Network-based integration of systems genetics data reveals pathways associated with lignocellulosic biomass accumulation and processing

Eshchar Mizrachi[a,b,1,2], Lieven Verbeke[c,d,1], Nanette Christie[a,b], Ana C. Fierro[c,d], Shawn D. Mansfield[e], Mark F. Davis[f], Erica Gjersing[f], Gerald A. Tuskan[g], Marc Van Montagu[d,h,2], Yves Van de Peer[b,d,h], Kathleen Marchal[a,c,d,2], and Alexander A. Myburg[a,b,2]

[a]Department of Genetics, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria 0028, South Africa; [b]Genomics Research Institute, University of Pretoria, Pretoria 0028, South Africa; [c]Department of Information Technology, Ghent University, iMinds, 9052 Gent, Belgium; [d]Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Gent, Belgium; [e]Department of Wood Science, University of British Columbia, Vancouver, BC, Canada V6T 1Z4; [f]Biosciences Center, National Renewable Energy Laboratory, Golden, CO 80401; [g]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831; and [h]Department of Plant Systems Biology, VIB, 9052 Gent, Belgium

Contributed by Marc Van Montagu, December 12, 2016 (sent for review October 7, 2016; reviewed by Kenneth Keegstra and Joachim Selbig)

As a consequence of their remarkable adaptability, fast growth, and superior wood properties, eucalypt tree plantations have emerged as key renewable feedstocks (over 20 million ha globally) for the production of pulp, paper, bioenergy, and other lignocellulosic products. However, most biomass properties such as growth, wood density, and wood chemistry are complex traits that are hard to improve in long-lived perennials. Systems genetics, a process of harnessing multiple levels of component trait information (e.g., transcript, protein, and metabolite variation) in populations that vary in complex traits, has proven effective for dissecting the genetics and biology of such traits. We have applied a network-based data integration (NBDI) method for a systems-level analysis of genes, processes and pathways underlying biomass and bioenergy-related traits using a segregating *Eucalyptus* hybrid population. We show that the integrative approach can link biologically meaningful sets of genes to complex traits and at the same time reveal the molecular basis of trait variation. Gene sets identified for related woody biomass traits were found to share regulatory loci, cluster in network neighborhoods, and exhibit enrichment for molecular functions such as xylan metabolism and cell wall development. These findings offer a framework for identifying the molecular underpinnings of complex biomass and bioprocessing-related traits. A more thorough understanding of the molecular basis of plant biomass traits should provide additional opportunities for the establishment of a sustainable bio-based economy.

systems genetics | lignocellulosic biomass | cell wall | bioenergy | network-based data integration

Wood (secondary xylem tissue) from trees represents a significant proportion of global carbon sequestration (1, 2), while also providing raw materials, be it in the form of timber, paper, or other biomaterials and value-added derivatives, such as cellulose (3) and lignin (4). Fast-growing trees such as poplars and eucalypts with short rotation times, small genome sizes (500–600 Mbp), large genetic diversity, and established breeding populations are widely cultivated as woody biomass crops and are well suited for biotechnological improvement (5–8). A tree's amenability to bioprocessing for pulp, paper, cellulose, and other bioproducts is dependent on the aggregate of its wood properties, which are a function of cellular architecture and the chemistry and ultrastructure of the secondary cell walls (SCWs) of wood fiber cells that compose the bulk of woody biomass (9). These properties are determined by overlapping developmental programs and pathways that have to be coordinated during secondary xylem development (xylogenesis) (10). Some of these pathways, SCW polysaccharide and lignin biosynthesis in particular, represent a strong, irreversible carbon sink and directly or indirectly use core metabolites (glucose/UDP-glucose and fructose) that are also used for growth and physiological/cellular homeostasis (energy metabolism, production of amino acids, etc.). These metabolic interdependencies and the multitude of biological processes involved result in woody biomass traits having complex genetic architectures, especially in highly outbred organisms such as forest trees (11).

Functional genetics approaches based mainly on single-gene perturbations have been informative in revealing components of the SCW producing system in model plants (12). However, they have not necessarily yielded insight into the complex interactions of the genes that naturally affect cell wall chemistry and ultrastructure in ways that do not interfere with normal plant growth, form, and biomass accumulation. The characterization of natural genetic perturbations segregating in phenotypically wild-type individuals offers an attractive alternative tool to study properties that emerge from permissible genetic variation (11, 13). With the availability of high-throughput genotyping and high-resolution linkage maps in *Populus* (14, 15) and *Eucalyptus* (16–19), genetic approaches such as quantitative trait locus (QTL) (14, 20, 21) and LD-based association mapping (22–25) are becoming feasible to

Significance

Carbon fixation and accumulation as lignocellulosic biomass is of global ecological and industrial importance and most significantly occurs in the form of wood development in trees. Traits of importance in biomass accumulation are highly complex and, aside from environmental factors, are affected by many pathways and thousands of genes. We have applied a network-based data integration method for a systems genetics analysis of genes, processes, and pathways underlying biomass and bioenergy-related traits using segregating *Eucalyptus* hybrid tree populations. We could link biologically meaningful sets of genes to complex traits and at the same time reveal the molecular basis of trait variation. Such a holistic view of the biology of wood formation will contribute to genetic improvement and engineering of plant biomass.

Author contributions: E.M., M.V.M., Y.V.d.P., K.M., and A.A.M. designed research; E.M., L.V., N.C., S.D.M., M.F.D., E.G., and G.A.T. performed research; E.M., L.V., N.C., and A.C.F. analyzed data; and E.M., L.V., N.C., S.D.M., M.F.D., E.G., G.A.T., M.V.M., Y.V.d.P., K.M., and A.A.M. wrote the paper.

Reviewers: K.K., Michigan State University; and J.S., Max Planck Institute of Molecular Plant Physiology.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the NCBI Sequence Read Archive (accession no. SUB2087452).

[1]E.M. and L.V. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: mamon@psb.ugent.be, eshchar.mizrachi@fabi.up.ac.za, kathleen.marchal@intec.ugent.be, or zander.myburg@fabi.up.ac.za.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1620119114/-/DCSupplemental.

Table 1. Comparison of using NBDI-transformed versus using nontransformed expression values in prioritizing reference genes for 13 traits

Trait	NBDI enrichment *p*	Nontransformed enrichment *p*
DBH (over bark)	6.38×10^{-20}	2.81×10^{-11}
DBH (under bark)	5.55×10^{-15}	1.38×10^{-10}
Bark thickness	4.61×10^{-19}	1.38×10^{-10}
Wood density	6.52×10^{-10}	5.49×10^{-12}
Lignin content	2.83×10^{-01}	9.28×10^{-01}
Total C5 sugar in walls	1.02×10^{-01}	2.83×10^{-01}
Total C6 sugar in walls	1.02×10^{-01}	9.75×10^{-01}
Glucose released	5.49×10^{-12}	3.03×10^{-04}
Percent of max glucose released	3.29×10^{-14}	1.03×10^{-04}
Xylose released	1.88×10^{-13}	1.02×10^{-01}
% of max xylose release	2.21×10^{-03}	7.19×10^{-01}
Glucose + xylose released	2.81×10^{-11}	3.03×10^{-04}
Percent of max sugar released	3.29×10^{-14}	2.21×10^{-03}

For each trait, 300 genes were selected. The enrichment score corresponds to the *P* value of a hypergeometric enrichment test.

study woody biomass traits in long-lived perennial plants. While such association mapping approaches facilitate the delimitation of genome positions harboring causal variation (13), they provide little information as to how these genes and their variants act together in biological pathways to influence trait variation.

Complementing genetic information with molecular phenotypes (e.g., transcript levels) can contribute to a better mechanistic understanding of trait variation (24, 26, 27). Expression QTL (eQTL) analysis (28) allows the identification of genomic loci associated with variation in molecular phenotypes. In contrast to QTL or association analyses, eQTL analysis also identifies the genes that are affected by this variation and thus potentially contribute to the complex trait. However, because complex traits are subject to the combined effect of multiple genetic loci, each with a small effect on the trait, there is generally low power to detect statistically significant associations when relying on single-gene analysis. In addition, single gene associations do not show how genes and/or pathways interact to explain a complex trait (29).

To cope with the aforementioned limitations of the association problem, more integrated systems genetics approaches have been proposed (29, 30). Methods that use network models to represent molecular a priori knowledge on the organism/trait of interest (31) have been particularly successful to perform association analysis in clonal systems (32–35). In the context of outbred populations, network-based methods have been applied for gene prioritization (36) or to increase the reliability of eQTL association mapping itself (37) but not yet for integrative association analysis.

Here we applied a systems genetics approach to study the genomic loci and pathways affecting wood formation in *Eucalyptus*. We generated coupled genetics/genomics (linkage map and immature xylem transcriptome) data for 156 individuals segregating from an F2 pseudobackcross between a *Eucalyptus grandis* × *Eucalyptus urophylla* F1 interspecific hybrid tree and an unrelated *E. urophylla* tree and profiled traits representative of tree growth (diameter at breast height and bark thickness), wood properties (wood basic density and cell wall composition), and bioprocessing metrics (sugar release). Data were integrated using a unique network-based data integration (NBDI) approach that allows combining genotyping, expression profiling, and prior network information to prioritize genes and molecular mechanisms associated with complex wood formation traits.

Results

Network-Based Gene–Trait Association. One hundred fifty-six *E. grandis* × *E. urophylla* F2 interspecific backcross trees were profiled for transcript abundance in immature xylem and for woody biomass traits that relate to growth, wood density, cell wall composition, and sugar extractability (Table 1 and *Materials and Methods*). Genotyping, expression profiling, inferred eQTL associations, and prior network information were simultaneously used to prioritize genes and molecular mechanisms associated with each of the measured traits. To this end we developed an integration approach that makes use of a gene interaction network model in which nodes are genes and edges represent two types of information. If derived from prior information, edges reflect relations between genes and gene products, derived from Kyoto Encyclopedia of Genes and Genomes (KEGG) (*SI Materials and Methods*). If derived from eQTL associations, edges reflect that the connected genes share the same eQTL and thus are likely functionally related and coregulated.

This network model is then used to propagate on a per sample basis the expression signals of genes to a local network neighborhood. This propagation transforms the original gene expression data to network-diffused gene expression data (Fig. 1 and *Materials and Methods*). In the network-diffused expression matrix, each data point can be interpreted as the original expression signal of a gene in a sample, modulated by the expression of the genes that are close neighbors in the network (i.e., that are likely found in the same pathways or to share eQTLs, etc.). Modulation implies that if nodes in the local neighborhood of a gene are

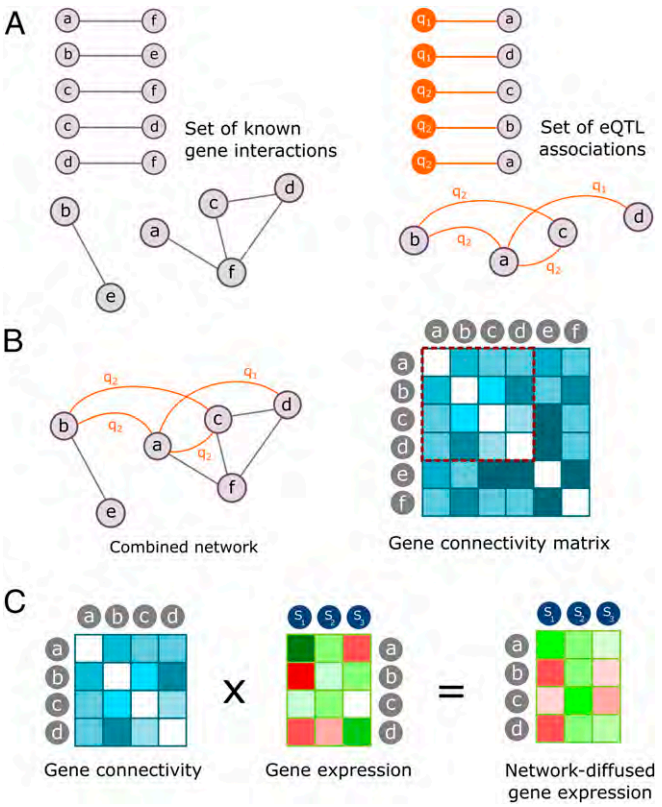

Fig. 1. An overview of the NBDI approach. Letters a–f represent genes, q_1–q_3 are eQTL, and S_1–S_3 are samples (xylem samples from individual trees). (*A*) Prior knowledge and eQTL associations are converted to a prior and functional interaction network, respectively. (*B*) The two networks are integrated to construct a combined network, and a gene–gene connectivity matrix is calculated (bright and dark colors show high and low connectivity, respectively). (*C*) The connectivity matrix (genes not present in the expression data are omitted) is used to diffuse gene expression data through the network.

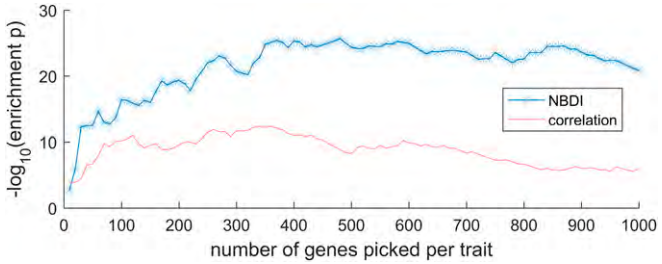

Fig. 2. Performance comparison of using NBDI-transformed (blue starred line) versus using nontransformed expression correlation data (red line) in prioritizing trait-associated genes. Enrichment probability (*y* axis; *P* value obtained using a hypergeometric enrichment test) of the combined gene selection (the union of all of the gene sets that were obtained per trait or, alternatively, the set of genes linked to at least one trait) is shown as a function of the number of genes prioritized per individual trait (*x* axis).

also expressed, the expression signal of the gene is confirmed and its relevance is increased; otherwise, its relevance is decreased. Using diffused gene expression is comparable to module-based analysis, where before further analysis, groups of genes with correlating gene expression (modules) are identified, under the assumption that these genes participate in the same biological processes or belong to the same pathway. The network neighborhood used to modulate the expression of a gene can be viewed as an implicit module of functionally related genes, with the main advantage of our approach that each individual gene can be prioritized or deprioritized for a particular trait based on the integrated information from network neighbors.

Once the gene expression values are diffused through the network to obtain the transformed expression values (referred to as NBDI-transformed values), genes relevant for each trait under study are identified by correlation analysis. Correlating NBDI-transformed values to each of the traits allows for ranking, per trait, of genes of potential relevance to that trait (for gene selections for each trait, see Dataset S1). To benchmark the performance of the NBDI approach, we compiled a literature-based set of reference genes with known biological roles in SCW biosynthesis (the best characterized and a central biological process in wood formation; Dataset S2) and calculated the extent to which gene sets associated with the traits were enriched for these reference genes. To illustrate the added value of using NBDI-transformed values, we also performed the same associations but using the nontransformed gene expression values when correlating gene expression variation to trait variation. The difference in prioritization performance between our approach and that of using nontransformed expression values is illustrated as a function of the number of reference genes that is prioritized for either all traits together (Fig. 2) or for each individual trait (Fig. S1). Compared with using nontransformed expression data, NBDI-transformed data result in higher enrichment of reference genes, regardless of how many genes per trait were selected (Fig. 2). This higher enrichment can be explained by the fact that for most traits the NBDI transformation results in assigning a higher prioritization rank to reference genes than when using the nontransformed values (Fig. S2). In addition, with NBDI-transformed genes we were able to prioritize genes that associated to several traits as well as genes that were trait specific (Fig. S3). In the remainder of the analyses, we selected the 300 genes (Dataset S1) of which the NBDI-transformed expression correlated best with the trait. With this threshold a significant enrichment for reference genes (Table 1) was obtained, while still yielding results that were unique enough to explain trait differences. Applying these criteria resulted in a gene selection of 1,529 nonredundant genes (Dataset S1) that were linked to at least one of the traits and of which 102 were found in the reference set ($p_{\mathrm{NBDI}} = 1.38 \times 10^{-21}$).

Genes and Pathways of Relevance to the Traits Under Study. The NBDI approach combines genetic and prior information with gene expression variation to prioritize, per trait, relevant genes/pathways influenced by genetic variation in the population. This combined information is captured in the two complementary views, an eQTL (Fig. 3) and a network view (Fig. 4).

First, because of its network model, NBDI implicitly imposes that genes prioritized for a trait should also share an eQTL at one or more loci in the genome. Most eQTLs for genes prioritized by the NBDI approach for a particular trait should therefore cluster together in hot spots rather than being randomly scattered along the genome. Fig. 3 shows this is indeed the case.

Second, as the NBDI approach favors genes that are connected in the interaction network, at least some of the genes associated with a trait, when projected on a gene interaction network, should cluster together and constitute a molecular subnetwork underlying the trait providing a complementary network view. Because this network of curated gene interactions (derived from KEGG) is sparser than the network used for the NBDI analysis that also includes eQTL overlap relations, by definition only a subset of the gene selections and relations can be visualized in subnetworks. To visualize eQTL relations for genes that could be connected through KEGG, eQTL overlap relations were overlaid on the identified subnetworks. For each trait, the largest connected component of the obtained network is

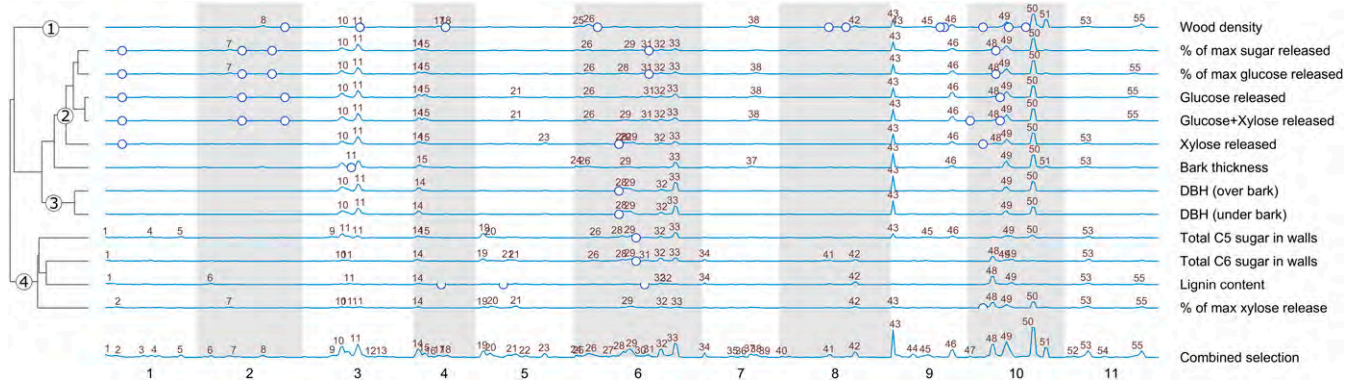

Fig. 3. Location of eQTLs shared by the genes selected for each individual trait. Height of the peaks represents the number of genes that have an eQTL in the same 5-cM-wide region (bin). Adjacent bins overlap 50%. Peak numbers correspond to the peaks identified for the combined gene selection (bottom row). Genes with eQTL in each peak are listed in Dataset S3. Blue circles indicate previously identified trait QTLs. Traits are clustered based on the eQTL density profile (*SI Materials and Methods*). *x* axis labels are chromosome numbers. The white and gray vertical bars represent chromosomes. Circled numbers on the tree on the left correspond to the major groups of traits discussed in the main text.

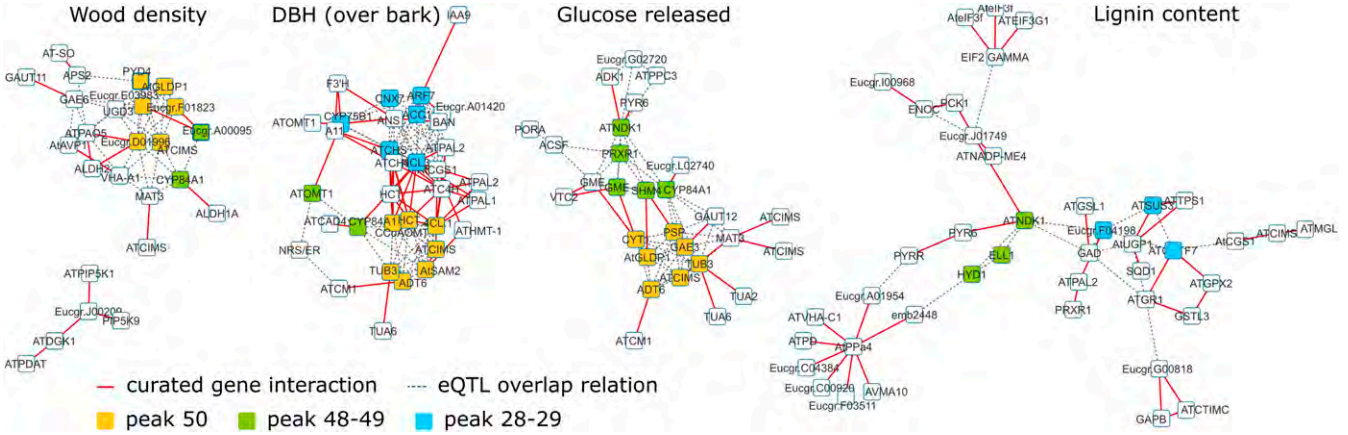

Fig. 4. Molecular networks associated with wood density, DBH, glucose released, and lignin content, representing the four trait groups (Fig. 3). Interactions present in KEGG are shown as red edges. Additional eQTL-based edges (gray dotted lines) were added when present. Due to the fact that some *Arabidopsis* genes have multiple homologs in the *Eucalyptus* genome (and vice versa), gene symbols may be present multiple times. Colors represent the genes with eQTL in the indicated frequency peaks (Fig. 3).

extracted, containing the prioritized genes that can be connected in the network through direct edges. Representative subnetworks of the broad trait classes (Fig. 3) are presented in Fig. 4. The networks corresponding to wood density, diameter at breast height (DBH), glucose released, and lignin content contain 20, 39, 28, and 48 genes, respectively, from the original 300 selected genes per trait and were highly significant (the probability of obtaining a connected component with the observed size purely by chance was smaller than 10^{-10}, 10^{-23}, 10^{-17}, and 10^{-32}, respectively).

As shown in Figs. 3 and 4, the traits under study depend, at least partially, on similar processes and shared eQTLs. This is expected given that all traits reflect wood-related properties and will be, to different extents, phenotypically related. However, trait-specific differences can be identified. When the traits are clustered based on the eQTLs shared by the genes in their gene selections (*SI Materials and Methods*), four broad groups of traits can be identified (Fig. 3): (1) wood density, (2) sugar release (bioprocessing) metrics, (3) growth traits, and (4) C5 and C6 sugar in cell walls together with lignin content and the percent of maximum xylose release. Below, the biological functions of the genes associated with the shared eQTL peaks found for traits and trait groups are discussed in more detail (Fig. S4 and Dataset S3).

The biological functions of the genes associated with traits in groups 2 (sugar release) and 3 (growth traits) are characterized by shared cell wall related processes (cell wall organization and bioprocessing, hemicellulose metabolic processes, xylan biosynthesis, glucuronxylan biosynthesis, and lignin biosynthesis) involving genes with eQTL located at peaks 43, 48, 49, and 50, and by processes that are shared by all traits at eQTL peaks 10 and 11 (but for which no functional overrepresentation could be assigned) (Fig. S4). All growth related traits (bark thickness and DBH) and all bioprocessing (sugar release) metrics (except percent maximum xylose release) belong to this group.

Growth-related traits (group 3) seem to be dominated, in addition to the cell wall related processes mentioned above, by anthocyanin related processes (eQTL peak 33). This is also illustrated in the DBH subnetwork that is representative for this trait group (Fig. 4). The DBH subnetwork contains a considerable number of lignin-related genes. Most bona fide lignin biosynthesis genes that are highly expressed in developing xylem (8, 38) (Dataset S3) were indeed associated with growth (DBH and bark thickness traits). In addition to these lignin-related genes, the DBH network contains several genes involved in hormone signaling (IAA9 and ARF7) and flavonoid biosynthesis (ANS, BAN, and A11). The relationship between lignin and flavonoid biosynthesis has been shown in *Arabidopsis*, where silencing of hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (HCT) resulted in decreased plant growth and redirection of the metabolic flux into flavonoid production through chalcone synthase (39).

In contrast, sugar release traits (trait group 2) and wood density (trait group 1) differ from the growth-related traits (group 3) in their larger involvement of photosynthesis-related processes (related to eQTL peak 46; Dataset S3). Because secondary xylem tissue is the main carbohydrate sink in woody plants, its composition is expected to be affected by the availability of fixed carbon resulting from photosynthesis (40). A representative subnetwork of this group of sugar release traits (percent glucose released; Fig. 4) is much less dominated by prior interactions (they might be less known for the processes involved) but clearly shows the presence of two clusters of genetic associations involving, among others, genes encoding TUBULIN ALPHA-2 CHAIN (TUA2), COBALAMIN-INDEPENDENT METHIONINE SYNTHASE/METHIONINE SYNTHESIS 1 (ATCIMS/MS1), and IRREGULAR XYLEM 8/GALACTURONOSYLTRANSFERASE 12 (IRX8/GAUT12). homologs of which have previously been identified as associated with hemicellulose and lignin related properties in *Populus* (24). IRX8/GAUT12 is known to be involved in xylan structure (41), and modulation of its expression in poplar results in improved sugar release efficiency (42). More indirect arguments can be made for TUA2 and ATCIMS/MS1. Tubulins affect cortical microtubule arrangement and thus cellulose microfibril angle (MFA) (43, 44). TUA2 specifically is known to be a target of SND1, the master regulator of secondary cell wall deposition (45, 46), and is one of two alpha-tubulins that is significantly up-regulated in *Eucalyptus* during tension wood formation, during which MFA is one of the main physical changes in the wood (47). Given that MFA is thought to affect wood ultrastructure and stiffness (48, 49), it is interesting to find this tubulin associated with bioprocessing-related traits such as sugar release in this study. Several associations were identified (mainly with sugar release efficiency) for genes involved in cysteine and methionine metabolism, including ATCIMS/MS1. The roles of these genes in biomass formation are becoming increasingly revealed, being linked directly to either lignin (50, 51) or hormone-mediated growth regulation (52). In addition, SCW polysaccharide biosynthesis genes known to be expressed in developing xylem (8) were mainly associated with variation in wood density and glucose release efficiency (Dataset S3). In the latter case, the majority were xylan modification genes, affecting patterns of acetylation, and glucuronic acid and methyl-glucuronic acid decoration of the xylan backbone (53, 54) (Dataset S3).

Trait group 4 contains traits related to the total cell wall sugar content and, surprisingly, also lignin content and percent maximum xylose released. These traits are characterized by the relatively smaller effect of major eQTL peaks that dominate most of the other biomass traits. The lignin subnetwork (Fig. 4) in general lacks most of the genes related to lignin biosynthesis itself. Indeed, few association between the variation of expression of SCW biosynthetic genes (cellulose, xylan, and lignin pathway genes) and the final C5 and C6 sugar and lignin content of the cell wall were apparent (Dataset S3). Several genes highly associated with variation in lignin content code for enzymes involved in carbon metabolism, including phosphoenolpyruvate (phosphoenolpyruvate carboxykinase enolase), pyruvate (plastidial pyruvate kinase 3 and malate dehydrogenase), and acetyl-CoA (pyruvate dehydrogenase E1 α subunit) metabolism, as well as pathways producing UDP-glucose and fructose (UDP-glucose pyrophosphorylase and sucrose synthase; Dataset S3). Genes involved in mitochondrial energy metabolism were also associated with lignin content. Very distinctive for this group of traits are also the genes that relate to abiotic stress (eQTL peak 34) and RNA modification (a process associated with a very distinctive eQTL peak 48).

As an additional external validation, we overlaid previously identified QTLs (blue circles in Fig. 3) for the same complex traits (20) with the obtained eQTL frequency peaks (Fig. 3). These results show that at least some of these trait QTLs are in close proximity to eQTL frequency peaks (especially peaks 28, 29, 48, and 49, located on chromosomes 6 and 10), providing additional evidence that the gene selection is relevant for the trait under study. Several eQTL peaks cannot be directly mapped to trait QTLs. These might represent polymorphisms that only have detectable effects on molecular subcomponents of a trait but cannot be directly associated with the phenotype itself. Given that complex traits are affected by different molecular traits in epistatic and nonlinear ways, a direct link between molecular traits and phenotypic traits is not always expected or the effects are too numerous and small to detect at the level of trait QTLs given the relatively small size of experimental population ($n = 156$).

Discussion

The observed variability of woody biomass traits in this study is explained by the variation of combinations of genes or sets of closely interacting pathways influenced by genetic variation segregating in this particular interspecific backcross population. Because of this, linking a quantitative trait to the expression of individual genes might fail or be incomplete if the trait under study is influenced by variation in the expression of large numbers of genes that in turn can be influenced by the expression of other genes, etc. If this is the case, then any method that captures only the marginal effect of a gene on a trait might render only a partial view of the genes that are involved in the underlying biological processes. To cope with this statistical issue we have developed a network-based data integration approach (NBDI) that combines genotyping, expression profiling, and prior network information to prioritize genes and molecular mechanisms associated with measured traits.

This NBDI approach is based on a network model in which connections between genes reflect interactions derived from either prior molecular interaction information or from eQTL information. In the latter case it is assumed that if two genes share an eQTL, they are connected in the network because of a shared coregulation mechanism. Even though incidental overlap of eQTLs is possible, for instance, through the action of separate polymorphisms in tightly linked but unrelated genes, we assumed that the majority of the overlapping trans-eQTLs can be treated as evidence of a shared regulatory polymorphism, as reflected by the shared functional annotations observed for the associated genes. Gene expression signals are then propagated through the network model to obtain an integrated signal that is used to explain the variation in the external traits.

We applied the NBDI approach to study the genomic loci and pathways affecting wood formation in *Eucalyptus*. The experimental setup used [with high linkage disequilibrium (LD) and large effect QTLs segregating in a single family] is complementary to low-LD studies (with high resolution but typically small effect associations) in populations of unrelated individuals (e.g., refs. 24, 55).

Using our integrative systems genetics approach allowed for prioritizing genes contributing to woody biomass traits and identifying the putative regulatory loci with which these genes and traits are predominantly associated. Based on this analysis, a clear distinction could be made between growth and sugar release (bioprocessing) related traits and traits related to the total cell wall sugar content. Unexpectedly, we noticed little association between the variation of expression of SCW biosynthetic genes (cellulose, xylan, and lignin pathway genes) and the final C5 and C6 sugar and lignin content of the cell wall. Rather, most bona fide lignin biosynthesis genes were associated with growth-related traits (DBH and bark thickness), and most SCW polysaccharide biosynthesis genes were associated with variation in wood density and glucose release efficiency. Several genes highly associated with variation in lignin content code for enzymes involved in carbon metabolism and in mitochrondrial energy metabolism. As a result, we hypothesize that variation in the expression of SCW biosynthetic genes has an effect on the growth and ultrastructure and resultant processability of the secondary cell wall, whereas the quantity of sequestered carbon in the cell wall (in the form of polysaccharides and phenolics) is more related to variation in primary carbon metabolism pathways and hence precursor availability. This assumption further establishes the strong link between physiological/cellular homeostasis and secondary processes such as SCW polysaccharide and lignin biosynthesis that represent a strong, irreversible carbon sink in woody plants.

Materials and Methods

Experimental Population, Transcriptome, and Complex Trait Analysis. The F2 backcross population was generated from a cross between an *E. grandis* × *E. urophylla* F1 interspecific hybrid parent (GUSAP1, Sappi Forest Research, South Africa) and an unrelated *E. urophylla* parent (USAP1) (18). At 3 y old, immature xylem tissue was harvested from 156 individuals as previously described (56). Samples were collected from 3-y-old trees over a 7.5-h period between 0900 and 1630 hours for 3 d. Total RNA was isolated (57) and used for RNA-Seq expression profiling (30 million; Illumina PE50, BGI Hong Kong). Gene expression values (FPKM) were calculated per gene model using TopHat version 1.3 and Cufflinks version 1.0.3 (bias correction and quartile normalization was enabled for the FPKM calculation) (58, 59). Diameter (cm) at breast height (DBH) of the main stem was assessed as described previously (20). Bark thickness was calculated as the difference between over-bark and under-bark DBH measurements. A wood disk taken at breast height (1.35 m) was used to determine wood basic density using the water displacement method (www.tappi.org/content/SARG/T258.pdf). Chemical wood properties were assessed using different analytical methods, including pyrolysis molecular beam mass spectrometry (pyMBMS).

Trait QTL mapping, eQTL mapping, and eQTL classification are described in *SI Materials and Methods*.

NBDI Association Analysis. First, a hybrid gene interaction network was constructed using curated gene interactions downloaded from KEGG and eQTL overlap relations (Fig. 1*A*). For the latter, we investigated for pairs of genes whether these genes had overlapping eQTL intervals (Fig. S5). If this is the case, a connection in the hybrid network is added. Once the network is constructed, a graph node kernel was calculated (the Laplacian exponential diffusion kernel; 60) to quantify how well each node in the network connects to all other nodes (Fig. 1*B*). The resulting connectivity matrix was then multiplied with the gene expression matrix to obtain the diffused or transformed gene expression matrix (Fig. 1*C*). Genes in the network connectivity matrix that were not present in the gene expression matrix were removed and vice versa. The transformed gene expression was finally linked to the measured traits by calculating the absolute value of the Pearson correlation between the transformed gene expression and the measured traits. After ranking, the top 300 genes exhibiting the highest correlation were selected for further analysis. For details of the eQTL overlap procedure, network construction, connectivity calculation, and association analysis, see *SI Materials and Methods* and Figs. S5 and S6.

ACKNOWLEDGMENTS. This work was supported by the Department of Science and Technology (Strategic Grant for the *Eucalyptus* Genomics Platform) and National Research Foundation of South Africa (Bioinformatics and Functional Genomics Programme, Grants 86936 and 97911 to A.A.M.), Sappi South Africa and the Technology and Human Resources for Industry Programme (Grant 80118) through the Forest Molecular Genetics Programme at the University of Pretoria (to A.A.M.), Ghent University Multidisciplinary Research Partnership from nucleotides to networks (Project 01MR0410W to Y.V.d.P. and K.M.), the European Union (FP7/2007-2013) under ERC Advanced Grant Agreement 322739–DOUBLEUP (to Y.V.d.P.), the Fonds Wetenschappelijk Onderzoek - Vlaanderen (Projects 3G042813, G.0A53.15N, and SBO-NEMOA to K.M.), and the BioEnergy Science Center, a US Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research on the Department of Energy Office of Science (G.A.T.). Finally, the authors acknowledge Sappi Forest Research for the plant materials and growth and wood property data used in the study.

1. Bonan GB (2008) Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science* 320(5882):1444–1449.
2. Crowther TW, et al. (2015) Mapping tree density at a global scale. *Nature* 525(7568): 201–205.
3. Mizrachi E, Mansfield SD, Myburg AA (2012) Cellulose factories: Advancing bioenergy production from forest trees. *New Phytol* 194(1):54–62.
4. Ragauskas AJ, et al. (2014) Lignin valorization: Improving lignin processing in the biorefinery. *Science* 344(6185):1246843.
5. Hinchee M, et al. (2009) Short-rotation woody crops for bioenergy and biofuels applications. *In Vitro Cell Dev Biol Plant* 45(6):619–629.
6. Sederoff R, Myburg A, Kirst M (2009) Genomics, domestication, and evolution of forest trees. *Cold Spring Harb Symp Quant Biol* 74:303–317.
7. Séguin A (2011) How could forest trees play an important role as feedstock for bioenergy production? *Curr Opin Environ Sustain* 3(1-2):90–94.
8. Myburg AA, et al. (2014) The genome of *Eucalyptus grandis*. *Nature* 510(7505):356–362.
9. Mansfield SD (2009) Solutions for dissolution–engineering cell walls for deconstruction. *Curr Opin Biotechnol* 20(3):286–294.
10. Hussey SG, Mizrachi E, Creux NM, Myburg AA (2013) Navigating the transcriptional roadmap regulating plant secondary cell wall deposition. *Front Plant Sci* 4:325.
11. Mizrachi E, Myburg AA (2016) Systems genetics of wood formation. *Curr Opin Plant Biol* 30:94–100.
12. Vanholme R, et al. (2012) A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis*. *Plant Cell* 24(9):3506–3529.
13. Ingvarsson PK, Street NR (2011) Association genetics of complex traits in plants. *New Phytol* 189(4):909–922.
14. Muchero W, et al. (2015) High-resolution genetic mapping of allelic variants associated with cell wall chemistry in *Populus*. *BMC Genomics* 16(1):24.
15. Geraldes A, et al. (2013) A 34K SNP genotyping array for *Populus trichocarpa*: Design, application to the study of natural populations and transferability to other *Populus* species. *Mol Ecol Resour* 13(2):306–323.
16. Bartholomé J, et al. (2015) High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. *New Phytol* 206(4):1283–1296.
17. Hudson CJ, et al. (2012) A reference linkage map for *Eucalyptus*. *BMC Genomics* 13(1):240.
18. Kullan ARK, et al. (2011) High-density genetic linkage maps with over 2,400 sequence-anchored DarT markers for genetic dissection in an F2 pseudo-backcross of *Eucalyptus grandis* × *E. urophylla*. *Tree Genet Genomes* 8(1):163–175.
19. Silva-Junior OB, Faria DA, Grattapaglia D (2015) A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytol* 206(4):1527–1540.
20. Kullan AR, et al. (2012) Genetic dissection of growth, wood basic density and gene expression in interspecific backcrosses of *Eucalyptus grandis* and *E. urophylla*. *BMC Genet* 13:60.
21. Thumma BR, et al. (2010) Quantitative trait locus (QTL) analysis of wood quality traits in *Eucalyptus nitens*. *Tree Genet Genomes* 6(2):305–317.
22. McKown AD, et al. (2014) Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. *New Phytol* 203(2):535–553.
23. Evans LM, et al. (2014) Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nat Genet* 46(10):1089–1096.
24. Porth I, et al. (2013) Network analysis reveals the relationship among wood properties, gene expression levels and genotypes of natural *Populus trichocarpa* accessions. *New Phytol* 200(3):727–742.
25. Wegrzyn JL, et al. (2010) Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. *New Phytol* 188(2):515–532.
26. Du Q, et al. (2016) Genetic architecture of growth traits in Populus revealed by integrated quantitative trait locus (QTL) analysis and association studies. *New Phytol* 209(3):1067–1082.
27. Thavamanikumar S, Southerton S, Thumma B (2014) RNA-Seq using two populations reveals genes and alleles controlling wood traits and growth in *Eucalyptus nitens*. *PLoS One* 9(6):e101104.
28. Jansen RC, Nap JP (2001) Genetical genomics: The added value from segregation. *Trends Genet* 17(7):388–391.
29. Feltus FA (2014) Systems genetics: A paradigm to improve discovery of candidate genes and mechanisms underlying complex traits. *Plant Sci* 223:45–48.
30. Baute J, et al. (2016) Combined large-scale phenotyping and transcriptomics in maize reveals a robust growth regulatory network. *Plant Physiol* 170(3):1848–1867.
31. Proost S, Mutwil M (2016) Tools of the trade: Studying molecular networks in plants. *Curr Opin Plant Biol* 30:143–150.
32. De Maeyer D, Weytjens B, De Raedt L, Marchal K (2016) Network-based analysis of eQTL data to prioritize driver mutations. *Genome Biol Evol* 8(3):481–494.
33. Verbeke LP, et al. (2015) Pathway relevance ranking for tumor samples through network-based data integration. *PLoS One* 10(7):e0133503.
34. Ding L, Wendl MC, McMichael JF, Raphael BJ (2014) Expanding the computational toolbox for mining cancer genomes. *Nat Rev Genet* 15(8):556–570.
35. Shi K, Gao L, Wang B (2016) Discovering potential cancer driver genes by an integrated network-based approach. *Mol Biosyst* 12(9):2921–2931.
36. Verbeke LP, Cloots L, Demeester P, Fostier J, Marchal K (2013) EPSILON: An eQTL prioritization framework using similarity measures derived from local networks. *Bioinformatics* 29(10):1308–1316.
37. Jia P, Zhao Z (2014) Network.assisted analysis to prioritize GWAS results: Principles, methods and perspectives. *Hum Genet* 133(2):125–138.
38. Carocha V, et al. (2015) Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis*. *New Phytol* 206(4):1297–1313.
39. Hoffmann L, et al. (2004) Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell* 16(6):1446–1465.
40. Fatichi S, Leuzinger S, Körner C (2014) Moving beyond photosynthesis: From carbon source to sink-driven vegetation modeling. *New Phytol* 201(4):1086–1095.
41. Peña MJ, et al. (2007) *Arabidopsis irregular xylem8* and *irregular xylem9*: Implications for the complexity of glucuronoxylan biosynthesis. *Plant Cell* 19(2):549–563.
42. Biswal AK, et al. (2015) Downregulation of *GAUT12* in *Populus deltoides* by RNA silencing results in reduced recalcitrance, increased growth and reduced xylan and pectin in a woody biofuel feedstock. *Biotechnol Biofuels* 8:41.
43. Spokevicius AV, et al. (2007) β-tubulin affects cellulose microfibril orientation in plant secondary fibre cell walls. *Plant J* 51(4):717–726.
44. Qiu D, et al. (2008) Gene expression in *Eucalyptus* branch wood with marked variation in cellulose microfibril orientation and lacking G-layers. *New Phytol* 179(1):94–103.
45. Ko J-H, Yang SH, Park AH, Lerouxel O, Han K-H (2007) ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in *Arabidopsis thaliana*. *Plant J* 50(6):1035–1048.
46. Hussey SG, et al. (2011) *SND2*, a NAC transcription factor gene, regulates genes involved in secondary cell wall development in *Arabidopsis* fibres and increases fibre cell area in *Eucalyptus*. *BMC Plant Biol* 11:173.
47. Mizrachi E, et al. (2015) Investigating the molecular underpinnings underlying morphology and changes in carbon partitioning during tension wood formation in *Eucalyptus*. *New Phytol* 206(4):1351–1363.
48. Evans R, Ilic J (2001) Rapid prediction of wood stiffness from microfibril angle and density. *Forest Prod J* 51(3):53–57.
49. Mansfield SD, et al. (2009) Revisiting the transition between juvenile and mature wood: A comparison of fibre length, microfibril angle and relative wood density in lodgepole pine. *Holzforschung* 63(4):449–456.
50. Shen B, Li C, Tarczynski MC (2002) High free-methionine and decreased lignin content result from a mutation in the *Arabidopsis S*-adenosyl-L-methionine synthetase 3 gene. *Plant J* 29(3):371–380.
51. Li X, Weng J-K, Chapple C (2008) Improvement of biomass through lignin modification. *Plant J* 54(4):569–581.
52. Mao D, et al. (2015) FERONIA receptor kinase interacts with S-adenosylmethionine synthetase and suppresses S-adenosylmethionine production and ethylene biosynthesis in Arabidopsis. *Plant Cell Environ* 38(12):2566–2574.
53. Rennie EA, Scheller HV (2014) Xylan biosynthesis. *Curr Opin Biotechnol* 26:100–107.
54. Busse-Wicher M, Grantham NJ, Lyczakowski JJ, Nikolovski N, Dupree P (2016) Xylan decoration patterns and the plant secondary cell wall molecular architecture. *Biochem Soc Trans* 44(1):74–78.
55. Porth I, et al. (2013) Genome-wide association mapping for wood characteristics in *Populus* identifies an array of candidate single nucleotide polymorphisms. *New Phytol* 200(3):710–726.
56. Ranik M, Creux NM, Myburg AA (2006) Within-tree transcriptome profiling in wood-forming tissues of a fast-growing *Eucalyptus* tree. *Tree Physiol* 26(3):365–375.
57. Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Report* 11(2):113–116.
58. Kim D, et al. (2013) TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.
59. Trapnell C, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7(3):562–578.
60. Fouss F, Francoisse K, Yen L, Pirotte A, Saerens M (2012) An experimental investigation of kernels on graphs for collaborative recommendation and semisupervised classification. *Neural Netw* 31:53–72.
61. Basten CJ, Weir BS, Zeng Z-B (2004) *QTL Cartographer, version 1.17* (Department of Statistics, North Carolina State University, Raleigh, NC), p 188.
62. Doerge RW, Churchill GA (1996) Permutation tests for multiple loci affecting a quantitative character. *Genetics* 142(1):285–294.
63. Okuda S, et al. (2008) KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Res* 36(Web Server issue):W423-6.