RESEARCH

Genetic markers and tree properties predicting wood biorefining potential in aspen (*Populus tremula*) bioenergy feedstock

Sacha Escamez[1†], Kathryn M. Robinson[1†], Mikko Luomaranta[1], Madhavi Latha Gandla[2], Niklas Mähler[1], Zakiya Yassin[3], Thomas Grahn[3], Gerhard Scheepers[3], Lars-Göran Stener[4], Stefan Jansson[1], Leif J. Jönsson[2], Nathaniel R. Street[1] and Hannele Tuominen[1,5*]

Abstract

Background Wood represents the majority of the biomass on land and constitutes a renewable source of biofuels and other bioproducts. However, wood is recalcitrant to bioconversion, raising a need for feedstock improvement in production of, for instance, biofuels. We investigated the properties of wood that affect bioconversion, as well as the underlying genetics, to help identify superior tree feedstocks for biorefining.

Results We recorded 65 wood-related and growth traits in a population of 113 natural aspen genotypes from Sweden (https://doi.org/10.5061/dryad.gtht76hrd). These traits included three growth and field performance traits, 20 traits for wood chemical composition, 17 traits for wood anatomy and structure, and 25 wood saccharification traits as indicators of bioconversion potential. Glucose release after saccharification with acidic pretreatment correlated positively with tree stem height and diameter and the carbohydrate content of the wood, and negatively with the content of lignin and the hemicellulose sugar units. Most of these traits displayed extensive natural variation within the aspen population and high broad-sense heritability, supporting their potential in genetic improvement of feedstocks towards improved bioconversion. Finally, a genome-wide association study (GWAS) revealed 13 genetic loci for saccharification yield (on a whole-tree-biomass basis), with six of them intersecting with associations for either height or stem diameter of the trees.

Conclusions The simple growth traits of stem height and diameter were identified as good predictors of wood saccharification yield in aspen trees. GWAS elucidated the underlying genetics, revealing putative genetic markers for bioconversion of bioenergy tree feedstocks.

Keywords Biorefining, Feedstock recalcitrance, Bioenergy, Forest feedstocks, Saccharification, Biomass

†Sacha Escamez and Kathryn M. Robinson have contributed equally to this work

*Correspondence:
Hannele Tuominen
hannele.tuominen@slu.se
Full list of author information is available at the end of the article

Introduction

Lignocellulosic woody biomass represents the majority of biomass on land [2]. This biomass contains mostly three types of natural polymers: cellulose, hemicelluloses and lignin, each of which can be converted into precursors for biofuels and other bioproducts [41]. However, the processes for deconstructing these polymers into usable units remain costly due to structural and chemical hindrance, a problem known as biomass recalcitrance [32].

Overcoming biomass recalcitrance requires the identification of less recalcitrant feedstocks as well as knowledge on the biological basis of lignocellulose recalcitrance [33, 59, 62, 65, 68]. Fast growing trees from the *Populus* genus (poplars, aspens and hybrids) represent promising feedstocks [35] on account of their lignocellulose composition [50], advanced domestication and efficient cultivation techniques [7]. Furthermore, genomes of numerous *Populus* species have been sequenced [12, 24, 28, 30, 44, 49, 57, 61, 67], which enables investigation of the genetics underlying lignocellulose properties for a better understanding of the biochemistry behind and breeding for less recalcitrance.

Our knowledge of the genetic basis for plant traits has greatly advanced owing to genome-wide association studies (GWAS), which relate variation in traits to variation in the sequence of the genomes of different individuals, down to single nucleotide resolution. These variations of nucleotide composition at single loci, also known as single nucleotide polymorphisms (SNPs), can reveal genetic markers for quantitative variation in traits, or even reveal involvement of genes in shaping a quantitative trait [39].

In a striking example, GWAS of the timing of bud set identified a single locus explaining the majority of local adaptation along a latitudinal gradient in a Swedish population of European aspen *Populus tremula* [61]. However, individual loci found by GWAS usually explain only a fraction of the total trait variance, and often a large portion of the genetically heritable variance remains undetermined by significant associations [8, 39]. Nevertheless, finding SNPs associated with only a fraction of the variation in traits of interest could still lead to progress through marker-assisted selection (MAS) or genomics-assisted selection (GAS) for beneficial wood properties [8].

In *Populus trichocarpa,* GWAS revealed SNPs and genes significantly associated with four wood chemical composition traits [20]. Furthermore, associations were discovered between SNPs and 16 wood chemical composition and wood structure traits in *P. trichocarpa* [43]. Wood chemical composition traits were also linked to SNPs by GWAS in *P. nigra* [21] and *P. deltoides* [11]. Xie et al. [66] re-evaluated previous associations in *P. trichocarpa* [36, 43] by focusing on a chromosome known to harbour quantitative trait locus (QTL) for lignin composition, resulting in the identification and characterization of a new transcriptional regulator of lignin biosynthesis. Using both single and multi-trait GWAS, 7 SNPs were identified in association to wood anatomical properties of a *P. trichocarpa* natural population [5].

Advances in genome (re)sequencing and statistical methods for finding associations in GWAS have facilitated these recent findings [8, 28]. Yet, the emerging picture of the genetics underlying highly quantitative or complex traits, such as wood properties and bioconversion potential, remains limited, in part due to our limited precision in the quantifications of these traits for entire tree populations [8, 58]. For example, lignin is composed of different types of monomers, and measurement of only the total amount of lignin in wood obscures the influence and the regulation of the abundance of the different types of lignin monomers [58]. Increasing the number of analysed traits and the depth of the analyses is likely needed for GWAS analyses of especially the complex traits [8, 25, 58]. Extensive phenotyping also allows better characterization of the relationships between traits, for example to identify which wood chemical composition and structure traits determine wood bioconversion potential.

Here, we present a large-scale phenotyping effort, monitoring 65 traits related to wood properties, tree growth, and wood saccharification in a common garden trial comprising a collection of natural aspen (*Populus tremula*) genotypes (the so-called SwAsp collection) collected across Sweden [29]. Through genetic correlation and multivariate analyses, we identified wood chemical composition and structural traits correlating with recalcitrance as well as whole stem bioconversion potential. Through GWAS, we identified several novel genetic loci linked to both tree growth and whole stem bioconversion potential.

Results

Natural variation in 65 growth, wood, and biorefinery traits in aspen

Natural variation in wood and biorefinery traits was investigated in 113 clonally replicated aspen trees of the SwAsp collection. After ten years of growth in a common garden in southern Sweden, we measured stem height and diameter, wood chemical composition (20 traits), wood structural and anatomical properties (17 traits), as well as recovery of monosaccharides from wood saccharification with or without acidic pretreatment (25 traits), amounting to 64 traits (Fig. 1a, Additional file 1). Finally, we estimated total wood glucose yield (TWG; Additional file 1). While glucose release provides information about biomass recalcitrance to saccharification, TWG provides a proxy for overall tree performance.

All traits showed phenotypic variation among the genotypes (Fig. 1a, Additional file 1). Around 30% of the total variation was explained by the two first components in a principle component analysis (Fig. 1b), with these largely being influenced by variation in the saccharification traits (Fig. 1c). Indeed, the saccharification traits, such as glucose release after enzymatic hydrolysis with pretreatment
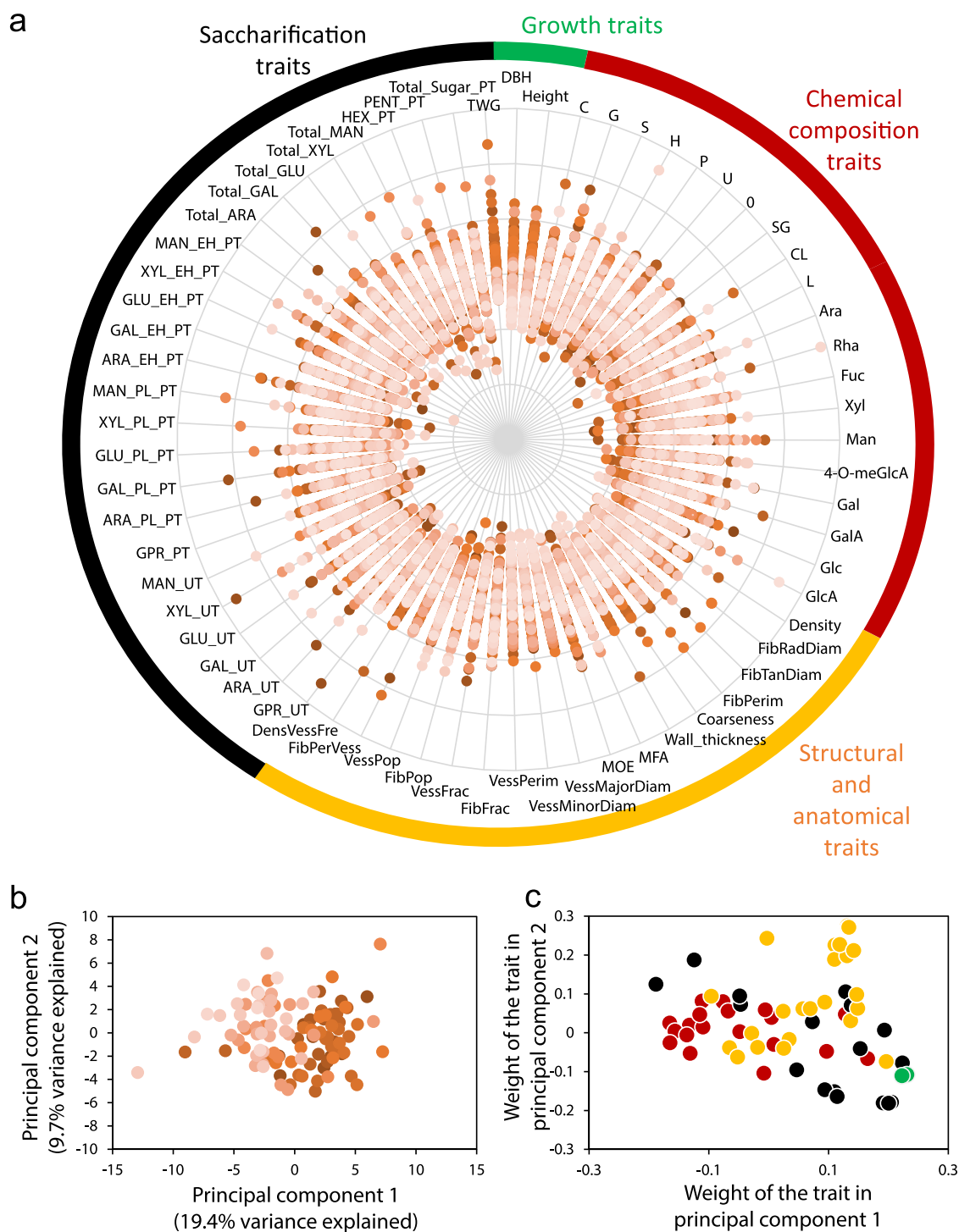
Fig. 1 The SwAsp natural variants display a wide range of phenotypic variation. **a** 65 traits related to tree growth, wood chemical composition, wood structure and anatomy, and saccharification traits of woody biomass in 10-year-old aspen trees from 113 genotypes. Each point represents the median scaled and centred measurement of a trait for one genotype (z-transformation across the tree population for each trait). Coloured labels around the plot indicate categories of traits (chemical composition; structure and anatomy; saccharification and growth). Abbreviations are defined in Additional file 1. **b** Principal component analysis (PCA) scatter plot showing that the SwAsp genotypes differ from each other based on their wood properties and saccharification. The broad-sense heritability for PC1 was 0.515 (95% confidence interval 0.404 to 0.623) and for PC2 0.189 (95% confidence interval 0.08 to 0.314). Colours indicate the 12 different locations of origin for the different genotypes in Sweden. **c** Coefficients scatter plot of traits. Each point corresponds to a trait; while the colours indicate which trait category they belong to (as in **a**)

and total wood glucose yield, displayed almost 50% increase from the lowest to the highest yielding genotype (Fig. 2, Additional file 1). Lignin traits, such as total lignin content and the ratio between the syringyl (S) and guaiacyl (G) type lignin (SG), that are central in determining feedstock recalcitrance, also varied substantially among the different genotypes (Fig. 2, Additional file 1). Tree growth varied most of all traits. Two genotypes (47 and 76) stood out as having remarkably high stem height and diameter (Additional file 1). These genotypes had also the highest TWG.

We estimated the broad-sense heritability (H^2) of the different traits (Additional file 2). Some traits, such as those linked to wood xylose units and xylose released by saccharification, showed nearly no heritability, while traits related to tree growth and wood anatomy showed moderate to high heritability ($H^2 > 0.5$). Wood chemical composition traits showed varying heritability; generally lower for wood monosaccharide units and higher for lignin composition traits, especially the S-type and G-type lignin content (Additional file 2).

Next, genetic correlations were estimated among the different traits. The tree growth traits (height and diameter) correlated positively with wood density, xylem cell wall thickness, xylem cell diameters and wood carbohydrate content (Fig. 3, Additional file 3). The correlations for the saccharification traits varied somewhat depending on the sugar analysed, but the release of sugars having the highest abundance in wood, glucose (GLUEHPT) and xylose (XYLEHPT), correlated positively with the growth traits of the trees and negatively with lignin content. Another striking result was that the glucose release GLUEHPT correlated negatively with the wood content of all hemicellulose sugar units (Ara, Fuc, Gal, GalA, GlcA, Man, 4-O-meGlcA, Rha, Xyl) (Fig. 3, Additional file 3). A slight positive correlation was present between GLUEHPT and the S-to-G lignin ratio (SG).

In a phenological study of the SwAsp population, the timing of bud set was shown to correlate with the geographical origin of the genotypes [61]. On the other hand, studies of secondary metabolites or leaf shape in that same population showed no correlation between these traits and the geographical origin of the genotypes [26, 37]. These previous observations raise the question of whether wood and biorefinery traits display a geographical cline. The growth traits, stem height and diameter, showed an expected, clear relationship to the geographical origin of the genotypes (Fig. 2, Additional file 1). Even though the traits related to wood chemical composition, wood anatomy and structure, and saccharification did not show clear geographic clines on a population level (Fig. 2), correlation analysis on the clonal basis showed significant effect of the geographic origin for

several traits. For instance, relative carbohydrate content, cell wall thickness, vessel diameter and wood density correlated negatively with the latitude of the clonal origin, while relative lignin content, content of hemicellulose sugar units and S/G lignin ratio correlated positively with the latitude (Additional file 1). Since the heritabilities for these traits were high (Additional file 2), these results suggest that aspen genotypes from northern Sweden had, on average, more total lignin, S-type lignin and hemicelluloses, and less carbohydrates, lower vessel diameter and wood density than genotypes from southern Sweden. Decreased wood density towards the north is surprising since in several tree species wood density normally increases with decreased volume growth [4].

Identification of traits that influence wood recalcitrance

To better characterize the traits influencing wood recalcitrance to bioprocessing, we performed multivariate analyses for the glucose release from saccharification, as well as for the TWG. We employed orthogonal projections to latent structures (OPLS; [55], which considers all traits simultaneously, to get an overview of the relationships between wood properties and glucose release or TWG (Fig. 4). OPLS models were created that explained high proportion of the variation for both the glucose release after enzymatic hydrolysis with pretreatment (GLUEHPT) and for TWG ($R^2 = 0.56$ and 0.52, respectively), but the predictivity of the model was not strong ($Q^2 = 0.17$ and 0.29, respectively). The OPLS models supported negative contribution of wood hemicellulose sugar units and lignin on both GLUEHPT and TWG, while several wood anatomy traits, such as diameter of the fibres and the vessels, the ratio of fibres to vessels (FibPerVess) and coarseness (weight of fibres over a certain length of wood), contributed positively to the models of both traits (Fig. 4).

Genetic polymorphisms are significantly associated with the total wood glucose yield

To further decipher the genetics underlying wood properties and amenability to improvements in bioprocessing, we performed a genome-wide association study (GWAS). In this analysis, latitude of origin of each SwAsp genotype was included as a covariate due to the presence of the latitudinal clines (Additional file 1). An FDR cutoff of 0.1 was selected to identify putative associations [52].

Single nucleotide polymorphisms (SNPs) with FDR value less than 0.1 (q-value < 0.1) were identified in 17 loci for five traits (Table 1, Additional file 4). Most of the SNPs were located in intergenic regions or upstream/ downstream of gene coding regions.

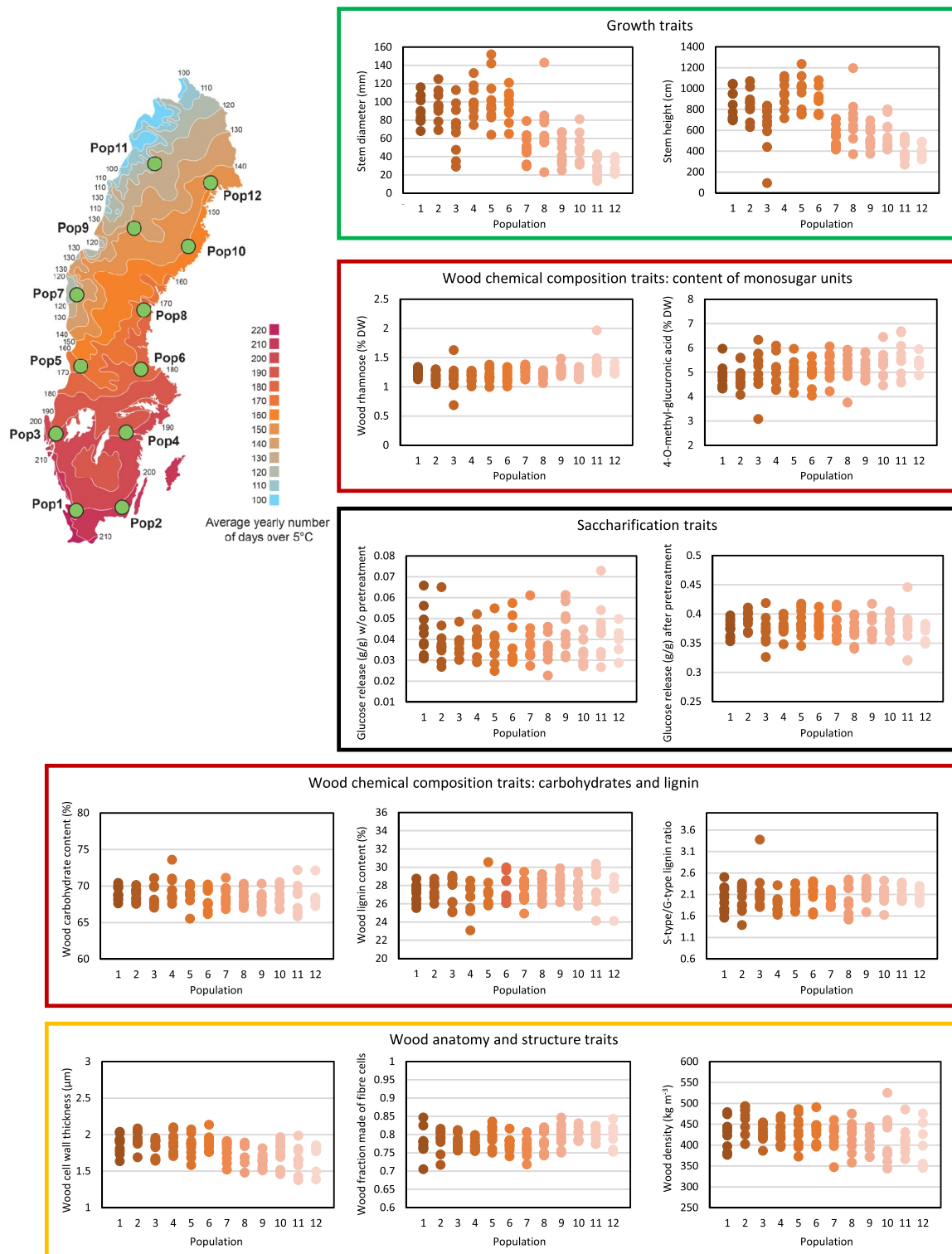No significant associations were observed for glucose release rates, but 11 associations for TWG with

Fig. 2 Wood and biorefining traits and geographical origin of the SwAsp trees. The average values are shown for key representative traits for growth, wood chemical composition, wood anatomy and structure, and traits related to saccharification for the 113 different genotypes of the SwAsp collection. The values are grouped according to the geographic origin of the genotypes in 12 locations across Sweden. The locations for the different geographic origins (Pop1–Pop12) are illustrated on the map in the upper left corner
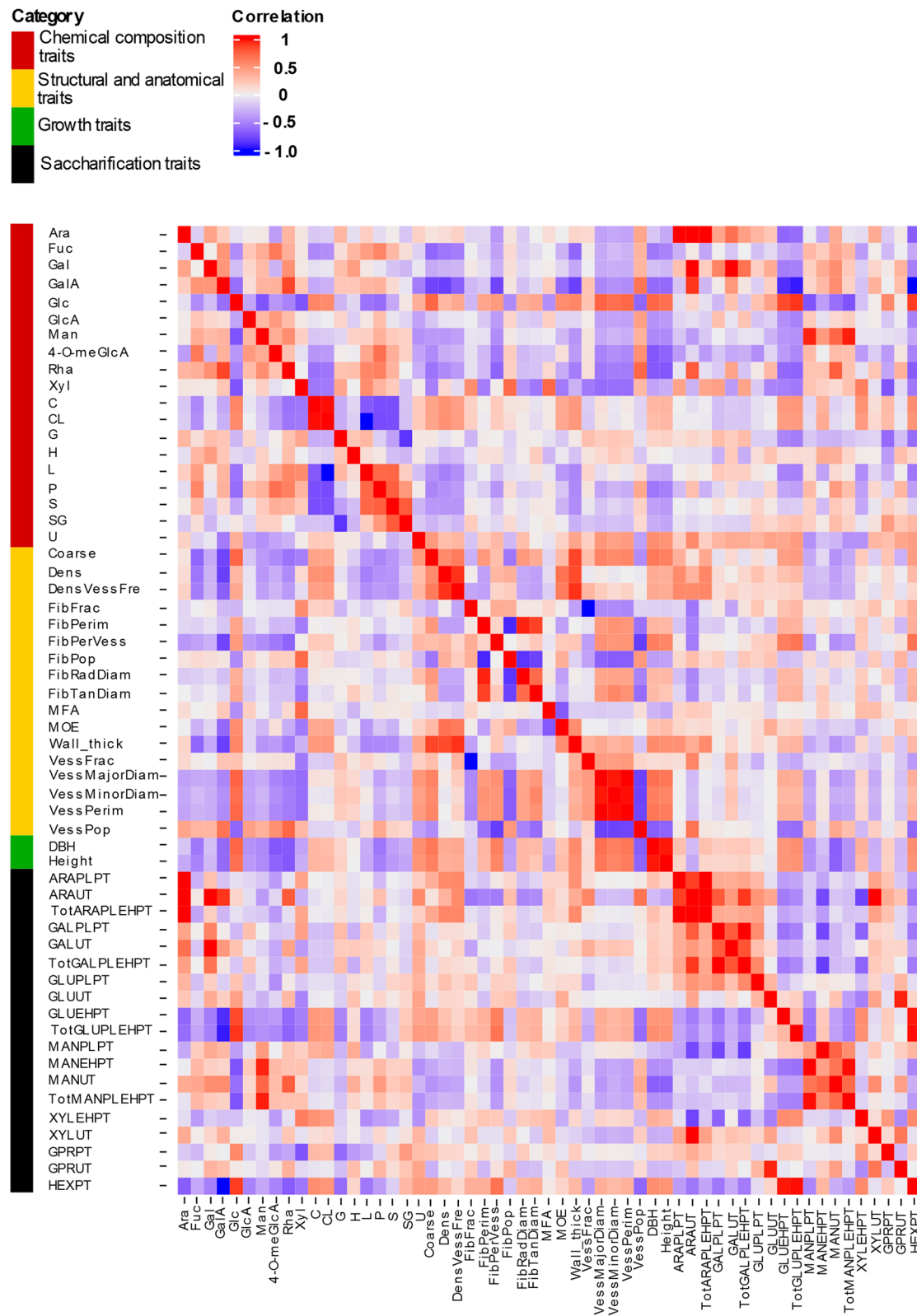
Fig. 3 Pairwise genetic correlations between 58 traits in SwAsp trees. The vertical sidebar represents the four categories of traits: wood chemical composition (red), wood structure and anatomy (yellow), growth (green) and wood saccharification (black). Six analysed SwAsp traits were omitted from the correlation analyses due to very low heritabilities. Trait abbreviations are defined in Additional file 1
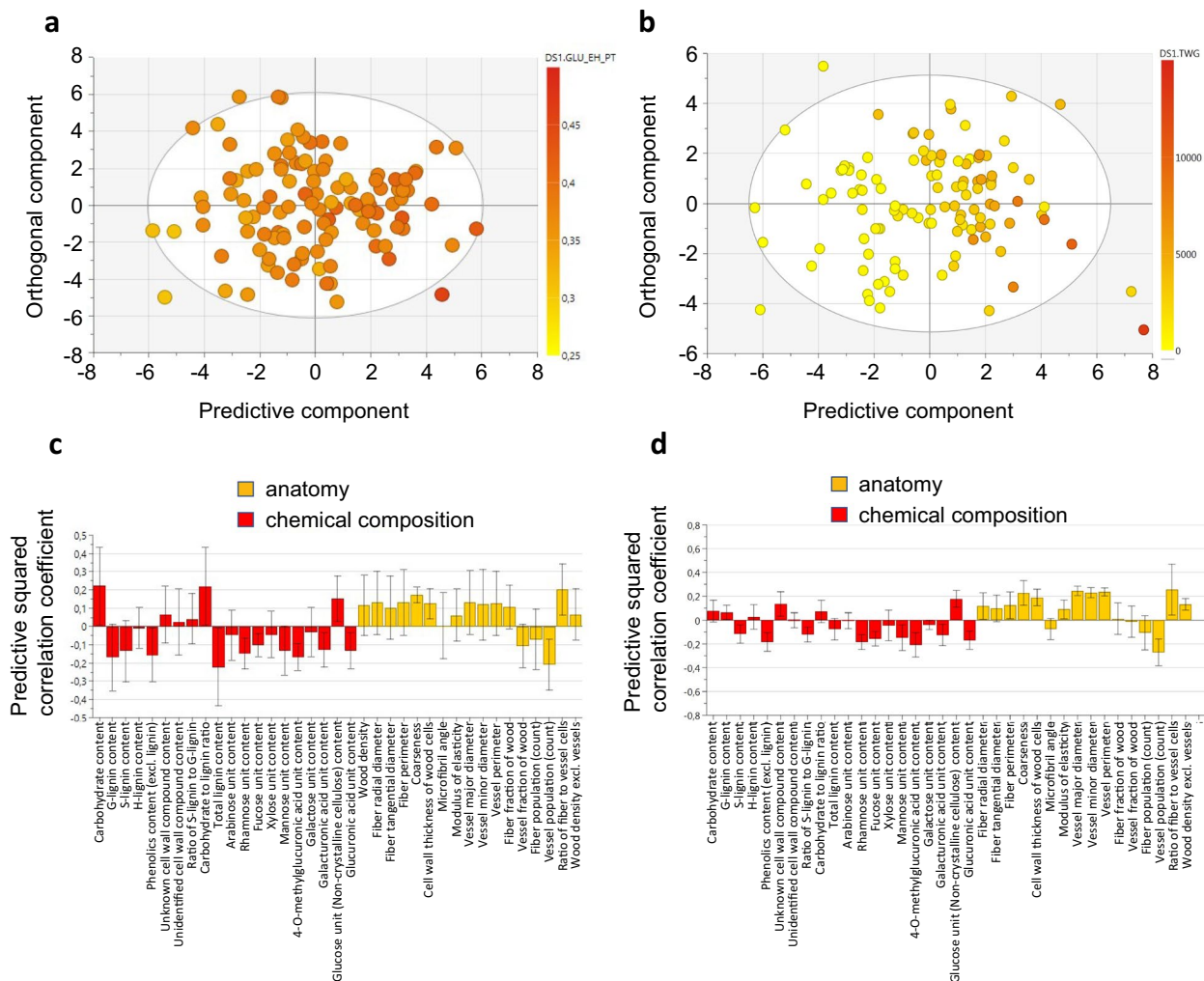
Escamez *et al. Biotechnology for Biofuels and Bioproducts*   (2023) 16:65

Page 7 of 16

Fig. 4 Multivariate analysis of the potential relationships between wood properties and glucose release or TWG. **a**, **b** Orthogonal Projection to Latent Structure (OPLS) scatter plot showing separation in glucose release after pretreatment **(a)** and total wood glucose yield **(b)**. The points on the scatter plot correspond to SwAsp genotypes, while their colour indicates the median for the trait in each genotype. The predictive component separates the lines along the X-axis of the scatter plot, while separation along the Y-axis is not predictive. **c**, **d** OPLS loadings plot for glucose release after pretreatment **(c)** and total wood glucose yield **(d)** in relation to wood chemical composition and wood anatomy traits. The bars indicate the coefficient ("weight") of each trait in the OPLS model. The traits with positive values correlate positively and the traits with negative values negatively with glucose release after pretreatment **(c)** and total wood glucose yield **(d)**. Predictive squared correlation coefficient (Q^2) scores over 0.5 indicate significant predictivity of a model

q-value < 0.1 were identified on 11 chromosomes, with each individual SNP explaining 22 to 26% of phenotypic variation (Table 1, Additional file 4). Six of the loci for TWG associations intercepted with loci containing SNPs for either stem diameter at breast height (DBH) or stem height (Height) (Table 1, Additional file 4). These six loci were all intergenic except for chr9_2882991_T_C which was located 362 bp upstream from Potra2n9c19049 (Oxa2A membrane insertase) and 757 bp downstream from Potra2n9c19048 (hypothetical protein) (Table 1, Additional file 4). The chr9_2882991_T_C SNP also showed statistically significant differences in

the phenotypes between the SwAsp genotype groups with homozygous and heterozygous alleles for not only TWG but also DBH and height (Fig. 5b–d). While the proportion of phenotypic variation explained by chr9_2882991_T_C was 0.26, the minor allele frequency was low (0.05) such that there was only one SwAsp genotype with a homozygous minor allele for this SNP (Fig. 5d, Additional file 4). In addition to the six loci intercepting with the DBH and/or height, GWAS revealed a locus with 12 TWG-SNP associations with q-value < 0.1 (Fig. 5a) in an intergenic region in the chromosome 1, in a region spanning 3327 base-pairs and including 38 SNPs

Table 1 Genes and genomic features associated with SNPs at *q*-value < 0.1 in the SwAsp genome-wide association study of 65 traits monitored in the Swedish aspen collection

Gene[1]	Feature[2]	Description(s)	The number of SNPs associated with each trait					
			DBH	Wall thickness	FibFrac	Height	VessFrac	TWG
7c16159; 7c16161	Intergenic		1					1
15c29048; 15c29049	Intergenic		1			1		1
9c19049; 9c19048	Upstream; downstream	Oxa2A membrane insertase; hypothetical protein	1					1
3c8002	Downstream	MATE efflux family protein	1					
1c2354; 1c2355	Intergenic					18		12
4c9179; 4c9180	Intergenic					2		2
5c11372; 5c11373	Intergenic					1		1
10c20558	Exonic	2-Oxoglutarate dehydrogenase, E1						1
10c21196; 10c21197	Upstream; downstream	Geranyl diphosphate synthase 1; hypothetical protein						1
16c29591	UTR3	L-Galactono-1,4-lactone dehydrogenase						1
17c31917; 17c31919	Intergenic							1
18c33227; 18c33228	Intergenic							1
10c22263; 10c22264	Intergenic							1
14c26895	Intronic	AGAMOUS-like 20						1
3c6553; 3c6554	Intergenic			6				
12c24540;12c24542	upstream; UTR3-downstream	DnaJ homolog subfamily B member; Major facilitator superfamily protein			5		5	
17c31483	Downstream	NA					1	

[1] The full name of the *P. tremula* gene models includes "Potra2n" in front of the "gene"

[2] The feature upstream and downstream indicates location of the SNPs within 2 kbp from the coding region, while the feature intergenic indicates location of the SNPs further than 2 kbp from the coding region

with R^2 values > 0.2, considered to be in linkage disequilibrium with the 12 significant SNPs (Additional file 4). The most significant SNP by *P*-value and *q*-value in this locus was chr1_28056992_G_A (with a major allele frequency of 0.132 and PVE of 0.26). The SwAsp genotype groups with homozygous and heterozygous alleles for the chr1_28056992_G_A significantly partitioned the variance of TWG as well as DBH and height (Fig. 5e–g). Out of the remaining putative associations for TWG, the only SNP that resided in the coding region of a gene, chr10_2830421_T_G, corresponded to Potra2n10c20558

(E1 subunit of 2-oxoglutarate dehydrogenase). Although this result is based on only two SwAsp genotypes (47 and 76) with the homozygous recessive allele (Fig. 5j, Additional file 4), it also showed statistically significant differences in the height and DBH phenotypes among the allele groups (Fig. 5h and i).

SNPs with *q*-value < 0.1 were found for the fraction of the wood made of fibres (FibFrac) and vessels (VessFrac) downstream of Potra2n12c24540 (DnaJ homolog subfamily member) and upstream of Potra2n12c24542 (Major facilitator superfamily protein member)

(See figure on next page.)
Fig. 5 Genome-wide association analysis of total wood glucose yield (TWG). **a** Manhattan plot for total wood glucose yield (TWG). Each point indicates location of a SNP along the 19 chromosomes of *Populus tremula*. The blue horizontal line indicates the *q*-value level of 0.1. The least significant SNPs (*P*-values > 0.05) have been omitted for plot clarity. **b**–**j** Tree height, diameter at breast height (DBH) and TWG in relation to their SNP genotype for the three most significant associations for TWG. Boxplots show phenotypic values of height, diameter and TWG amongst allele classes for the SNPs with the smallest p-values in the association tests; chr1_28056992_G_A had the highest statistical significance among the SNPs in the chromosome 1 GWAS hotspot for TWG (see also Additional file 4). The jittered points around each box represent median phenotypic values of the SwAsp clonal replicates. Analysis of variance F-ratios and *P*-values are reported, where the dependent variable is the phenotype and the independent variable is the SNP genotype class The significances are indicated at. < 0.1, * < 0.05, and ** < 0.01. The three traits are shown for the SNPs chr9_2882991_T_C (**b**–**d**), chr1_28056992_G_A (**e**–**g**) and chr10_2830421_T_G (**h**–**j**)
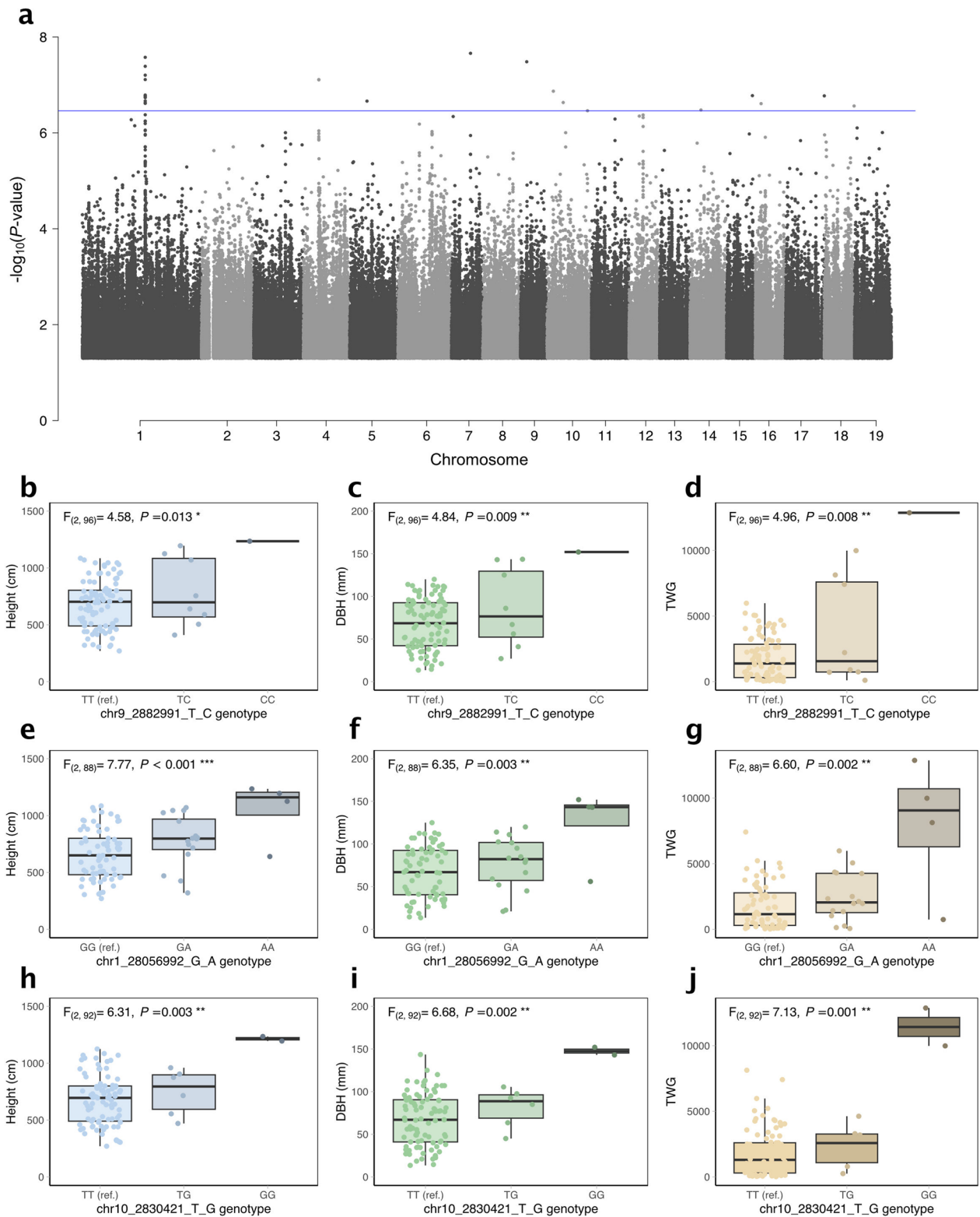
Fig. 5 (See legend on previous page.)

(Table 1, Additional file 4). Furthermore, six SNPs with *q*-value < 0.1 were identified for cell wall thickness of the wood cells (wall thickness) in an intergenic region.

Discussion

Wood biomass from fast growing trees represents a promising source of biofuels and other bioproducts in the forthcoming transition away from fossil fuels [41, 47]. The high cost of deconstructing woody biomass, however, hinders wood biorefining [32]. To overcome this biomass recalcitrance, it is necessary to understand how wood properties relate to wood recalcitrance. We report here the phenotyping of a population of aspen genotypes for 65 traits related to tree growth, wood anatomy and structure, wood cell wall chemical composition, and wood bioprocessing yield.

Using genetic correlations and multivariate modelling, we identified a set of wood traits that correlate with the glucose yield from saccharification. Lignin content and especially G-lignin content had a negative influence on the glucose yield after enzymatic hydrolysis with pretreatment, which is in line with a positive effect of S/G ratio in our earlier analysis of 40 transgenic *Populus* lines [10], as well as in the analyses of *P. trichocarpa* [53, 68] and *Salix viminalis* [40] natural variants. A negative effect of S/G ratio was reported in a small selection of natural *P. trichocarpa* variants, and it was proposed that S/G ratio might instead influence xylose release after enzymatic hydrolysis [33]. We could not confirm this as no correlation was found between xylose release after enzymatic hydrolysis and S/G ratio in our dataset (Fig. 3). In addition to lignin, a consistent negative influence on sugar (glucose) yields after saccharification with pretreatment was imposed by the hemicellulose sugars (Figs. 3 and 4). This is most probably related to the fact that the pretreatment was adjusted to a rather mild level of severity, resulting in part of the hemicelluloses remaining intact in the feedstock: the wood xylose unit content was 0.2–0.3 g/g DW depending on the clone, while 0.12–0.16 g/g DW xylose was retrieved from the biomass into the pretreatment liquid (Additional file 1). Hemicelluloses are, in addition to lignin, the most important wood recalcitrance factors [31]. It is therefore likely that the hemicelluloses retained in the wood after the pretreatment limited the saccharification efficiency. Furthermore, the contents of the hemicellulose sugars also correlated in a similar, negative fashion with the relative carbohydrate content of wood (Fig. 2), which could also contribute to the negative influence of the hemicellulose sugars on the glucose release.

Identifying the genetics underlying wood properties that foster bioprocessing potential help with selecting or creating superior biorefinery feedstocks [11]. GWAS has frequently been used to identify single nucleotide polymorphisms associated with wood properties [5, 43, 58]. GWAS for saccharification traits is rare in forest tree species, but a notable association was found in *Salix viminalis* for glucose release in a non-coding region of the genome [40]. We did not find any associations for glucose release, but several loci of putative associations for the total wood glucose yield (Table 1). One of these was located in the coding region of E1 subunit of 2-oxoglutarate dehydrogenase (E1-OGDH) (Potra2n10c20558) which participates in the mitochondrial TCA cycle to provide reducing power for oxidative phosphorylation and carbon skeletons for various metabolic pathways. Since variation in E1-OGDH was not only linked to variation in TWG but also tree height and diameter (Fig. 5h–j), it is possible that the association between E1-OGDH and TWG was caused by variation in tree growth. This is supported by the well-known relationship between mitochondrial metabolism and growth both in plants and animals [38, 45]. Also TCA cycle has been linked to plant productivity [69], and OGDH has been proposed as one of the enzymes controlling the flux through the TCA cycle [1]. Work in *Arabidopsis* has shown that mutations in the two E1-OGDH genes (AT3G55410 and AT4G26910) resulted in impaired photosynthesis, reduced levels of chlorophyll and nitrate, reduced fitness and reduced growth [6]. The exact mechanism underlying the observed reductions in the different growth traits was, however, not clarified. The OGDH substrate 2-oxoglutarate is important for several different metabolic processes including biosynthesis of some amino acids. It is therefore possible that natural variation in E1-OGDH contributes to the activity of the enzyme to control not only the flux of the TCA cycle, but perhaps also entry of 2-oxoglutarate into the various metabolic pathways, such as biosynthesis of amino acids that are needed for tree growth. Interestingly, the Oxa2A membrane insertase (Potra2n9c19049) and the geranyl diphosphate synthase (Potra2n10c21196), located in close proximity to a SNP for TWG, are both mitochondrial proteins [9, 27], supporting the link between mitochondrial function and TWG yield.

TWG is a composite trait consisting of glucose yields after saccharification on a whole-tree-biomass basis. Consequently, we identified six polymorphic loci that intercepted for TWG and the biomass-related parameters of tree height and stem diameter (Table 1), pointing out the importance of tree biomass yields on TWG. For breeding purposes, it is an interesting question which one is more important for saccharification yields on a whole tree basis; biomass production or saccharification efficiency (per g material). In an earlier study of transgenic hybrid aspen trees, we

found that increased biomass production compensated for the loss of glucose release (per gram) caused by the transgene expression [15]. Vice versa, gains from increased glucose release after saccharification of lignocellulosic feedstocks have frequently been offset by decreased biomass production of trees [3, 22, 59, 60]. It therefore seems that saccharification yields can be efficiently increased simply by increasing biomass production of the trees. Biomass production is influenced by tree volume and wood density, and it was interesting that in the currently investigated population of aspen trees it was the tree volume-related traits of height and diameter that correlated better than wood density with the glucose release after saccharification with pretreatment (Fig. 3), and that tree height and diameter were the traits that intercepted with TWG in the GWAS analysis (Fig. 5). Furthermore, the two genotypes with the highest stem volume had also the highest TWG (Additional file 1). These results imply that the simple measurements of tree height and diameter might be sufficient to predict saccharification yields on a whole-tree basis. Earlier studies have seldom approached this question since saccharification has traditionally been defined on a process basis, resulting in identification of chemical composition as the most important factor influencing sugar yields. However, a very similar conclusion was drawn in a recent analysis of field grown *P. trichocarpa* trees where stem diameter was identified as the main driver for ethanol yield on a whole-field basis [23]. This leads to the question of what is the impact of tree volume on the other traits determining biomass production or saccharification. Notably, positive genetic correlation existed in our population between the tree volume traits and wood density as well as glucose release, which both act to increase the TWG (Fig. 3). Furthermore, negative correlation existed between tree growth and lignin content and hemicellulose sugars, implying that breeding efforts towards increased tree volume production might, similar to the currently investigated material, suppress the accumulation of wood chemical properties that have negative influence on glucose release rate and TWG.

Conclusions

We identified significant natural variation in growth and wood-related traits in aspen, which allowed identification of chemical and genetic markers for bioprocessing purposes of lignocellulosic feedstocks. Our data indicate that whole tree saccharification yields can be improved, at least in *Populus* feedstocks, by simply breeding for increased tree volume growth without a negative impact on wood parameters, such as wood density or the content of lignin and hemicelluloses, that also influence saccharification efficiency and yield. An outstanding example of this was the two genotypes that had the largest stem volume as well as the highest saccharification yields within the SwAsp population. Interestingly, polymorphism in a mitochondrial TCA cycle enzyme OGDH associated with the variation in both the volumetric tree traits and the saccharification yield, and is hence an interesting candidate for a genetic marker linked to stem volume and saccharification yield.

Materials and methods

Plant material

The Swedish Aspen (SwAsp) collection consists of 113 *Populus tremula* aspen genotypes from 12 locations across Sweden [29]. The genotypes represent potential sub-populations (Fig. 2), but whole-genome sequencing and sequence comparisons have shown that these genotypes are mostly unrelated [61].

The genotypes were clonally propagated in 2003 from root cuttings and grown in a randomized block experiment in a plantation in southern Sweden (Ekebo, 55.9°N). Three to five trees per genotype were successfully established in 2004 [29, 61].

After ten years of growth, tree height and diameter at breast height (DBH) were measured, and wood samples were collected from the stem. At 79 cm above ground, a 1-cm-thick section of the stem was collected, and the south-western facing quarter of the stem section was aliquoted for wood chemical composition analyses. In addition, 80–90 cm above ground, another piece of stem was harvested for analysis of wood anatomical and structural properties from the south-western facing quarter of the stem section. We obtained a full set of successful phenotypic measurements for a total of 418 trees (Additional file 1).

Analyses of wood chemical composition

The wood quarters selected for compositional analyses were manually debarked, cut into roughly match-stick-sized wood pieces and freeze dried (CoolSafe Pro 110–4, LaboGene A/S, Denmark). This material was homogenized by coarse milling (Retsch ZM 200 centrifugal mill, Retsch GmbH, Germany) and sieved (Retsch AS 200) into two particle size fractions. The fraction of particle size between 0.1 mm and 0.5 mm was aliquoted for subsequent saccharification experiments (see below), while the fraction of particle size under 0.1 mm was aliquoted for pyrolysis coupled with gas chromatography followed by mass spectrometry analysis (pyrolysis-GC/MS) and monosaccharide composition analysis. Both analyses were performed as technical duplicates for each tree.

Carbohydrate content, lignin content, lignin composition, and content of other phenolics were determined by pyrolysis-GC/MS as previously described [17]. Briefly, 40 μg—80 μg of homogenized wood powder was loaded into an autosampler (PY-2020iD and AS-1020E, Frontier Labs, Japan), allowing a sub-sample (∼1 μg) into the pyrolizer of the GC/MS apparatus (Agilent, 7890A/5975C, Agilent Technologies AB, Sweden). Following pyrolysis, the samples were separated along a DB-5MS capillary column (30 m × 0.25 mm i.d., 0.25-μm-film thickness, J&W, Agilent Technologies), and scanned by the mass spectrometer along the m/z range 35–250. The GC/MS data were processed as previously described [16]. Results were normalized by expressing the area of each peak as a percentage of the total peak area considering all peaks.

Cell wall monosaccharide units were quantified following the acidic methanolysis and trimethylsilyl (TMS) derivatization method as described previously [14]. Briefly, wood powder was washed with HEPES buffer (4 mM, pH 7.5) containing 80% ethanol, as well as methanol:chloroform 1:1 (V:V) and acetone to generate alcohol-insoluble residues (AIRs) which were then dried. To avoid contamination with glucose from starch, the AIRs were treated with 1 unit per AIR mg of type I α-amylase (Roche 10102814001, Roche GmbH, Germany). The de-starched AIRs, and inositol as an internal standard, were methanolysed using 2 M HCl/MeOH at 85 °C for 24 h. Following repeated washes with methanol, the samples and standard were silylated using Tri-sil reagent (3–3039, SUPELCO, Sigma-Aldrich, Germany) at 80 °C for 20 min. The solvent was evaporated under a stream of nitrogen and pellets were dissolved in 1 mL hexane and filtered through glass wool. The filtrates were evaporated until 200 μL remained, of which 0.5 μL were analysed by GC/MS (7890A/5975C; Agilent Technologies AB, Sweden) according to Sweeley et al. [54]. The levels of the sugars and sugar acids are presented in the hydrous form.

Saccharification assays and total wood glucose yield (TWG)

Saccharification assays without or with acid pretreatment of the biomass were performed following an established methodology [14]. In short, 50 mg of dry wood powder (moisture measured with an HG63 moisture analyser, Mettler-Toledo, USA) with particle size between 0.1 mm and 0.5 mm were pretreated with 1% (w/w) sulphuric acid (fraction of sulphuric acid based on the mass of the whole reaction mixture) during 10 min at 165 °C in a single-mode microwave system (Initiator Exp, Biotage, Sweden), or remained untreated. The pretreated samples were centrifuged to separate the solid fraction from the pretreatment liquid. The solid fraction was washed with

ultrapure water and sodium citrate buffer (50 mM, pH 5.2). The washed, pretreated solid fraction as well as the untreated samples were enzymatically hydrolysed 72 h at 45 °C under agitation, using 25 mg of a 1:1 (w/w) mixture of liquid enzyme preparations Celluclast 1.5 L (measured CMCase activity of 480 units per gram of liquid enzyme preparation, following Ghose [18] and Novozyme 188 (measured β-glucosidase activity of 15 units per gram liquid enzyme preparation, following Mielenz [34] (Sigma-Aldrich). Sodium citrate buffer (50 mM, pH 5.5) was added to reach 1 g of final reaction mixture. During enzymatic saccharification, samples were collected at 2 h and 72 h. Glucose production rates were determined at 2 h using an Accu-Chek ®Aviva glucometer (Roche Diagnostics Scandinavia AB, Sweden). Monosaccharide (arabinose, galactose, glucose, xylose and mannose) yields in pretreatment liquids and enzymatic hydrolysates collected at 72 h were determined using high-performance anion-exchange chromatography with pulsed amperometric detection (Ion Chromatography System ICS-5000, Dionex, USA) as previously described [63]. Saccharification was performed on technical duplicates for each tree.

The total-wood glucose yield from an entire tree trunk (TWG) was calculated using the formula $TWG = 1/3 \times \pi \times height \times (diameter/2)^2 \times wood \; density \times glucose \; release_{(AFTER \; PRETREATMENT)}$, as previously described [10], assuming a conical shape of the tree stem.

Anatomical and structural characterization

Anatomical and structural features were determined on parallelepipedal wood pieces across the stem diameter using the SilviScan® instrument (CSIRO, Australia) which consists of three separate units: (i) a cell scanner with a video microscope for measurement of the numbers and sizes of fibres and vessels; (ii) a density scanner recording X-ray absorption images for measuring wood density; and (iii) a diffraction scanner recording X-ray diffraction images for measuring the microfibril angle. The measurements on these parallelepipedal wood pieces were then projected onto the entire wood section to reflect the average values for the entire wood section of each tree. Full description of the different traits from the SilviScan measurements can be found in Additional file 1.

Statistical estimations of the genetic parameters

The genetic parameters for each trait were estimated statistically based on measurements on individual trees for each genotype according to the model $Y_{ijk} = \mu + b_i + c_j + e_{ijk}$ where Y_{ijk} is the observation k in block i for clone j, μ is the mean of the trait in this trial, b_i is the fixed effect of block i, c_j is the random effect of clone j (normally and independently distributed

with mean 0 and variance V_c; NID[0,V_c]), and e_{ijk} is the random error term for observation ijk (NID[0,V_e]). The variances V_c and V_e were estimated for each trait according to the Restricted Maximum Likelihood (REML) method using the ASREML software [19]. To estimate genetic parameters, we considered that V_c is equal to V_G (the genotypic variance among clones for the trait) and V_e is equal to V_E (the environmental variance for the trait). Correlation analysis was not performed for traits ARA_EH_PT, GAL_EH_PT, Total_SUGAR_PT, Total_XYL_PL + EH_PT, XYL_PL_PT and PENT_PT due to very low heritabilities.

For each trait, broad-sense heritability (H^2) was estimated by dividing genotypic variance (V_G) by the total variance of this trait V_T where $V_T = V_G + V_E$. The genotypic coefficient of variation (CV_G) for a trait was calculated by dividing the genotypic standard deviation of the trait $\sqrt{V_G}$ by the mean value of the trait ($\bar{x}$), and multiplying the result by 100. The genetic correlation (r_G) between trait 1 with genotypic variance V_{G1} and trait 2 with genotypic variance V_{G2} was calculated by dividing the genotypic genetic covariance (cov_{G1G2}) between these traits by the square root of the product of their individual genetic variances; $r_G = cov_{G1G2} / \sqrt{V_{G1} \times V_{G2}}$.

Multivariate analyses

Multivariate analyses using all wood traits to predict glucose release by saccharification, or total wood glucose yield (TWG), were performed using Orthogonal Projections to Latent Structures (OPLS) regression [55], with 1 + 3 components.

Genome-wide association study (GWAS)

Phenotypic data were subjected to a scripted pipeline, comprising a set of quality control steps and the estimation of a best linear unbiased predictor (BLUP) phenotypic value for each SwAsp genotype and for each trait in the GWAS. The pipeline is described in [49] and scripts are available at https://github.com/sarawestman/Genome_paper. Briefly, phenotypic outliers were removed using the 'OutlierTest' function of the 'car' package in R [46],version 3.0.10; [13], phenotypes were tested with the Shapiro–Wilk test and any non-normally distributed random effects or error terms were transformed using an Ordered Quantile normalization in the 'bestNormalize' package in R (version 1.6.1, [42]). Subsequently, a BLUP with a restricted maximum likelihood approach was

used to estimate the genotypic effect of a given phenotype, as detailed in Wang et al. [61] using the model $z_{jkl} = u + b_j + g_k + e_{jkl}$ where z_{jkl} is the phenotype of the *lth* individual in the *jth* block from the *kth* genotype, u is the grand mean and e_{jkl} is the residual error term. The genotype and residual terms were considered random effects, and field block was considered a fixed effect.

Details of the SwAsp DNA sequencing and SNP calling, filtering and functional annotation have been described previously [48, 49], resulting in 99 unrelated individual genotype sequences for GWAS with 6,806,717 bi-allelic SNPs. Similar to Mähler et al. [37], SNPs were considered as intergenic if they laid further than 2 kbp away from a gene, while SNPs within 2 kbp of a gene were considered associated with that gene.

Genome-wide association mapping was conducted using GEMMA [70] with univariate Linear Mixed Models (LMMs), which are association tests between SNP markers and phenotypic BLUP values. Two covariates were included in the GWAS model: although relatedness in the SwAsp collection was weak, the first covariate was a relatedness matrix of all individuals in the study that was centre-scaled in GEMMA (using the parameter "-gk 1") as previously described [61]; the second was the latitude of origin of each SwAsp genotype, which was applied to eliminate any spurious associations resulting from size differences of the trees that result from the latitudinal sampling cline that influences seasonality-determined growth in the SwAsp collection [29]. False discovery rate (FDR) of each association was calculated as the "*q*-value" using R [51] following the principle of the Benjamini–Hochberg procedure [52]. The percentage of phenotypic variance explained (PVE) by each SNP, for each trait, was also estimated using the formula described previously [61]. GWAS results were visualized using Manhattan plots generated in the 'qqman' package in R [56]. Allele boxplots were generated using the 'ggplots2' package in R [64]. The distributions of the phenotypic data were tested for normality using Shapiro–Wilk tests, transformed with ordered quartile normalization (described above), and homogeneity of variances tested with a Bartlett test in R prior to analyses of variance amongst SNP genotype groups. The 'anova' function was applied in R to a linear model where the phenotype was the dependent variable and the SNP allele class the independent variable. Boxplots were plotted using the ggplots2 package in R [64].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13068-023-02315-1.

Additional file 1: The median value of each trait for each of the SwAsp *Populus* lines and correlation of each trait with the latitude of genotype origin.

Additional file 2: Statistical estimation of broad-sense heritability H^2 and genetic coefficient of variation CV(G) for all the monitored traits.

Additional file 3: Genetic correlations between the tree growth, wood and saccharification traits.

Additional file 4: Significant (*q*-value <0.1) single nucleotide polymorphisms (SNPs) identified in the genome-wide association study (GWAS) of 65 traits recorded in the Swedish aspen common garden.

Author contributions
HT designed the study, with help from KMR, SJ, GS, LGS, LJJ, NRS, and SE. Phenotypic characterization of the trees was performed by SE, ML, MLG, KMR, and TG, with supervision by LJJ, GS and HT. OPLS modelling of traits was performed by SE. LGS performed analyses of broad-sense heritability and genetic correlations. Genome-wide association study for identification of SNPs was performed by KMR, NM and NRS. SE and HT wrote the manuscript, with assistance from all co-authors. All authors read and approved the final manuscript.

Availability of data and materials
The phenotypic dataset on tree growth, wood properties and saccharification traits in the SwAsp population is available in the Dryad repository, https://doi.org/10.5061/dryad.gtht76hrd. All other data used for analyses in this manuscript are either displayed in the additional files, or available upon request to the corresponding author.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors consent to the publication of this manuscript.

Competing interests
HT, NRS and SJ are shareholders in Woodheads AB. LJJ is inventor of patents in the area biomass processing.

Author details
[1]Department of Plant Physiology, Umeå Plant Science Centre (UPSC), Umeå University, 901 87 Umeå, Sweden. [2]Department of Chemistry, Umeå University, 901 87 Umeå, Sweden. [3]RISE AB, Drottning Kristinas Väg 61 B, 114 28 Stockholm, Sweden. [4]The Forestry Research Institute of Sweden, Ekebo, 268 90 Svalöv, Sweden. [5]Present Address: Department of Forest Genetics and Plant Physiology, Umeå Plant Science Centre (UPSC), Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden.

References
1. Araújo WL, Martins AO, Fernie AR, Tohge T (2014) 2-Oxoglutarate: linking TCA cycle function with amino acid, glucosinolate, flavonoid, alkaloid, and gibberellin biosynthesis. Front Plant Sci. 15:552
2. Bar-On YM, Phillips R, Milo R. The biomass distribution on earth. Proc Natl Acad Sci. 2018;2018(115):6506–11.
3. Bonawitz ND, Chapple C. Can genetic engineering of lignin deposition be accomplished without an unacceptable yield penalty? Curr Opin Biotech. 2013;24:336–43.
4. Carlquist S. Ecological strategies of xylem evolution. Berkley: University of California Press; 1975.
5. Chhetri HB, Furches A, Macaya-Sanz D, Walker AR, Kainer D, Jones P, et al. Genome-wide association study of wood anatomical and morphological traits in *Populus trichocarpa*. Front Plant Sci. 2020;11:1391.
6. Condori-Apfata JA, Batista-Silva W, Medeiros DB, Vargas JR, Valente LML, Heyneke E, et al. The Arabidopsis E_1 subunit of the 2-oxoglutarate dehydrogenase complex modulates plant growth and seed production. Plant Mol Biol. 2019;101:183–202.
7. Dickmann DI. Silviculture and biology of short-rotation woody crops in temperate regions: then and now. Biomass Bioenerg. 2006;30:696–705.
8. Du Q, Lu W, Quan M, Xiao L, Song F, Li P, et al. Genome-wide association studies to improve wood properties: challenges and prospects. Front Plant Sci. 2018;9:1912.
9. Ducluzeau AL, Wamboldt Y, Elowski CG, Mackenzie SA, Schuurink RC, Basset GJ. Gene network reconstruction identifies the authentic trans-prenyl diphosphate synthase that makes the solanesyl moiety of ubiquinone-9 in Arabidopsis. Plant J. 2012;69:366–75.
10. Escamez S, Gandla ML, Derba-Maceluch M, Lundqvist S-O, Mellerowicz EJ, Jönsson LJ, et al. A collection of genetically engineered *Populus* trees reveals wood biomass traits that predict glucose yield from enzymatic hydrolysis. Sci Rep. 2017;7:15798.
11. Fahrenkrog AM, Neves LG, Resende MF, Vazquez AI, Campos G, Dervinis C, et al. Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*. New Phytol. 2017;213:799–811.
12. Faivre-Rampant P, Zaina G, Jorge V, Giacomello S, Segura V, Scalabrin S, et al. New resources for genetic studies in *Populus nigra*: genome-wide SNP discovery and development of a 12k infinium array. Mol Ecol Resour. 2016;16:1023–36.
13. Fox J, Weisberg S. An R companion to applied regression. 3rd ed. Thousand Oaks: Sage; 2019.
14. Gandla ML, Derba-Maceluch M, Liu X, Gerber L, Master ER, Mellerowicz EJ, et al. Expression of a fungal glucuronoyl esterase in *Populus*: Effects on wood properties and saccharification efficiency. Phytochemistry. 2015;112:210–20.
15. Gandla ML, Mähler N, Escamez S, Skotare T, Obudulu O, Möller L, et al. Overexpression of vesicle-associated membrane protein *PttVAP27-17* as a tool to improve biomass production and the overall saccharification yields in *Populus* trees. Biotechnol Biofuels. 2021;14:43.
16. Gerber L, Eliasson M, Trygg J, Moritz T, Sundberg B. Multivariate curve resolution provides a high-throughput data processing pipeline for pyrolysis-gas chromatography/mass spectrometry. J Anal Appl Pyrol. 2012;95:95–100.
17. Gerber L, Öhman D, Kumar M, Ranocha P, Goffner D, Sundberg B. High-throughput microanalysis of large lignocellulosic sample sets by pyrolysis-gas chromatography/mass spectrometry. Physiol Plantarum. 2016;156:127–38.
18. Ghose TK. Measurement of cellulase activities. Pure Appl Chem. 1987;59:257–68.
19. Gilmour A, Thompson R, Cullis B, Welham S. ASReml [computer program]. 1997. NSW Agriculture, Orange, Australia. https://vsni.co.uk/software/asreml.

20. Guerra FP, Suren H, Holliday J, Richards JH, Fiehn O, Famula R, et al. Exome resequencing and GWAS for growth, ecophysiology, and chemical and metabolomic composition of wood of *Populus trichocarpa*. BMC Genom. 2019;20:1–14.

21. Guerra FP, Wegrzyn JL, Sykes R, Davis MF, Stanton BJ, Neale DB. Association genetics of chemical wood properties in black poplar (*Populus nigra*). New Phytol. 2013;197:162–76.

22. Ha CM, Rao X, Saxena G, Dixon RA. Growth-defense trade-offs and yield loss in plants with engineered cell walls. New Phytol. 2021;231:60–74.

23. Happs RM, Bartling AW, Doeppke C, Harman-Ware AE, Clark R, Webb EG, et al. Economic impact of yield and composition variation in bioenergy crops: *Populus trichocarpa*. Biofuel Bioprod Biorefin. 2021;15:176–88.

24. Hou Z, Li A, Zhang J. Genetic architecture, demographic history, and genomic differentiation of *Populus davidiana* revealed by whole-genome resequencing. Evol Appl. 2020;13:2582–96.

25. Ingvarsson PK, Street NR. Association genetics of complex traits in plants. New Phytol. 2011;189:909–22.

26. Keefover-Ring K, Ahnlund M, Abreu IN, Jansson S, Moritz T, Albrectsen BR. No evidence of geographical structure of salicinoid chemotypes within *Populus tremula*. PLoS ONE. 2014;9: e107189.

27. Kolli R, Soll J, Carrie C. OXA2b is crucial for proper membrane insertion of COX2 during biogenesis of complex IV in plant mitochondria. Plant Physiol. 2019;179:601–15.

28. Lin Y-C, Wang J, Delhomme N, Schiffthaler B, Sundström G, Zuccolo A, et al. Functional and evolutionary genomic inferences in *Populus* through genome and population sequencing of American and European aspen. Proc Natl Acad Sci USA. 2018;115:10970–8.

29. Luquez V, Hall D, Albrectsen BR, Karlsson J, Ingvarsson P, Jansson S. Natural phenological variation in aspen (*Populus tremula*): the SwAsp collection. Tree Genet Genomes. 2008;4:279–92.

30. Ma T, Wang J, Zhou G, Yue Z, Hu Q, Chen Y, et al. Genomic insights into salt adaptation in a desert poplar. Nat Commun. 2013;4:1–9.

31. Martín C, Dixit P, Momayez F, Jönsson LJ. Hydrothermal pretreatment of lignocellulosic feedstocks to facilitate biochemical conversion. Frontiers Bioeng Biotechnol. 2022;10: 846592.

32. McCann MC, Carpita NC. Biomass recalcitrance: a multi-scale, multi-factor and conversion-specific property. J Exp Bot. 2015;66:4109–18.

33. Meng X, Pu Y, Yoo CG, Li M, Bali G, Park DY, et al. An in-depth understanding of biomass recalcitrance using natural poplar variants as the feedstock. Chemsuschem. 2017;10:139–50.

34. Mielenz JR. Biofuels: methods and protocols, methods in molecular biology, vol. 581. New York: Humana Press; 2009.

35. Mola-Yudego B, Arevalo J, Díaz-Yáñez O, Dimitriou I, Haapala A, Ferraz Filho AC, et al. Wood biomass potentials for energy in northern Europe: forest or plantations? Biomass Bioenerg. 2017;106:95–103.

36. Muchero W, Guo J, DiFazio SP, Chen J-G, Ranjan P, Slavov GT, et al. High-resolution genetic mapping of allelic variants associated with cell wall chemistry in *Populus*. BMC Genom. 2015;16:1.

37. Mähler N, Schiffthaler B, Robinson KM, Terebieniec BK, Vučak M, Mannapperuma C, et al. Leaf shape in *Populus tremula* is a complex, omnigenic trait. Ecol Evol. 2020;10:11922–40.

38. Møller IM, Rasmusson AG, Van Aken O. Plant mitochondria—past, present and future. Plant J. 2021;108:912–59.

39. Nordborg M, Weigel D. Next-generation genetics in plants. Nature. 2008;456:720–3.

40. Ohlsson JA, Hallingbäck HR, Jebrane M, Harman-Ware AE, Shollenberger T, Decker SR, et al. Genetic variation of biomass recalcitrance in a natural Salix viminalis (L.) population. Biotechnol Biofuels. 2019;12:135.

41. Percival Zhang YH. Next generation biorefineries will solve the food, biofuels, and environmental trilemma in the energy–food–water nexus. Energy Sci Eng. 2013;1:27–41.

42. Peterson RA, Cavanaugh JE. Ordered quantile normalization: a semiparametric transformation built for the cross-validation era. J Appl Stat. 2020;47:2312–27.

43. Porth I, Klapšte J, Skyba O, Hannemann J, McKown AD, Guy RD, et al. Genome-wide association mapping for wood characteristics in *Populus* identifies an array of candidate single nucleotide polymorphisms. New Phytol. 2013;200:710–26.

44. Qiu D, Bai S, Ma J, Zhang L, Shao F, Zhang K, et al. The genome of *Populus alba* x *Populus tremula* var. *glandulosa* clone 84K. DNA Res. 2019;26:423–31.

45. Quéméneur JB, Danion M, Cabon J, Collet S, Zambonino-Infante J-L, Salin K. The relationships between growth rate and mitochondrial metabolism varies over time. Sci Rep. 2022;12:16066.

46. R Core Team. R. A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2022.

47. Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, et al. Lignin valorization: improving lignin processing in the biorefinery. Science. 2014;344:1246843.

48. Rendón-Anaya M, Wilson J, Sveinsson S, Fedorkov A, Cottrell J, Bailey MES, et al. Adaptive introgression facilitates adaptation to high latitudes in European aspen (*Populus tremula* L.). Mol Biol Evol. 2021;38:5034–50.

49. Robinson KM, Schiffthaler B, Liu H, Westman SM, Kalman TA, Rendón-Anaya M, Canovi C, Bernhardsson C, Delhomme N, Jenkins J, Wang J, Mähler N, Richau KH, Stokes V, A'Hara S, Cottrell J, Coeck K, Diels T, Vandepoele K, Mannapperuma C, Park E-J, Plaisance S, Jansson S, Ingvarsson PK, Street NR. An improved chromosome-scale genome assembly and population genetics resource for *Populus tremula*. BioRxiv. 2023. https://doi.org/10.1101/805614v1.full.

50. Sannigrahi P, Ragauskas AJ, Tuskan GA. Poplar as a feedstock for biofuels: a review of compositional characteristics. Biofuel Bioprod Biorefin. 2010;4:209–26.

51. Storey JD, Bass AJ, Dabney A, Robinson D. qvalue: Q-value estimation for false discovery rate control. 2021; R package version 2.24.0.

52. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci USA. 2003;100:9440–5.

53. Studer MH, DeMartini JD, Davis MF, Sykes RW, Davison B, Keller M, et al. Lignin content in natural *Populus* variants affects sugar release. Proc Natl Acad Sci USA. 2011;108:6300–5.

54. Sweeley CC, Elliott W, Fries I, Ryhage R. Mass spectrometric determination of unresolved components in gas chromatographic effluents. Anal Chem. 1966;38:1549–53.

55. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). J Chemom. 2002;16:119–28.

56. Turner S. qqman: Q-Q and Manhattan Plots for GWAS Data. 2017; R package version 0.1.4. https://CRAN.R-project.org/package=qqman. Accessed 19 April 2021.

57. Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr & Gray). Science. 2006;313:1596–604.

58. Tuskan GA, Muchero W, Tschaplinski TJ, Ragauskas AJ. Population-level approaches reveal novel aspects of lignin biosynthesis, content, composition and structure. Curr Opin Biotechnol. 2019;56:250–7.

59. Van Acker R, Leplé J-C, Aerts D, Storme V, Goeminne G, Ivens B, et al. Improved saccharification and ethanol yield from field-grown transgenic poplar deficient in cinnamoyl-CoA reductase. Proc Natl Acad Sci USA. 2014;111:845–50.

60. Vanholme R, Cesarino I, Rataj K, Xiao Y, Sundin L, Goeminne G, et al. Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. Science. 2013;341:1103–6.

61. Wang J, Ding J, Tan B, Robinson KM, Michelson IH, Johansson A, et al. A major locus controls local adaptation and adaptive life history variation in a perennial plant. Genome Biol. 2018;19:72.

62. Wang Z, Pawar PM, Derba-Maceluch M, Hedenström M, Chong S-L, Tenkanen M, et al. Hybrid aspen expressing a carbohydrate esterase family 5 acetyl xylan esterase under control of a wood-specific promoter shows improved saccharification. Front Plant Sci. 2020;11:380.

63. Wang Z, Winestrand S, Gillgren T, Jönsson LJ. Chemical and structural factors influencing enzymatic saccharification of wood from aspen, birch and spruce. Biomass Bioenergy. 2018;109:125–34.

64. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.

65. Wilkerson C, Mansfield S, Lu F, Withers S, Park J-Y, Karlen S, et al. Monolignol ferulate transferase introduces chemically labile linkages into the lignin backbone. Science. 2014;344:90–3.

66. Xie M, Muchero W, Bryan AC, Yee KL, Guo H-B, Zhang J, et al. A 5-enolpyruvylshikimate 3-phosphate synthase functions as a transcriptional repressor in *Populus*. Plant Cell. 2018;30:164–1660.

67. Yang W, Wang K, Zhang J, Ma J, Liu J, Ma T. The draft genome sequence of a desert tree *Populus pruinosa*. GigaScience. 2017;6:gix075.

Escamez *et al. Biotechnology for Biofuels and Bioproducts*   (2023) 16:65

Page 16 of 16

68. Yoo CG, Yang Y, Pu Y, Meng X, Muchero W, Yee KL, et al. Insights of bio-mass recalcitrance in natural *Populus trichocarpa* variants for biomass conversion. Green Chem. 2017;19:5467–78.
69. Zhang Y, Fernie AR. On the role of the tricarboxylic acid cycle in plant productivity. J Integr Plant Biol. 2018;60:1199–216.
70. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44:821–4.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.