Determination of Larch taxa with Near Infrared Spectroscopy

Vincent Segura[1], Jean-Paul Charpentier[1,2], Kévin Ader[1,2] ,Corinne Buret[1], Nathalie Boizot[1,2], Luc E. Pâques[1]

In the context of **forest tree breeding**, a strong **quality control** is required especially when the improved varieties derive from **interspecific hybridization**. This is the case of **Larch** whose improved varieties are **hybrids between European and Japanese species**. In this case, the quality control faces two challenges: (1) **verifying the parental species** in seed orchards, and (2) **attesting the hybrid status** of seedlings in nurseries. In both cases, a **cost effective high-throughput technique** is required to identify the **genetic origin** of the trees.

In this context, **near infrared spectroscopy (NIRS)** is a potentially interesting alternative to more traditional **barcoding** approaches based on **morphological** or **molecular markers** which are typically tedious and/or costly. The use of NIRS for barcoding relies on the assumption that some **chemical** and/or **physical properties** are able to **discriminate** the **genetic origin** of the analyzed samples (Cruickshank and Munck, 2011[a]). A recent study in Pine has successfully tested this hypothesis for leaf samples (Meder *et al.*, 2014[b]), underlining **the potential of NIRS for quality control in forest tree breeding** programs. The present study aims at evaluating this approach on a total of 350 leaf samples that have been collected over three consecutive years on European, Japanese and hybrid Larches that grow on the same site.

Samples used in the study

Year	N	Species		
		European Larch	Hybrid Larch	Japanese Larch
2011	36	16	13	7
2012	157	68	33	56
2013	157	67	34	56

Ability of NIRS to predict the genetic origin of samples

Partial least square discriminant analysis (PLS-DA) calibrations were carried out by year of sampling and the predictive ability of the models was assessed through a cross-validation repeated 1,000 times.

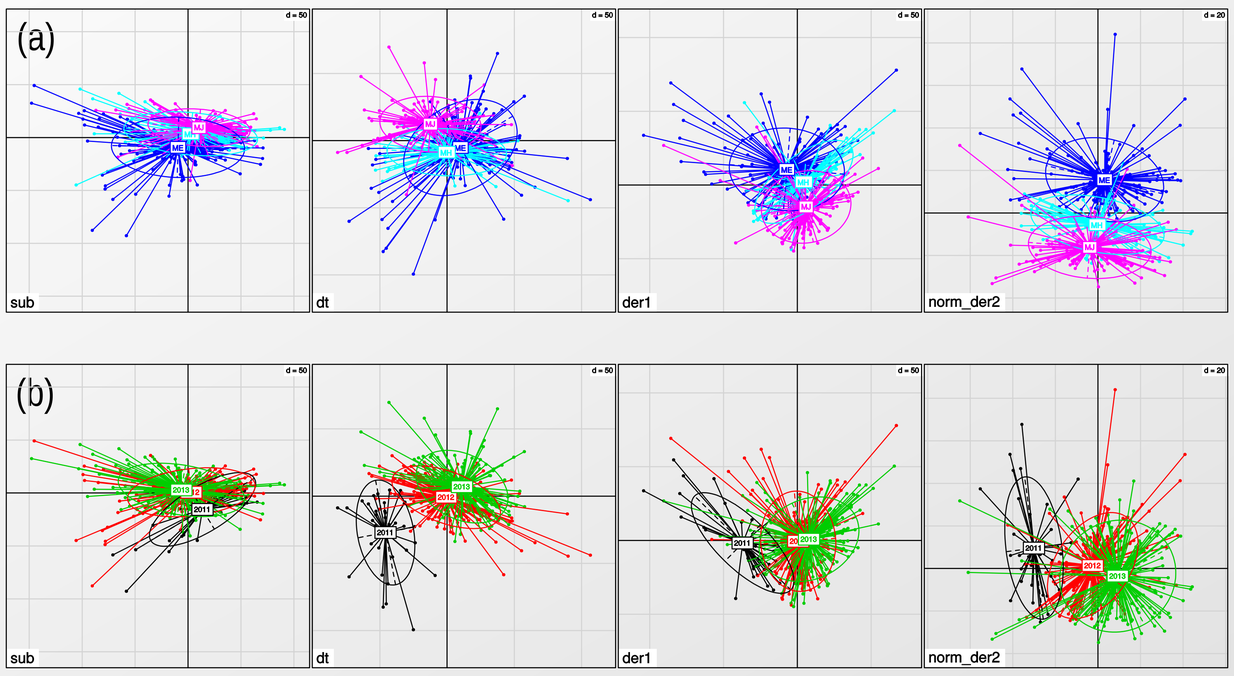
Year	N	Pretreatment	# LV	Accuracy	% Samples	
					≥ 0.95	< 0.80
2011	36	norm_der2	4	0.80	61.1	30.6
2012	157	der1	6	0.95	89.2	5.7
2013	157	der2	14	0.95	87.9	7.0

Validations across the sampling years

Accuracy of within-year PLS-DA models across sampling years.

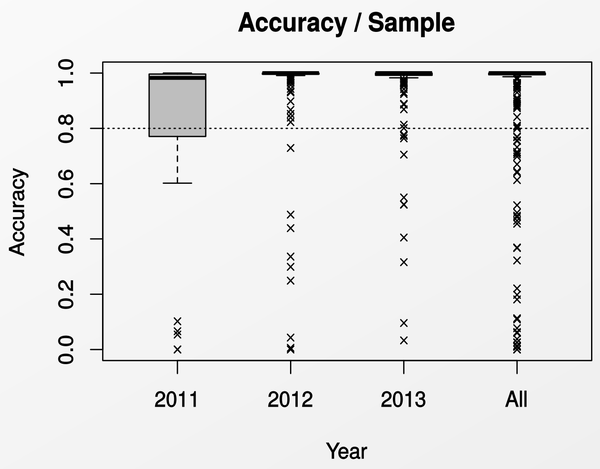
		Validation		
		2011	2012	2013
Calibration	2011		0.76	0.65
	2012	0.89		0.69
	2013	0.64	0.80	

A **principal component analysis** of the spectral dataset revealed some **grouping** according to the **species** (a) and **sampling year** (b), suggesting that **NIRS is able to capture both useful (genetic) and undesirable (year) information**. In addition, the ability of NIR spectra to capture the useful information varied depending on the statistical pretreatment of the spectra.



Accuracy of PLS-DA models by sample

Several samples were **systematically wrongly predicted** in the **cross-validation**, suggesting some mistakes in their genetic origin. Such **mistakes** have likely lead to a **downward bias** in our **prediction accuracy estimates**. Some molecular analyses using DNA barcoding are thus currently under progress in order to test this hypothesis and update our prediction accuracy.



PLS-DA calibrations across sampling years

Calibration (cross-validation)							Validation	
Years	N	Pretreatment	# LV	Accuracy	% Samples		N	Accuracy
					≥ 0.95	< 0.80		
2011-2012	129	der1	6	0.94	85.3	7.0	64	0.94
2012-2013	130	norm_der1	7	0.88	70.0	19.2	63	0.89
2011-2013	209	der2	14	0.92	80.4	12.9	105	0.90
2011-2012-2013	234	der1	11	0.91	81.6	11.5	116	0.88

Our study is the first to evaluate to **ability of NIRS to determine the taxa in Larch**. A recent study highlighted a similar approach in pine using leaf samples collected at a single time point within a particular trial (Meder *et al.*, 2014[b]). In comparison our study includes some **environmental variability** through **multiple harvests** (3 consecutive years) within a particular trial. Our results demonstrate that **this variability affects the predictive ability of the models** and should thus be accounted for in order to **increase the robustness** of the model, a prerequisite to routinely use them in a breeding program.

References

[a]Cruickshank RH and Munck L (2011) *Zootaxa* 2933: 55-56 ; [b]Meder R *et al.* (2014) *JNIRS* 22: 337-345

Acknowledgements

We are grateful to the French ministry of food agriculture and forest for funding the study through the program “DGPAAT Evaluation des variétés forestières”.

Author details

[1] INRA, Centre de recherche Val de Loire, UR0588 AGPF

[2] INRA, Plateforme Régionale GénoBois, Orléans, France

* **vincent.segura@orleans.inra.fr**