analysis provides the first evidence that *L. mutabilis* cultivation in Europe can generate a consistent amount of biomass for the European biorefinery industry and delivers the first genetic tools developed for targeted breeding of this crop.

Cultivation of *L. mutabilis* in Europe can produce between 27 and 54 t/ha of agricultural residues, respectively in Mediterranean and North-European conditions. Notably, we find that the biomass composition of these residues is comparable with the composition of other current sources of biomass, such as miscanthus and sorghum, and superior to the one of other lupin species (Hodgson et al., 2010; Ţîţei, 2020; Zhao et al., 2009). We report cellulose yields in *L. mutabilis* up to 500 g/kg DM, which outperform average yield of common biomass feedstock such as miscanthus (400 g/kg DM, Hodgson et al., 2010) and sweet sorghum (180 g/kg DM, Zhao et al., 2009). Glucose (54%) emerges

as the main sources of fermentable sugars in *L. mutabilis*, consistently with previous results reported for the analysis of biomass quality in *L. rotundiflorus* and *L. nootkatensis* (Kamm et al., 2006; Radillo et al., 2011). In these studies, hydrolysis of plant material yielded between 48 and 55% of total fermentable carbohydrates, from glucose and xylose. Further studies should investigate the presence of xylose also in *L. mutabilis*. However, taking into account the similar cell wall content of glucose between those lupin species, the conversion of *L. mutabilis* biomass into fermentable sugars appears possible and promising. In particular, accessions LIB001, LIB027 and LIB067 emerge in this panel for their higher cellulose content across the different environments. LIB215, instead, stands out only in the Netherlands, indicating the better adaptation of this accession to North-European conditions.

The broad variation in yield and composition of biomass encountered in the GWAS panel confirms the high potential of germplasm collections as important breeding resources for the improvement of biomass quality. Large ranges of variation in biomass yield across environments can be attributed to the effect of different environmental conditions on *L. mutabilis* development (Gulisano et al., 2022). This is confirmed by the dissection of phenotypic variation for biomass components across trials which sees the effect of the environment as the largest on these traits (Table 1). The heritability values scored for these traits across environments (H^2 =0.45–0.79) are consistent with similar studies on other crops (Li, Wang, et al., 2016; Petit et al., 2020) and suggest a relevant effect of genetic factors on cell wall composition. GWAS analyses have proven to be a powerful tool to detect genetic components controlling quantitative traits. This tool has been used in many studies to approach complex traits such as cell wall composition, leading to the identification of QTL and candidate genes in important biomass crops (Li, Wang, et al., 2016; Li, Yunqiao, et al., 2016; Nguyen et al., 2020; Slavov et al., 2013) but never in lupin. In the present study, we successfully used GWAS to link variation in the yield of biomass fractions and in cell wall composition of *L. mutabilis* to putative candidate genes. In the absence of a genome sequence for *L. mutabilis*, the use of the closely related *L. angustifolius* genome proved suitable. Furthermore, the inclusion of population structure and of the origin of the accessions as covariate into the mixed model of the GWAS, ensured a good control over false positive associations, as shown by the QQ plots for the trait assessed (Figure S3).

A total of 46 QTL were detected in association with biomass quality traits across all chromosomes, except for chromosomes 13 (Figure 5). Overall, between 3 and 10 QTL were identified for all the traits evaluated (Figure 6). The phenotypic variation explained per trait by each locus ranged from 2.7% to 15.9%, supporting the idea that the genetic basis of biomass quality traits is mainly controlled by several minor effect quantitative trait genes. Similarly, large number of QTL for biomass composition have been reported in miscanthus (van der Weijde et al., 2017), maize (Li, Wang, et al., 2016; Torres et al., 2013) and many other species. The identification of single SNPs in association with different traits contributes also to highlight the complexity of the cell wall and the high interconnection between different cell wall components. For instance, co-localization of SNPs for arabinose and galacturonic acid on chromosome 20 are supported by the high correlation (r = 0.8) between these cell wall components. Furthermore, large correlations between cell wall components imply also that QTL affecting one trait can be identified in association with other strictly correlated traits. This is for example the case for the detection of a CESA4 gene involved in cellulose biosynthesis in relation with lignin (ADL) content. Negative correlations between cellulose and lignin content have been reported in many studies, and even though to a lower extent, they are also reported in our study (r = −0.4).

From our identification of SNPs associated to changes in cell wall composition emerges a list of QTL and putative candidate genes that could play a key role in the genetic improvement of *L. mutabilis*. It should, however, be noted that these results are based on the use of *L. angustifolius* reference genome. Therefore, future work is required to confirm the exact localization of the putative candidate genes in *L. mutabilis* genome, once available, and to carry out the functional studies required to confirm their roles in cell wall biosynthesis.

4.1 | Potential candidate genes and implications for the genetic improvement of *L. mutabilis* biomass

Identifying genomic regions and genes associated with cell wall composition is crucial for improving biomass quality through pathway engineering. Among the candidate genes found in the present study, we discuss the presence of some of the most relevant genes involved in cellulose production and deposition, pectin degradation and regulation of monolignol biosynthesis (see Table 2). Changes in these processes have been shown to be key to the determination of biomass quality, as they can lead to improvements in biomass production and saccharification but also significantly reduce biomass recalcitrance.

The common analysis of all environments under study highlighted the presence of QTL associated across environments with variation in the content of NDF (chr 8 and 9), HEM (chr 2, 3, 11), CEL (chr 14), ADL (chr 12),

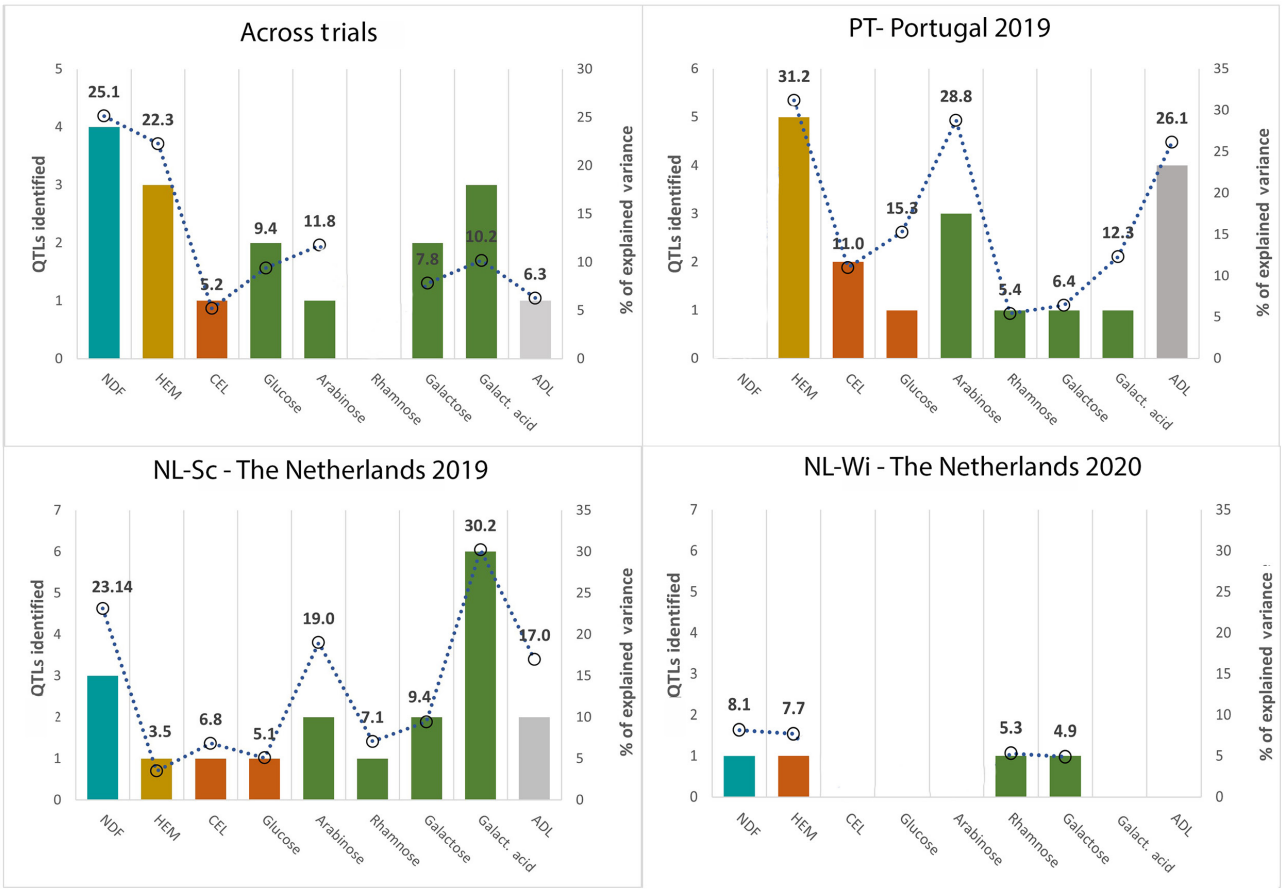

FIGURE 6 Survey of the quantitative trait loci (QTL) identified for biomass quality traits in a collection of 223 *Lupinus mutabilis* accessions, across two different cropping conditions in Europe. Colored bars indicate total number of QTL identified per trait in our analyses. Circular points refer to the cumulative proportion of variance explained by all identified single-nucleotide polymorphisms for a given trait.

Glucose (chr 6), Galactose (chr 18), Arabinose (chr 20), and Gal. acid (chr 8 and 20). Candidate genes of potential interest were identified in correspondence of the majority of these QTL and are reported in Table 2. In particular, on chromosome 6 we report two SNPs linked with significant variation on the content of glucose (4%–5.3%), the building block of cellulose. In regions flanking these SNPs, we identify a transcription factor of the WRKY family, *WRKY15*, acting as a negative regulator of cell wall lignification (Ge et al., 2020) and a GDP-D-mannose transporter, *GONST2*, responsible for the glycosylation of Glycosylinositolphosphorylceramides (GIPCs, a predominant class of lipids in the plasma membrane) and linked to changes in cellulose content (Jing et al., 2018). Increasing cellulose content and decreasing cell wall lignification are two major features of "ideal" biomass, leading to higher yields of fermentable sugars, hence these QTL can represent interesting targets for biomass improvement. Furthermore, two more SNPs were associated with glucose content on chromosome 6, in the specific location NL-Sc, confirming the presence of interesting regions for biomass improvement on this chromosome. Similarly, on chromosome 12 we report the presence of another transcription factor of the WRKY family, *WRKY33*, in correspondence of a SNP detected in association with ADL. *WKRY33* is a direct transcriptional repressor of *CESA8*, a non-redundant cellulose synthase which, together with *CesA4* and *CesA7*, is responsible for the synthesis of the main bulk of cellulosic biomass deriving from SCW. The single knockout of *WRKY33* has led to large increases in cell wall thickness, stem biomass density and aboveground biomass, suggesting targeting of this gene as a possible strategy to generate additional cell wall biomass in the stems of bioenergy crops without affecting the health and growth habit of the plant (Wang et al., 2010, 2013).

On chromosome 8, we report the identification of multiple QTL in association with NDF and pectin related monosaccharides (such as arabinose, galactose, rhamnose and galacturonic acid) both across location and in the single environments. Of particular interest, we report the finding of *RAV1* gene across environments, a negative regulator of pectin degradations, and of *pectin esterases*

involved in the methyl-esterification of pectin in PT and NL-Wi locations. The recent identification of RAV1 pectate lyase 18 in Poplar has proposed the presence of an important link between ethylene signaling in plant and pectin degradation, by showing that *RAV1* interacts with pectinases during early events of cell wall degradation (Tavares et al., 2019). As suggested by the same authors, these results open up the possibilities to gain control on the process of cell wall degradation and facilitate cell wall hydrolysis for production of fermentable sugars (Tavares et al., 2019). The pectinesterases found act, instead, on the methyl-esterification of pectin. Inhibition of *pectin esterases* has been shown to reduce the amount of de-esterification of pectin in Arabidopsis, tobacco and wheat and to efficiently increase the efficiency of enzymatic saccharification, leading to higher release of glucose after enzymatic hydrolysis (Lionetti et al., 2010). However, on chromosome 11, we report the finding of *WAT1* gene in association with HEM, whose regulation can result in strong alteration of lignin content and increased sugar extractability, according to Ranocha et al. (2010). Notably, we report the finding of *WAT1* also in association with ADL in PT, on chromosome 18, suggesting the presence of duplicated versions of this gene in *L. mutabilis* genome and an important function in SCW formation.

Finally, the identification of a SNP co-localizing for arabinose and galacturonic acid on chromosome 20, revealed the presence of a *SPP1* gene encoding a sucrose phosphatase 1 in this region, which catalyzes the last step of sucrose synthesis where sucrose-6-phosphate (Suc6P) is dephosphorylated and released as sucrose. Overexpression of sucrose-phosphate synthase has been shown to improve biomass production and saccharification (De Vega et al., 2021; Falter & Voigt, 2016), probably not only by increasing cellulose but also by indirectly reducing the content of pectic and hemicellulosic components as shown by our findings, and could find similar application in *L. mutabilis*. LRR receptor-like serine/threonine-protein kinases, like the one we identify on chromosome 18 in association with galactose, can also have an active role in triggering sucrose synthase production at the phloem level, which stimulates cellulose synthases resulting in cellulose overproduction (Ghosh et al., 2013).

Moreover, our study emphasizes that the association of genomic regions with phenotypic traits can also be location specific, thus different QTL can be detected in different environments, which is in agreement with the idea that the genetic control of agronomic traits might differ in response to different environmental conditions. Within the location specific QTL identified and the potential candidate genes detected around their confidence interval, we highlight the finding of relevant genes involved in cellulose and lignin production for both Mediterranean and North European environments. In particular, in the Mediterranean environment (PT) we report the finding of a *F-box SKIP31* protein (*SKIP31*, chr 18) and a Phosphatidylinositol-4-phosphate 5-kinase signal transduction gene (*PIP5K6*, chr 6) in association with ADL, and of a gene encoding a protein IQ-DOMAIN 32-like (*IQD32*) in positive association with glucose content on chromosome 10. Instead, in North-European conditions (NL-Sc and NL-Wi), we identify a *CesA4* gene on chromosome 19, in correspondence with a decrease in lignin content, and a *cinnamoyl-CoA reductase-like SNL6* gene (*CCR-SNL6*, chr 5) and *peroxidases* encoding gene (chr 19) in association respectively with HEM and ADL.

Genes like *CesA4* and *IQD32* can play an important role to improve cellulose production. Overexpression of SCW *CesA4* has resulted in greater cellulose content and increased total polysaccharide content, for example, in alfalfa and shrub willow (Serapiglia et al., 2012; Yang et al., 2010). However, attempts in other crops have resulted in decreased cellulose content and reduced growth suggesting that direct translation of an engineering strategy from a crop to another is not always straightforward (Brandon & Scheller, 2020). Instead, the calmodulin-binding protein IQ-DOMAIN 32-like has been identified as tightly co-expressed with secondary CESA genes in Poplar and highly upregulated during phases of enhanced cellulose biosynthesis (Badmi et al., 2018).Conversely, *SKIP31*, *PIP5K6, CCR-SNL6 and peroxidases* encoding genes can play an important role in decreasing biomass recalcitrance through alterations of the lignin pathway. Genes *SKIP31* and *PIP5K6* have been previously reported in association with the biosynthesis of monolignol components (Dauwe et al., 2007; Hodgson-Kratky et al., 2021), while *CCR-SNL6* and peroxidase genes have been associated with changes in total lignin content. Several studies have demonstrated the role of F-box proteins in the post-translational regulation of phenylpropanoids, key enzymes in the lignin biosynthesis pathway (Yu et al., 2019; Zhang et al., 2013). In particular, the gene encoding for F-box SKIP31 protein was found to be upregulated (2.7-fold changes) in sugarcane genotypes characterized by a high syringyl/guaiacyl (S/G) monolignol ratios (Hodgson-Kratky et al., 2021). Shifts in the S/G ratio have resulted in increased ethanol yields in many studies (Fu et al., 2011; Ho-Yue-Kuang et al., 2016; Jung et al., 2013), suggesting the modification of the monolignol ratio as a viable option for improving the efficiency of enzymatic hydrolysis of biomass. Moreover, such an approach has been shown to not have negative consequences on plant development (Franke et al., 2000; Ho-Yue-Kuang et al., 2016; Stewart et al., 2009), while an overall reduction of lignin content has been generally associated with yield penalty (Benjamin et al., 2013; Jung et al., 2013; Masarin

TABLE 2 Putative candidate genes in QTL associated to biomass quality of *Lupinus mutabilis* with a role in cell wall composition.

Location	Trait	SNP name	Chr	SNP pos. (bp)	LOD	PVE	Gene ID	Gene name	Description
All	NDF	M7437	8	7839123	4.59	5.71	LOC109354462		Extensin-2-like
							LOC109353869	RAV1	AP2/ERF and B3 domain-containing transcription factor At1g51120-like
	HEM	M9798	11	21441300	4.08	4.73	LOC109359103	WAT1	WAT1-related protein At2g37460-like
	CEL	M11863	14	10434333	4.23	4.48	LOC109327136	EXPA8	Expansin-A8-like
	ADL	M10606	12	3492591	4.28	6.28	LOC109361641	WRKY33	Probable WRKY transcription factor 33
	Galactose	M15279	18	14096121	4.58	4.13	LOC109332916		Probable LRR receptor-like serine/threonine-protein kinase At1g12460
		M15280	18	14096122	4.06	3.69	LOC109333139	KINUA	Kinesin-like protein KIN-UA
	Glucose	M5759	6	19564231	4.42	4.09	LOC109350121	WRKY15	Probable WRKY transcription factor 15
	Glucose	M5931	6	23471229	5.2	5.28	LOC109350274	GONST2	GDP-mannose transporter GONST2-like
	Ara/Gal. acid	M16431	20	11887436	5.18/4.7	13.2/5	LOC109336456	SPP1	Sucrose-phosphatase 1-like
PT	ADL	M5620	6	14743057	4.12	6.17	LOC109350828	EXPA7	Expansin-A7-like
							LOC109352006	PIP5K6	Phosphatidylinositol 4-phosphate 5-kinase 6-like
	ADL	M15249	18	13811328	4.25	6.02	LOC109332980	SKIP31	F-box protein SKIP31
							LOC109332164	WAT1	WAT1-related protein At4g19185-like
	HEM	M2124	2	19114587	4.39	7.75	LOC109337735	SRF7	Protein STRUBBELIG-RECEPTOR FAMILY 7-like
							LOC109337703	PE	Pectinesterase-like
	Glucose	M8427	10	4374335	7.26	15.26	LOC109358624	IQD32	Protein IQ-DOMAIN 32-like
NL-Sc	ADL	M15721	19	8815933	4.35	6.86	LOC109334350	CESA4	Cellulose synthase A catalytic subunit 4 [UDP-forming]
							LOC109354492		Pyruvate kinase isozyme A, chloroplastic-like
							LOC109333967	PER7	Peroxidase 7-like
	HEM	M7490	8	12066409	4.14	3.49	LOC109354794		Beta-glucosidase BoGH3B-like
	Gal. acid/Ara	M7413	8	7680541	4.18	5.49/11.99	LOC109354452	CRK10	Cysteine-rich receptor-like protein kinase 10
							LOC109353935		Extensin-like
							LOC109353651	NDR1	Protein NDR1-like
	Gal. acid	M1040	1	25847112	4.53	2.87	LOC109358630	FPP3	Filament-like plant protein 3
		M12931	16	6337087	4.71	10.36	LOC109330343	EXPA4	Expansin-A4-like
	Arabinose	M3320	3	21772968	4.05	7.01	LOC109343654	INO1	Inositol-3-phosphate synthase-like
							LOC109343658	RLM1	Transcription factor RLM1-like

TABLE 2 (Continued)

Location	Trait	SNP name	Chr	SNP pos. (bp)	LOD	PVE	Gene ID	Gene name	Description
NL-Wi	NDF	M2112	2	18998440	4.19	8.15	LOC109337693	PE3	Pectinesterase 3-like
							LOC109337703	PE	Pectinesterase-like
	HEM	M4404	5	378011	4.12	7.68	LOC109347281	CCR-SNL6	Cinnamoyl-CoA reductase-like SNL6

Note: Putative candidate genes were identified in a QTL confidence interval of ±80 Kpb from the associated SNP.

et al., 2011). Regulation of *Cinnamoyl-CoA reductase* (CRR) genes and peroxidases genes have instead resulted in strong alteration of lignin content and increased sugar extractability (Bart et al., 2010; Goujon et al., 2003; Jones et al., 2001; Ruel et al., 2009). Finally, our analysis also identifies extensins (chr 8) and expansins (chr 6, 14, 16) responsible respectively for creating and loosening cross-linked networks in the cell wall, both across location and in specific-locations models.

5 | CONCLUSION

The present study is the first one to provide insights into *L. mutabilis* biomass quality and its genetic architecture. The results of this investigation reveal that *L. mutabilis* presents a valuable source of high-quality biomass, comparable or even superior to other common biomass feedstocks. The panel of 223 accessions selected for this study brings to light an unexploited wealth of heritable variation for cell wall compositional traits, which is key to the improvement of biomass quality. The GWAS analysis resulted in the identification of 46 QTL for biomass quality and proposed relevant candidate genes associated with some of these QTL. Genes involved in cellulose and sucrose synthesis could play an important role in tailoring cellulose content. Conversely, genes involved in the regulation of monolignol biosynthesis and in the degradation of pectin are good candidates for decreasing biomass recalcitrance and increasing the efficiency of enzymatic saccharification. After validation of the exact localization of QTL and candidate genes in *L. mutabilis* genome, the implementation of these findings in molecular breeding programs will unquestionably accelerate the development and introduction of *L. mutabilis* as a multipurpose crop for Europe. Furthermore, outstanding genotypes with high cellulose content have been identified and are valuable material to be included in breeding programs for biomass quality. Finally, the use of this *L. mutabilis* panel for the first genome-wide study on this species, has indicated that this collection of accessions includes not only an extensive phenotypic variation but also genetic. Hence, this panel can also be valuable for further genetic studies of *L. mutabilis* traits that are still poorly understood.

AUTHOR CONTRIBUTIONS

Agata Gulisano and Luisa M. Trindade conceived and designed the experiment. Agata Gulisano performed all the analysis, interpreted the data and wrote the manuscript. Luisa M. Trindade wrote the proposal and revised the manuscript. Maria-João Paulo provided guidance in the statistical analysis and genome-wide association study and revised the manuscript. Annemarie Dechesne helped

performing the biochemical analysis. All authors read and approved the manuscript.

ACKNOWLEDGMENTS

We thank Eibertus N. van Loo for providing technical assistance in bioinformatics and for helping with the SNP calling of the sequenced data. We also thank Tim Koorevaar and Win van der Slikke for their help in preprocessing of biomass material.

FUNDING INFORMATION

This project has received funding from the BioBased Industries Joint Undertaking under the European Union's Horizon 2020 research and innovation program under grant agreement No. 720726 (LIBBIO).

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could constitute a potential conflict of interest.

DATA AVAILABILITY STATEMENT

The genotypic data used for the genome wide association study and generated by genotyping by restriction site-associated DNA sequencing (RAD-seq) are openly available at the 4TU. Research Data repository (http://doi.org/10.4121/21334146.v1). All other data are available on request.

ORCID

Agata Gulisano https://orcid.org/0000-0002-7643-963X
Luisa M. Trindade https://orcid.org/0000-0003-1541-2094

REFERENCES

Badmi, R., Payyavula, R. S., Bali, G., Guo, H.-B., Jawdy, S. S., Gunter, L. E., Yang, X., Winkeler, K. A., Collins, C., & Rottmann, W. H. (2018). A new calmodulin-binding protein expresses in the context of secondary cell wall biosynthesis and impacts biomass properties in *Populus*. *Frontiers in Plant Science*, *9*, 1669.

Bart, R. S., Chern, M., Vega-Sánchez, M. E., Canlas, P., & Ronald, P. C. (2010). Rice Snl6, a cinnamoyl-CoA reductase-like gene family member, is required for NH1-mediated immunity to Xanthomonas oryzae pv. Oryzae. *PLoS Genetics*, *6*(9), e1001123.

Benjamin, Y., Cheng, H., & Görgens, J. F. (2013). Evaluation of bagasse from different varieties of sugarcane by dilute acid pretreatment and enzymatic hydrolysis. *Industrial Crops and Products*, *51*, 7–18.

Brandon, A. G., & Scheller, H. V. (2020). Engineering of bioenergy crops: Dominant genetic approaches to improve polysaccharide properties and composition in biomass. *Frontiers in Plant Science*, *11*, 282.

Carvajal-Larenas, F. E., Linnemann, A. R., Nout, M. J. R., Koziol, M., & van Boekel, M. A. J. S. (2016). *Lupinus mutabilis*: Composition, uses, toxicology, and debittering. *Critical Reviews in Food Science and Nutrition*, *56*(9), 1454–1487. https://doi.org/10.1080/10408398.2013.772089

Chen, Y., Liu, C., Chang, P. R., Cao, X., & Anderson, D. P. (2009). Bionanocomposites based on pea starch and cellulose nanowhiskers hydrolyzed from pea hull fibre: Effect of hydrolysis time. *Carbohydrate Polymers*, *76*(4), 607–615.

Dauwe, R., Morreel, K., Goeminne, G., Gielen, B., Rohde, A., Van Beeumen, J., Ralph, J., Boudet, A.-M., Kopka, J., & Rochange, S. F. (2007). Molecular phenotyping of lignin-modified tobacco reveals associated changes in cell-wall metabolism, primary metabolism, stress metabolism and photorespiration. *The Plant Journal*, *52*(2), 263–285.

De Vega, J. J., Peel, N., Purdy, S. J., Hawkins, S., Donnison, L., Dyer, S., & Farrar, K. (2021). Differential expression of starch and sucrose metabolic genes linked to varying biomass yield in miscanthus hybrids. *Biotechnology for Biofuels*, *14*(1), 1–15.

Doyle, J. J., & Doyle, J. L. (1987). *A rapid DNA isolation procedure for small quantities of fresh leaf tissue*.

Falter, C., & Voigt, C. A. (2016). Improving biomass production and saccharification in *Brachypodium distachyon* through overexpression of a sucrose-phosphate synthase from sugarcane. *Journal of Plant Biochemistry and Biotechnology*, *25*(3), 311–318.

Franke, R., McMichael, C. M., Meyer, K., Shirley, A. M., Cusumano, J. C., & Chapple, C. (2000). Modified lignin in tobacco and poplar plants over-expressing the Arabidopsis gene encoding ferulate 5-hydroxylase. *The Plant Journal*, *22*(3), 223–234.

Fu, C., Mielenz, J. R., Xiao, X., Ge, Y., Hamilton, C. Y., Rodriguez, M., Chen, F., Foston, M., Ragauskas, A., & Bouton, J. (2011). Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(9), 3803–3808.

Garcia, A., Gandini, A., Labidi, J., Belgacem, N., & Bras, J. (2016). Industrial and crop wastes: A new source for nanocellulose biorefinery. *Industrial Crops and Products*, *93*, 26–38.

Ge, S., Han, X., Xuwen, X., Shao, Y., Zhu, Q., Liu, Y., Juan, D., Juan, X., & Zhang, S. (2020). WRKY15 suppresses tracheary element differentiation upstream of VND7 during xylem formation. *The Plant Cell*, *32*(7), 2307–2324.

Ghosh, J. S., Chaudhuri, S., Dey, N., & Pal, A. (2013). Functional characterization of a serine-threonine protein kinase from *Bambusa balcooa* that implicates in cellulose overproduction and superior quality fiber formation. *BMC Plant Biology*, *13*(1). https://doi.org/10.1186/1471-2229-13-128

Gifford, J. M., Chae, W. B., Swaminathan, K., Moose, S. P., & Juvik, J. A. (2015). Mapping the genome of Miscanthus sinensis for QTL associated with biomass productivity. *GCB Bioenergy*, *7*(4), 797–810.

Gladstones, J. (1970). Lupins as crop plants. *Field Crop Abstracts*, *23*, 123–147.

Goujon, T., Ferret, V., Mila, I., Pollet, B., Ruel, K., Burlat, V., Joseleau, J.-P., Barriere, Y., Lapierre, C., & Jouanin, L. (2003). Down-regulation of the AtCCR1 gene in *Arabidopsis thaliana*: Effects on phenotype, lignins and cell wall degradability. *Planta*, *217*(2), 218–228.

Gulisano, A., Alves, S., Rodriguez, D., Murillo, A., Van Dinter, B. J., Torres, A. F., Gordillo-Romero, M., de Lourdes Torres, M., Neves-Martins, J., Paulo, M. J., & Trindade, L. M. (2022).

Diversity and agronomic performance of Lupinus mutabilis germplasm in European and Andean environments. *Frontiers in Plant Science*, *13*, 1718.

Gulisano, A., Alves, S., Martins, J. N., & Trindade, L. M. (2019). Genetics and breeding of Lupinus mutabilis: An emerging protein crop. *Frontiers in Plant Science*, *10*, 1385.

Hane, J. K., Ming, Y., Kamphuis, L. G., Nelson, M. N., Garg, G., Atkins, C. A., Bayer, P. E., Bravo, A., Bringans, S., Cannon, S., Edwards, D., Foley, R., Gao, L.-l., Harrison, M. J., Huang, W., Hurgobin, B., Li, S., Liu, C.-W., McGrath, A., … Singh, K. B. (2017). A comprehensive draft genome sequence for lupin (*Lupinus angustifolius*), an emerging health food: Insights into plant–microbe interactions and legume evolution. *Plant Biotechnology Journal*, *15*(3), 318–330. https://doi.org/10.1111/pbi.12615

Hodgson, E. M., Lister, S. J., Bridgwater, A. V., Clifton-Brown, J., & Donnison, I. S. (2010). Genotypic and environmentally derived variation in the cell wall composition of miscanthus in relation to its use as a biomass feedstock. *Biomass and Bioenergy*, *34*(5), 652–660.

Hodgson-Kratky, K., Perlo, V., Furtado, A., Choudhary, H., Gladden, J. M., Simmons, B. A., Botha, F., & Henry, R. J. (2021). Association of gene expression with syringyl to guaiacyl ratio in sugarcane lignin. *Plant Molecular Biology*, *106*(1), 173–192.

Ho-Yue-Kuang, S., Alvarado, C., Antelme, S., Bouchet, B., Cézard, L., Le Bris, P., Legée, F., Maia-Grondard, A., Yoshinaga, A., & Saulnier, L. (2016). Mutation in Brachypodium caffeic acid O-methyltransferase 6 alters stem and grain lignins and improves straw saccharification without deteriorating grain quality. *Journal of Experimental Botany*, *67*(1), 227–237.

Jing, B., Ishikawa, T., Soltis, N., Inada, N., Liang, Y., Murawska, G., Andeberhan, F., Pidatala, R., Yu, X., Baidoo, E., & Kawai-Yamada, M. (2018). GONST2 transports GDP-mannose for sphingolipid glycosylation in the Golgi apparatus of Arabidopsis. *BioRxiv*, 346775.

Jones, L., Roland Ennos, A., & Turner, S. R. (2001). Cloning and characterization of irregular xylem4 (irx4): A severely lignin-deficient mutant of Arabidopsis. *The Plant Journal*, *26*(2), 205–216.

Jung, J. H., Vermerris, W., Gallo, M., Fedenko, J. R., Erickson, J. E., & Altpeter, F. (2013). RNA interference suppression of lignin biosynthesis increases fermentable sugar yields for biofuel production from field-grown sugarcane. *Plant Biotechnology Journal*, *11*(6), 709–716.

Kamm, B., Kamm, M., Schmidt, M., Starke, I., & Kleinpeter, E. (2006). Chemical and biochemical generation of carbohydrates from lignocellulose-feedstock (*Lupinus nootkatensis*)—Quantification of glucose. *Chemosphere*, *62*(1), 97–105.

Kang, H. M., Ye, C., & Eskin, E. (2008). Accurate discovery of expression quantitative trait loci under confounding from spurious and genuine regulatory hotspots. *Genetics*, *180*(4), 1909–1925.

Kumar, P., Barrett, D. M., Delwiche, M. J., & Stroeve, P. (2009). Methods for pretreatment of lignocellulosic biomass for efficient hydrolysis and biofuel production. *Industrial & Engineering Chemistry Research*, *48*(8), 3713–3729.

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

Li, K., Wang, H., Xiaojiao, H., Liu, Z., Yujin, W., & Huang, C. (2016). Genome-wide association study reveals the genetic basis of stalk cell wall components in maize. *PLoS One*, *11*(8), e0158906.

Li, M., Yunqiao, P., & Ragauskas, A. J. (2016). Current understanding of the correlation of lignin structure with biomass recalcitrance. *Frontiers in Chemistry*, *4*, 45.

Lionetti, V., Francocci, F., Ferrari, S., Volpi, C., Bellincampi, D., Galletti, R., D'Ovidio, R., De Lorenzo, G., & Cervone, F. (2010). Engineering the cell wall by reducing de-methyl-esterified homogalacturonan improves saccharification of plant tissues for bioconversion. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(2), 616–621.

López-Malvar, A., Butron, A., Malvar, R. A., McQueen-Mason, S. J., Faas, L., Gómez, L. D., Revilla, P., Figueroa-Garrido, D. J., & Santiago, R. (2021). Association mapping for maize Stover yield and saccharification efficiency using a multiparent advanced generation intercross (MAGIC) population. *Scientific Reports*, *11*(1), 1–9.

Malik, P. L., Janss, L., Nielsen, L. K., Borum, F., Jørgensen, H., Eriksen, B., Schjoerring, J. K., & Rasmussen, S. K. (2019). Breeding for dual-purpose wheat varieties using marker–trait associations for biomass yield and quality traits. *Theoretical and Applied Genetics*, *132*(12), 3375–3398.

Masarin, F., Gurpilhares, D. B., Baffa, D. C. F., Barbosa, M. H. P., Carvalho, W., Ferraz, A., & Milagres, A. M. F. (2011). Chemical composition and enzymatic digestibility of sugarcane clones selected for varied lignin content. *Biotechnology for Biofuels*, *4*(1), 1–10.

Méchin, V., Argillier, O., Hébert, Y., Guingo, E., Moreau, L., Charcosset, A., & Barriere, Y. (2001). Genetic analysis and QTL mapping of cell wall digestibility and lignification in silage maize. *Crop Science*, *41*(3), 690–697.

Mikić, A., Ćupina, B., Mihailović, V., Krstić, Đ., Antanasović, S., Zorić, L., Đorđević, V., Perić, V., & Srebrić, M. (2013). Intercropping white (*Lupinus albus*) and Andean (*Lupinus mutabilis*) lupins with other annual cool season legumes for forage production. *South African Journal of Botany*, *89*, 296–300.

Mousavi-Derazmahalleh, M., Nevado, B., Bayer, P. E., Filatov, D. A., Hane, J. K., Edwards, D., Erskine, W., & Nelson, M. N. (2018). The western Mediterranean region provided the founder population of domesticated narrow-leafed lupin. *Theoretical and Applied Genetics*, *131*(12), 2543–2554.

Nguyen, D. T., Gomez, L. D., Harper, A., Halpin, C., Waugh, R., Simister, R., Whitehead, C., Oakey, H., Nguyen, H. T., & Nguyen, T. V. (2020). Association mapping identifies quantitative trait loci (QTL) for digestibility in rice straw. *Biotechnology for Biofuels*, *13*(1), 1–16.

Nordborg, M., Borevitz, J. O., Bergelson, J., Berry, C. C., Chory, J., Hagenblad, J., Kreitman, M., Maloof, J. N., Noyes, T., & Oefner, P. J. (2002). The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics*, *30*(2), 190–193.

Pancaldi, F., & Trindade, L. M. (2020). Marginal lands to grow novel bio-based crops: A plant breeding perspective. *Frontiers in Plant Science*, *11*, 227.

Pennells, J., Cruickshank, A., Chaléat, C., Godwin, I. D., & Martin, D. J. (2021). Sorghum as a novel biomass for the sustainable production of cellulose nanofibers. *Industrial Crops and Products*, *171*, 113917.

Petit, J., Gulisano, A., Dechesne, A., & Trindade, L. M. (2019). Phenotypic variation of cell wall composition and stem morphology in hemp (*Cannabis sativa* L.): Optimization of

methods. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.00959

Petit, J., Salentijn, E. M. J., Paulo, M.-J., Thouminot, C., Jan, B., van Dinter, G., Magagnini, H.-J. G., Tang, K., Amaducci, S., & Wang, S. (2020). Genetic variability of morphological, flowering, and biomass quality traits in hemp (*Cannabis sativa* L.). *Frontiers in Plant Science*, *11*, 102.

Radillo, J. J., Vargas, M. A., Ruiz-López, R. R., Macías, L. B., Ramírez, P. M., García-López, P. M., & Toral, F. A. L. D. (2011). Fermentable sugars from lupinus rotundiflorus biomass by hydrochloric acid hydrolysis. *BioResources*, *6*(1), 344–355.

Ranocha, P., Denancé, N., Vanholme, R., Freydier, A., Martinez, Y., Hoffmann, L., Köhler, L., Pouzet, C., Renou, J.-P., & Sundberg, B. (2010). Walls are thin 1 (WAT1), an Arabidopsis homolog of Medicago truncatula NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. *The Plant Journal*, *63*(3), 469–483.

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovich, S., Goodman, M. M., & Buckler, E. S. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(20), 11479–11484.

Renaud, E. N., Lammerts van Bueren, E. T., Myers, J. R., Paulo, M. J., Van Eeuwijk, F. A., Zhu, N., & Juvik, J. A. (2014). Variation in broccoli cultivar phytochemical content under organic and conventional management systems: Implications in breeding for nutrition. *PLoS One*, *9*(7), e95683.

Robbins, M. D., Sim, S.-C., Yang, W., Van Deynze, A., van der Knaap, E., Joobeur, T., & Francis, D. M. (2011). Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato. *Journal of Experimental Botany*, *62*(6), 1831–1845.

Ruel, K., Berrio-Sierra, J., Derikvand, M. M., Pollet, B., Thévenin, J., Lapierre, C., Jouanin, L., & Joseleau, J.-P. (2009). Impact of CCR1 silencing on the assembly of lignified secondary walls in *Arabidopsis thaliana*. *New Phytologist*, *184*(1), 99–113.

Schnitzer, M., Monreal, C. M., & Powell, E. E. (2014). Wheat straw biomass: A resource for high-value chemicals. *Journal of Environmental Science and Health, Part B*, *49*(1), 51–67.

Serapiglia, M. J., Cameron, K. D., Stipanovic, A. J., & Smart, L. B. (2012). Correlations of expression of cell wall biosynthesis genes with variation in biomass composition in shrub willow (*Salix* spp.) biomass crops. *Tree Genetics & Genomes*, *8*(4), 775–788.

Slavov, G., Allison, G., & Bosch, M. (2013). Advances in the genetic dissection of plant cell walls: Tools and resources available in miscanthus. *Frontiers in Plant Science*, *4*, 217. https://doi.org/10.3389/fpls.2013.00217

Stewart, J. J., Akiyama, T., Chapple, C., Ralph, J., & Mansfield, S. D. (2009). The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. *Plant Physiology*, *150*(2), 621–635.

Tavares, E. Q., De Souza, A. P., Romim, G. H., Grandis, A., Plasencia, A., Gaiarsa, J. W., Grima-Pettenati, J., de Setta, N., Van Sluys, M. A., & Buckeridge, M. S. (2019). The control of endopolygalacturonase expression by the sugarcane RAV transcription factor during aerenchyma formation. *Journal of Experimental Botany*, *70*(2), 497–506.

Tian, S.-Q., Zhao, R.-Y., & Chen, Z.-C. (2018). Review of the pretreatment and bioconversion of lignocellulosic biomass from wheat straw materials. *Renewable and Sustainable Energy Reviews*, *91*, 483–489.

Tiekstra, S., & Adriaanse, M. (2014). *Tomato packaging—Packing tomatoes in a box made from tomato side streams, closed loop approach*. PTS Symposium Innovative Packaging Munich, 2014.

Ţîţei, V. (2020). Some biological features and biomass quality of *Lupinus albus* and *Lupinus luteus* in Moldova. *Lucrări Ştiinţifice USAMV- Iaşi Seria Agronomie*, *63*(1), 19–26.

Torres, A. F., van der Weijde, T., Dolstra, O., Visser, R. G. F., & Trindade, L. M. (2013). Effect of maize biomass composition on the optimization of dilute-acid pretreatments and enzymatic saccharification. *Bioenergy Research*, *6*(3), 1038–1051.

Turner, S. D. (2018). Qqman: An R package for visualizing GWAS results using Q-Q and Manhattan plots. *Journal of Open Source Software*, *3*(25), 731. https://doi.org/10.21105/joss.00731

van der Weijde, T., Dolstra, O., Visser, R. G. F., & Trindade, L. M. (2017). Stability of cell wall composition and saccharification efficiency in miscanthus across diverse environments. *Frontiers in Plant Science*, *7*, 2007. https://doi.org/10.3389/fpls.2016.02004

van Rossum, B. J., Kruijer, W., van Eeuwijk, F., Boer, M., Malosetti, M., Bustos-Korts, D., Millet, E., Paulo, J., Verouden, M., & Wehrens, R. (2020). *Package 'statgenGWAS'*. R package version, 1(7).

Van Raden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414–4423.

Van Soest, P. J., Robertson, J. B., & Lewis, B. A. (1991). Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *Journal of Dairy Science*, *74*(10), 3583–3597. https://doi.org/10.3168/jds.S0022-0302(91)78551-2

Vos, P. G., Paulo, M. J., Voorrips, R. E., Visser, R. G., van Eck, H. J., & van Eeuwijk, F. A. (2017). Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theoretical and Applied Genetics*, *130*(1), 123–135. https://doi.org/10.1007/s00122-016-2798-8

Wang, H., Avci, U., Nakashima, J., Hahn, M. G., Chen, F., & Dixon, R. A. (2010). Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(51), 22338–22343.

Wang, X., Bojing, D., Liu, M., Sun, N., & Qi, X. (2013). Arabidopsis transcription factor WRKY33 is involved in drought by directly regulating the expression of CesA8. *American Journal of Plant Sciences*, *4*(6A), 21–27.

Yang, S. S., Wayne Wenzhong, X., Tesfaye, M., Lamb, J. A. F. S., Jung, H.-J. G., VandenBosch, K. A., Vance, C. P., & Gronwald, J. W. (2010). Transcript profiling of two alfalfa genotypes with contrasting cell wall composition in stems using a cross-species platform: Optimizing analysis by masking biased probes. *BMC Genomics*, *11*(1), 1–18.

Yu, S.-i., Kim, H., Yun, D.-J., Suh, M. C., & Lee, B.-h. (2019). Post-translational and transcriptional regulation of phenylpropanoid biosynthesis pathway by Kelch repeat F-box protein SAGL1. *Plant Molecular Biology*, *99*(1), 135–148.

Zhang, X., Gou, M., & Liu, C.-J. (2013). Arabidopsis Kelch repeat F-box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonia-lyase. *The Plant Cell*, *25*(12), 4994–5010.

Zhao, Y. L., Dolat, A., Steinberger, Y., Wang, X., Osman, A., & Xie, G. H. (2009). Biomass yield and changes in chemical composition

of sweet sorghum cultivars grown for biofuel. *Field Crops Research*, *111*(1–2), 55–64.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gulisano, A., Dechesne, A., Paulo, M.-J., & Trindade, L. M. (2023). Investigating the potential of Andean lupin as a lignocellulosic feedstock for Europe: First genome-wide association study on *Lupinus mutabilis* biomass quality. *GCB Bioenergy*, *15*, 38–57. https://doi.org/10.1111/gcbb.13006