Genetic Determinants for Enzymatic Digestion of Lignocellulosic Biomass Are Independent of Those for Lignin Abundance in a Maize Recombinant Inbred Population[1][W][OPEN]

Bryan W. Penning, Robert W. Sykes, Nicholas C. Babcock, Christopher K. Dugard, Michael A. Held[2], John F. Klimek, Jacob T. Shreve[3], Matthew Fowler, Angela Ziebell, Mark F. Davis, Stephen R. Decker, Geoffrey B. Turner, Nathan S. Mosier, Nathan M. Springer, Jyothi Thimmapuram, Clifford F. Weil, Maureen C. McCann, and Nicholas C. Carpita*

Departments of Biological Sciences (B.W.P., M.C.M., N.C.C.), Botany and Plant Pathology (C.K.D., M.A.H., J.F.K., N.C.C.), and Agronomy (N.C.B., C.F.W.), Laboratory of Renewable Resources Engineering and Agricultural and Biological Engineering (N.S.M.), and Bioinformatics Core (J.T.S., J.T.), Purdue University, West Lafayette, Indiana 47907; National Bioenergy Center, National Renewable Energy Laboratory, Golden, Colorado 80401 (R.W.S., M.F., A.Z., M.F.D., S.R.D., G.B.T.); and Department of Plant Biology, University of Minnesota, St. Paul, Minnesota 55108 (N.M.S.)

ORCID ID: 0000-0003-0770-314X (N.C.C.).

Biotechnological approaches to reduce or modify lignin in biomass crops are predicated on the assumption that it is the principal determinant of the recalcitrance of biomass to enzymatic digestion for biofuels production. We defined quantitative trait loci (QTL) in the Intermated B73 × Mo17 recombinant inbred maize (*Zea mays*) population using pyrolysis molecular-beam mass spectrometry to establish stem lignin content and an enzymatic hydrolysis assay to measure glucose and xylose yield. Among five multiyear QTL for lignin abundance, two for 4-vinylphenol abundance, and four for glucose and/or xylose yield, not a single QTL for aromatic abundance and sugar yield was shared. A genome-wide association study for lignin abundance and sugar yield of the 282-member maize association panel provided candidate genes in the 11 QTL of the B73 and Mo17 parents but showed that many other alleles impacting these traits exist among this broader pool of maize genetic diversity. B73 and Mo17 genotypes exhibited large differences in gene expression in developing stem tissues independent of allelic variation. Combining these complementary genetic approaches provides a narrowed list of candidate genes. A cluster of *SCARECROW-LIKE9* and *SCARECROW-LIKE14* transcription factor genes provides exceptionally strong candidate genes emerging from the genome-wide association study. In addition to these and genes associated with cell wall metabolism, candidates include several other transcription factors associated with vascularization and fiber formation and components of cellular signaling pathways. These results provide new insights and strategies beyond the modification of lignin to enhance yields of biofuels from genetically modified biomass.

Gaining genetic control of cell wall composition and architecture is essential for optimizing the efficiencies of biofuels conversion processes without compromising plant vigor and biomass yield (Carpita and McCann, 2008). Lignin is a key determinant of the recalcitrance of the plant cell wall to enzymatic digestion of cellulose, xylans, and other noncellulosic glycans to monosaccharides (saccharification yield) that may be used for biofuels production (Chen and Dixon, 2007; Himmel et al., 2007; Novaes et al., 2010). The genes involved in lignin biosynthesis from Phe to hydroxycinnamates and monolignol substrates are known, and modulating enzyme activities in this pathway alters the abundance of these substrates and the composition of polymerized lignin in the cell wall and increases the yield of sugars by enzymatic digestion (Bonawitz and Chapple, 2010; Vanholme et al., 2010). Forward genetic screens have also led to the discovery of visual and lignin-related phenotypes, such as brown-midrib maize (*Zea mays*) and sorghum (*Sorghum bicolor*), which show improved digestion by ruminant animals and increased sugar yield by enzyme hydrolysis

[1] This work was supported by the National Science Foundation Hy-Bi, an Emerging Frontiers in Research and Innovation program (grant no. 0938033), by the U.S. Department of Energy Feedstock Genomics Program, Office of Biological and Environmental Research, Office of Science (grant no. DE–FOA–0000598), and by the Center for Direct Catalytic Conversion of Biomass to Biofuels, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences (grant no. DE–SC0000997).

[2] Present address: Department of Chemistry and Biochemistry, The Ohio University, 350 West State Street, Athens, OH 45701.

[3] Present address: Department of Biochemistry and Molecular Biology and Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802.

* Address correspondence to carpita@purdue.edu.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Nicholas C. Carpita (carpita@purdue.edu).

[W] The online version of this article contains Web-only data.

[OPEN] Articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.114.242446

(Cherney et al., 1991; Marita et al., 2003; Vermerris et al., 2007). However, reducing lignin content either in mutant or transgenic lines can result in reduced biomass yields (Li and Chapple, 2010; Simmons et al., 2010; Fu et al., 2011a, 2011b). This strategy is also undesirable for pyrolytic or chemical catalytic conversion processes that may utilize lignin for hydrocarbon fuels and aromatic coproducts (Venkatakrishnan et al., 2014).

The composition of polymerized lignin in the cell wall strongly impacts saccharification yields in poplar (*Populus* spp.) and Arabidopsis (*Arabidopsis thaliana*; Bonawitz and Chapple, 2010). Increasing expression of the poplar ferulate-5-hydroxylase gene results in lignin that is almost entirely syringyl (S) lignin (Stewart et al., 2009) and substantially enhances kraft pulping efficiency in poplar (Huntley et al., 2003) and increases the effectiveness of a liquid hot water treatment to improve Glc yield via enzymatic digestion (Li et al., 2010b). Improvements in processing efficiencies are also observed by exploiting the natural variation in lignin composition in poplar. Studer et al. (2011) noted that sugar release was significantly higher in certain poplar genotypes with normal lignin content, inferring that cell wall characteristics other than lignin influence recalcitrance. Suggested factors include cellulose crystallinity and cross-linking of polysaccharides (Himmel et al., 2007).

In this study, we provide evidence for candidate genes unrelated to lignin abundance that could be utilized to impact saccharification yields. The identification of candidate genes and alleles underlying biomass recalcitrance is critically impacted by characteristics of the specific conversion process, including the nature and extent of pretreatments. In fact, in Arabidopsis, the impacts of genetic modifications of lignin are sometimes unseen without pretreatment (Li and Chapple, 2010). In switchgrass (*Panicum virgatum*), acid pretreatment increased saccharification yield in transgenic lines with reduced cinnamyl alcohol dehydrogenase (Fu et al., 2011b). Thermochemical pretreatments of biomass such as steam or ammonia explosion that reach temperatures above the range for lignin phase transition cause lignin to coalesce into molten bodies that redistribute within the biomass (Donohoe et al., 2008). We treated postharvest stem (stover) cell walls isolated from members of maize recombinant inbred lines (RILs) with steam explosion (Selig et al., 2011) and subsequently determined that quantitative trait loci (QTL) for lignin abundance and those for saccharification yield of Glc or Xyl do not overlap under these conditions.

Pyrolysis molecular-beam mass spectrometry (PyMBMS) gives the relative abundance of mass-to-charge ratio (*m/z*) fragments diagnostic for monomeric units of phenylpropanoid, hydroxycinnamic acid, and cell wall polysaccharides (Evans and Milne, 1987; Boon, 1989; Tuskan et al., 1999; Yin et al., 2010). Grass species have distinctive type II cell wall compositions characterized by primary and secondary walls enriched in xylan and phenylpropanoid networks in addition to cellulose and lignin (Carpita and Gibeaut, 1993; Carpita, 1996). We validated diagnostic fragments for guaiacyl (G) and S lignin as reliable measures of abundance in maize cell walls (Penning et al., 2014). In parallel, we used a high-throughput protocol to quantify saccharification yields of Glc and Xyl using a cocktail of polysaccharide-hydrolyzing enzymes (Selig et al., 2009, 2011).

Because of the genetic colinearity of grasses (Devos, 2005), gene discovery in any one of them enables the rapid identification of putative orthologs in annual and perennial C4 energy grasses, such as *Miscanthus* spp., switchgrass, sorghum, and tropical varieties of maize (Carpita and McCann, 2008; Dweikat et al., 2012). The maize Intermated B73 × Mo17 (IBM) population is an immortalized set of lines derived from two elite inbreds, B73 and Mo17, that underwent four rounds of intermating at the F2 stage to increase recombination and, thus, genetic resolution (Sharopova et al., 2002). Five intermatings led to a 3.8-fold increase in size of the genetic map compared with an equivalent F2 population (Lee et al., 2002). While QTL provide valuable information for marker-assisted breeding, each QTL interval consists of 95 to 510 candidate genes, of which only one or a few contribute to a trait. In addition to RILs, genetic resources in maize include the genotype-structured association panel (AP) of 282 lines covering 80% of the total genetic diversity (Bradbury et al., 2007), suitable for genome-wide association studies (GWAS). GWAS takes advantage of linkage disequilibrium generated by ancestral recombinations in populations with wide variation (Tian et al., 2011; Li et al., 2012). By applying GWAS to the AP and accounting for kinship, we screened 10 genomic regions captured within our 11 QTL for candidate genes contributing to our traits of interest.

Because phenotype is the eventual product of the expressed transcriptome, we present a comparison of transcript levels in developing maize B73 and Mo17 stems using RNA deep sequencing. We used differential expression to identify candidate genes in each QTL interval that exhibit a more than 5-fold difference in transcript abundance between the Mo17 and B73 parental lines consistent with the polarity of the trait. We found single-nucleotide polymorphisms (SNPs) and insertions/deletions (INDELs) in the sequences of differentially expressed genes to discriminate alleles potentially useful for plant breeding from indirect effects of trans-acting factors in the genome. Combining different genetic approaches provides new perspectives on identifying biomass-relevant traits beyond lignin abundance or structure and streamlines gene discovery to optimize feedstocks for conversion to a broad range of biofuels and bio-based products.

RESULTS

PyMBMS Is a High-Throughput Screen to Determine Lignin Composition in the Maize IBM Population

Cell walls were isolated from 263 IBM RILs and their two parental lines, B73 and Mo17, from 3 years of field-grown, postharvest stover. We identified QTL with multiyear support to reduce contributions to phenotypic variance from environmental variation. Accumulative

growing degree days were similar in all 3 years, but accumulated rainfall was higher in the final year of the study (Supplemental Fig. S1). Insufficient data were available to conduct a genotype × environment study. In each year, three to five biological replicates of each line were pooled, with technical replicates for all PyMBMS experiments. Fragment ions diagnostic for G lignin (*m/z* 124, 137, 138, and 151) and S lignin (*m/z* 154, 167, 168, and 194) were validated previously as quantitative traits for maize (Penning et al., 2014). An example of PyMBMS spectra from two stover IBM RIL lines, one relatively low in lignin (IBM205) and one relatively high in lignin (IBM063), is shown in Figure 1A. A digital subtraction spectrum shows that these two RILs have different ion abundances of diagnostic fragments for derivatives of pentose, hexose, hydroxycinnamates, and G and S lignin (Fig. 1B). Up to 1.8-fold differences in the relative abundance of fragments diagnostic of G and S lignin were found across the IBM population in all 3 years (Fig. 2, A and C); values far greater than those of the two parents indicate significant transgressive segregation. An ANOVA confirmed a genetic basis (genotype) for differences in all lignin traits as well as differences in lignin traits due to environment (year) and genotype × environment interactions (Supplemental Table S1).

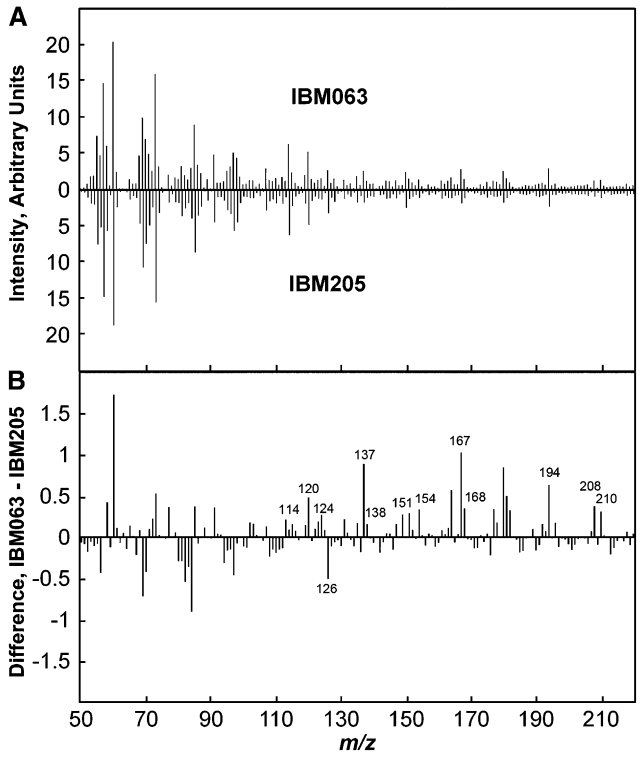

Figure 1. Example of PyMBMS spectra for two IBM RILs and their digital subtraction spectrum. A, PyMBMS spectra from the high-lignin IBM063 (top) and the low-lignin IBM205 (bottom). B, Digital subtraction of IBM205 from IBM063. Assigned fragments are as follows: Xyl, *m/z* 114; 4-vinylphenol, *m/z* 120; hexose, *m/z* 126; G lignin, *m/z* 124, 137, 138, and 151; S lignin, *m/z* 154, 167, 168, and 194 (Penning et al., 2014). IBM063 has many lignin-related fragment ions in higher abundance than does IBM205.

Each ion was normally distributed in the population by the Shapiro-Wilk *W* test, shown as quantile-quantile plots (Supplemental Fig. S2, A and C). The *m/z* 114 assigned to pentose + hexose and the *m/z* 126 assigned to hexose were not correlated with cellulose, noncellulosic glucan, or xylan content in maize (Penning et al., 2014). The fragment *m/z* 120 assigned to 4-vinylphenol, and previously assigned as derived from *p*-coumaric acid (Evans and Milne, 1987; Boon, 1989), was not correlated with the abundance of saponifiable *p*-coumaric acid (Penning et al., 2014). We identified two QTL for *m/z* 120 (4-vinylphenol) abundance.

A High-Throughput Saccharification Assay Gives Relative Glc and Xyl Yield in the Maize IBM Population

Two years of the equivalent cell wall materials used in PyMBMS were available for the studies of saccharification yield by a semiautomated, high-throughput saccharification screen (Selig et al., 2009, 2011). Values for Glc and Xyl varied 1.5- to 2.2-fold over the population in each year. An ANOVA confirmed a genetic basis (genotype) for differences in Glc and Xyl release as well as differences in Glc release due to environment (year) and genotype × environment interactions of Glc and Xyl release (Supplemental Table S1). Environment effects for Xyl release were not found to be significantly different.

Glc and Xyl yield were normally distributed in the population by the Shapiro-Wilk *W* test, shown as frequency distributions (Fig. 2, B and D) and quantile-quantile plots (Supplemental Fig. S2, B and D). Using Pearson's analysis, we found a weak correlation of cellulose content with saccharification yield of Glc and Xyl in a single year only, 0.62 and 0.60, respectively ($P < 0.0001$). Proportions of G and S lignin and noncellulosic Glc and Xyl did not correlate with sugar release from pretreated stover in a large multiyear subset of cell wall samples from the IBM populations (Fig. 3).

QTL Analysis across Multiple Years Reveals No Overlapping QTL for Lignin Abundance and Saccharification Yield

Using a 95% confidence threshold, we obtained 81 QTL in 50 genomic locations for G and S lignin, their sum, and the ratio of S to G lignin across three single years (Supplemental Table S2; Supplemental Fig. S3). Fourteen QTL in five genomic locations (QTL5, QTL6, QTL8, QTL9, and QTL10) overlapped in multiple years (Table I; Fig. 4). Also, QTL for 4-vinylphenol overlapped in two genomic locations (QTL4 and QTL6′) across 2 years of data (Fig. 4). As a control, we also tested four fragments (*m/z* 30, 33, 34, and 35) over 2 years that are not assigned to cell wall-related molecules and registered only one significant QTL over the threshold likelihood of

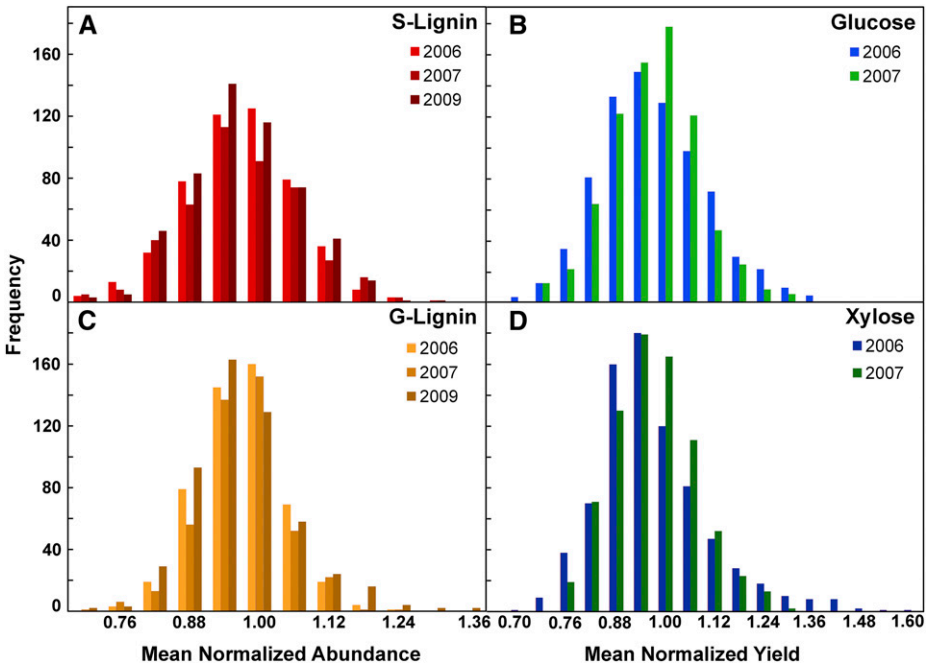

Figure 2. Frequency distributions for the relative abundance of G and S lignin and the relative yield of Glc and Xyl from saccharification. A, Abundance of S lignin (sum of *m/z* 154, 167, 168, and 194). B, Glc yield from saccharification. C, Abundance of G lignin (sum of *m/z* 124, 137, 138, and 151). D, Xyl yield from saccharification.

odds (LOD) score in eight tests, appropriate for a 5% false-positive rate.

We observed 115 QTL in 81 genomic locations for Glc and/or Xyl saccharification yield across two single years (Supplemental Table S2; Supplemental Fig. S3). Eight QTL in four genomic locations (QTL1, QTL2, QTL3, and QTL7) overlapped in multiple years (Table I; Fig. 4). Only QTL6 and QTL6′ on chromosome 7 showed multiyear overlap between lignin and 4-vinylphenol abundance (Fig. 4). In summary, we established seven QTL for aromatic compound abundance and four for saccharification yield in the IBM population that were consistent across multiple years of field testing (Fig. 4; Table I).

Candidate Genes within the QTL Are Identified by a GWAS

Cell walls were also isolated from the 282-line maize AP and assayed by PyMBMS and for saccharification yield from stover. The much greater recombination frequency of this population gave significant locations to single-gene resolution in GWAS analysis. We used flanking markers of the intervals established by our QTL analysis in the IBM population to locate the physical map range of each multiyear QTL. Using a confidence threshold of $P = 0.001$, we selected 116 chromosomal markers for the lignin, 4-vinylphenol, and saccharification yield-related traits (Supplemental Table S3). Some candidate genes identified by GWAS likely result from alleles of genotypes other than B73 and Mo17 contributing to the trait. Therefore, we examined all candidate genes for SNPs or INDELs between Mo17 and B73 from MaizeGDB (http://www.maizegdb.org) (Supplemental Table S4; Lawrence et al., 2007). From our GWAS across the 11 QTL, 23 of the 106 gene models do not have SNPs or INDELs between B73 and Mo17 promoter or coding regions (Table I). Among the remaining 83 are genes encoding several transcription factors, components of cellular signaling pathways, cell wall substrate generation, synthesis and glycosyl transfer, and hydrolysis (Supplemental Table S4).

Differential Transcript Profiling between Mo17 and B73 Identifies Additional Candidate Genes within QTL

We deep sequenced RNA transcripts from the rind tissues of internodes of greenhouse-grown Mo17 and B73 plants at a physiological age where elongation continues in upper internodes and elongation has ceased and secondary wall formation predominates in lower internodes; for this study, we averaged the relative expression of each gene in lower internodes 4 and 5. Remarkable differences in expression fold difference were found for many genes in each QTL interval (Supplemental Table S5). Summed across the QTL intervals, 462 genes showed expression differences 5-fold or greater; we found 237 with expression differences of 10-fold or greater and 17 with greater than 100-fold difference (Supplemental Table S5). Among genes displaying a 5-fold difference in expression were at least 16 transcription factors, 27 signaling-related genes, and 40 cell wall-related genes (Supplemental Table S5).

DISCUSSION

QTL for Phenylpropanoid Abundance and Saccharification Yield Are Different

PyMBMS has been validated as a valuable, high-throughput method to assay the relative abundance

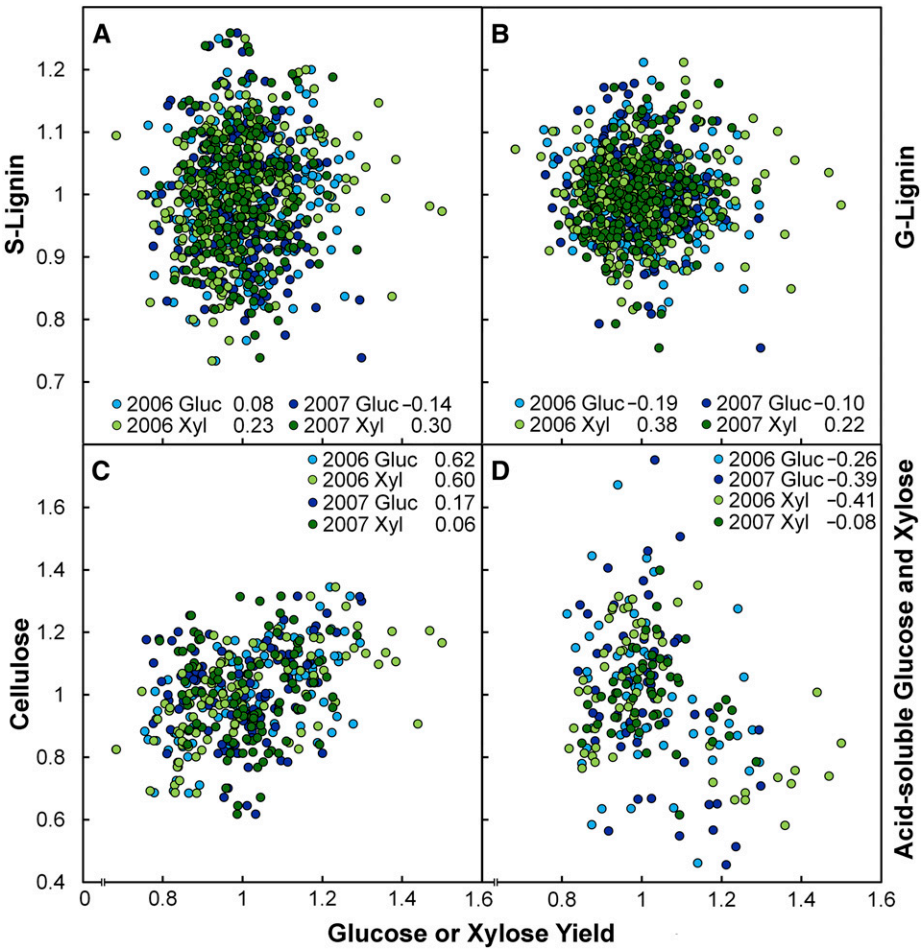

Figure 3. Comparison of lignin and cell wall polysaccharide abundance with saccharification yield in the IBM population. A and B, S (A) and G (B) lignins calculated from PyMBMS *m/z* ion abundances are compared with relative Glc and Xyl release in standardized saccharification yield assays conducted in 2 years. S lignin is the sum of the relative abundance of mass fragments *m/z* 154, 167, 168, and 194; G lignin was calculated from PyMBMS ion abundances *m/z* 124, 137, 138, and 151. C and D, Cellulose content (mg mg^{-1} cell wall; C) and cell wall acid-soluble Glc and Xyl (mg mg^{-1} cell wall; D) are compared with Glc and Xyl release in standardized saccharification yield assays. Year, sugar, and Pearson's correlation coefficients are shown. All values are mean-normalized relative abundances.

of lignin- and hydroxycinnamic acid-derived aromatic residues in woody tissues (Evans and Milne, 1987; Boon, 1989) and grasses (Penning et al., 2014). Because ANOVA found significant environment and genotype × environment effects on our traits (Supplemental Table S1), we considered only QTL with multiple year support to contain genes whose contributions to phenotype are more likely to be based on genotype.

Despite the improvements in saccharification yield made possible by the alteration of lignin content and composition (Vanholme et al., 2010; Fu et al., 2011a), none of our QTL for lignin and 4-vinylphenol abundance that have multiple year support overlap those for saccharification yield. In contrast, all but one of our QTL for saccharification yield (QTL3) map to at least one of the chromosomal regions associated with proxies of lignin abundance identified in other studies of maize, such as Klason lignin determination (total insoluble lignin) and artificial ruminant digestibility assays (Supplemental Table S6; Barrière et al., 2010; Lorenz et al., 2010; Lorenzana et al., 2010; Thomas et al., 2010; Truntzler et al., 2010). Saccharification yields vary in untreated materials but to a much lesser extent following pretreatments (Mosier et al., 2005). Use of a pretreatment that minimizes lignin interference enables the identification of different QTL underlying novel recalcitrance factors, such as polysaccharide and lignin nanoscale and mesoscale architecture, biopolymer side group substitutions, or other types of interactions. The identification of causal genes within these QTL represents new opportunities to improve enzymatic saccharification yield beyond that provided by pretreatments or the production of advanced biofuels from chemical catalysis and/or pyrolysis.

Limitations of GWAS in the Selection of Candidate Genes from RILs

A pervasive problem in QTL studies is the difficulty in identifying causative genes among the numbers of genes within large physical intervals. GWAS can be used to refine the search by taking advantage of natural genetic variation in SNPs and INDELs among a great many genotypes of a species. Maize underwent a genome duplication event approximately 11.4 million years ago (Gaut and Doebley, 1997). Random loss of genes generated great diversity in the derived inbreds and landraces (Schnable et al., 2009), and this diversity is captured in the AP. In general, the smaller the number of genes controlling a trait, the more likely a GWAS will define

Table I. *Summary of total and candidate genes within 11 QTL regions*

QTL	Chromosome	Trait	Maximum LOD[a]	Effect on Trait[b]	Genetic Location	Physical Location	No. of Genes			
							Total Genes	GWAS Total[c]	GWAS SNP/INDEL[d]	Δ Exp B73/Mo17 ≥5-Fold[e]
					cM	*Mb*				
1	1	Glc yield	8.4/3.8	−2.2/−1.3	345–348	224–236	296	11	11	35
2	2	Glc yield	13/8.6	+3.3/+2.0	396–414	165–190	510	30	16	68
3	4	Glc yield	3.5/15	−1.7/−1.8	35–77	0.6–5.1	187	22	20	32
4	4	4-Vinylphenol	8.7/5.0	−2.8/−2.4	276–301	46–86	443	2	2	54
5	6	S lignin	6.1/6.1	−1.9/−2.3	348–357	148–151	95	8	8	17
6	7	S/G lignin	6.1/3.4	−1.6/−1.5	226–248	100–118	217	0	0	28
6′	7	4-Vinylphenol	5.4/4.7	−1.9/−2.2	226–248	100–118	217	5	2	28
7	8	Xyl yield	14/9.0	+3.5/+2.0	180–190	16–40	359	11	10	53
8	8	G lignin	4.1/7.4	+2.6/+3.1	356–365	128–142	324	8	7	29
9	9	S and G lignin	5.0/15	−1.9/−3.6	91–97	11–15	107	4	3	13
10	10	S/G lignin	5.6/12	+1.9/+2.6	185–191	55–83	346	5	4	43

[a]Largest LOD score for the trait in the QTL region for year 1/year2. [b]Percentage of trait increase explained by location; positive values are B73 dominant, and negative values are Mo17 dominant. [c]No. of significant identified markers by GWAS locations using the AP. [d]No. of candidate genes with SNP/INDELs in B73 versus Mo17 identified by GWAS. [e]No. of genes with 5-fold or greater differential expression between B73 and Mo17 parental lines.

genes associated with that trait. For example, GWAS revealed an SNP in a single allele of complement factor H contributing to age-related macular degeneration in humans (Klein et al., 2005). Similarly, a GWAS of large and small dog breeds was able to identify a gene encoding an insulin-like growth factor, IGF1, within a major QTL for size (Sutter et al., 2007). However, GWAS can be compromised when a great many gene products contribute to a complex trait and when kinship relationships are not taken into account. In maize, GWAS was used to find a strong correlation between the *Dwarf8* locus and flowering time using a general linear model (Thornsberry et al., 2001), but the association was found to be spurious by association with population structure (Larsson et al., 2013). Nevertheless, models that correct for kinship have allowed the discovery of genes controlling many agronomic traits in rice (*Oryza sativa*; Huang et al., 2010) and sorghum (Morris et al., 2013).

Taking account of kinship, GWAS gave several strong candidate genes with $P \leq 10^{-8}$ within the QTL intervals displaying significant SNP/INDEL support and consistent with allelic variation between Mo17 and B73 (Supplemental Fig. S4, A–D). For saccharification yield, we identify a prefoldin gene and three genes of unknown function in QTL1 (Supplemental Fig. S4A) and a particularly large number in QTL3 (Fig. 5), with genes encoding three *SCARECROW9-LIKE* transcription factors and one *SCARECROW14-LIKE* transcription factor, a thaumatin-like protein, a bell-like homeodomain protein (*BL4*), a brassinolide-signaling kinase, a powdery mildew resistance protein, a cyclin D7, and four genes of unknown function (Supplemental Table S4). For lignin and 4-vinylphenol, candidate genes with $P \leq 10^{-8}$ within the QTL intervals displaying SNP/INDEL support are an inositol phosphoceramide synthase1 and a harpin-related protein in QTL5 (Supplemental Fig. S4B) and two Leu-rich repeat proteins and a protein of unknown function in QTL8 (Fig. 5). Several significant GWAS locations within the QTL but with no SNP/INDEL support in Mo17 and B73 indicate that the maize genome diversity holds many more trait-determining genes in the AP than are present in the narrowed diversity represented in these two parents. For example, a xylosidase in QTL3 is a GWAS candidate gene but has no polymorphisms between B73 and Mo17. While the activity of the β-xylosidase might contribute to a saccharification phenotype in the IBM population, it is a potential target for selective breeding only with other genotypes in the AP.

Phenotype Variation Can Arise from Differential Expression of Common Alleles

Tiling arrays demonstrated that B73, Mo17, and many other inbreds exhibit significant copy number variation as a result of diploidization following genome duplication, such that inbreds may have two, one, or even zero copies of an ancestral gene at any locus (Springer et al., 2009). Therefore, one might expect to observe significant variation in levels of expression based on gene dosage.

Approximately 25% of the duplicated genes in the maize genome remain since the most recent tetraploid event (Schnable et al., 2009). In some families, a substantially higher number of duplicates were found. For example, all duplicates of the 10 cellulose synthase CesA family members were retained (Penning et al., 2009). Because conventional algorithms only map sequences to unique genes, the high degree of duplication in maize led to an unexpectedly large number of unmapped sequences. To overcome this problem, we developed a custom Perl script to identify those reads with only one valid alignment, with exactly two valid alignments, and with more than two valid alignments (Supplemental Table S7). Those with two valid alignments were invariably from one or

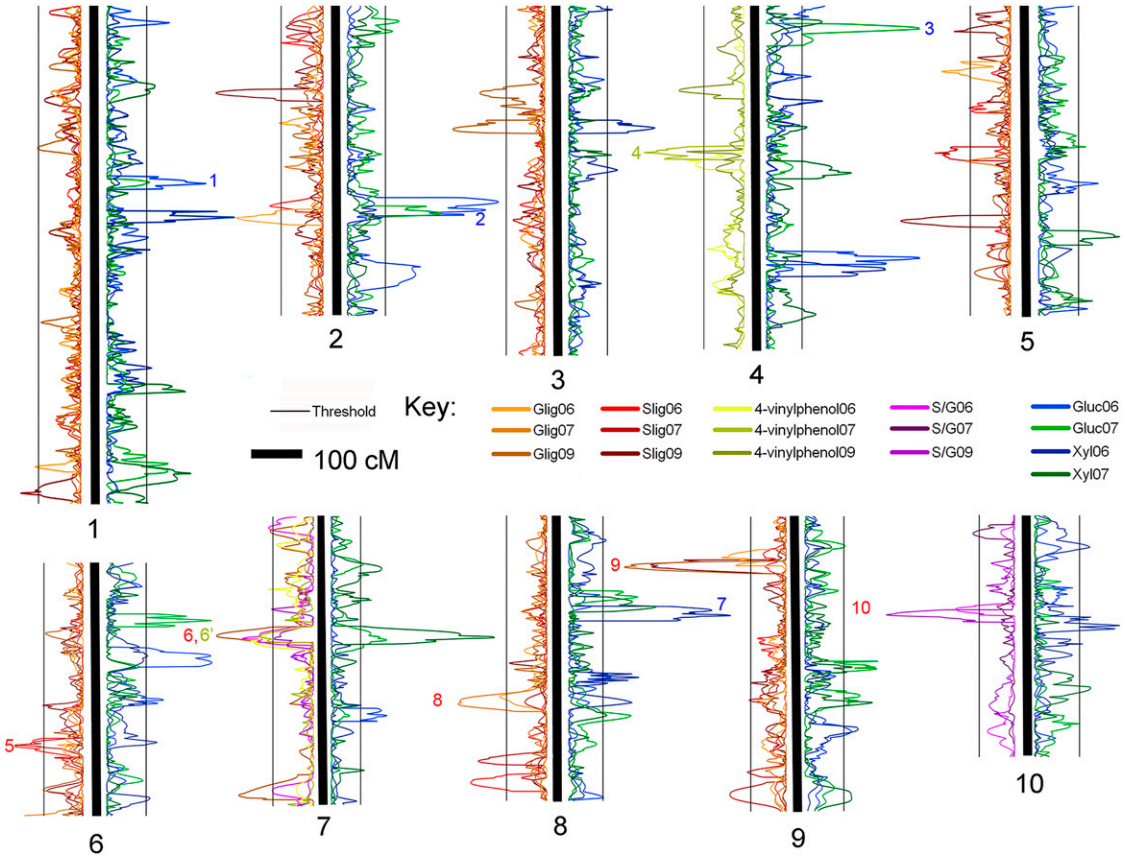

Figure 4. Lignin abundance, 4-vinylphenol abundance, and saccharification yield QTL. QTL for 4-vinylphenol, S and G lignin, and saccharification yield are shown on the 10 maize chromosomes. The 95% experiment-wise confidence cutoff is shown as a thin gray bar. QTL with multiple year overlap for each trait are numbered with color indicating type (see key). QTL1, QTL2, and QTL3 are Glc release, and QTL7 is Xyl release. QTL4 and QTL6′ are 4-vinylphenol abundance. QTL5 is S lignin, QTL6 and QTL10 are S-G lignin ratio, QTL8 is G lignin, and QTL9 is both S and G lignin. Bar = 100 cM.

both duplicate genes. Those with more than two alignments were discarded. This analysis resulted in an average alignment of approximately 80% of the total reads.

Paschold and colleagues (2012) found that nearly 70% of the genes of Mo17 and B73 exhibited at least 2-fold differences in expression in primary root tissues. Close to 500 genes within our 11 QTL intervals change by at least 5-fold. In the two QTL that we examine in detail (Fig. 5), only about 10% of differentially expressed genes have no polymorphisms between the two parents in the promoter and coding sequences (Supplemental Tables S3 and S7). Only seven of the 72 genes with $P < 0.001$ from the GWAS study exhibit greater than 5-fold difference in expression. Thus, the genotype-specific differential expression studies represent a valuable complementary approach to identifying candidate genes within QTL contributing to a trait independent of discrimination by SNPs or INDELs.

Several previous expression and QTL studies have focused on genes of lignin biosynthesis (Lorenz et al., 2010; Lorenzana et al., 2010; Thomas et al., 2010; Truntzler et al., 2010; Bosch et al., 2011). Consistent with these studies, we find homologs for almost every gene in the phenylpropanoid synthesis pathway present within at least one of our lignin abundance- or saccharification yield-related QTL regions (Supplemental Table S6). However, the effect of our pretreatment removes lignin recalcitrance by melting and partially removing it from the cell wall (Donohoe et al., 2008). Even though the B73 inbred line is a derived stiff-stalk line (Liu et al., 2003) and has slightly more stem lignin than does Mo17 (Penning et al., 2014), QTL were established with both Mo17 and B73 as the dominant parent (Supplemental Table S6). Overall, for our lignin/4-vinylphenol QTL, 15 genes involved in monolignol synthesis or polymerization were differentially expressed at least 5-fold in B73 as compared with Mo17; only five of the monolignol genes within the saccharification yield QTL differed by 5-fold or more (Supplemental Table S6). Alteration of the activity of any of these genes, either through mutation or reduced expression of the genes by RNA interference, can impact lignin abundance (Ma, 2007; Fu et al., 2011a; Saathoff et al., 2011; Tamasloukht et al., 2011; Saballos et al., 2012).

In addition to the coordinated control of secondary wall formation, the lignin, cellulose, and possibly xylan secondary cell wall pathways are also regulated independently (Zhou et al., 2009; Demura and Ye, 2010; Zhong

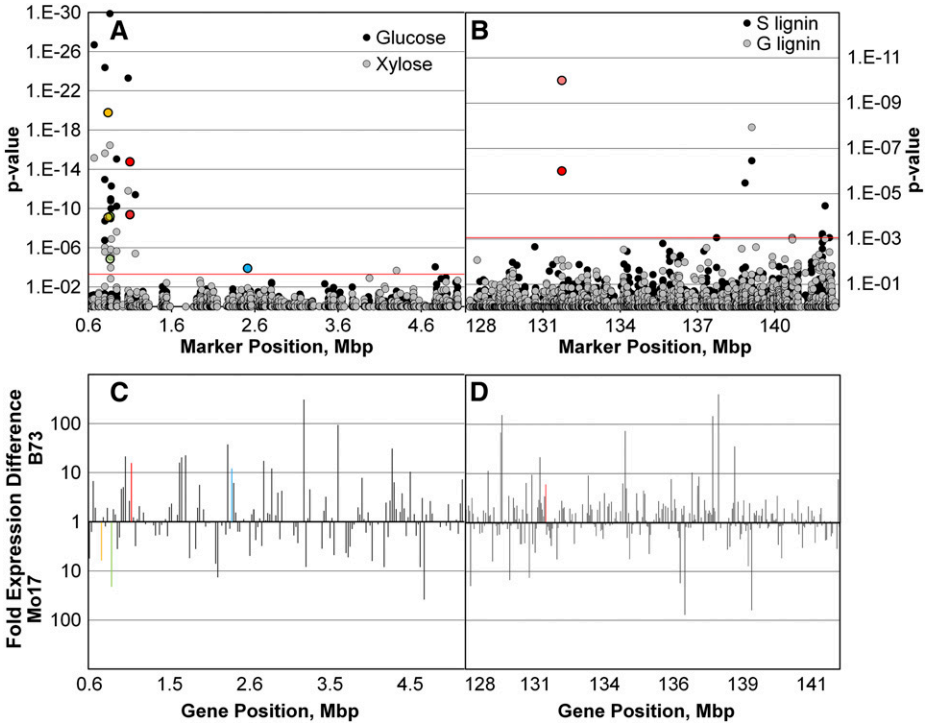

Figure 5. Marker and gene positions as identified by GWAS and differential expression, respectively, in QTL3 and QTL8. Physical map locations in Mb (A and B) were identified using flanking markers, and a threshold value of 0.001 is indicated by red lines. Differential expression fold change (C and D) is shown in logarithmic scale, with higher B73 fold change upward and higher Mo17 fold change downward. GWAS predictions for QTL3 for Glc yield (A) are paired with expression fold differences between B73 and Mo17 parents across the QTL interval (C). GWAS predictions for QTL8 for G and S lignin (B) are paired with expression fold differences between B73 and Mo17 parents across the QTL interval (D). Markers flanking the QTL in cM were used to match the closest markers with physical map positions (www.maizegdb.org). Expression analysis (RNA sequencing) was performed on complementary DNA populations derived from developing internodes 4 and 5 of greenhouse-grown B73 and Mo17 parents 63 d after planting during peak secondary wall formation. Color codes for symbols in A and C, all Glc yield related, are as follows: orange, gene of unknown function; green, *SCARECROW9-LIKE* transcription factor; red, *BR-SIGNALING KINASE1*; and blue, *β-XYLOSIDASE*. Dark red symbols (S lignin) and light red symbols (G lignin) both indicate the same candidate gene of unknown function in B and D.

et al., 2010; Kim et al., 2013). MYB genes serve as inducers or repressors of secondary wall synthesis, and their down-regulation is coordinated with the up-regulation of the No apical meristem/Arabidopsis transcription activation factor/Cup-shaped cotyledon (NAC)/NAC secondary wall thickening promoting factor (NST) genes (Zhao and Dixon 2011; Zhong et al., 2011). For lignin/4-vinylphenol traits, expression differences of 5-fold or greater are found with a NAC domain and a MYB gene in QTL5, with MYB30 in QTL6, and with MYB61 in QTL8 (Supplemental Table S5). We find genes encoding other transcription factor types, such as bZIP and two types of zinc RING finger proteins in QTL5, three WRKY proteins in QTL6, and a homolog of the Arabidopsis *RESPONSE REGULATOR-LIKE12* gene in QTL9 (Supplemental Fig. S5).

In contrast to the predominance of differentially expressed lignin biosynthetic enzymes in lignin-related QTL, saccharification yield QTL contain 13 carbohydrate-related genes and exhibit 5-fold or greater expression difference, compared with only seven genes with a 5-fold expression for the lignin-related QTL. Saccharification-related QTL include genes that encode cellulose- and xylan-related enzymes such as UDP-glucuronosyltransferases, *CELLULOSE SYNTHASE2*, the nucleotide interconversion pathway gene *UDP-XYLOSE EPIMERASE1* (*MUR4*), and *IRREGULAR XYLEM14* (Supplemental Table S5). *IRREGULAR XYLEM14* coexpressed with *IRREGULAR XYLEM9* represent synthases of secondary wall xylan backbones (Lee et al., 2012). In addition, signaling-related genes such as wall-associated kinases and homeobox transcription factor regulators and several different types of transcription factors are also overexpressed in the dominant parent, notably, a hecate transcription factor (*HEC1*), a homeobox-leucine zipper gene (*HAT14*), and a cell wall-associated kinase (*WAK1*; Supplemental Table S5). We have identified a WRKY-like sequence highly expressed in Mo17 as a strong candidate in QTL7 for Xyl yield in saccharification (Supplemental Fig. S4C). Mutations in a dicot WRKY transcription factor that represses secondary wall development in pith cells are associated with ectopic

deposition of cellulose and xylan in these cells, increasing biomass density by almost one-half (Wang et al., 2010).

Narrowing the Candidate Gene List

The full range of candidate genes using GWAS and differential expression as independent criteria is shown in Supplemental Figure S4 and Supplemental Tables S3 and S4. We illustrate our criteria for generating a combined candidate gene list from the GWAS analysis and the differentially expressed gene list for QTL3, a Glc yield trait, and for QTL8, a G lignin abundance trait (Fig. 5; Supplemental Table S8). From a total of 187 genes in QTL3, we identified 22 candidate genes by GWAS, 20 of which show polymorphisms in the promoter and/or coding regions between B73 and Mo17 (Table I). We also found SNP or INDEL polymorphisms within the gene sequences of 27 of the 32 genes that gave 5-fold or greater expression differences (Supplemental Table S8). Four genes for Glc yield in QTL3, a rhomboid family protein, BR-SIGNALING KINASE1, a SCARECROW9-LIKE transcription factor, and a gene of unknown function, not only fulfill the criteria that they are above the threshold in GWAS analysis and exhibit polymorphisms between B73 and Mo17 genotypes but also are differentially expressed to 5-fold or greater between the two parents (Fig. 5; Supplemental Table S7). For QTL8 for G lignin abundance, only seven of 324 genes are identified by GWAS as having polymorphisms between Mo17 and B73 (Supplemental Table S4), and 29 are differentially expressed 5-fold or greater (Supplemental Table S5). A single gene of unknown function satisfies all three criteria (Supplemental Table S8).

GWAS and genotype-specific differential expression are independent approaches to identify genes contributing to trait. GWAS primarily reveals genes in which SNPs might alter allele strength across a diverse population of genotypes, irrespective of transcript abundance, whereas differential expression of the two RIL parents provides candidates based on expression level, irrespective of polymorphisms that might affect activity. Genes with no polymorphisms might be expressed differently as an indirect effect, for example, of mutation in a trans-acting factor elsewhere in the genome. Integrating several genetic tools such as GWAS and differential expression will prove useful to resolving those genes for which expression levels contribute to phenotype from candidates for which SNP/INDEL differences indicate more discrete targets for breeding.

MATERIALS AND METHODS

Maize Tissue and Stover Collection for Cell Walls and RNA

The maize (*Zea mays*) Mo17 and B73 parental lines and the IBM RIL population were grown at the University of Illinois field station in Champaign-Urbana. An AP of 282 inbreds and landraces was grown at the Purdue University Agricultural Center for Research and Education. Weather data for total weekly growing degree days and accumulated precipitation were gathered from the Illinois State Water Survey (http://www.isws.illinois.edu/atmos/statecli/cuweather/). Degree growing days were calculated as the weekly average of the daily maximum and minimum temperatures minus the maize base growth temperature of 50°F. The lower 70 cm of field-dried maize stover (internodes 4 and/or 5) was collected and warm air dried for 1 to 2 weeks at about 60°C. After drying, approximately 5-cm sections of the internodes of three to five plants were sampled and ground to less than 0.5-mm particle size in a Wiley Mill with a 40-mesh screen. About 3 g of ground stover was washed twice with 10 mL of 50% (v/v) ethanol and water at 70°C, followed by five washes with 10 mL of distilled water at room temperature. Samples were centrifuged in a swinging-bucket rotor at 2,500 rpm for 3 to 5 min, and the supernatant was removed. In between centrifuge steps, the samples were resuspended by vortex mixing. For the final wash, the water was left and the samples were freeze dried.

For differential gene expression studies, B73 and Mo17 were grown in a greenhouse at Purdue University. Plant materials were harvested at 63 d after planting at a stage when upper internodes were elongating and lower internodes had ceased elongation and exhibited maximal secondary wall formation. Rind tissues from a minimum of three plants were excised asceptically from internodes 4 through 7 and frozen in liquid N_2. Frozen tissues were pulverized individually in a mortar and pestle under liquid N_2, and about 2 mg of the powder was incubated with 1 mL of ice-cold TRIzol reagent (Invitrogen, Life Technologies). RNA was extracted and purified according to the manufacturer's directions and dissolved in 100 μL of diethyl pyrocarbonate-treated nanopure water. RNA quality and concentration were determined spectrophotometrically. RNA samples from three biological replicates for each internode were pooled, and frozen samples were delivered to Purdue's Genomics Core facility (http://www.genomics.purdue.edu/services/htl.shtml) for Illumina sequencing. The relative abundance of genes expressed predominantly during secondary wall formation was similar for internodes 4 and 5, and these data were averaged.

PyMBMS

PyMBMS was carried out at the National Renewable Energy Laboratory in an Extreal model TQMS C50 mass spectrometer using a modified protocol from Sykes et al. (2008). Samples (approximately 4 mg) in duplicate or triplicate were placed in 80-μL stainless-steel cups and placed in trays in a Frontier model PY-2020 iD autosampler. Samples were inserted into the pyrolysis oven with a helium flow of 0.9 L min^{-1} at standard pressure and temperature. The autosampler furnace was maintained at 500°C, with the interface set at 350°C. The autosampler pyrolysis unit transferred the samples through a one-quarter-inch line maintained at approximately 350°C to the molecular-beam mass spectrometer. Total pyrolysis time was 30 s, with pyrolysis reaction completed in 20 s or less. A Merlin data acquisition system with 17 eV electron ionization was used to gather mass spectral data from *m/z* 30 to 450. Acceptable variance among replicate samples was 10% or less, or the samples were rerun.

High-Throughput Saccharification Assay

Semiautomated procedures were optimized from those described previously (Selig et al., 2009). Briefly, each sample of cell wall material was loaded in special spring-loaded hoppers and autodispensed robotically in three 5 ± 0.2 mg aliquots into 0.3-mL wells of stainless-steel plates (96-well format), and 250 μL of water was added by an automated micropipet. The trays were sealed in Teflon tape (3M), placed in bolted metal containers, and subjected to a pretreatment of 180°C for 17.5 min, then cooled to ambient temperature before the tape was removed. Enzymatic saccharification was initiated by the addition of 40 mL of a 1:24 dilution of Ctec2 enzyme stock preparation (Novozyme) in 1 M sodium citrate, pH 5, supplemented with 0.2% (w/v) sodium azide. Plates were sealed again in Teflon tape, and reactions were maintained 50°C for 70 h, then cooled to ambient temperature.

For assay of sugars, plates were centrifuged at 1,700 rpm for 20 min to pellet undigested wall material, and 20 μL of supernatant liquid was transferred in triplicate into 200 μL of water in wells of spectrophotometry plates. Glc and Xyl were assayed enzymatically with Glc oxidase and Xyl dehydrogenase as described (Selig et al., 2011) and quantified by comparison with standards treated similarly. All assays were validated statistically. Samples with a triplicate coefficient of variance greater than 15% were rerun. Acceptable standard curves required an r^2 of 0.995 or greater.

Carbohydrate Analyses

All analyses were run in triplicate with acceptable coefficients of variance of 10% or less. Samples (5 ± 0.2 mg) of isolated cell walls were hydrolyzed in

1 mL of 2 M trifluoroacetic acid containing 0.5 μmol of myoinositol (internal standard) at 120°C for 90 min in a conical centrifuge tube, cooled to ambient temperature, and undigested cellulose was pelleted by centrifugation. The cellulosic pellet was washed several times with water and suspended in 1 mL of water. Samples were taken for determination of Glc equivalents by phenol-sulfuric assay (Dubois et al., 1956).

The trifluoroacetic acid in the supernatant liquid was evaporated in a stream of nitrogen gas after the addition of 1 mL of liquefied *tert*-butyl alcohol. Alditol acetate derivatives were prepared from a subset of these samples as described previously (Gibeaut and Carpita, 1991). Derivatives were separated by gas-liquid chromatography on a 0.25-mm × 30-m SP-2330 column (Supelco). Temperature was held at 80°C during injection, then ramped quickly to 170°C at 25°C min^{-1}, and then to 240°C at 5°C min^{-1}, with a 10-min hold at the upper temperature. Helium flow was 1 mL min^{-1} with splitless injection. Electron-impact mass spectrometry was performed with an Agilent Technologies MSD mass spectrometer at 70 eV and a source temperature of 250°C.

Statistical Analyses

Two to three replicates of PyMBMS trait data (*m/z* ions) previously associated with S or G lignin (Evans and Milne, 1987; Tuskan et al., 1999) were averaged together and mean normalized to correct for skewing as a result of instrument changes between pyrolysis runs in 2009 and weather differences to growth. ANOVA was performed with the PROC GLM command in SAS software version 9.2. The general linear model was used instead of the ANOVA command as the genotype, and year-independent variables had a large difference in treatment number, leading to an unbalanced design. Genotype and year were the independent variables for each trait from the QTL study (dependent variables) in the model statement. Genotype × year interactions were calculated for all traits. The data set consisted of the same trait data used in the QTL analysis. Effects were considered significant at $P \leq 0.001$ for the case of getting a larger F value if the independent variable had no effect on the dependent variable.

Pearson's correlations were carried out using the SAS proc corr procedure. Ions with strong Pearson's correlations to each other (*m/z* 124, 137, 138, and 151 for G lignin and *m/z* 154, 167, 168, and 194 for S lignin) were added together for a measure of S and G lignin. Three replicates of the Glc and Xyl release values from the saccharification assay were averaged together and mean normalized to compare with the PyMBMS data. The S and G lignin traits were graphed versus Glc and Xyl release in Excel using an *x-y* scatterplot. Frequency distributions were produced in Excel using the countif statement for increments of approximately one-tenth of the difference between the highest and lowest sample values for each trait. Normality was tested in SAS using the proc univariate procedure and reporting the Shapiro-Wilks test for normality, as it was appropriate for the sample sizes tested.

QTL Analysis

Genotype data for the 263 IBM RILs are available at the Maize Genetics Database (http://www.maizegdb.org/ibm302scores.html). A total of 2,175 markers were used to construct a genetic map covering the 10 maize chromosomes at an average distance of 1.5 centimorgan (cM) using published maize map data based on these RILs to ensure proper marker order (Davis et al., 1999; Sharopova et al., 2002). Genotype files were converted for use by QTL Cartographer (Basten et al., 1999), with the inclusion of the trait data before it was mean normalized. We tested mean-normalized phenotypes in the QTL analysis and found no difference in LOD traces from the non-normalized data. The S and G lignin monomers, the S-G ratio, the total lignin, and the Xyl and Glc release values were run as traits in a QTL analysis.

QTL analysis was performed in QTL Cartographer for Windows (Basten et al., 1999) using composite interval mapping with a step size of 2 cM and a window size of 10 cM. QTL cofactors were found by forward and backward stepwise regression with $P \leq 0.01$. Threshold LOD scores of 95% experiment-wise confidence levels for each trait were found by permutation analysis of 1,000 trials of randomly assorted genotypes for each trait measured.

Locating Genes within QTL Regions

Markers flanking the QTL in cM were used to find the closest markers with physical map positions on the IBM 2008 neighbors map at www.maizegdb.org (Schaeffer et al., 2011). General feature format (GFF) files containing all the genes within the region were obtained using the physical map positions from MaizeGDB.org with the table browser application available at http://genomaize.org (Karolchik et al., 2004). A custom Perl script was used to parse the GFF file data and find the orientation, start, and stop locations for each gene and place them into a new gene location database for each flanking QTL region to be used for further database queries.

All maize gene sequences were found in a maize sequence database based on the B73 5a.WGS translations database from www.maizegdb.org (Lawrence et al., 2007) using the fastacmd program from the National Center for Biotechnology Information (Altschul et al., 1990). The closest Arabidopsis (*Arabidopsis thaliana*) gene and descriptions to the maize genes were found by protein-protein BLAST (Altschul et al., 1990) against an Arabidopsis gene sequence version 9 housed at The Arabidopsis Information Resource (www.arabidopsis.org; Lamesch et al., 2012). The BLAST results were parsed by a custom Perl script to find the closest related Arabidopsis gene name, description, and expect value and placing them in a Microsoft Excel-readable file format with the maize gene name and start and stop locations (Supplemental Table S7).

Genome-Wide Association Screening

GWAS was carried out on the maize AP (Flint-Garcia et al., 2005; Hufford et al., 2012) using TASSEL version 3 (Bradbury et al., 2007), with all markers on the imputed haplotype data of entire single chromosomes from http://www.panzea.org/index.html as described by Chia et al. (2012). Imputation was used to prevent bias due to missing genotype data by providing substituted data from the nearest marker. Haplotypes were filtered at $P = 0.1$ and merged to the average of three replicates for each trait using the 282 individuals of the maize AP. A kinship matrix was calculated in TASSEL on the filtered gene haplotype data set with the kinship command. A mixed linear model analysis was performed using kinship, haplotype, and trait data. Each chromosome was run as a separate analysis. P value results for locations in the QTL region are shown in the graphs with a significance threshold of 0.001. Genes for GWAS were located using a custom Perl script that used the GWAS location in bp and queried the GFF files generated from www.maizegdb.org of the QTL regions as detailed above to find any genes within 1,000 bp of the GWAS location.

Differential Expression Analysis

RNA samples from three biological replicates for each internode were pooled, and frozen samples were delivered to Purdue's Genomics Core facility. RNA sequencing was conducted at the Purdue Genomics Core facility using an Illumina HiSeq 2000 to process 100- × 100-bp libraries with approximately 400-bp size inserts. The libraries were barcoded, pooled, and sequenced in equivalent proportions across three lanes of a single flow cell to reduce intralane bias. After the sequencing run was complete, the core facility decoded the multiplexed sample data, removed sequencing adapters, and provided the resulting sequence files to the Bioinformatics Core at Purdue for further preprocessing. Each internode sample represented an average of approximately 70 million total reads.

FastQC version 0.10.0 (Andrews, 2010) was used for quality assessment and trimming for removal of bases with PHRED33 scores of less than 30 using FASTX-Toolkit version 0.0.13.2 (Hannon, 2010). The reads were mapped to the maize genome using Bowtie2 version 2.0.0 (Langmead and Salzberg, 2012) with default parameters and processed with Sequence Alignment/Map (SAM) tools version 0.1.16 (Li et al., 2009) and HTSeq version 0.5.3p3 (Simon, 2010) to generate counts per gene. Because of the high degree of gene duplication in maize, reads mapping to two locations were parsed by a custom Perl script and split between the two locations. Mapping locations for genes were determined using a custom Perl script and the maize GFF file from PlantGDB (http://www.plantgdb.org).

We selected Bowtie2 over Tophat for mapping for several reasons. Tophat is built upon Bowtie2 with the added functionality of sensitively mapping across and identifying intron/exon splice sites. The algorithm is especially tuned to mammalian genomes but of far less relevance in plant genome transcriptome analysis. We obtained an average mapping rate of approximately 80%, indicating few reads lost due to nonstandard splice junction variants. However, the principal reason was that the –k3 feature of Bowtie2 also permitted the use of our gene duplication detection custom Perl script, which was the major factor in achieving the high mapping rate. Inspection of the mapping files in SAM format was conducted using the FLAG variable to filter out any mapped reads that were not uniquely positioned and properly paired in mapping. The –k3

parameter was used to force Bowtie2 to report up to thee valid alignments for each read. This allowed our custom Perl scripts to parse the SAM file to identify those reads with only one valid alignment for digital gene expression analysis, those reads with exactly two alignments considered separately for gene duplication, and those reads with more than two alignments that were rejected as uninformative (Supplemental Table S7).

One count per million or greater was used as a threshold for the detection of transcript (Li et al., 2010a; Chang et al., 2012). Several well-known secondary cell wall genes showed approximately equal high expression in internodes 4 and 5. A custom Perl script was used to parse the full maize gene expression file within the QTL regions, and the counts for each tissue were appended to an Excel readable file with maize gene names, locations, and closest Arabidopsis gene by sequence with description and expect value (Supplemental Table S7). Differential analysis was determined using a custom Perl script that averaged internodes 4 and 5 and compared B73 with Mo17. Values of zero in either B73 or Mo17 were set to one count for the determination of fold change.

SNP/INDEL Analysis

SNPs and INDELs were determined for the entire open reading frame (ORF) of genes and a region 1,000 bp in front of the gene start position to represent the promoter region of the ORF at www.maizegdb.org (Schaeffer et al., 2011). Each gene was searched using the MaizeGDB Genome Browser by gene name with the diversity track: Mo17 SNPs and INDELs (from Joint Genome Institute). Each SNP or INDEL for the ORF or promoter region was counted separately.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Comparison of weekly degree growing days and accumulative rainfall for the 3 years of IBM population growth seasons.

Supplemental Figure S2. Quantile-quantile plots for population variance in lignin abundance and saccharification yields from the cell walls of the IBM population.

Supplemental Figure S3. LOD scores of all QTL for lignin and 4-vinylphenol abundance and saccharification.

Supplemental Figure S4. Marker and gene positions as identified by GWAS and differential expression for the remaining QTL.

Supplemental Table S1. ANOVA by genotype and year for lignin estimated by pyrolysis and saccharification potential.

Supplemental Table S2. All significant QTL for lignin and 4-vinylphenol abundance and saccharification yield.

Supplemental Table S3. Marker positions in QTL regions identified by GWAS of the AP.

Supplemental Table S4. Significant GWAS locations with SNPs or INDELs in genes of Mo17 compared with those of B73.

Supplemental Table S5. Significant differential expression locations within QTL regions and the closest Arabidopsis or rice gene.

Supplemental Table S6. Comparison of QTL identified in this study and previous studies of related traits using approximately 20-cM intervals (bins).

Supplemental Table S7. Useful Perl scripts for QTL, GWAS, and differential expression analysis in maize.

Supplemental Table S8. SNPs and INDELs in genes of Mo17 compared with those of B73 within QTL3 and QTL8 that are differentially expressed.

ACKNOWLEDGMENTS

We thank Steve Moose (University of Illinois) for making the maize stover from the IBM lines available for collection, for helpful discussions during the course of this project, and for review of the article. We thank Dr. Phillip SanMiguel (Purdue University Core Genomics Facility) for helpful advice, creating libraries, and performing the transcriptome analyses. We also thank Jerry Tuskan (Oak Ridge National Laboratory) and Ron Sederoff (North Carolina State University) for their review of an earlier draft of this article.

Received May 6, 2014; accepted June 24, 2014; published June 27, 2014.

LITERATURE CITED

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **215:** 403–410

Andrews S (2010) FastQC. http://www.bioinformatics.babraham.ac.uk/projects/fastqc (July 9, 2014)

Barrière Y, Mechin V, Denoue D, Bauland C, Laborde J (2010) QTL for yield, earliness, and cell wall quality traits in topcross experiments of the F838 × F286 early maize RIL progeny. Crop Sci **50:** 1761–1772

Basten CJ, Weir BS, Zeng ZB (1999) QTL Cartographer, Version 1.16. North Carolina State University Press, Raleigh. http://statgen.ncsu.edu/qtlcart/manual/ (July 9, 2014)

Bonawitz ND, Chapple C (2010) The genetics of lignin biosynthesis: connecting genotype to phenotype. Annu Rev Genet **44:** 337–363

Boon JJ (1989) An introduction to pyrolysis mass spectrometry of lignocellulosic material: case studies on barley straw, corn stem, and Agropyron. *In* A Chesson, ER Ørskov, eds, Physicochemical Characterization of Plant Residues for Industrial Feed Use. Elsevier Applied Science, Amsterdam, pp 25–49

Bosch M, Mayer CD, Cookson A, Donnison IS (2011) Identification of genes involved in cell wall biogenesis in grasses by differential gene expression profiling of elongating and non-elongating maize internodes. J Exp Bot **62:** 3545–3561

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics **23:** 2633–2635

Carpita NC (1996) Structure and biogenesis of the cell walls of grasses. Annu Rev Plant Physiol Plant Mol Biol **47:** 445–476

Carpita NC, Gibeaut DM (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. Plant J **3:** 1–30

Carpita NC, McCann MC (2008) Maize and sorghum: genetic resources for bioenergy grasses. Trends Plant Sci **13:** 415–420

Chang YM, Liu WY, Shih ACC, Shen MN, Lu CH, Lu MYJ, Yang HW, Wang TY, Chen SCC, Chen SM, et al (2012) Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. Plant Physiol **160:** 165–177

Chen F, Dixon RA (2007) Lignin modification improves fermentable sugar yields for biofuel production. Nat Biotechnol **25:** 759–761

Cherney JH, Cherney DJR, Akin DE, Axtell JD (1991) Potential of brown-midrib, low-lignin mutants for improving forage quality. Adv Agron **46:** 157–198

Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, et al (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet **44:** 803–807

Davis GL, McMullen MD, Baysdorfer C, Musket T, Grant D, Staebell M, Xu G, Polacco M, Koster L, Melia-Hancock S, et al (1999) A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. Genetics **152:** 1137–1172

Demura T, Ye ZH (2010) Regulation of plant biomass production. Curr Opin Plant Biol **13:** 299–304

Devos KM (2005) Updating the 'crop circle.' Curr Opin Plant Biol **8:** 155–162

Donohoe BS, Decker SR, Tucker MP, Himmel ME, Vinzant TB (2008) Visualizing lignin coalescence and migration through maize cell walls following thermochemical pretreatment. Biotechnol Bioeng **101:** 913–925

Dubois M, Gilles KA, Hamilton JK, Rebers PA, Smith F (1956) Colorimetric method for determination of sugars and related substances. Anal Chem **28:** 350–356

Dweikat I, Weil C, Moose S, Kochian L, Mosier N, Ileleji K, Brown P, Peer W, Murphy A, Taheripour F, et al (2012) Envisioning the transition to a next-generation biofuels industry in the US Midwest. BioFPR **6:** 376–386

Evans RJ, Milne TA (1987) Molecular characterization of the pyrolysis of biomass. 1. Fundamentals. Energy Fuels **1:** 123–137

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize

association population: a high-resolution platform for quantitative trait locus dissection. Plant J **44:** 1054–1064

Fu C, Mielenz JR, Xiao X, Ge Y, Hamilton CY, Rodriguez M Jr, Chen F, Foston M, Ragauskas A, Bouton J, et al (2011a) Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. Proc Natl Acad Sci USA **108:** 3803–3808

Fu CX, Xiao XR, Xi YJ, Ge YX, Chen F, Bouton J, Dixon RA, Wang ZY, et al (2011b) Downregulation of cinnamyl alcohol dehydrogenase (CAD) leads to improved saccharification efficiency in switchgrass. Bioenerg Res **4:** 153–164

Gaut BS, Doebley JF (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. Proc Natl Acad Sci USA **94:** 6809–6814

Gibeaut DM, Carpita NC (1991) Tracing cell wall biogenesis in intact cells and plants: selective turnover and alteration of soluble and cell wall polysaccharides in grasses. Plant Physiol **97:** 551–561

Hannon GJ (2010) FASTX-Toolkit. http://hannonlab.cshl.edu/fastx_toolkit/ (July 9, 2014)

Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science **315:** 804–807

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, et al (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet **42:** 961–967

Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, et al (2012) Comparative population genomics of maize domestication and improvement. Nat Genet **44:** 808–811

Huntley SK, Ellis D, Gilbert M, Chapple C, Mansfield SD (2003) Significant increases in pulping efficiency in C4H-F5H-transformed poplars: improved chemical savings and reduced environmental toxins. J Agric Food Chem **51:** 6178–6183

Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ (2004) The UCSC Table Browser data retrieval tool. Nucleic Acids Res **32:** D493–D496

Kim WC, Ko JH, Kim JY, Kim J, Bae HJ, Han KH (2013) MYB46 directly regulates the gene expression of secondary wall-associated cellulose synthases in Arabidopsis. Plant J **73:** 26–36

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, et al (2005) Complement factor H polymorphism in age-related macular degeneration. Science **308:** 385–389

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res **40:** D1202–D1210

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods **9:** 357–359

Larsson SJ, Lipka AE, Buckler ES (2013) Lessons from Dwarf8 on the strengths and weaknesses of structured association mapping. PLoS Genet **9:** e1003246

Lawrence CJ, Schaeffer ML, Seigfried TE, Campbell DA, Harper LC (2007) MaizeGDB's new data types, resources and activities. Nucleic Acids Res **35:** D895–D900

Lee C, Zhong R, Ye ZH (2012) Arabidopsis family GT43 members are xylan xylosyltransferases required for the elongation of the xylan backbone. Plant Cell Physiol **53:** 135–143

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the Intermated B73 × Mo17 (IBM) population. Plant Mol Biol **48:** 453–461

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics **25:** 2078–2079

Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, et al (2010a) The developmental dynamics of the maize leaf transcriptome. Nat Genet **42:** 1060–1067

Li X, Chapple C (2010) Understanding lignification: challenges beyond monolignol biosynthesis. Plant Physiol **154:** 449–452

Li X, Ximenes E, Kim Y, Slininger M, Meilan R, Ladisch M, Chapple C (2010b) Lignin monomer composition affects Arabidopsis cell-wall degradability after liquid hot water pretreatment. Biotechnol Biofuels **3:** 27

Li X, Zhu C, Yeh CT, Wu W, Takacs EM, Petsch KA, Tian F, Bai G, Buckler ES, Muehlbauer GJ, et al (2012) Genic and nongenic contributions to natural variation of quantitative traits in maize. Genome Res **22:** 2436–2444

Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics **165:** 2117–2128

Lorenz AJ, Coors JG, Hansey CN, Kaeppler SM, deLeon N (2010) Genetic analysis of cell wall traits relevant to cellulosic ethanol production in maize (*Zea mays* L.). Crop Sci **50:** 842–852

Lorenzana RE, Lewis MF, Jung HJG, Bernardo R (2010) Quantitative trait loci and trait correlations for maize stover cell wall composition and glucose release for cellulosic ethanol. Crop Sci **50:** 541–555

Ma QH (2007) Characterization of a cinnamoyl-CoA reductase that is associated with stem development in wheat. J Exp Bot **58:** 2011–2021

Marita JM, Vermerris W, Ralph J, Hatfield RD (2003) Variations in the cell wall composition of maize brown midrib mutants. J Agric Food Chem **51:** 1313–1321

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, et al (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci USA **110:** 453–458

Mosier N, Wyman C, Dale B, Elander R, Lee YY, Holtzapple M, Ladisch M (2005) Features of promising technologies for pretreatment of lignocellulosic biomass. Bioresour Technol **96:** 673–686

Novaes E, Kirst M, Chiang V, Winter-Sederoff H, Sederoff R (2010) Lignin and biomass: a negative correlation for wood formation and lignin content in trees. Plant Physiol **154:** 555–561

Paschold A, Jia Y, Marcon C, Lund S, Larson NB, Yeh CT, Ossowski S, Lanz C, Nettleton D, Schnable PS, et al (2012) Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. Genome Res **22:** 2445–2454

Penning BW, Hunter CT, Tayengwa R, Eveland AL, Dugard CK, Olek AT, Vermerris W, Koch KE, McCarty DR, Davis MF, et al (2009) Genetic resources for maize cell wall biology. Plant Physiol **151:** 1703–1728

Penning BW, Sykes RW, Babcock NC, Dugard CK, Klimek JF, Gamblin D, Davis MF, Filley TR, Mosier NS, Weil CF, et al (January 30, 2014) Validation of PyMBMS for quantitation of phenylpropanoid abundance in lignocellulosic biomass from a maize recombinant inbred population. Bioenerg Res http://dx.doi.org/10.1007/s12155-014-9410-3

Saathoff AJ, Sarath G, Chow EK, Dien BS, Tobias CM (2011) Downregulation of cinnamyl-alcohol dehydrogenase in switchgrass by RNA silencing results in enhanced glucose release after cellulase treatment. PLoS ONE **6:** e16416

Saballos A, Sattler SE, Sanchez E, Foster TP, Xin Z, Kang C, Pedersen JF, Vermerris W (2012) Brown midrib2 (Bmr2) encodes the major 4-coumarate:coenzyme A ligase involved in lignin biosynthesis in sorghum (Sorghum bicolor (L.) Moench). Plant J **70:** 818–830

Schaeffer ML, Harper LC, Gardiner JM, Andorf CM, Campbell DA, Cannon EKS, Sen TZ, Lawrence CJ (2011) MaizeGDB: curation and outreach go hand-in-hand. Database (Oxford) **2011:** bar022

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science **326:** 1112–1115

Selig MJ, Adney WS, Himmel ME, Decker SR (2009) The impact of cell wall acetylation on corn stover hydrolysis by cellulolytic and xylanolytic enzymes. Cellulose **16:** 711–722

Selig MJ, Tucker MP, Law C, Doeppke C, Himmel ME, Decker SR (2011) High throughput determination of glucan and xylan fractions in lignocelluloses. Biotechnol Lett **33:** 961–967

Sharopova N, McMullen MD, Schultz L, Schroeder S, Sanchez-Villeda H, Gardiner J, Bergstrom D, Houchins K, Melia-Hancock S, Musket T, et al (2002) Development and mapping of SSR markers for maize. Plant Mol Biol **48:** 463–481

Simmons BA, Loqué D, Ralph J (2010) Advances in modifying lignin for enhanced biofuel production. Curr Opin Plant Biol **13:** 313–320

Simon A (2010) HTSeq: analysing high-throughput sequencing data with python. http://www-huber.embl.de/users/anders/HTSeq/doc/index.html (July 9, 2014)

Springer NM, Ying K, Fu Y, Ji T, Yeh CT, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, et al (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet **5:** e1000734

Stewart JJ, Akiyama T, Chapple C, Ralph J, Mansfield SD (2009) The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. Plant Physiol **150:** 621–635

Studer MH, DeMartini JD, Davis MF, Sykes RW, Davison B, Keller M, Tuskan GA, Wyman CE (2011) Lignin content in natural Populus variants affects sugar release. Proc Natl Acad Sci USA **108:** 6300–6305

Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, Padhukasahasram B, Karlins E, Davis S, Jones PG, et al (2007) A single IGF1 allele is a major determinant of small size in dogs. Science **316:** 112–115

Sykes R, Kodrzycki B, Tuskan G, Foutz K, Davis M (2008) Within tree variability of lignin composition in Populus. Wood Sci Technol **42:** 649–661

Tamasloukht B, Wong Quai Lam MS, Martinez Y, Tozo K, Barbier O, Jourda C, Jauneau A, Borderies G, Balzergue S, Renou JP, et al (2011) Characterization of a cinnamoyl-CoA reductase 1 (CCR1) mutant in maize: effects on lignification, fibre development, and global gene expression. J Exp Bot **62:** 3837–3848

Thomas J, Guillaumie S, Verdu C, Denoue D, Pichon M, Barriére Y (2010) Cell wall phenylpropanoid-related gene expression in early maize recombinant inbred lines differing in parental alleles at a major lignin QTL position. Mol Breed **25:** 105–124

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES IV (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet **28:** 286–289

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet **43:** 159–162

Truntzler M, Barrière Y, Sawkins MC, Lespinasse D, Betran J, Charcosset A, Moreau L (2010) Meta-analysis of QTL involved in silage quality of maize and comparison with the position of candidate genes. Theor Appl Genet **121:** 1465–1482

Tuskan G, West D, Bradshaw HD, Neale D, Sewell M, Wheeler N, Megraw B, Jech K, Wiselogel A, Evans R, et al (1999) Two high-throughput techniques for determining wood properties as part of a molecular genetics analysis of hybrid poplar and loblolly pine. Appl Biochem Biotechnol **77-79:** 55–65

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. Plant Physiol **153:** 895–905

Venkatakrishnan VK, Degenstein J, Delgass WN, Agrawal R, Ribeiro FH (2014) High pressure fast-pyrolysis, fast-hydropyrolysis and catalytic hydrodeoxygenation of cellulose: production of liquid fuel from biomass. Green Chem **16:** 792–802

Vermerris W, Saballos A, Ejeta G, Mosier NS, Ladisch MR, Carpita NC (2007) Molecular breeding to enhance ethanol production from corn and sorghum stover. Crop Sci **47:** S142–S153

Wang H, Avci U, Nakashima J, Hahn MG, Chen F, Dixon RA (2010) Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. Proc Natl Acad Sci USA **107:** 22338–22343

Yin T, Zhang X, Gunter L, Priya R, Sykes R, Davis M, Wullschleger SD, Tuskan GA (2010) Differential detection of genetic loci underlying stem and root lignin content in Populus. PLoS ONE **5:** e14021

Zhao Q, Dixon RA (2011) Transcriptional networks for lignin biosynthesis: more complex than we thought? Trends Plant Sci **16:** 227–233

Zhong R, Lee C, McCarthy RL, Reeves CK, Jones EG, Ye ZH (2011) Transcriptional activation of secondary wall biosynthesis by rice and maize NAC and MYB transcription factors. Plant Cell Physiol **52:** 1856–1871

Zhong R, Lee C, Ye ZH (2010) Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis. Trends Plant Sci **15:** 625–632

Zhou J, Lee C, Zhong R, Ye ZH (2009) MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in *Arabidopsis*. Plant Cell **21:** 248–266