Industrial Crops & Products 107 (2017) 159–171

Contents lists available at ScienceDirect

Industrial Crops & Products

journal homepage: www.elsevier.com/locate/indcrop

Research paper

Near-infrared spectroscopy enables the genetic analysis of chemical properties in a large set of wood samples from *Populus nigra* (L.) natural populations

Mesfin Nigussie Gebreselassie[a], Kévin Ader[a,b], Nathalie Boizot[a,b], Frédéric Millier[a,b], Jean-Paul Charpentier[a,b], Ana Alves[c], Rita Simões[c], José Carlos Rodrigues[c], Guillaume Bodineau[d], Francesco Fabbrini[e,f], Maurizio Sabatti[e], Catherine Bastien[a], Vincent Segura[a,*]

[a] *INRA, UR588 Amélioration, Génétique et Physiologie Forestières, Orléans, France*
[b] *INRA, Plateforme régionale Génobois, Orléans, France*
[c] *Centro de Estudos Florestais, Instituto Superior de Agronomia, 1349-017 Lisboa, Portugal*
[d] *INRA, UE995 Génétique et Biomasse Forestières, Orléans, France*
[e] *Department for Innovation in Biological, Agro-food and Forest systems, University of Tuscia, 01100 Viterbo, Italy*
[f] *Alasia Franco Vivai s.s., Strada Solerette 5/A, 12038 Savigliano, Italy*

ARTICLE INFO

Keywords:
Populus nigra L.
Near-infrared spectroscopy
Cell wall composition
Genetic variation
Clonal repeatability
Genotype × Environment interaction

ABSTRACT

High-throughput techniques for the compositional analysis of lignocellulosic biomass are essential to allow the genetic analysis and genetic improvement of bioenergy feedstocks. In this study, we investigated the feasibility of using near-infrared (NIR) spectroscopy for rapid assessment of wood chemical traits in a large sample of *Populus nigra* L. individuals evaluated in clonal trials at two contrasting sites. Spectra were acquired from 5799 wood samples collected in 3 harvests corresponding to two coppice rotations at one site and one coppice rotation at the second. Calibrations were developed and validated using 120 reference samples, representing spectral and chemical variations in the samples. The resulting global and site specific calibrations for most of the traits were at least good enough for ranking of genotypes, demonstrating the usefulness of NIR analysis for phenotyping the studied population. Clonal repeatability (H_c^2) estimates of the studied traits based on all samples were moderate to high (H_c^2 ranging from 0.57 to 0.89 in the 3 harvests). When data were pooled over coppice rotations or sites, the genotype × environment interaction was more evident across sites than across rotations. However, the interaction was smaller than the genotype main effect for all traits, except for glucose and extractives contents. Importantly, the interaction resulted mainly from re-ranking of a few genotypes leaving a substantial amount of stable and performant genetic material, which may encourage breeding for improved main wood components. Optimization of the NIR analysis for assessing clonal trials would facilitate the exploitation of standing genetic variation of energy or chemical related traits in tree breeding program.

1. Introduction

There is currently a considerable interest in moving to alternative and sustainable sources of energy because of the increasing global energy demand, depletion of fossil fuel reserves, fossil fuel-derived climate change and energy related geopolitical tensions. To circumvent some of the prevailing challenges, special focus has recently been given to the production of biofuels from lignocellulosic biomass. Lignocellulosic ethanol is expected to provide a large share of global transportation fuel needs with much less adverse effects than fossil fuels (Schubert, 2006; Sticklen, 2008). However, realizing this potential will require the synchronized occurrence of genetically improved material, suitable biomass production systems and bioconversion technologies that efficiently convert biomass into bioethanol (Ragauskas et al., 2006; Rubin, 2008).

Candidate biomass feedstocks for the production of second generation bioethanol comprise perennial grasses (e.g., switchgrass and *Miscanthus*) and forest trees (e.g., poplars, *Eucalyptus* and willow) (Abramson et al., 2010). Comparative advantages of poplars (*Populus* spp. and hybrids) in the impending green economy include their rapid growth rates (Bradshaw et al., 2000), good coppicing ability (Ceulemans and Deraedt, 1999) and favourable cell wall chemistry

* Corresponding author.
E-mail address: vincent.segura@inra.fr (V. Segura).

http://dx.doi.org/10.1016/j.indcrop.2017.05.013
Received 6 January 2017; Received in revised form 5 May 2017; Accepted 7 May 2017

(Guerra et al., 2013; Porth et al., 2013; Wegrzyn et al., 2010). In particular, *Populus nigra* possesses many important characteristics such as adaptability, rooting ability of stem cuttings and resistance to diseases that make it attractive as parent in several hybrid breeding programs in Europe (Cagelli and Lefevre, 1995; Frison et al., 1994).

The major source of lignocellulosic biomass is the plant cell wall, a heterogeneous complex mainly composed of cellulose, hemicelluloses and lignin, with the cellulose microfibrils and the hemicellulosic chains being embedded in lignin (Rubin, 2008). For bioethanol production, the polysaccharides (cellulose and hemicelluloses) are of particular interest because their enzymatic hydrolyses release fermentable monomeric sugars during saccharification. Poplars show substantial variability in cell wall composition, with cellulose content ranging from 42 to 49%, hemicellulose from 16 to 23%, and lignin from 21 to 29% (Sannigrahi et al., 2010). More recently, substantial genetic variation in cell wall chemical traits has been reported for black cottonwood (*P. trichocarpa*) (Guerra et al., 2016; Porth et al., 2013; Wegrzyn et al., 2010) and black poplar (*P. nigra*) (Guerra et al., 2013).

A critical bottleneck in efficient and cost-effective biomass saccharification for bioethanol production is the natural recalcitrance of plant cell walls to enzymatic hydrolysis (Rubin et al., 2007). The most obvious way to reduce biomass recalcitrance is through genetic improvement of trees for wood chemical composition. Poplar breeding for bioenergy can take advantage of past improvements in growth and disease resistance. However, current poplar clonal varieties have not been selected and bred for the qualitative characteristics of the biomass. Thus, there is a need to explore the potential for improvement of cell wall composition to release fermentable sugars and subsequently integrate biorefinery related selection criteria into poplar tree breeding programs. More specifically, development of dedicated bioenergy poplar for future biorefineries requires an understanding of the genetic architecture (extent of genetic variation and covariation, degree of genetic control, underlying polymorphisms/alleles) of both biomass production and biomass composition. This, in turn, accelerates the selection or development of new clones that produce high biomass yields, which are more amenable to bioconversion.

A recent approach to dissect the genetic architecture of “hard-to-measure” complex traits, such as lignocellulosic biomass quality, is to combine high-throughput phenotyping and genomics (Yang et al., 2014). The discovery and analysis of genetic information have been facilitated by the advances in high-throughput sequencing and genotyping platforms together with the availability of reference genome sequences for model forest tree species (Neale and Kremer, 2011). However, high-throughput phenotyping is lagging behind genomics (Araus and Cairns, 2014). Standard methods, such as wet chemistry, used for assessing the chemical composition of wood are costly and low-throughput, which limit their use for assaying of large number of samples as required in genetic studies and breeding programs. As a consequence, the genetic analysis and genetic improvement of cell wall composition may be hindered.

Near-infrared (NIR) spectroscopy is a high-throughput technology that can be applied towards the rapid characterization of a large number of lignocellulosic biomass samples with minimal cost. It is an indirect method based on multivariate statistical analysis to establish relationship between NIR absorbance spectra and reference values of properties of interest using a representative sample set. NIR spectroscopy has been successfully used to predict wood chemical traits in many forest tree species (Tsuchikawa and Kobori, 2015), including *Populus* (Robinson and Mansfield, 2009; Zhang et al., 2014; Zhou et al., 2011), *Eucalyptus* (Alves et al., 2012, 2011; Baillères et al., 2002; Poke and Raymond, 2006; Raymond and Schimleck, 2002), and *Pinus* (Alves et al., 2006; Jiang et al., 2014; Schwanninger et al., 2011a,b; Schwanninger and Hinterstoisser, 2011). Indeed, some studies have utilized NIR predictions for estimating genetic parameters of wood properties, mainly in *Pinus* (Da Silva Perez et al., 2007; Gaspar et al., 2011; Isik et al., 2011) and *Eucalyptus* (Costa e Silva et al., 2008; Hamilton et al., 2009; Kube et al., 2001; Poke et al., 2006; Raymond et al., 2001; Raymond and Schimleck, 2002; Schimleck et al., 2004; Stackpole et al., 2011, 2010).

To our knowledge, the evaluation of calibration models covering standing genetic variation available in natural and breeding populations of poplar is limited. NIR calibration is useful in genetic studies and selection/breeding activities because such applications require assessment of phenotypes in a large number of samples collected in multi-site environments. In this context, development of calibrations mainly depends on the range of variation of the traits of interest within and across environments. For poplars, this range may be defined not only by the genetic composition of the study population but also by the environmental conditions of the plantation site, short rotation coppice (SRC) management and the age of the tree at sampling time. The purpose of this study was to develop NIR calibration models to predict wood chemical properties, with the aim of applying the predictions to evaluate their genetic variability in natural populations of European black poplar covering the range of the species in Western Europe. Also, the resulting calibrations could be used for rapid screening of elite *P. nigra* clones from natural populations to be used in breeding programs. More specifically, this paper addresses the following objectives: (1) to develop and evaluate calibrations for predicting phenotypes of wood chemical traits in a large sample size (n = 5799) based on NIR spectra, (2) to estimate genetic variation in wood chemical properties of young trees and the degree of their genetic control, and (3) to quantify the magnitude and investigate the nature of genotype × environment (G × E) interaction of the same traits measured across coppice rotations as well as across sites.

2. Materials and methods

2.1. Wood samples and sample preparation

Clonally replicated trials of a *P. nigra* association population were established in 2008 at two contrasting sites located in central France (Orléans, ORL) and northern Italy (Savigliano, SAV) under a SRC system. At each site, a randomized complete block design (RCBD) was used, with a single tree per block and six replicates per genotype. The *P. nigra* population assayed in this study represent the natural range of the species in Western Europe, as it was composed of a diverse set of 1160 cloned genotypes (hereafter, each cloned genotype referred to as genotype) sampled in 14 natural metapopulations across 11 river catchments of four European countries (Table 1). More details concerning the experimental design, site characteristics (soil, climate) and plantation management practices can be found in Guet et al. (2015).

For the analysis of wood chemical properties, a total of 5799 wood samples were taken at 1 m above the ground from 2-yr-old trees in three different harvests (rotations/sites): (i) 289 genotypes in 3 blocks resulting in 795 samples harvested in ORL in March 2010 (end of first coppice cycle, 2008–2009) (hereafter referred to as ORL2010); (ii) 1066 genotypes in 3 blocks resulting in 2805 samples harvested in ORL in February 2012 (end of second coppice cycle, 2010–2011) (hereafter referred to as ORL2012); and (iii) 777 genotypes in 3 blocks resulting in 2199 samples harvested in SAV in January 2011(end of second coppice cycle, 2009–2010) (hereafter referred to as SAV2011). Circumference at 1 m was measured on all trees of the two sites just before harvest. For each harvest, the final number of biological replicates per genotype ranged between 2 and 3 because of mortality. The samples collected in ORL in 2010 and 2012 have been harvested during two successive 2-yr rotations of the same stool. The wood samples were oven dried at 30 °C for several days until a constant weight was reached, shredded into small pieces with a big cutter and milled using RETSCH SM 2000 cutting mills (SM2000, Retsch, Haan, Germany) to pass through a 1 mm metal sieve in order to get biomass powders onto which NIR spectra were collected. The wood samples were not debarked and both NIR measurements and biochemical analysis were made on non-debarked wood samples.

Table 1
Location, river management and number of studied genotypes in ORL and SAV for the 14 *P. nigra* metapopulations. Where metapopulations were represented by individual trees sampled in different stands distributed along one river, a range of latitudes, longitudes and altitudes is given. Metapopulations were ordered by country according to the latitude of origin. Altitude is expressed in metres a.s.l.

Country	River catchment	Metapopulation	Latitude	Longitude	Altitude	Cohorts[a]	River management[b]	Number of studied genotypes		
								ORL	SAV	Common
France	Adour	Adour	42°53′N–43°23′N	0°02′W–00°56′W	52–902	Mature	Partially regulated	62	52	49
Italy	Basento	Basento	40°24′N – 40°38′N	15°56′E – 16°39′E	37–286	Juvenile/ mature	Partially regulated	26	15	14
France	Dranse	Dranse	46°23′N	06°30′E	374	Juvenile/ mature	Dynamic	40	42	39
France	Durance	Durance	43°51′N	04°59′E	60	Juvenile/ mature	Partially regulated	14	8	1
Germany	Kuhkopf	Kuhkopf	49°49′N	08°30′E	91	Juvenile/ mature	Regulated	53	46	37
France	Loire	Loire	47°00′N – 47°51′N	00°44′W – 02°58′E	29–154	Juvenile/ mature	Dynamic	215	197	165
Netherlands	NL	NL	50°31′N – 52°37′N	03°35′E – 06°23′E	0 – 287	Mature	Regulated	47	42	37
France	Nohèdes	Nohede	42°37′N	02°17′E	820	Mature	Dynamic	43	38	35
Italy	Paglia	Paglia	42°45′N–42°52′N	11°45′E–11° 55′E	235–358	Juvenile/ mature	Dynamic	47	42	41
France	Drôme	Ramieres	44°41′N–44°45′N	04°55′E–05°24′E	145	Juvenile/ mature	Dynamic	178	99	91
France	Rhin	Rhin	48°16′N–48°37′N	07°41′E–07°49′E	135–160	Mature	Regulated	66	50	48
Italy	Stura	Stura	44°17′N – 44°23′N	06°56′E – 07°12′E	825–1699	Juvenile/ mature	Dynamic	25	29	25
Italy	Ticino	Ticino	45°12′N–45°16′N	08°59′E–09°04′E	60–70	Juvenile/ mature	Dynamic	103	78	62
France	Allier	ValAllier	46°24′N	03°19′E	220	Juvenile/ mature	Dynamic	147	39	39

[a] Juvenile trees were defined as non-reproductive trees.
[b] Regulated if water flows have been regulated to facilitate navigation or to prevent floods; dynamic if water flows are not regulated and allow some flooding events.

2.2. NIR spectra collection, pretreatment and selection of reference samples

Once established, NIR calibration models can be an inexpensive and high-throughput method for accurate estimation of wood chemical properties. However, their initial development involves several steps, including spectral data collection, spectral data pretreatment, selection and analysis of reference samples, application of a multivariate calibration method, model selection and model validation. The NIR spectra of 5799 wood powder samples were measured with a spectrometer Spectrum 400 (Perkin Elmer, Waltham, MA, USA) over 45 days between the end of April 2015 and the beginning of July 2015. Prior to analysis, samples were stabilized in a climatized chamber (20%RH) at 24 °C for a minimum of 1 day. Samples in quartz cups were placed in a rotating device above the integration sphere window and spectra acquired in a temperature controlled room (24 °C). All measurements were done in diffuse-reflectance mode and the obtained spectra were computed as Log (1/R) and expressed in absorbance. The scanning range for all samples was from 10,000 cm^{-1} to 4000 cm^{-1} (1000–2500 nm) with a spectral resolution of 8 cm^{-1} and a zero filling factor of 4 resulting in a number of data points at every 2 cm^{-1}. For each wood sample, 64 scans were acquired and averaged. Background was carried out regularly using Spectralon® as reference.

Undesirable sources that likely affect the quality of spectral data include sample moisture content, particle size, temperature and humidity of the spectrometer laboratory, batch effects (e.g., date of spectral data collection) and so on. Before applying multivariate analysis methods such as partial least squares (PLS) regression, it is important to reduce or remove undesired variations in the recorded sample spectra to reduce noise and enhance calibrations. For this reason, several common spectral pretreatment techniques (normalization, detrend, first and second derivatives on raw or normalized spectra) were applied to the raw spectra for comparisons or to find the best combination. Absorption spectra were first restricted to the wavenumber range of 8000–4000 cm^{-1} since spectra recorded within 10,000–8000 cm^{-1} has mainly noise. For illustration, plot of raw spectra of wood powder samples from ORL2012 harvest is shown in Fig. S1. The spectra pretreatment was performed with R software (R Core Team, 2015). The R packages *prospectr* (Stevens and Ramirez-Lopez, 2013) and *signal* (signal developers, 2013) were used to perform detrend and derivations, respectively. Since spectral data pretreatment can improve exploratory analysis, principal component analysis (PCA) was performed on the resulting 7 spectra modalities (raw, pretreated) to explore the data for potential outlying spectra and clustering of the samples according to genotypes, date of spectral data collection, temperature and humidity of the spectrometer laboratory and operators (not shown). In this initial exploratory analysis, no samples were removed as outliers.

Careful selection of representative samples (reference samples) is a prerequisite to develop NIR spectra based calibrations. We chose to select 120 reference samples based on spectral data because the NIR spectra basically contains information about several properties of a wood sample for which the calibration is carried out. These samples should therefore be selected in order to represent most of the spectral variation of a large population of wood samples (n = 5799) collected from a multi-environment experiment. Also, the reference samples should best represent the sources of variation likely to occur in future samples such as plantation site, coppice rotation and genotype, which could enhance the robustness of the resulting calibrations. To do so, we first calculated the mean spectrum for each genotype within harvest. PCA was then performed on the resulting genotypic spectra across all harvests. The results obtained provided two types of information. First, compared to other spectra modalities, first derivative spectra (first derivative on raw spectra) showed a more uniform distribution of the genotypic spectra on the first 2 PCs (Fig. S2) and were thus chosen to be used for the selection of reference samples. Second, the genotypic spectra showed clear clusters in the space of the first 2 PCs according to harvests (Fig. S2). We thus decided to select an equal number of genotypes from each harvest to constitute the reference sample set. Euclidean distances were computed between the genotypic spectra within harvests. Subsequently, a representative subset of genotypes was

selected within each harvest following the Kennard-Stone algorithm which allows to select samples with a uniform distribution over the predictor space (Kennard and Stone, 1969). A total of 45 genotypes (i.e., 14–16 genotypes per harvest) were selected in order to reach a total of 120 samples when considering the 2–3 biological replicates of each genotype in each harvest.

2.3. Wood chemical analysis of reference samples

This section is described in detail in Supplementary Information Text (*SI Text*). The 120 selected samples were analyzed for chemical composition following standard analytical methods (wet chemical analysis, HPLC, analytical pyrolysis) to generate reference values used to develop dedicated calibrations to predict wood chemical traits in all the samples (n = 5799). Wood chemical traits included: (i) extractives content; (ii) lignin content (Klason lignin, acid-soluble lignin); and (iii) the content of the two most abundant cell wall sugars (glucose, xylose). Analytical pyrolysis was used to assess lignin composition [relative proportion of p-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) units] according to Rodrigues et al. (2001, 1999) and Alves et al. (2006). Except for analytical pyrolysis, at least two technical replicates were performed per sample. For analytical pyrolysis, technical replicates were done only for a few samples to estimate the root mean square error (RMSE) of the method for further comparison with the RMSE of the corresponding NIR calibration.

2.4. Development of NIR calibration models using partial least squares (PLS) regression

R software was used for PLS regression model development (R Core Team, 2015). To perform the calibrations, we used the R package *pls* (Mevik and Wehrens, 2007). Various home-made functions were also used to carry out the calibrations with PLS regression in a cross-validation scheme with an optional detection of potential outlier observations. Moreover, the function "carspls_LOO" was used for automatically selecting a subset of wavenumbers to be included in the PLS regression as proposed by Li et al. (2009). The selection is based on an iterative exclusion of wavenumbers according to their weight in a PLS regression and following an exponential decreasing function. Consequently, the selected wavenumbers are specific to the trait being calibrated. More details about this method are given in Li et al. (2009).

Prior to a final calibration step, we detected potential outlying observations within the 120 reference samples using either box-and-whisker plots or *P-value* thresholds of z-tests on the cross-validation residuals of PLS calibrations. The final calibration step involved splitting of the 120 reference samples into a calibration set (n = 99, ~5/6) and a validation set (n = 21, ~1/6) using Kennard-Stone algorithm (Kennard and Stone, 1969) per harvest on first derivative spectra. Next, outliers detected in the previous step were removed from both calibration and validation data sets. The resulting calibration set was then used to build the model with a leave-one-out (LOO) cross-validation with or without automatic wavenumber selection using the CARS algorithm (Li et al., 2009). The optimal number of components in the PLS regression model was optimized within the cross-validation using Wold's criterion (Li et al., 2002), which was set up at 1. The following statistics were calculated for each model both within the training (cross-validation) and validation sets:

- The coefficient of determination defined as $R^2 = 1 - \left({}^{RSS}/_{TSS}\right)$, where RSS is the residual sum of squares (sum of squares of differences between observed and predicted values), and TSS is the total sum of squares (sum of squares of differences between observations and their mean);
- The root mean square error defined as $RMSE = \sqrt{\left({}^{RSS}/_{n}\right)}$, where RSS is defined as above and n is the number of observations;
- The ratio of prediction to standard deviation defined as $RPD = {}^{SD}/_{RMSE}$, where SD is the standard deviation of the observations, and RMSE is defined as above.

The models with best statistics were selected and, when validated, used to predict all samples (n = 5799) included in the study.

2.5. Estimation of genetic parameters for NIR-predicted wood chemical traits

The NIR-predicted wood chemical traits were all approximately normally distributed and data transformations were not considered necessary prior to genetic analysis. In order to estimate variance components of traits, linear mixed models (Henderson, 1984) involving spatial effects were fitted using breedR package (Muñoz and Sanchez, 2015) in software R for the analysis of all predicted traits within harvests. Both block and spatial effects account for the environmental variation within the experimental field. Block effects account for global field variations, while spatial effects capture the environmental heterogeneity not accounted by the block effects because of the relatively large size of each block. Furthermore, spectra data have been collected according to the ordered field positions of the trees. So spectra collection date is likely to contribute to the so called spatial variation revealed by the variograms. Accounting for the date effect could help to interpret the spatial effects, if necessary.

For each of the traits, the following mixed model was fitted:

$$y = X\beta + Zu + Rb + Nd + e \quad (1)$$

where y is a vector of individual tree data for a predicted wood chemical trait, β is a vector of fixed effects (over all mean or intercept), u is a vector of random effects of genotypes (genetic effects of genotypes or genotypic values), b is a vector of random effects of blocks, d is a vector of random effects of the dates of NIR spectra collection and e is a vector of residuals. X, Z, R, and N are known incidence matrices relating the observations to the fixed effects in vector β and random effects in vectors u, b, and d, respectively, assuming $u \sim N(0, \sigma_G^2 I)$, $b \sim N(0, \sigma_b^2 I)$, $d \sim N(0, \sigma_d^2 I)$, $e \sim N(0, R)$, where $\sigma_G^2$is the genotypic variance, σ_b^2 is the block variance, σ_d^2 is the date variance, R is the residual covariance matrix, and I is an identity matrix. A spatial residual structure was implemented in order to decompose e into spatially dependent (ξ) and spatially independent (η) residuals (Dutkowski et al., 2002), leading to the following decomposition of R:

$$R = \sigma_\xi^2 [AR1(\rho_{col}) \otimes AR1(\rho_{row})] + \sigma_\eta^2 I \quad (2)$$

where σ_ξ^2 is the spatially dependent residual variance, σ_η^2 is the spatially independent residual variance, I is an identity matrix and $AR1(\rho)$ is a first-order autoregressive correlation matrix.

The mixed model described in model 1 was compared with a model without decomposition of the residual term into spatially dependent and independent effects based on the Akaike information criterion (AIC) and was found to have a lower AIC (i.e., better performance) in all data sets (i.e., ORL2012 and SAV2011 harvests) for all predicted phenotypes. However, spatial trends were not modelled for ORL2010 harvest because the number of genotypes per harvested block was not large enough to capture the within block spatial variation. Moreover, the level of sampling within each block induced heterogeneity in the spatial distribution of the corresponding samples, so estimation of spatial effects over the trial could be biased.

Within each harvest and for each phenotype, reduced models (dropping block or spectra collection date effect) were also fitted and compared to the corresponding full models based on the AIC. Finally, the model yielding the best fit (lowest AIC) was selected for variance component estimation and to adjust the phenotype for non-genetic random effects (block, date, spatially dependent residuals).

Variance components from the selected mixed model were further used to estimate broad-sense heritabilities of the NIR-predicted wood

chemical traits within harvests. Individual tree broad-sense heritability (H_i^2) was calculated using the following equation:

$$H_i^2 = \sigma_G^2 \Big/ (\sigma_G^2 + \sigma_e^2) \quad (3)$$

where σ_G^2 and σ_e^2 are the genotypic and residual variance components, respectively. Clonal mean broad-sense heritability or clonal repeatability (H_c^2) was calculated as:

$$H_c^2 = \sigma_G^2 \Big/ \left(\sigma_G^2 + {}^{\sigma_e^2}\!/_r\right) \quad (4)$$

where r is the average number of replicates per genotype for a trait under consideration for a given harvest. Standard errors for heritability were calculated using the Delta method (Lynch and Walsh, 1998). Standard errors were multiplied by 1.96 to construct the 95% confidence interval (CI) for heritability.

Finally, we used genotypes shared between coppice rotations or sites for fair comparisons of genetic parameter estimates and for assessing stability of genetic parameters between rotations as well as between sites or characterizing G × E interaction. A set of 289 genotypes were shared between rotations within Orleans' trial (ORL2010 vs ORL2012 harvests), while 683 genotypes were shared between sites (ORL2012 vs SAV2011 harvests).

For analysis within harvests based on shared genotypes, the following model was fit:

$$y = X\beta + Zu + e \quad (5)$$

Where y is a vector of individual tree data for a predicted wood chemical trait that was adjusted for block, date and spatially dependent effects with the previously selected mixed model (model 1) and the remaining parameters were assigned as described in model 1.

In order to test and evaluate the extent of G × E interaction across rotations and sites, the following G × E mixed model was fitted:

$$y = X\beta + Zu + Mp + e \quad (6)$$

where y is a vector of individual tree data for a predicted wood chemical trait that was adjusted for block, date and spatially dependent effects with model 1, β is a vector of fixed effects (over all mean and rotations or sites), p is a vector of random effects of G × E interaction. X, Z and M are incidence matrices relating the observations to the fixed effects in vector β and random effects in vectors u and p, respectively, assuming, $p \sim N(0, \sigma_{G\times E}^2 I)$, where $\sigma_{G\times E}^2$ is the G × E interaction variance. The remaining parameters were assigned as described in model 1. Likelihood ratio tests (LRT) between the full model and a reduced model without G × E interaction effect were performed to test the significance of G × E interaction effect. Correlations between adjusted genotype means were also estimated using the Spearman's rank correlation to further characterize the stability of genotype means between rotations or sites. Finally, when the extent of G × E variance was found to be more than 50% of the genotypic variance (Shelbourne, 1972), a further decomposition of the G × E interaction was carried out following method 1 of Muir et al. (1992). Such decomposition enables to partition the G × E sum of squares into scaling effect and genotype rank change.

3. Results

3.1. Variability in wood chemical properties of reference samples

The descriptive statistics for traits analyzed in the laboratory of the 120 reference samples are presented in Table 2. The range of variation in most of the traits analyzed was considerable, and provided the potential to develop reliable calibrations. For example, Klason lignin content ranged from 16.8% to 26.5%, whereas glucose content ranged from 30.6% to 50.3%. Overall, the range of wood chemical traits within our reference data set was 5–21 times the RMSE of the standard methods, making this reference data set acceptable for building near-infrared multivariate calibration models (Table 2).

Table 2
Descriptive statistics for lignin monomers (H, G, S), lignin composition (H/G, S/G), lignin content (Klason lignin, Py-lignin, acid-soluble lignin), cell wall sugars (xylose, glucose, xylose/glucose, C5/C6) and extractives analyzed by standard laboratory methods for 120 reference wood samples.

Trait	Unit	RMSE	Min.	Max.	Mean
H-lignin	% Lignin	0.60	2.80	11.0	5.00
G-lignin	% Lignin	0.84	41.20	53.10	46.40
S-lignin	% Lignin	1.01	39.50	54.90	48.60
H/G	fold	0.01	0.05	0.25	0.11
S/G	fold	0.03	0.86	1.34	1.06
Klason lignin	% CWR	1.61	16.80	26.5	21.50
Py-lignin	% CWR	1.49	20.20	27.00	23.10
Acid-soluble lignin	% CWR	0.32	4.60	7.00	6.10
Xylose	% CWR	1.15	13.10	18.70	15.30
Glucose	% CWR	1.87	30.60	50.30	40.40
Xylose/Glucose	fold	0.03	0.29	0.48	0.38
C5/C6	%	1.14	17.9	29.30	23.90
Extractives	% DW	0.54	6.20	17.70	10.40

DW: dry weight; CWR: cell wall residue (extractives-free dry weight); RMSE: root mean square error of the standard methods for replicate analysis.

3.2. Calibration, validation and prediction

The absorption spectra modalities (with or without pretreatment) and reference values of the reference samples were used to develop NIR calibration models at a global scale for a majority of the traits, except for lignin contents, where site specific models showed higher predictive performance than the global ones (Table 3). The reference values in the calibration and validation data sets for the wood chemical traits of black poplar were comparable (i.e., had similar means and ranges), which means that reliable models can be developed and effectively verified (Fig. S3).

Summary statistics that demonstrate the performance of the models in calibration and validation data sets are reported in Table 3 and plots of the predicted versus measured component values for selected calibrations are shown in Fig. S4. Pretreated spectral data provided better calibrations than raw spectra. Automatic wavenumber selection improved model performance, for some of the traits, compared with full range.

Global calibration models developed for the prediction of H-lignin, lignin H/G and S/G ratios, xylose/glucose, C5/C6 and extractives were good, with coefficients of determination (R^2) ranging from 0.75–0.91 and 0.72–0.83 in calibration and validation data sets, respectively (Table 3, Fig. S4). The R^2 values for calibration and validation sets of glucose were 0.76 and 0.64, respectively. The models for G-lignin and S-lignin had moderate performance in cross-validation (R^2_{cv} = 0.68 and 0.64, respectively), while the model for S-lignin showed a higher accuracy of prediction (R^2_{val} = 0.77). The model for xylose showed inadequate fit in the calibration data set (R^2_{cv} = 0.48) as well as a poor prediction performance in the validation data set (R^2_{val} = 0.29).

On the other hand, global models for lignin content (Klason lignin, Py-lignin and acid-soluble lignin) were very specific (i.e., they were good in predicting samples included in the models but very poor in predicting samples of an independent validation set) (Table 3). Therefore, we followed the site specific approach to model these characteristics (Table 3, Fig. S4). For Klason lignin, the model developed for Orleans site (R^2_{cv} = 0.78, R^2_{val} = 0.60) had a better fit, whereas good models for Py-lignin and acid-soluble lignin were obtained at Savigliano (R^2_{cv} = 0.79, R^2_{val} = 0.73 and R^2_{cv} = 0.77, R^2_{val} = 0.79, respectively).

With the exception of Klason lignin, Py-lignin, acid-soluble lignin and xylose global models, the remaining global models described in Table 3 and Fig. S4 were used to predict wood chemical properties for

Table 3
NIR calibration models (leave-one-out cross-validation) and validation statistics for wood chemical properties evaluated on 120 reference samples. For trait abbreviations see the caption of Table 2.

Trait	Model type	Calibration set (n = ~5/6)								Validation set (n = ~1/6)				
		nblambda	Pretreatment	nbcomp	R^2_{cv}	$RMSE_{cv}$	RPD_{cv}	nobs	nb. outliers	R^2_{val}	$RMSE_{val}$	RPD_{val}	nobs	nb. outliers
H-lignin	Global	Full range	der2	9	0.75	0.80	2.0	91	8	0.80	0.89	2.3	21	0
G-lignin	Global	29	der2	5	0.68	1.26	1.8	94	5	0.51	1.33	1.5	20	1
S-lignin	Global	Full range	norm-der2	13	0.64	1.25	1.7	90	9	0.77	1.02	2.2	20	1
H/G	Global	652	der2	8	0.82	0.02	2.4	92	7	0.83	0.02	2.5	19	2
S/G	Global	947	norm-der2	12	0.84	0.03	2.5	91	8	0.72	0.04	2.0	21	0
Klason lignin	Global	Full range	der1	5	0.61	1.17	1.6	91	8	0.27	1.44	1.2	20	1
	Site: ORL	Full range	dt	6	0.78	0.94	2.2	56	10	0.60	1.31	1.6	13	1
	Site: SAV	Full range	der2	5	0.25	1.43	1.2	33	0	−4.33	1.22	0.5	7	0
Py-lignin	Global	Full range	norm-der2	7	0.75	0.59	2.0	97	2	−0.18	0.78	1.0	21	0
	Site: ORL	Full range	norm-der2	7	0.72	0.65	1.9	65	1	−1.43	0.87	0.7	14	0
	Site: SAV	Full range	der1	8	0.79	0.39	2.2	26	7	0.73	0.35	2.1	6	1
Acid-soluble lignin	Global	Full range	der2	7	0.61	0.23	1.6	92	7	0.35	0.29	1.3	18	3
	Site: ORL	Full range	norm-der2	6	0.49	0.25	1.4	61	5	0.21	0.29	1.2	13	1
	Site: SAV	Full range	norm-der2	8	0.77	0.16	2.1	30	3	0.79	0.15	2.4	6	1
Xylose	Global	Full range	der1	6	0.48	0.73	1.4	88	11	0.29	0.85	1.2	21	0
Glucose	Global	46	norm-der1	6	0.76	1.49	2.0	94	5	0.64	1.25	1.7	18	3
Xylose/Glucose	Global	28	norm	8	0.79	0.02	2.2	90	9	0.75	0.02	2.0	20	1
C5/C6	Global	129	der2	6	0.85	0.89	2.6	91	8	0.81	1.08	2.4	21	0
Extractives	Global	326	der1	9	0.91	0.74	3.3	91	8	0.74	1.03	2.0	20	1

nblambda: number of selected wavenumbers; nbcomp: number of PLS components; R^2_{cv}: coefficient of determination of cross-validation; $RMSE_{cv}$: root mean square error of cross-validation; RPD_{cv}: ratio of performance to deviation of cross-validation; nobs: number of samples statistically analyzed; R^2_{val}: coefficient of determination of validation; $RMSE_{val}$: root mean square error of validation; RPD_{val}: ratio of performance to deviation of validation; norm: normalized spectra; dt: detrending spectra; der1: first derivative spectra; der2: second derivative spectra; norm-der1: first derivative on normalized spectra; norm-der2: second derivative on normalized spectra; Full range spectrum: 8000–4000 cm^{-1}.

the entire sample set (n = 5799) to study phenotypic variability, degree of genetic control and G × E interaction. Klason lignin at Orleans and Py-lignin and acid-soluble lignin at Savigliano were predicted with the site specific models because of the poor prediction performance of their corresponding global models. On the other hand, Klason lignin at Savigliano and Py-lignin and acid-soluble lignin at Orleans were not predicted since their corresponding site specific models had poor performances, especially in validation sets. Xylose was not predicted since the quality of the model was considered poor as mentioned before.

Boxplots of the distributions of NIR-predicted wood chemical traits (without adjustment for micro-environmental effects) are presented in Fig. S5. The range of phenotypic variation in most predicted wood traits was considerable. For example, the predicted Klason lignin content ranged from 16.1% to 27.9%, whereas predicted glucose content ranged between 30.2% and 49.7%. All the predicted values were in line with the results obtained for the reference data set (i.e., they were pretty much close to the limits or within the range of variation observed for the reference data set) (Table 2). Moreover, based on comparisons of the RMSE of the models (Table 3) to the RMSE of the standard methods (Table 2), the uncertainties associated with the predictions can be regarded as acceptable. It is worth mentioning, however, that the predicted S/G values (0.69–1.43) didn't fall within the range of values reported for other populations of *P. nigra* (1.3–2.1) (Guerra et al., 2013) and *P. trichocarpa* (1.5–2.4) (Guerra et al., 2016) despite variations of almost the same magnitude.

3.3. Variance components and broad-sense heritability of wood chemical traits within harvests

Due to some unbalance in genotype representation across harvests, the genetic analysis of individual harvests using only the shared genotypes was done to ensure fair comparisons of genetic parameters for the same traits. Thus, a set of 289 genotypes were shared between ORL2010 and ORL2012 harvests (i.e., between coppice rotations), while 683 genotypes were shared between ORL2012 and SAV2011 harvests (i.e., between sites).

Based on analysis using 289 genotypes, high clonal repeatability (H_c^2) values were found for lignin monomers (H, G, S) (0.74 ± 0.05 to 0.81 ± 0.04), lignin composition (H/G, S/G) (0.75 ± 0.05 to 0.81 ± 0.04), Klason lignin (0.75 ± 0.05 to 0.80 ± 0.04) and cell wall sugars (0.72 ± 0.06 to 0.80 ± 0.04) in the 2 rotations (Fig. 1, Table S1). The exception was extractives content, for which the H_c^2 values were moderate to high (0.57 ± 0.09 to 0.72 ± 0.06). Using 683 genotypes, high H_c^2 values were found for all traits except extractives, namely, lignin monomers (0.74 ± 0.04 to 0.88 ± 0.02), lignin composition (0.77 ± 0.03 to 0.89 ± 0.01) and cell wall sugars (0.70 ± 0.04 to 0.81 ± 0.02) in the two sites (Fig. 2, Table S2). For extractives, the H_c^2 values were moderately high (0.62 ± 0.05 to 0.70 ± 0.04). At Savigliano, where trees grew more rapidly, the genetic control over all the chemical traits was generally stronger with the exception of C5/C6. For example, H_c^2 for lignin S/G ratio was higher (0.89 ± 0.01) at Savigliano compared to H_c^2 = 0.77 ± 0.03 at Orleans. These differences in heritability of the same traits between the two sites can be explained by scale effects (i.e., both increased expression of genetic variation and decreased residual variation at Savigliano as compared to Orleans) (Table S2). In comparison to site, rotation effects on H_c^2 were rather low for all traits with the exception of extractives, suggesting differences in magnitude of G × E interaction between rotations and sites. For extractives content, H_c^2 varied more between rotations than between sites. Next, in order to reflect the genetic variation in wood traits that could be present in the entire population, we repeated the genetic analysis using all genotypes available in each harvest.

To estimate genetic parameters for the NIR-predicted wood chemical traits for each single harvest using all genotypes available, data were analyzed with a full mixed model accounting for spatial effect.

Overall, the clonal repeatability estimates were highly comparable between the two analyses, i.e., all genotypes versus shared genotypes (Figs. 1–2, Table S1 and S3). We conclude that the genotypes shared between the harvests were adequate enough to capture the genetic variation existing in the entire population. This is interesting because we would not miss important information when analysing the G × E interaction both across rotations and sites using the shared genotypes.

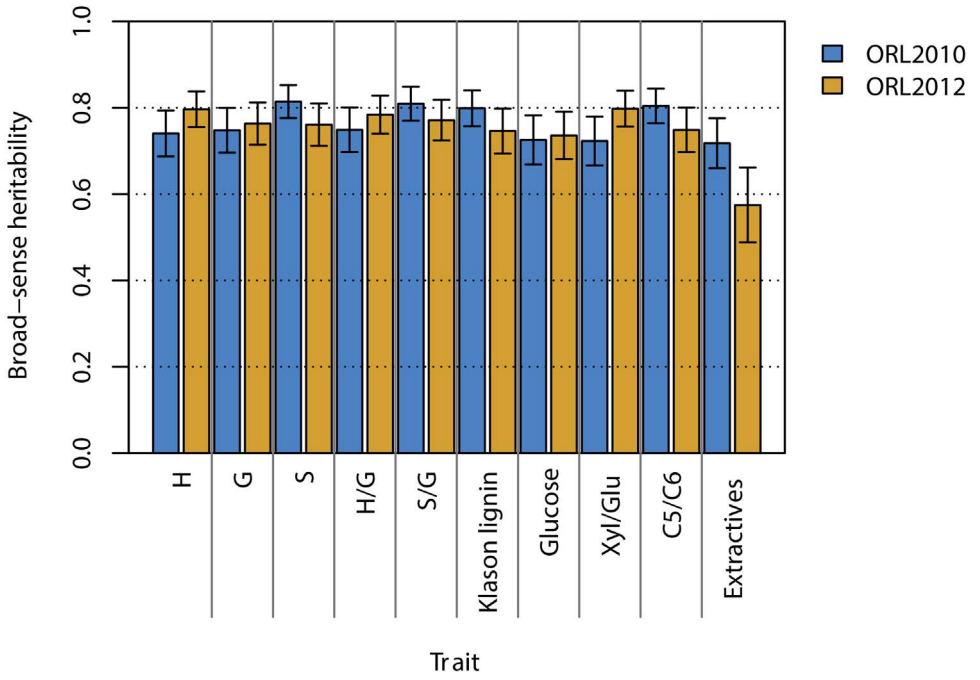

Fig. 1. Estimated clonal mean broad-sense heritability (clonal repeatability) (H_c^2) with error bars corresponding to the 95% confidence intervals for NIR-predicted wood chemical traits evaluated over two successive 2-yr rotations (ORL2010 and ORL2012) in a clonal trial at Orleans (France). For trait abbreviations see the caption of Table 2.

3.4. Genotype × environment (G × E) interaction effect on wood chemical traits

To assess the stability of genetic parameters or characterize G × E interaction for the NIR-predicted wood chemical traits, genotypes shared between rotations (ORL2010 vs ORL2012 harvests) or between sites (ORL2012 vs SAV2011 harvests) were used. Two strategies were adopted to assess G × E interaction including estimation of variance components and correlation between environments.

Combined mixed model analysis of variance of 289 genotypes evaluated across rotations at Orleans showed that the G × E interaction effect was significant (LRT *P-values* < 0.001, 0.01) for all traits, except for H-lignin and lignin H/G ratio (Table S4). However, the magnitude of the G × E interaction variance was rather low, compared to the genotypic variance component for all traits. The G × E interaction variance component explained only 1–18% of the total phenotypic variance, whereas the genotype main effect accounted for 30–53% (Fig. 3, Table S4). Based on the 683 genotypes tested across sites, for all traits highly significant (LRT *P-value* < 0.001) G × E interaction was found and the GxE variance reached more than 50% of the correspond-

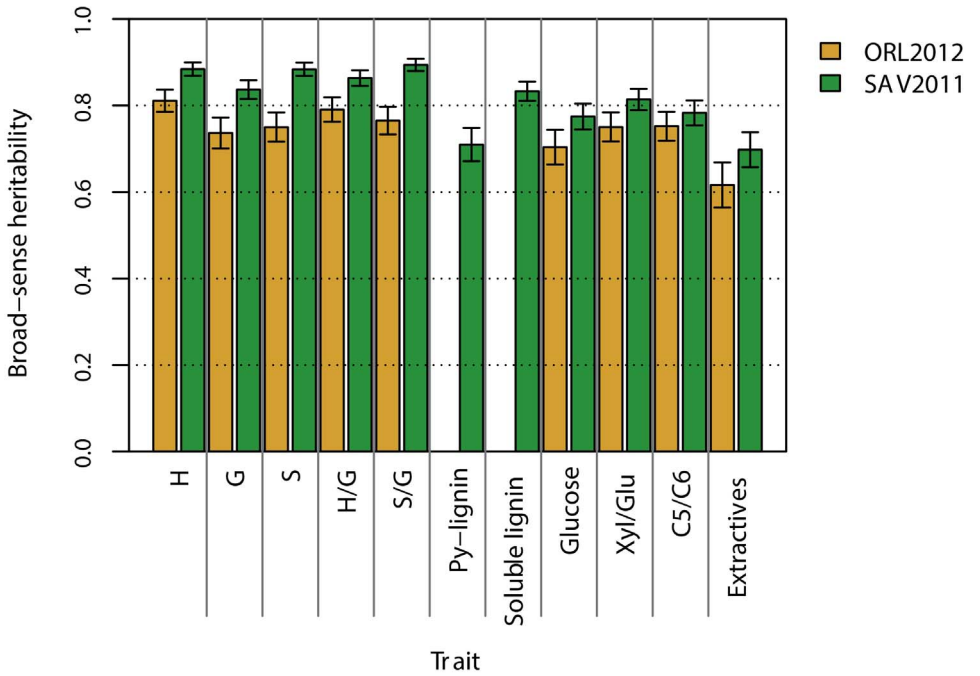

Fig. 2. Estimated clonal mean broad-sense heritability (clonal repeatability) (H_c^2) with error bars corresponding to the 95% confidence intervals for NIR-predicted wood chemical traits evaluated at two contrasting sites (Orleans, France: ORL2012; Savigliano, Italy: SAV2011). Trees were grown over two successive 2-yr rotations at Orleans (2008–2009, 2010–2011) and 1-yr and 2-yr rotations at Savigliano (2008, 2009–2010). Results are based on the data from the second rotations at the two sites. For trait abbreviations see the caption of Table 2.

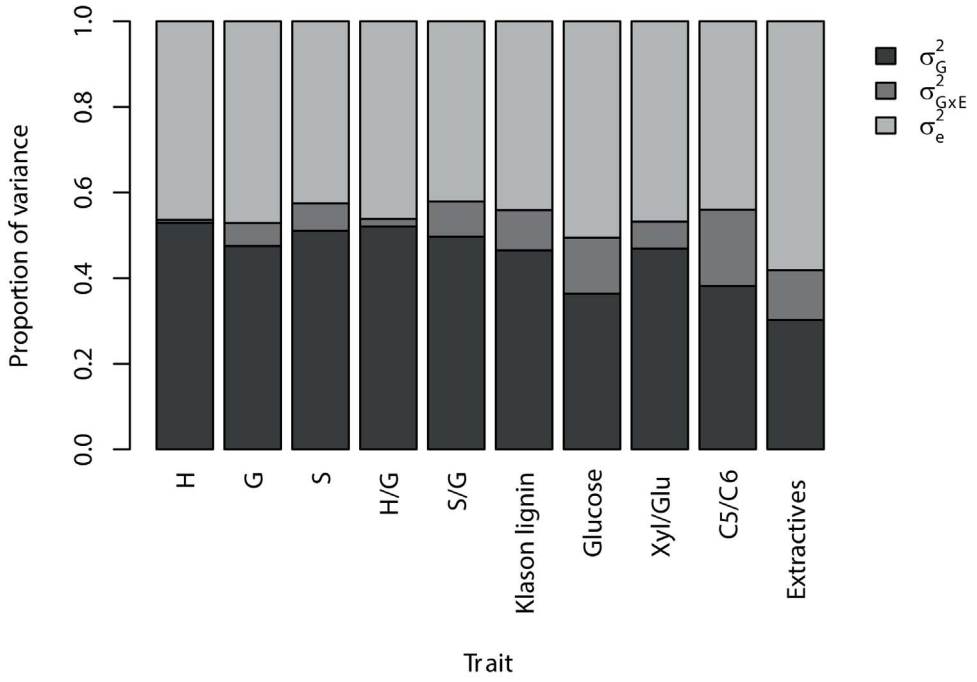

Fig. 3. Decomposition of total phenotypic variance for NIR-predicted wood chemical traits evaluated over two successive 2-yr rotations (ORL2010 and ORL2012) in a clonal trial at Orleans (France). Stacked barplot of the percentage of the total phenotypic variance explained by the genotype main effect (σ^2_G), genotype × environment (G × E) interaction effect ($\sigma^2_{G\times E}$) and residual effect (σ^2_ϵ) variance components for 10 wood chemical traits and using 289 shared genotypes. For trait abbreviations see the caption of Table 2.

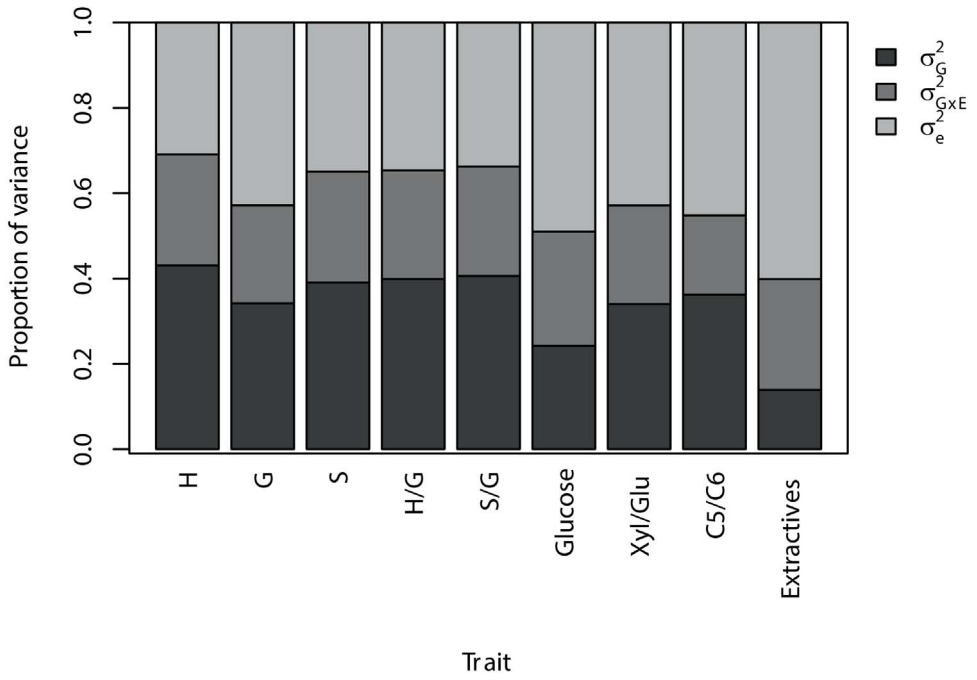

Fig. 4. Decomposition of total phenotypic variance for NIR-predicted wood chemical traits evaluated at two contrasting sites (ORL2012: Orleans, France; SAV2011: Savigliano, Italy). Stacked barplot of the percentage of the total phenotypic variance explained by the genotype main effect (σ^2_G), genotype × environment (G × E) interaction effect ($\sigma^2_{G\times E}$) and residual effect (σ^2_ϵ) variance components for 9 wood chemical traits and using 683 shared genotypes. For trait abbreviations see the caption of Table 2.

ing genetic variance (Fig. 4, Table S5). Importantly, for glucose and extractives, the G × E interaction variance component (27% and 26%, respectively) was even larger than the genetic variance component (24% and 14%, respectively) and this was somewhat mirrored in the relative values of the clonal repeatability, especially for extractives. Compared to extractives, glucose content is a key wood chemical trait in term of bioethanol production. During bioethanol production, glucose is released from cellulose in the plant cell walls via enzymatic hydrolysis of lignocellulosic biomass and could then be converted into bioethanol via fermentation. To assess if the observed G × E interaction for glucose in particular and all biochemical traits in general would have practical implications for poplar tree breeding for bioethanol production, it is noteworthy to further decompose the corresponding interaction variance components. Because the G × E interaction was in general more evident over sites than over rotations for all traits, we sought to zoom into the nature of the G × E interaction across sites.

Thus, G × E interaction was dissected according to the method 1 described by Muir et al. (1992). Results of the partitioning of the G × E

Table 4
Partitioning of genotype × environment (G × E) interaction sum of squares (SS G × E) for NIR-predicted wood chemical traits evaluated at two contrasting sites and using 683 shared genotypes according to Method 1 of Muir et al. (1992). Proportion of genotypes explaining 50% of the SS G × E was calculated according to their relative ecovalence (Lin et al., 1986). For trait abbreviations see the caption of Table 2.

Trait	$\sigma_{G\,ORL}$	$\sigma_{G\,SAV}$	$\sigma^2_G/\sigma^2_{G\times E}$	r_s	% SS G × E		Proportion of genotypes explaining 50% of G × E SS
					Scale effect	Re-ranking	
H-lignin	0.52	0.79	1.66	0.56	14.61	85.39	12.30
G-lignin	0.80	0.77	1.49	0.46	1.46	98.54	10.25
S-lignin	1.29	1.83	1.50	0.51	6.53	93.47	10.69
H/G	0.01	0.02	1.57	0.53	11.33	88.67	13.47
S/G	0.05	0.07	1.59	0.54	2.52	97.48	9.22
Glucose	1.24	1.26	0.91	0.34	0.23	99.77	10.25
Xylose/Glucose	0.01	0.01	1.47	0.45	0.02	99.98	11.71
C5/C6	0.76	0.77	1.95	0.49	0.01	99.99	11.13
Extractives	1.06	0.68	0.54	0.23	14.92	85.08	10.69

r_s: Spearman's rank correlation coefficient.

interaction sum of squares into sources due to scaling effects (heterogeneity of variances) and re-ranking indicated that the G × E interaction for all traits was dominated by changes in genotype ranking over the two sites (Table 4). Nevertheless, it appeared that only 9–13% of the genotypes, which correspond to the most interactive ones, were found to explain 50% of the G × E interaction sum of squares. Whereas the impact of G × E interaction seemed higher for glucose and extractives on the basis of the relative magnitude of their variance components, the proportion of interactive genotypes were found to be quite similar to those for other traits, suggesting less practical importance of the observed interaction. Extractives had some level of scale effect (14.9%) and still high re-ranking (85.1%), whereas glucose had little or no scale effect (0.23%) and high re-ranking (99.77%) (Table 4). Similarly, the genetic variances, expressed in terms of genetic standard deviation (σ_G), were not similar between the two sites for extractives, with higher variation at Orleans (i.e, some level of scale effect), whereas, for glucose, the genetic variances were more homogeneous between Orleans and Savigliano (1.24 and 1.26, respectively) (i.e., little or no scale effect). Although H-lignin and extractives had similar patterns of partitioning of G × E interaction sum of squares, the Spearman's rank correlation was much weaker for extractives, which was consistent with the relatively stronger G × E effect on this particular trait, as shown by the ratio of σ^2_G to $\sigma^2_{G\times E}$ (Table 4).

Furthermore, we assessed the stability of genotype ranking across rotations or sites on a genotypic mean basis for each wood trait using Spearman's rank correlation coefficients (r_s). Thus, (r_s) between the two rotations were stronger than 0.60 ($r_s = 0.64 - 0.71$) for all traits except C5/C6, glucose and extractives (Table S4), which was consistent with the relatively low level of G × E interaction observed across rotations for most traits (1–9%). For C5/C6, glucose and extractives, the correlations were in the range of 0.48–0.50, which was consistent with the relatively higher proportion of G × E interaction variances for these three traits (12–18%) (Fig. 3, Table S4). By contrast, the Spearman's rank correlations of genotypic means between the two sites were lower than 0.60 for most of the traits ($r_s = 0.45 - 0.56$), and has turned out to be much weaker for extractives ($r_s = 0.23$) and glucose ($r_s = 0.34$) (Table 4), which corroborated the relatively high level of G × E interaction observed across sites compared to across rotations (Fig. 4, Table S5).

4. Discussion

A study of this scale would not have been possible using the standard method of wood compositional analysis because of the high cost and time required. For example, to analyze the 120 reference samples in two technical replicates using the wet chemistry method, it took about two months. It means that, it would have taken around 8 years to analyze about the 6000 samples included in our study. A way to circumvent such technical limitation is to use an attractive technique that combines NIR spectroscopy with multivariate statistical analysis. NIR spectroscopy is an inexpensive and high-throughput technique for phenotyping large-scale wood samples required for the genetic analysis of biofuel related traits and, consequently, it can provide the opportunity to select or develop biofuel-type poplar clones. Nevertheless, NIR spectroscopy is an indirect method which is reliable only if calibration models are provided. In this study, chemical composition data from standard methods and NIR spectra of reference samples were used to develop and validate calibrations taking into consideration the phenotypic variation induced by multi-environment evaluation. The models based on a 120-samples reference set were then used to predict the composition of the 5799 black poplar samples covering the range of the species in Western Europe. Using NIR predictions, we evaluated their genetic variability and the extent of G × E interaction across coppice rotations and sites. To our knowledge, this is the first work to evaluate large-scale clonal trials of *P. nigra* for wood chemical traits using an indirect method of measurement.

4.1. Calibration reliability

When global calibration models developed for the prediction of 6 wood chemical traits (H-lignin, S/G, H/G, xylose/glucose, C5/C6, extractives) in samples of European black poplar were tested on an independent validation data set, they gave good fits, suggesting their potential use in genetic analysis of large data sets or for ranking of genotypes with respect to their predicted phenotypic performances in initial selection steps in breeding programs. RPD values of ~1.5 indicate that the models are acceptable as initial screening tools, whereas RPD greater than 2.5 suggest that the models are good for screening candidates in breeding programs (Yeh et al., 2005). Although site effects were apparent on some traits such as xylose/glucose ratio, we were able to develop non-site specific calibrations for such parameters.

By contrast, global models for lignin content (Klason lignin, Py-lignin, acid-soluble lignin) showed clearly poorer performance in the validation set than in the calibration set. There is no obvious explanation for these apparent differences in global model performance between calibration and validation sets. We further examined if model fit was better for site specific calibration than global ones for lignin content and found that the Klason lignin model had a good fit at Orleans, while Py-lignin and acid-soluble lignin models had good fits at Savigliano, which does explain why we could not have global models for these characteristics.

Only a few studies have investigated the efficiency of NIR calibration models for prediction of poplar wood composition and these focused on hybrid poplars instead of natural populations. Robinson and Mansfield (2009) used NIR spectra of 267 wild and transgenic

hybrid poplar samples coupled with a modified thioacidolysis protocol for predicting lignin monomer proportions (S, G, and H). The authors reported highly accurate calibrations with prediction R^2 values of 0.96, 0.96, and 0.71 for S, G and H, respectively. More recently, Zhou et al. (2011) used Fourier transform infrared spectroscopy (FTIR) and acetyl bromide method to develop a calibration model for predicting lignin content in hybrid poplar wood samples. They reported a strong calibration with cross-validation R^2 of 0.81 and prediction R^2 of 0.88. Our global model for H-lignin had higher prediction R^2 than the local model developed by Robinson and Mansfield (2009). However, we found lower prediction R^2 values for the predominant G and S lignin monomers. Compared with the lignin model developed by Zhou et al. (2011), our local models for Klason lignin, Py-lignin and acid-soluble lignin had slightly lower R^2 values. The differences between these previous studies and our work may be largely related to differences in study population and standard laboratory methods. However, in this study spectra were recorded on non-debarked and non-extracted wood samples because it is practically difficult to debark and extract a large amount of samples (n = ~6000). Although the presence of bark and extractives may disturb the spectra, we still attained sufficiently accurate models for a majority of the traits analyzed. Furthermore, the R^2 value may be misleading as it depends not only on the model error but also on the range of variation of the trait of interest within or across sites. For example, in this study the effect of site on the range of variation of some of the traits analyzed was quite high (Fig. S5). Consequently, different R^2 values can be obtained for calibrations for the same trait at the two sites while still having the same prediction error. Higher R^2 values can be obtained for calibrations at site that is more variable than other.

4.2. Variabilities, G × E interactions and broad-sense heritability of wood chemical properties

In this study, using NIR predictions of a large number of wood samples from *P. nigra* clonal trials, we assessed variabilities, G × E interaction and broad-sense heritability for wood chemical traits. The range of phenotypic variation in most NIR-predicted wood chemical traits in the black poplar populations studied was substantial. Guerra et al. (2013) used Pyrolysis molecular beam mass spectrometry (pyMBMS) to determine C6 sugars, total lignin content and S/G ratio in wood samples of 2-yr-old trees, representing 17 open-pollinated families of *P. nigra*. Porth et al. (2013) used wet laboratory approaches to determine xylose, glucose, Klason lignin and acid soluble lignin in wood samples of 9-yr-old trees, representing natural populations of *P. trichocarpa*. The range of variation observed for predicted glucose content (30.2–49.7%) in the present study is in accordance with that reported for C6 sugars (27.7–39.7%) by Guerra et al. (2013) and for glucose (40.7–61.7%) by Porth et al. (2013). We obtained predicted Klason lignin content (16.1–27.9%) that is well comparable to the results of total lignin content (Klason and soluble lignin) reported by Guerra et al. (2013) (19.5–26.5%) and Porth et al. (2013) (14.7–25.7%). The range of variation of the predicted lignin S/G ratio (0.69–1.43) described in this study doesn't mirror the range between 1.3 and 2.1 reported by Guerra et al. (2013) despite almost the same magnitude of variations, which might arise from the differences in the standard methods of lignin monomers determination.

The effects of harvest on some of the predicted wood chemical traits are evident in Fig. S5. This motivated us to ask whether there is a significant influence of G × E interaction on the wood chemical traits. Understanding the magnitude and nature of G × E interaction would be useful for establishing breeding objectives. To estimate the importance of G × E interaction, we examined variance contributions of G × E interaction for the wood traits and correlations of same traits between environments based on genotype means. In this study, significant G × E interaction was observed across rotations as well as across the two sites for a majority of the traits assessed, suggesting differential responses of genotypes to the environmental conditions. The G × E interaction variance component accounted for a lower proportion of the total variance across rotations than across sites and this was consistent with the rank correlations of genotype means between rotations and sites obtained for most traits examined. Together, it implies that genotype ranking was relatively more maintained between rotations than between sites. The observed differences in the magnitude of interactions between rotations and sites were not surprising, since the clonal trials were established at two contrasting sites, particularly in terms of soil fertility. Savigliano is characterized by a higher soil fertility compared to Orleans (Guet et al., 2015). Given the differences in edaphic factors between the trial sites, the significant G × E effect revealed for wood chemical traits across sites could result indirectly from the effects of edaphic factors on tree growth.

When the contributions of G × E interaction and genotype main effects to the total phenotypic variances of predicted wood traits were compared, all the traits had a higher percentage of variance due to genetic variance component, suggesting less consequences of interaction in poplar tree breeding for improved wood quality. The exceptions were the glucose and extractives contents across sites, for which the G × E variance components were larger than the genetic variance components, which was also consistent with the relatively lower rank correlations of genotype means between sites for these two traits. Nevertheless, the partitioning of the G × E sum of squares revealed that the G × E effect was mainly caused by a few interactive genotypes as for the other traits. This suggests that the interaction would have less consequences in the poplar tree breeding programs for biofuel production because there exists a high possibility to identify genotypes with stable wood quality across the two sites. To test this assumption, we have further computed the relative loss in genetic gain that would arise when selecting the best 5% genotypes for some relevant traits (S/G, H/G, glucose, xylose/glucose, C5/C6) on their genotype mean across the 2 sites instead of their genotype mean within each targeted site. We found that this loss would be fairly low relatively to the maximum expected gain in the two targeted sites (13.1 and 14.3% on average at Orléans and Savigliano, respectively).

To date, only a few studies have investigated the effect of G × E interaction on wood chemical properties, especially in poplars. Kačík et al. (2012) studied poplar hybrid clones and reported the presence of significant clone × site interaction for wood chemical traits (lignin content, cellulose, holocellulose, extractives, S/G ratio). However, the authors did not provide further information about the implications of the observed interaction for poplar tree breeding for wood quality. Similarly, Zhang et al. (2015) found significant clone × site interaction for lignin content and extractives in triploid hybrid clones of *P. tomentosa*. However, clone by site variance exceeded clonal variance only for holocellulose content, for which the authors did not detect significant interaction, and not for lignin content or extractives.

Consistent with the observed G × E interaction, extractives content showed relatively low within-site broad-sense heritability estimates in this study. Compared to the main wood components, extractives content may be of less interest as a direct selection trait in poplar breeding programs for biofuel production. Since chemical analysis was carried out on non-debarked wood samples in the present study, we wondered if such particular pattern of variation for extractives content would be somehow related to variation in bark proportion. To test this hypothesis, we sought to use the diameter of the samples as a proxy of bark proportion: samples with relatively large diameter are expected to have less bark, and consequently, less extractives. Clearly, extractives content tended to decrease with increasing tree diameter (not shown). We thus extended our G × E analyses to tree circumference at 1 m aboveground at harvest in order to check if it could explain the particular pattern of variation observed for extractives in comparison with the other wood chemical traits. Interestingly, we found that, albeit highly significant, the G × E interaction effect accounted for much less variation than the genotype main effect, resulting in G to G × E

variances ratio of 3.90 and 1.31, as well as rank correlations between genotype means of 0.68 and 0.53 across rotations and sites, respectively (Table S6). This pattern of G × E across rotations and sites was pretty much consistent with the pattern observed for all wood chemical traits, but did not explain the exceptionally interactive aspect of extractives. We thus conclude that, of all the traits evaluated in this study, extractives content was the most interactive trait with moderate heritability and we found no evidence for our hypothesis that the G × E effect on extractives is confounded by G × E effect on tree circumference. This result is also supported by the fact that we developed a good global calibration for extractives regardless of the differences in sample bark content between the two sites.

We also quantified the extent of genetic variation present within the European populations of black poplar in the clonal trials using the NIR predictions. Broad-sense heritability was estimated at both individual tree and clonal mean levels. For clonal selection, clonal mean broad-sense heritability (clonal repeatability) is more meaningful. Genetic analysis with NIR predictions revealed that the studied wood chemical traits were under moderate to high genetic control. However, care must be taken when interpreting the heritability estimates reported in the present study because they were estimated from phenotypic data that had been adjusted for within-site non-genetic random effects like block, date and spatially dependent residuals. Consequently, they were over-estimated to an extent that corresponds to an omission of non-genetic random variances in the denominator of the heritability ratio when estimated from the first model (using all genotypes within each harvest, as reported in Table S3). Still, our results suggested that satisfactory genetic gains could be realized in wood chemical traits through clonal selection using a fairly low number of replicates (2.7–2.8 per genotype on average) when NIR analysis is integrated in a breeding program to evaluate large sets of candidate clones. In this regard, the information produced in this research could be used for screening individuals with desirable traits from large-scale clonal trials as future potential parent trees for hybrid breeding programs aimed at cellulosic ethanol production. A general trend was observed for the studied traits in terms of clonal repeatability. Lignin monomers and lignin composition had the highest values, followed by lignin contents, cell wall sugars and extractives (Figs. 1 and 2, Tables S1 and S2). However, the estimated clonal repeatability differed more between sites than between rotations for the same traits, which was in agreement with the G × E interaction results. The higher clonal repeatability estimates obtained for most of the traits at Savigliano may be explained by the existence of a relatively favourable growth conditions for poplar trees at this site, which resulted in both increased expression of genetic variation and reduced residual variation. Savigliano could be a suitable growth site to apply the clonal evaluation as it provided the genotypes relatively suitable conditions for expressing their genetic potential compared to Orleans.

Using direct method of measurements, previous studies in *P. nigra* (Guerra et al., 2013) and *P. trichocarpa* (Guerra et al., 2016; Porth et al., 2013; Wegrzyn et al., 2010) have also shown that wood chemical properties are under moderate to high genetic control. For example, Guerra et al. (2013) studied 17 cloned open-pollinated families of *P. nigra* and reported individual broad-sense heritability (H_i^2) values of 0.46, 0.58 and 0.70 for C6 sugars, lignin and S/G, respectively. In the current report, the estimated H_i^2 values of 0.47 ± 0.05-0.55 ± 0.04, 0.52 ± 0.07-0.59 ± 0.06 and 0.55 ± 0.04–75 ± 0.03 for glucose, Klason lignin and S/G, respectively, compares favourably well with H_i^2 values reported by these previous authors (Figs. 1 and 2, Tables S1 and S2). More recently, Guerra et al. (2016) studied *P. trichocarpa* clones sampled in provenances and reported the clonal repeatability (H_c^2) estimates of 0.22, 0.33 and 0.81 for C6 sugars, lignin and S/G, respectively, with an average number of 3 biological replicates per clone. In comparison with the results of S/G reported by these authors, we found similar H_c^2 values for S/G (0.77 ± 0.03–0.89 ± 0.01) (Figs. 1 and 2, Table S1 and S2). Porth et al. (2013) studied the narrow-sense heritability of several wood properties in natural populations of *P. trichocarpa* using molecular markers to measure relatedness and reported values of 0.46, 0.66, 0.97 for glucose, Klason lignin and soluble lignin, respectively. We found higher clonal repeatability for glucose (0.70 ± 0.04-0.77 ± 0.03) and Klason lignin (0.75 ± 0.05–0.80 ± 0.04), but a lower value for acid-soluble lignin (0.83 ± 0.02), indicating that acid-soluble lignin may be under relatively lower genetic control in *P. nigra* than *P. trichocarpa* (Figs. 1 and 2, Table S1 and 2).

4.3. Adapting the NIR method to clonal trials

An initial step to harness the standing genetic variation in poplar is to evaluate natural populations in multi-site clonal trials. This allows to study the relative importance of genetic, environment and G × E interaction on important biomass production and biomass composition related traits. In parallel, screening good candidates from clonal trials as future parents would increase the genetic diversity available for breeding poplar trees for cellulosic ethanol production. The goal of bioenergy poplar breeding program is to simultaneously improve biomass production and biomass composition. To incorporate wood quality traits into breeding programs, however, tree breeders need low-cost and high-throughput techniques for determination of biomass composition. Standard methods for analysis of biomass composition such as wet chemistry are useful for evaluating small sample sets, but they have limitations to be used in tree breeding programs, where screening of a large number of samples is mandatory to identify those possessing desirable traits. Standard methods are laborious, costly and time consuming. An alternative way is to use NIR spectroscopy coupled with multivariate statistical approaches. NIR spectroscopy is a high-throughput technique for screening a large population. It is easy to operate, allows non-destructive analysis, needs little sample preparation, provides reliable information, requires less time and minimal cost for assessing large number of samples and captures multiple features of the samples with one operation (Lupoi et al., 2014).

The moderate to high heritability estimates and the detection of G × E interaction in this study are encouraging for NIR determination of wood chemical traits and for use in poplar breeding programs for cellulosic ethanol production. Integration of NIR analysis in multi-site clonal trials would allow simultaneous multi-trait evaluation and give access to identify potential trade-offs between biomass production and biomass composition, which in turn, supports poplar breeding programs to better monitor multi-trait selection and exploit the large variation present in natural gene pools. As a first check at the genotypic level, we have computed the correlations within each harvest between growth and wood properties and haven't found any adverse correlation within our dataset (Fig. S6). These results are encouraging towards the development of performing clones dedicated to biomass and biofuel production.

Despite its importance, optimal procedures for developing NIR calibrations for rapid prediction of wood composition in multi-site poplar clonal trials are not well established. In the present study, we developed NIR calibration models and successfully applied this indirect method to analyze the sources and extent of variability for wood chemical traits in large-scale clonal trials of *P. nigra*, which is the first work, as far as we know. Finally, future work on development of new calibration models would be useful to further establish the NIR calibration protocols for clonal trials. Some of the important points to consider will be the number of technical replicates for the reference samples to reduce the uncertainties associated with the standard methods and the number of biological replicates per genotype to reach enough accuracy on a clonal basis.

5. Conclusions

From our study of wood chemical traits in clonal trials of European black poplar at two contrasting sites, three important conclusions can

be drawn. (1) We successfully developed global and site specific NIR calibration models for predicting wood chemical traits in natural populations of European black poplar with reasonable accuracy. (2) We demonstrated the high throughput nature of the NIR method, by applying the calibrations to predict the wood chemical composition of the 5799 trees and by the analyses of these NIR predictions to estimate trait variance components and broad-sense heritabilities. (3) We further used the NIR predictions to test and evaluate the extent of G × E interaction across coppice rotations within a single site as well as across sites.

In this study, the moderate to high heritability estimates and the detection of G × E interaction suggests that the NIR-based technique can efficiently be used for dissecting the genetic basis of wood chemical properties in a multi-environment large-scale poplar clonal trials and for screening elite individuals from such trials as future parents for interspecific hybridization. Integration of such indirect method in poplar tree breeding programs would allow the exploitation of standing genetic variation in poplars for developing poplar genotypes that combine high biomass yield with superior wood quality for cellulosic ethanol production. Furthermore, the observed moderate to strong genetic control over the NIR-predicted wood chemical traits should pave the way for more detailed dissection of the genetic and molecular basis of the NIR-predicted wood compositional variation through molecular marker analysis of the NIR predictions. In particular, it would be useful to extend such analysis to association mapping aimed at identifying individual loci controlling the predicted phenotypic variation in the studied population of *P. nigra*.

Acknowledgements

The authors gratefully acknowledge the staff of the INRA-GBFOR experimental unit for the establishment and management of the experimental plantation in Orléans, wood sample collections in the two sites, and contribution to circumference measurements; Alasia Franco Vivai staff for the management of the experimental plantation in Savigliano. We would also like to thank Eduardo Cappa, Facundo Muñoz and Leopoldo Sanchez for useful discussion on the genetic analysis of data.

Funding: Establishment and management of the experimental sites until harvests were carried out with financial support from the NOVELTREE project (EU-FP7-211868). All analyses on wood samples were supported by the SYBIOPOP project funded by the French National Research Agency (ANR-13-JSV6-0001). M. N. G. was supported by a PhD grant jointly funded by the SYBIOPOP project (ANR-13-JSV6-0001) and the EFPA division of INRA.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.indcrop.2017.05.013.

References

Abramson, M., Shoseyov, O., Shani, Z., 2010. Plant cell wall reconstruction toward improved lignocellulosic production and processability. Plant Sci. 178, 61–72. http://dx.doi.org/10.1016/j.plantsci.2009.11.003.

Alves, A., Schwanninger, M., Pereira, H., Rodrigues, J., 2006. Calibration of NIR to assess lignin composition (H/G ratio) in maritime pine wood using analytical pyrolysis as the reference method. Holzforschung 60, 29–31. http://dx.doi.org/10.1515/HF.2006.006.

Alves, A., Simões, R., Stackpole, D., Vaillancourt, R., Potts, B., Schwanninger, M., Rodrigues, J., 2011. Determination of the syringyl/guaiacyl (S/G) ratio of *Eucalyptus globulus* Labill. wood lignin by NIR-based PLS-R models using analytical pyrolysis as the reference method. J. Near Infrared Spectrosc. 19, 343–348.

Alves, A.M.M., Simões, R.F.S., Santos, C.A., Potts, B.M., Rodrigues, J., Schwanninger, M., 2012. Determination of *Eucalyptus globulus* wood extractives content by near infrared-based partial least squares regression models: comparison between extraction procedures. J. Near Infrared Spectrosc. 20, 275–285. http://dx.doi.org/10.1255/jnirs.987.

Araus, J.L., Cairns, J.E., 2014. Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci. 19, 52–61. http://dx.doi.org/10.1016/j.tplants.2013.09.008.

Baillères, H., Davrieux, F., Ham-Pichavant, F., 2002. Near infrared analysis as a tool for rapid screening of some major wood characteristics in a eucalyptus breeding program. Ann. For. Sci. 59, 479–490. http://dx.doi.org/10.1051/forest:2002032.

Bradshaw, H.D., Ceulemans, R., Davis, J., Stettler, R., 2000. Emerging model systems in plant biology: poplar (*Populus*) as a model forest tree. J. Plant Growth Regul. 19, 306–313. http://dx.doi.org/10.1007/s003440000030.

Cagelli, L., Lefevre, F., 1995. The conservation of *Populus nigra* L. and gene flow with cultivated poplars in Europe. For. Genet. 2, 135–144.

Ceulemans, R., Deraedt, W., 1999. Production physiology and growth potential of poplars under short-rotation forestry culture. For. Ecol. Manage. 121, 9–23. http://dx.doi.org/10.1016/S0378-1127(98)00564-7.

Costa e Silva, J., Borralho, N.M.G., Araújo, J.A., Vaillancourt, R.E., Potts, B.M., Silva, J.C.E., Araujo, J.A., 2008. Genetic parameters for growth, wood density and pulp yield in *Eucalyptus globulus*. Tree Genet. Genomes 5, 291–305. http://dx.doi.org/10.1007/s11295-008-0174-9.

Da Silva Perez, D., Guillemain, A., Alazard, P., Plomion, C., Rozenberg, P., Carlos Rodrigues, J., Alves, A., Chantre, G., 2007. Improvement of Pinus pinaster Ait elite trees selection by combining near infrared spectroscopy and genetic tools. Holzforschung 61, 611–622. http://dx.doi.org/10.1515/HF.2007.118.

Dutkowski, G.W., Silva, J.C.E., Gilmour, A.R., Lopez, G.A., 2002. Spatial analysis methods for forest genetic trials. Can. J. For. Res. 32, 2201–2214. http://dx.doi.org/10.1139/x02-111.

Frison, E., Lefevre, F., de Vries, S., Turok, J., 1994. *Populus Nigra* Network: Report of the First Meeting 3–5 October 1994 Izmit, Turkey. IPGRI, Rome, Izmit, Turkey.

Gaspar, M.J., Alves, A., Louzada, J.L., Morais, J., Santos, A., Fernandes, C., Almeida, M.H., Rodrigues, J.C., 2011. Genetic variation of chemical and mechanical traits of maritime pine (*Pinus pinaster* Aiton). Correlations with wood density components. Ann. For. Sci. 68, 255–265. http://dx.doi.org/10.1007/s13595-011-0034-x.

Guerra, F.P., Wegrzyn, J.L., Sykes, R., Davis, M.F., Stanton, B.J., Neale, D.B., 2013. Association genetics of chemical wood properties in black poplar (*Populus nigra*). New Phytol. 197, 162–176. http://dx.doi.org/10.1111/nph.12003.

Guerra, F.P., Richards, J.H., Fiehn, O., Famula, R., Stanton, B.J., Shuren, R., Sykes, R., Davis, M.F., Neale, D.B., 2016. Analysis of the genetic variation in growth, ecophysiology, and chemical and metabolomic composition of wood of Populus trichocarpa provenances. Tree Genet. Genomes 12. http://dx.doi.org/10.1007/s11295-015-0965-8.

Guet, J., Fabbrini, F., Fichot, R., Sabatti, M., Bastien, C., Brignolas, F., 2015. Genetic variation for leaf morphology, leaf structure and leaf carbon isotope discrimination in European populations of black poplar (*Populus nigra* L.). Tree Physiol. 35, 850–863. http://dx.doi.org/10.1093/treephys/tpv056.

Hamilton, M.G., Raymond, C., Harwood, C., Potts, B., 2009. Genetic variation in *Eucalyptus nitens* pulpwood and wood shrinkage traits. Tree Genet. Genomes 5, 307–316. http://dx.doi.org/10.1007/s11295-008-0179-4.

Henderson, C.R., 1984. Applications of Linear Models in Animal Breeding Models 384 Univ. Guelphhttp://dx.doi.org/10.1002/9780470316856. (ch7).

Isik, F., Mora, C.R., Schimleck, L.R., 2011. Genetic variation in *Pinus taeda* wood properties predicted using non-destructive techniques. Ann. For. Sci. 68, 283–293. http://dx.doi.org/10.1007/s13595-011-0035-9.

Jiang, W., Han, G., Via, B.K., Tu, M., Liu, W., Fasina, O., 2014. Rapid assessment of coniferous biomass lignin-carbohydrates with near-infrared spectroscopy. Wood Sci. Technol. 48, 109–122. http://dx.doi.org/10.1007/s00226-013-0590-3.

Kačík, F., Ďurkovič, J., Kačíková, D., 2012. Chemical profiles of wood components of poplar clones for their energy utilization. Energies 5, 5243–5256. http://dx.doi.org/10.3390/en5125243.

Kennard, R., Stone, L.A., 1969. Computer aided design of experiments. Technometrics 11, 137–148.

Kube, P.D., Raymond, C.A., Banham, P.W., 2001. Genetic parameters for diameter, basic density, cellulose content and fibre properties for *Eucalyptus nitens*. For. Genet. 8, 285–294.

Li, B.B., Morris, J., Martin, E.B., 2002. Model selection for partial least squares regression. Chemom. Intell. Lab. Syst. 64, 79–89.

Li, H., Liang, Y., Xu, Q., Cao, D., 2009. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. Anal. Chim. Acta 648, 77–84. http://dx.doi.org/10.1016/j.aca.2009.06.046.

Lin, C.-S., Binns, M.R., Lefkovitch, L.P., 1986. Stability analysis: where do we stand? Crop Sci. 26, 894–900.

Lupoi, J.S., Singh, S., Simmons, B.A., Henry, R.J., 2014. Assessment of lignocellulosic biomass using analytical spectroscopy: an evolution to high-throughput techniques. Bioenergy Res. 7, 1–23. http://dx.doi.org/10.1007/s12155-013-9352-1.

Lynch, M., Walsh, B., 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates Sunderland, MA.

Mevik, B.H., Wehrens, R., 2007. The pls package: principal component and partial least squares regression in R. J. Stat. Softw. 18, 1–23. http://dx.doi.org/10.1159/000323281.

Muñoz, F., Sanchez, L., 2015. BreedR: Statistical Methods for Forest Genetic Resources Analysts.

Muir, W., Nyquist, W.E., Xu, S., 1992. Alternative partitioning of the genotype-by-environment interaction. Theor. Appl. Genet. 84, 193–200. http://dx.doi.org/10.1007/BF00224000.

Neale, D.B., Kremer, A., 2011. Forest tree genomics: growing resources and applications. Nat. Rev. Genet. 12, 111–122. http://dx.doi.org/10.1038/nrg2931.

Poke, F.S., Potts, B.M., Vaillancourt, R.E., Raymond, C.A., 2006. Genetic parameters for lignin, extractives and decay in *Eucalyptus globulus*. Ann. For. Sci. 63, 813–821.

http://dx.doi.org/10.1051/forest:2006080.
Porth, I., Klápště, J., Skyba, O., Lai, B.S.K., Geraldes, A., Muchero, W., Tuskan, G.A., Douglas, C.J., El-Kassaby, Y.A., Mansfield, S.D., 2013. Populus trichocarpa cell wall chemistry and ultrastructure trait variation, genetic control and genetic correlations. New Phytol. 197, 777–790. http://dx.doi.org/10.1111/nph.12014.
R Core Team, 2015. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing. R Core Team, Vienna, Austria. URL https://www.R-project.org/.
Ragauskas, A.J., Nagy, M., Kim, D.H., Eckert, C.A., Hallett, J.P., Liotta, C.L., 2006. From wood to fuels: integrating biofuels and pulp production. Ind. Biotechnol. 2, 55–65. http://dx.doi.org/10.1089/ind.2006.2.55.
Raymond, C.A., Schimleck, L.R., 2002. Development of near infrared reflectance analysis calibrations for estimating genetic parameters for cellulose content in *Eucalyptus globulus*. Can. J. For. Res. Can. Rech. For. 32, 170–176. http://dx.doi.org/10.1139/X01-174.
Raymond, C.A., Schimleck, L.R., Muneri, A., Michell, A.J., 2001. Genetic parameters and genotype-by-environment interactions for pulp yield predicted using near infrared reflectance analysis and pulp productivity in *Eucalyptus globulus*. For. Genet. 8, 213–224.
Robinson, A.R., Mansfield, S.D., 2009. Rapid analysis of poplar lignin monomer composition by a streamlined thioacidolysis procedure and near-infrared reflectance-based prediction modeling. Plant J. 58, 706–714. http://dx.doi.org/10.1111/j.1365-313X.2009.03808.x.
Rodrigues, J., Meier, D., Faix, O., Pereira, H., 1999. Determination of tree to tree variation in syringyl/guaiacyl ratio of *Eucalyptus globulus* wood lignin by analytical pyrolysis. J. Anal. Appl. Pyrolysis 48, 121–128. http://dx.doi.org/10.1016/S0165-2370(98)00134-X.
Rodrigues, J., Graça, J., Pereira, H., 2001. Influence of tree eccentric growth on syringyl/guaiacyl ratio in *Eucalyptus globulus* wood lignin assessed by analytical pyrolysis. J. Anal. Appl. Pyrolysis 58–59, 481–489. http://dx.doi.org/10.1016/S0165-2370(00)00121-2.
Rubin, E.M., Himmel, M.E., Ding, S., Johnson, D.K., Adney, W.S., 2007. Biomass Recalcitrance: Nat. 454, 804–807. http://dx.doi.org/10.1126/science.1137016.
Rubin, E.M., 2008. Genomics of cellulosic biofuels. Nature 454, 841–845. http://dx.doi.org/10.1038/nature07190.
Sannigrahi, P., Ragauskas, A.J., Tuskan, G.A., 2010. Poplar as a feedstock for biofuels: a review of compositional characteristics. Biofuels Bioprod. Biorefin. 4, 209–226. http://dx.doi.org/10.1002/bbb.206.
Schimleck, L.R., Kube, P.D., Raymond, C.A., 2004. Genetic improvement of kraft pulp yield in *Eucalyptus nitens* using cellulose content determined by near infrared spectroscopy. Can. J. For. Res. 34, 2362–2370. http://dx.doi.org/10.1139/X04-119.
Schubert, C., 2006. Can biofuels finally take center stage? Nat. Biotechnol. 24, 777–784. http://dx.doi.org/10.1038/nbt0706-777.
Schwanninger, M., Rodrigues, J.C., Gierlinger, N., Hinterstoisser, B., 2011a. Determination of lignin content in Norway spruce wood by Fourier transformed near infrared spectroscopy and partial least squares regression. Part 1. Wavenumber-selection and evaluation of the selected range. J. Near Infrared Spectrosc. 19, 319–329. http://dx.doi.org/10.1255/jnirs.945.
Schwanninger, M., Rodrigues, J., Hinterstoisser, B., 2011b. Determination of lignin content in Norway spruce wood by Fourier transformed near infrared spectroscopy and partial least squares regression analysis. Part 2: Development and evaluation of the final model. J. Near Infrared Spectrosc. 19, 331–341.
Shelbourne, C.J.A., 1972. Genotype-environment interaction: its study and its implications in forest tree improvement. Proc. IUFRO Genet.-SABARAO Joint Symp. Tokyo B-1 (I), 1–28.
Stackpole, D.J., Vaillancourt, R.E., Downes, G.M., Harwood, C.E., Potts, B.M., 2010. Genetic control of kraft pulp yield in *Eucalyptus globulus*. Can. J. For. Res. 40, 917–927. http://dx.doi.org/10.1139/X10-035.
Stackpole, D.J., Vaillancourt, R.E., Alves, A., Rodrigues, J., Potts, B.M., 2011. Genetic variation in the chemical components of Eucalyptus globulus Wood. G3: Genes Genomes Genet. G3 (1), 151–159. http://dx.doi.org/10.1534/g3.111.000372.
Stevens, A., Ramirez-Lopez, L., 2013. An introduction to the prospectr package.
Sticklen, M.B., 2008. Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. Nat. Rev. Genet. 9, 433–443. http://dx.doi.org/10.1038/nrg2336.
Tsuchikawa, S., Kobori, H., 2015. A review of recent application of near infrared spectroscopy to wood science and technology. J. Wood Sci. 61, 213–220. http://dx.doi.org/10.1007/s10086-015-1467-x.
Wegrzyn, J.L., Eckert, A.J., Choi, M., Lee, J.M., Stanton, B.J., Sykes, R., Davis, M.F., Tsai, C.J., Neale, D.B., 2010. Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. New Phytol. 188, 515–532. http://dx.doi.org/10.1111/j.1469-8137.2010.03415.x.
Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., Fang, W., Feng, H., Xie, W., Lian, X., Wang, G., Luo, Q., Zhang, Q., Liu, Q., Xiong, L., 2014. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nat. Commun. 5, 5087. http://dx.doi.org/10.1038/ncomms6087.
Yeh, T.F., Yamada, T., Capanema, E., Chang, H.M., Chiang, V., Kadla, J.F., 2005. Rapid screening of wood chemical component variations using transmittance near-infrared spectroscopy. J. Agric. Food Chem. 53, 3328–3332. http://dx.doi.org/10.1021/jf0480647.
Zhang, J., Novaes, E., Kirst, M., Peter, G.F., 2014. Comparison of pyrolysis mass spectrometry and near infrared spectroscopy for genetic analysis of lignocellulose chemical composition in *Populus*. Forests 5, 466–481. http://dx.doi.org/10.3390/f5030466.
Zhang, P., Wu, F., Kang, X., 2015. Chemical properties of wood are under stronger genetic control than growth traits in *Populus tomentosa* Carr. Ann. For. Sci. 72, 89–97. http://dx.doi.org/10.1007/s13595-014-0401-5.
signal developers, 2013. Signal: Signal Processing. URL: http://r-forge.r-project.org/projects/signal/.
Zhou, G., Taylor, G., Polle, A., 2011. FTIR-ATR-based prediction and modelling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars. Plant Methods 7, 1–10. http://dx.doi.org/10.1186/1746-4811-7-9.