Using Near Infrared Spectroscopy to explore geographical patterns of genetic diversity and predict quantitative phenotypes: application to natural populations of black poplar (Populus nigra).

Vincent Segura, Jean-Paul J.-P. Charpentier, Kévin Ader, Redouane El Malki, Justine Guet, Véronique V. Jorge, Patrick Poursat, Jean-Charles Bastien, Catherine Bastien

Using Near Infrared Spectroscopy to explore geographical patterns of genetic diversity and predict quantitative phenotypes: application to natural populations of black poplar (*Populus nigra*).

Vincent Segura[1], Jean-Paul Charpentier[1,2], Kévin Ader[1,2], Redouane ElMalki[1], Justine Guet[1,3], Véronique Jorge[1], Patrick Poursat[4], Jean-Charles Bastien[1] and Catherine Bastien[1].

[1] INRA, UR0588 Amélioration Génétique et Physiologie Forestières, Orléans, France

[2] INRA, Plateau technique Génobois, Orléans, France

[3] INRA, USC 1328 Laboratoire de Biologie des Ligneux et des Grandes Cultures, Orléans, France

[4] INRA, UE0995 Génétique et Biomasse Forestières, Orléans, France

segura@orleans.inra.fr

Background

Near Infrared Spectroscopy (NIRS) is a popular high-throughput technique for characterizing and predicting the physical and chemical properties of biological samples. In forestry research, it has mainly and successfully been applied to the prediction of multiple wood properties, including chemical composition and physical properties (reviewed by Tsuchikawa, 2007). More anecdotally, the discrimination potential of NIRS has also been explored and applied to wood and leaf samples in order to identify their corresponding species or provenances within species (Richardson *et al.*, 2003; Tsuchikawa, 2007; O'Reilly-Wapstra *et al.*, 2013), suggesting that it is able to capture genetic variation. Few recent studies have more directly confirmed this point by estimating the broad-sense heritability of NIRS and/or mapping corresponding genetic loci (Posada *et al.*, 2009; O'Reilly-Wapstra *et al.*, 2013). As a result, these genetic analyses suggested that NIRS signature is an interesting biological marker which, similarly to DNA-based markers, could be used to explore and analyze the structure of genetic diversity and to predict the heritable variation of quantitative traits. The present work aims at testing such hypotheses in natural populations of *P. nigra*.

Methods

NIR spectra have been collected on 791 wood samples corresponding to one to three replicates of 288 clones originated from 7 metapopulations that cover the native range of *P. nigra* in France and Northern Italy. The wood samples consisted in two-year-old stem sections harvested during winter 2009-2010 in three blocks of an experimental randomized complete block design located in Orléans, France. On the same trees, various phenotypes have been measured during two successive rounds of coppicing (2008-2009 and 2010-2011), including tree height, diameter, branching habit, bud flush, bud set and rust resistance. After an exploratory multivariate analysis of NIRS that aimed at studying the genetic structure of *P. nigra* populations, the genetic variability of NIRS has been evaluated by estimating the broad sense heritability (H²) together with the coefficient of genetic differentiation between metapopulations (Qst) from the application of mixed-linear models to each of the 2001 spectrum wave numbers. Then, the ability of NIRS to predict metapopulation membership has been investigated using multivariate discriminant analyses. Finally, we have evaluated the predictive ability and the accuracy of NIRS for predicting quantitative traits breeding values using partial least squares (PLS) regression.

Results

While Principal Component Analysis did not allow discriminating *P. nigra* metapopulations, Factorial Discriminant Analysis (FDA) highlighted interesting patterns that appeared consistent with metapopulations geographic origin. To assess the discrimination power of the FDA, we performed hierarchical clustering on a distance matrix computed from the projection of individuals onto all six FDA components and found that all individuals

belonging to a given metapopulation clustered together. These results suggested that NIRS is able to capture between-metapopulation genetic variation. We confirmed this point by partitioning the phenotypic variation along NIR spectra into genetics (including metapopulation and genotype nested into metapopulation) and residuals. This analysis showed that many NIR regions displayed significant between- and within-metapopulation genetic variation, with a broad sense heritability of the clonal mean of NIRS being on average equal to 0.41 and ranging from 0.01 to 0.84.

Given the fact that NIRS signature was able to capture a significant amount of genetic information, we were able to predict the metapopulation membership with a maximum overall accuracy of 69 % estimated through a 4-fold cross-validation repeated 100 times. The prediction performances varied greatly depending on the metapopulation, some being accurately predicted at more than 80%, suggesting that the metapopulation definition might not be optimal, and/or that admixture between some populations could be suspected.

We were also able to predict quantitative trait clonal means using the corresponding NIRS averaged by clone. Considering a NIRS global heritability of 1 (similarly to DNA-based markers), the prediction accuracy estimated through a 4-fold cross-validation repeated 100 times varied greatly depending on the phenotype, ranging on average from 0.11 for average branching angle in 2011 to 0.81 for diameter in 2009. Accounting for the estimated NIRS heritability weighted by the contribution of NIRS wave numbers to the phenotype prediction in the PLS models, the prediction accuracy significantly increased. For some phenotypes, these accuracies were even increased over 1 suggesting that NIRS was able to predict non genetic variability. This was noticeably the case for the phenotypes related to the morphology of the samples analyzed by NIRS, like height and diameter measured in 2009. In this case one could expect some correlation between the quantity and quality of biomass produced, which would increase the phenotypic prediction over the trait heritability. More surprisingly, we also got some prediction accuracies over 1 for traits related to phenology and rust resistance, underling the potential interest of NIRS for predicting phenotypes not related to wood physical or chemical properties.

Conclusions

In conclusion, the present study demonstrates that NIRS is a powerful high-throughput technique that could be used as a biological marker for exploring the genetic diversity and predicting quantitative phenotype of interest in breeding. In perspective, our results should and will be complemented in a near future by comparing the accuracy of NIRS predictions with those reached using DNA-based marker on the same dataset.

References

O'Reilly-Wapstra JM, Freeman JS, Barbour R, Vaillancourt RE, Potts BM (2013) Genetic analysis of the near-infrared spectral phenome of a global Eucalyptus species. Tree Genetics & Genomes, Online First.

Posada H, Ferrand M, Davrieux F, Lashermes P, Bertrand B (2009) Stability across environments of the coffee variety near infrared spectral signature. Heredity 102: 113-119

Richardson AD, Reeves III JB, Gregoire TG (2004) Multivariate analyses of visible/near infrared (VIS/NIR) absorbance spectra reveal underlying spectral differences among dried, ground conifer needle samples from different growth environments. New Phytologist 161: 291-301

Tsuchikawa S (2007) A Review of Recent Near Infrared Research for Wood and Paper. Applied Spectroscopy Reviews 42: 43-71

Figures / Tables

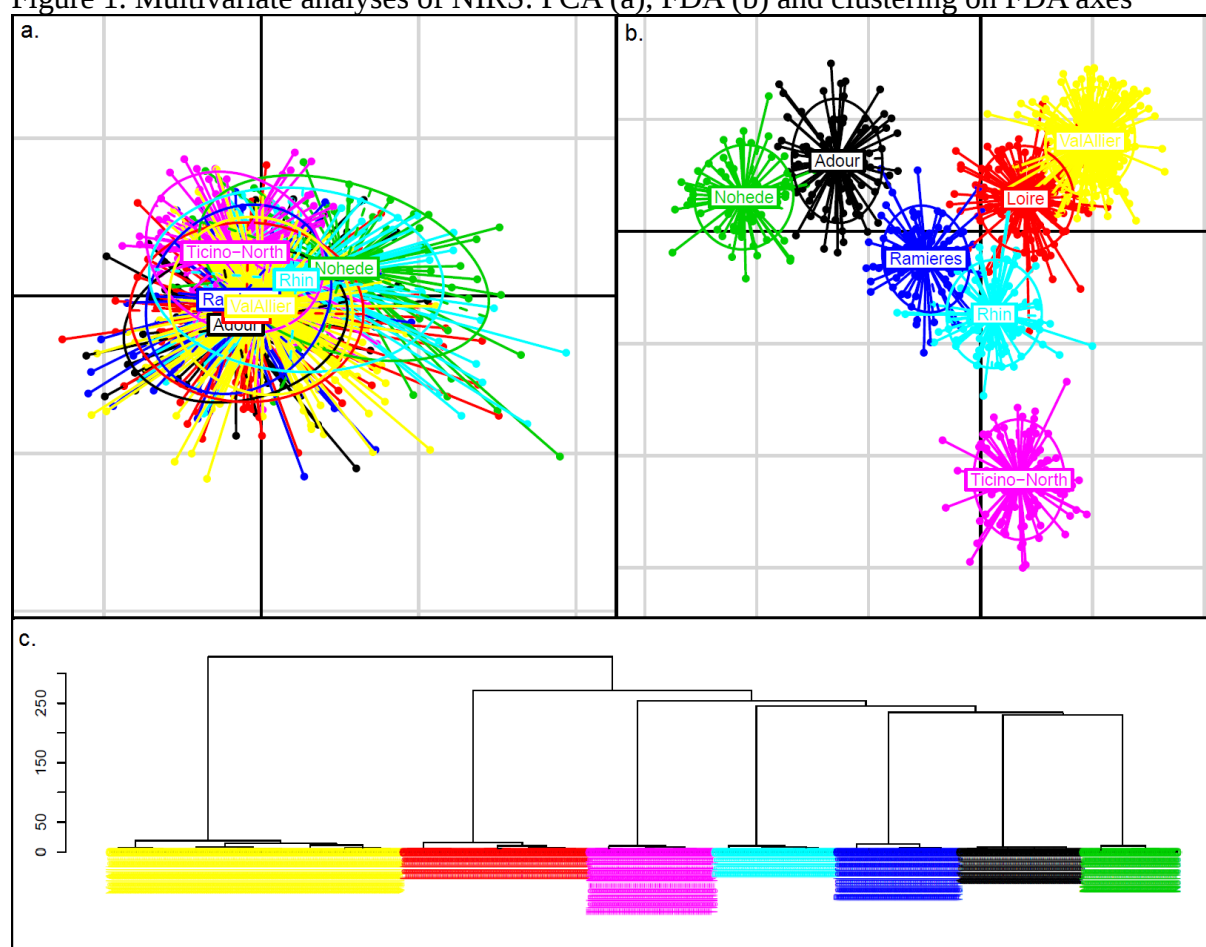Figure 1. Multivariate analyses of NIRS: PCA (a), FDA (b) and clustering on FDA axes

Figure 2. Partition of phenotypic variances along NIRS into metapopulation (red), clone nested into metapopulation (green) and residuals (blue). Corresponding NIRS (median and IQR) are also represented in light grey.
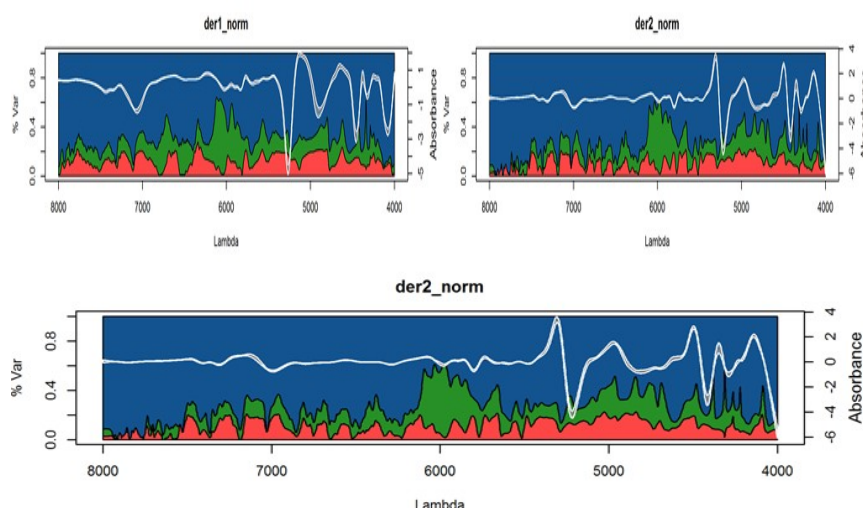

Table1. NIRS prediction accuracies (minimum / mean / maximum) for various quantitative traits. The column "Accuracy" gives the accuracy given a NIRS heritability of 1, while the column "Accuracy_H2_spec" account for NIRS estimated heritability (given in the column

"H2_spec_pls"). The broad sense heritability of the traits clonal mean of traits is indicated in the column ("H2_mean").

Trait	H2_mean	Accuracy	H2_spec_pls	Accuracy_H2_spec
HT08	0.50	0.49 / 0.56 / 0.63	0.13 / 0.45 / 0.56	0.67 / 0.88 / 1.36
HT09	0.72	0.67 / 0.76 / 0.82	0.16 / 0.43 / 0.56	1.02 / 1.19 / 1.68
CIR09	0.66	0.75 / 0.81 / 0.86	0.17 / 0.44 / 0.54	1.11 / 1.27 / 1.80
HT10	0.66	0.36 / 0.45 / 0.50	0.13 / 0.41 / 0.54	0.60 / 0.73 / 1.01
CIR10	0.66	0.41 / 0.51 / 0.56	0.17 / 0.42 / 0.52	0.64 / 0.81 / 1.13
HT11	0.77	0.43 / 0.53 / 0.58	0.16 / 0.43 / 0.54	0.65 / 0.85 / 1.35
CIR11	0.73	0.62 / 0.66 / 0.70	0.18 / 0.42 / 0.51	0.87 / 1.05 / 1.51
AGM09	0.67	0.26 / 0.47 / 0.54	0.19 / 0.46 / 0.53	0.60 / 0.70 / 0.74
AGM11	0.57	0.06 / 0.11 / 0.16	0.20 / 0.44 / 0.57	0.09 / 0.17 / 0.24
SYL09	0.71	0.28 / 0.40 / 0.49	0.23 / 0.40 / 0.47	0.59 / 0.64 / 0.75
SYL11	0.10	0.19 / 0.37 / 0.52	0.11 / 0.42 / 0.52	0.27 / 0.61 / 1.09
DEB09_104	0.94	0.58 / 0.64 / 0.66	0.10 / 0.45 / 0.56	0.84 / 1.05 / 2.05
DEB3_FITTED_DOY	0.97	0.61 / 0.65 / 0.67	0.23 / 0.49 / 0.59	0.84 / 0.96 / 1.40
BS09_254	0.80	0.35 / 0.47 / 0.55	0.19 / 0.43 / 0.54	0.51 / 0.75 / 1.09
DATE15_FITTED_DOY	0.81	0.43 / 0.47 / 0.53	0.21 / 0.43 / 0.53	0.60 / 0.74 / 0.96
DATE25_FITTED_DOY	0.55	0.14 / 0.30 / 0.45	0.22 / 0.43 / 0.50	0.30 / 0.45 / 0.64
NOTEMAX09	0.79	0.13 / 0.36 / 0.46	0.20 / 0.47 / 0.6	0.29 / 0.53 / 0.71
NOTEMAX10	0.82	0.36 / 0.46 / 0.55	0.14 / 0.44 / 0.55	0.50 / 0.74 / 1.19