Supplementary Materials: Green and Roasted Beans of Nicaraguan Coffea Arabica Varie-ties Processed with Different Post-Harvest practices

Gaia Meoni, Claudio Luchinat, Enrico Gotti, Alejandro Cadena and Leonardo Tenori

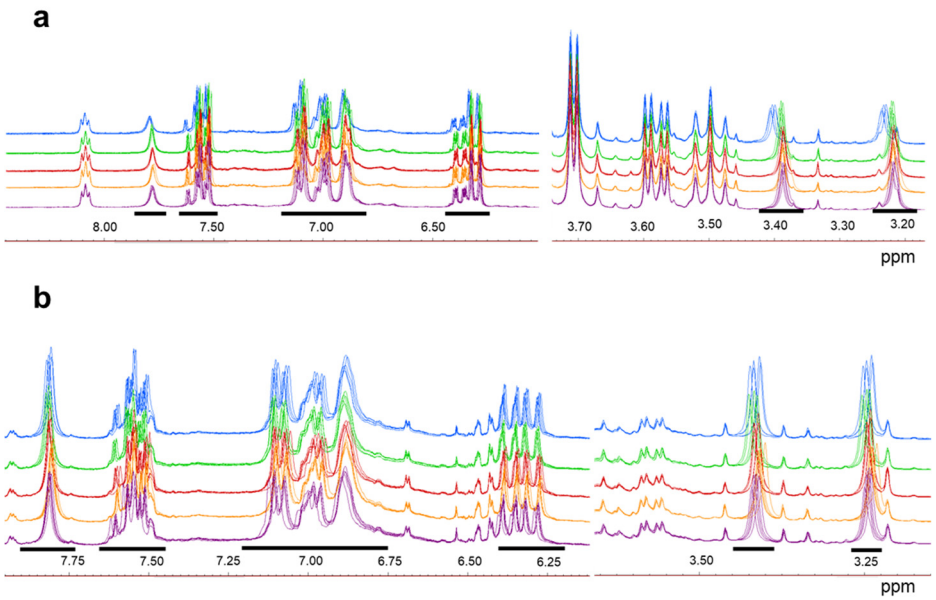

Figure S1. ^{1}H NMR spectra of coffee beans. a) green coffee beans; b) roasted coffee beans. Peaks underlined correspond to the resonance regions of caffeine and chlorogenic acid that, compared to most of the other signals present in coffee spectra, tend to shift a lot among the different spectra. These regions correspond to the variable size buckets used for the statistical analyses.

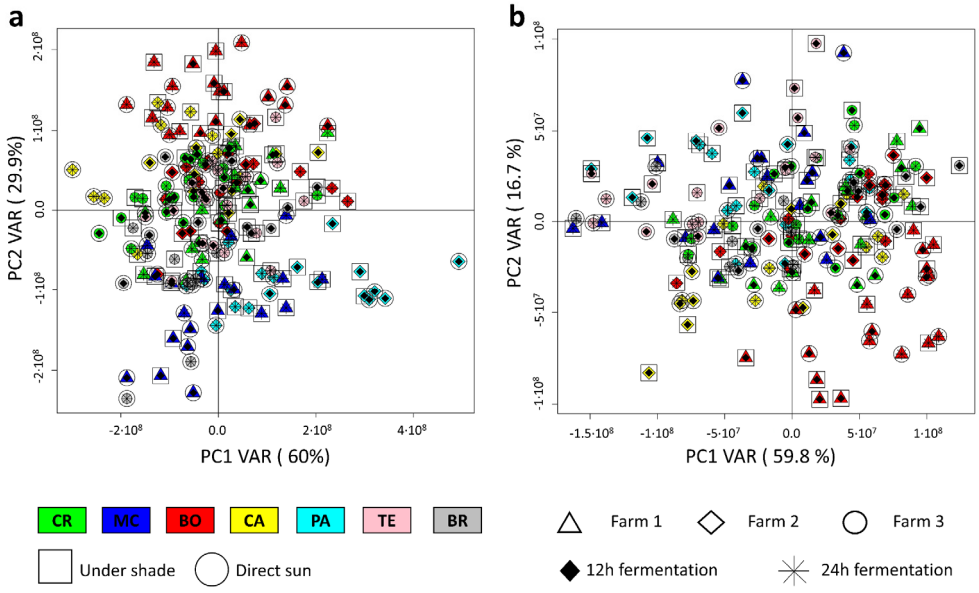

Figure S2. First two components of PCA score plots of ^{1}H NMR bucketed spectra of green (a) and roasted (b) coffee beans. Each point represents a spectrum: different colors indicate the seven varieties, symbols represent the three farms, symbols outside the colored points represents the two types of drying procedures (under shade and direct sun), and

symbols inside the colored points represent the two time points of fermentation (12h and 24) for each batch of samples. The percentage of variance for the two principal components is reported on the axes.

Table S1. Variety classification within the three farms through RF. For each cultivar is reported the class error resulted for the different RF models attempted. RF models have been created both for green and roasted coffee beans, starting from different input data: the matrices of bucketed spectra (Bucket.) and the matrices of quantified metabolites (Metab.). The overall predictive accuracy is reported for each model (overall pred.acc).

			Error for each class (%)							
Farm	**coffee**	**RF model**	**CR**	**MC**	**BO**	**CA**	**PA**	**TE**	**BR**	**overall pred.acc (%)**
1	**green**	**Bucket.**	0	0	5	/	/	/	/	98.3
		Metab.	5	5	0	/	/	/	/	96.7
	roasted	**Bucket.**	5	5	0	/	/	/	/	96.7
		Metab.	20	5	0	/	/	/	/	91.7
2	**green**	**Bucket.**	/	/	15	5	0	/	/	93.3
		Metab.	/	/	0	5	5	/	/	96.7
	roasted	**Bucket.**	/	/	0	0	0	/	/	100
		Metab.	/	/	25	20	0	/	/	85
3	**green**	**Bucket.**	10	/	/	/	/	20	10	86.7
		Metab.	5	/	/	/	/	10	15	90
	roasted	**Bucket.**	10	/	/	/	/	16	5	89.7
		Metab.	20	/	/	/	/	16	37	75.9

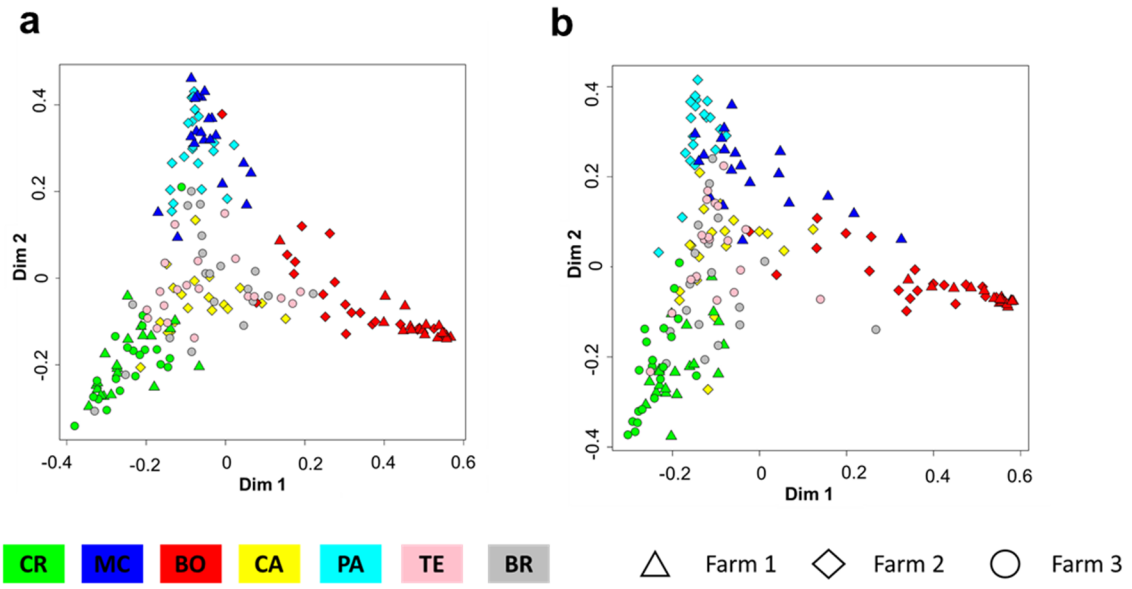

Figure S3. Random Forest multidimensional scaling (MDS) plots. a) RF plot of bucketed green coffee beans spectra; b) RF plot of bucketed roasted coffee beans spectra. Each point represents a bucketed spectrum: different colors indicate the seven cultivars; symbols represent the three farms. The overall predictive accuracy of RF model "a" is 87.2% and the overall predictive accuracy of RF model "b" is 86%.

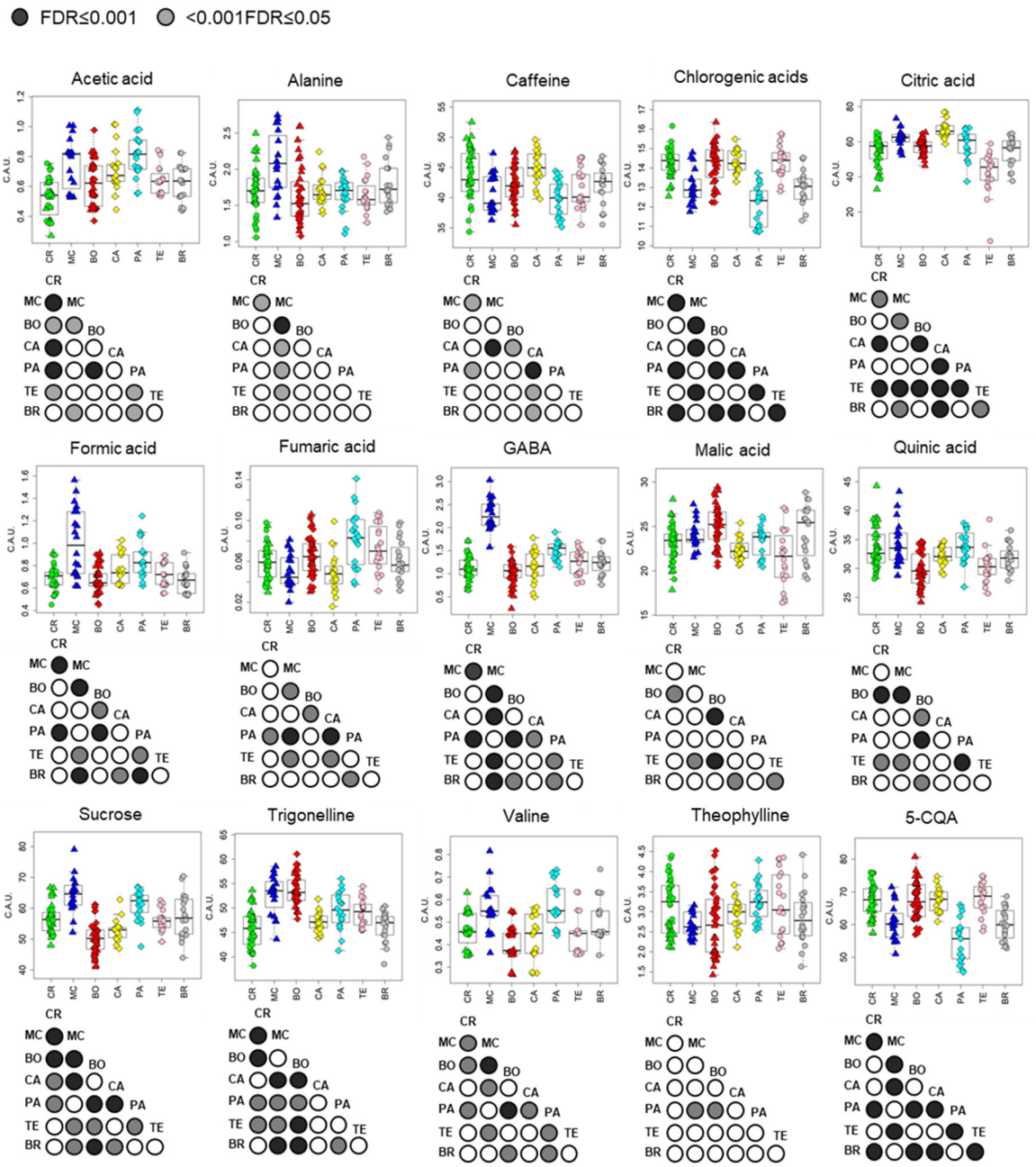

Figure S4. Univariate analysis on the identified metabolites: cultivars' comparison (green coffee beans) in. A boxplot panel is reported for each metabolite showing at least two statistically significant P-values among all the comparisons. Below each boxplot is reported a triangular matrix to represent the FDR p-value assessed through Kruskal-Wallis test followed by Dunn post-hoc analysis. FDR p-value ≤0.05 is considered to infer significant differences among independent samples from the seven distinct cultivars (catuai rojo, CR, green; maracaturra, MC, blue; bourbon, BO, red; caturra, CA, yellow; pacamara, PA, cyan; tekesic, TE, pink; boubon rojo, BR, gray)

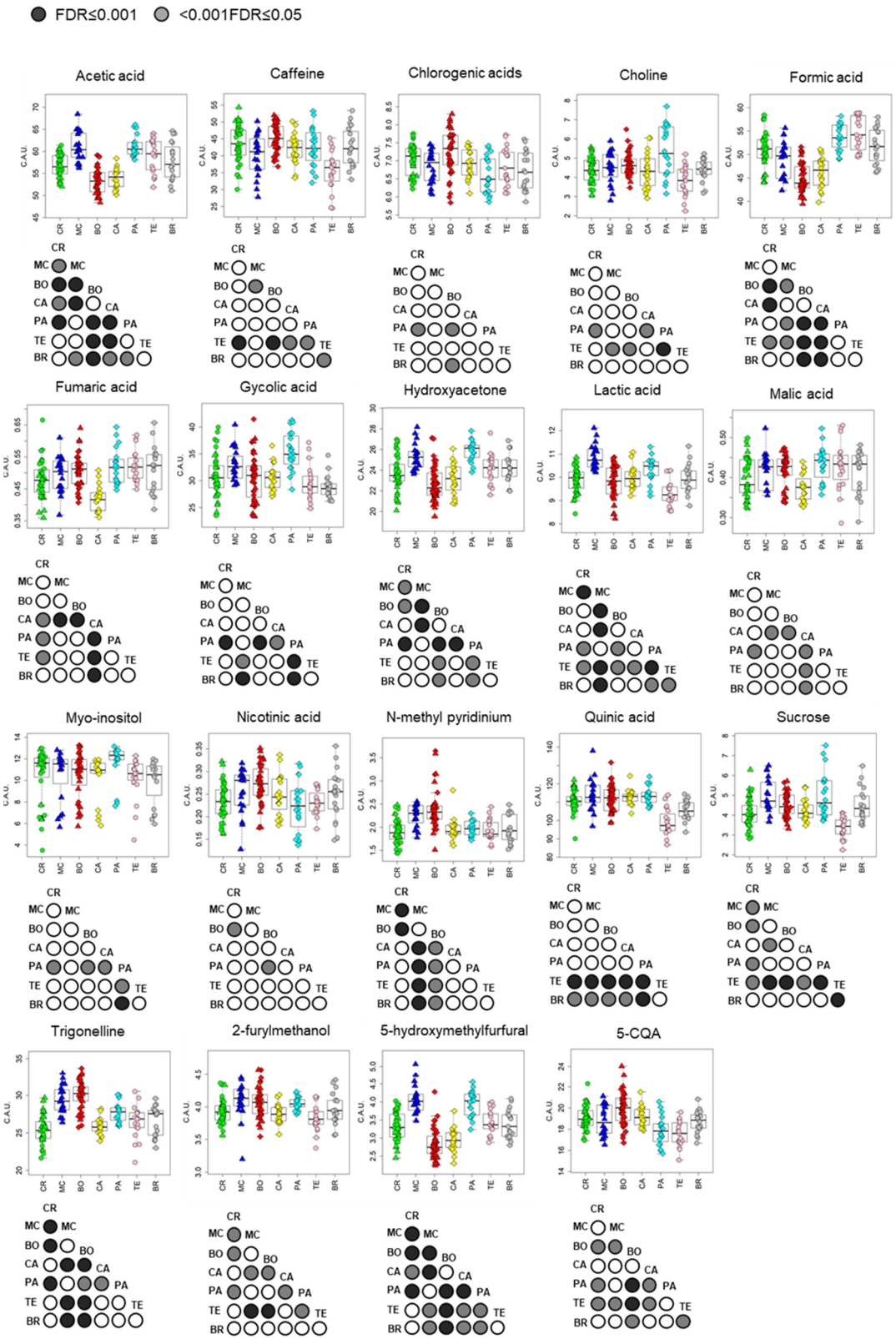

Figure S5. Univariate analysis on the identified metabolites: cultivars' comparison (roasted coffee beans) in. A boxplot panel is reported for each metabolite showing at least two statistically significant P-values among all the comparisons. Below each boxplot is reported a triangular matrix to represent the FDR p-value assessed through Kruskal-Wallis test followed by Dunn post-hoc analysis. FDR p-value ≤0.05 is considered to infer significant differences among independent samples from the seven distinct cultivars (catuai rojo, CR, green; maracaturra, MC, blue; bourbon, BO, red; caturra, CA, yellow; pacamara, PA, cyan; tekesic, TE, pink; boubon rojo, BR, gray)

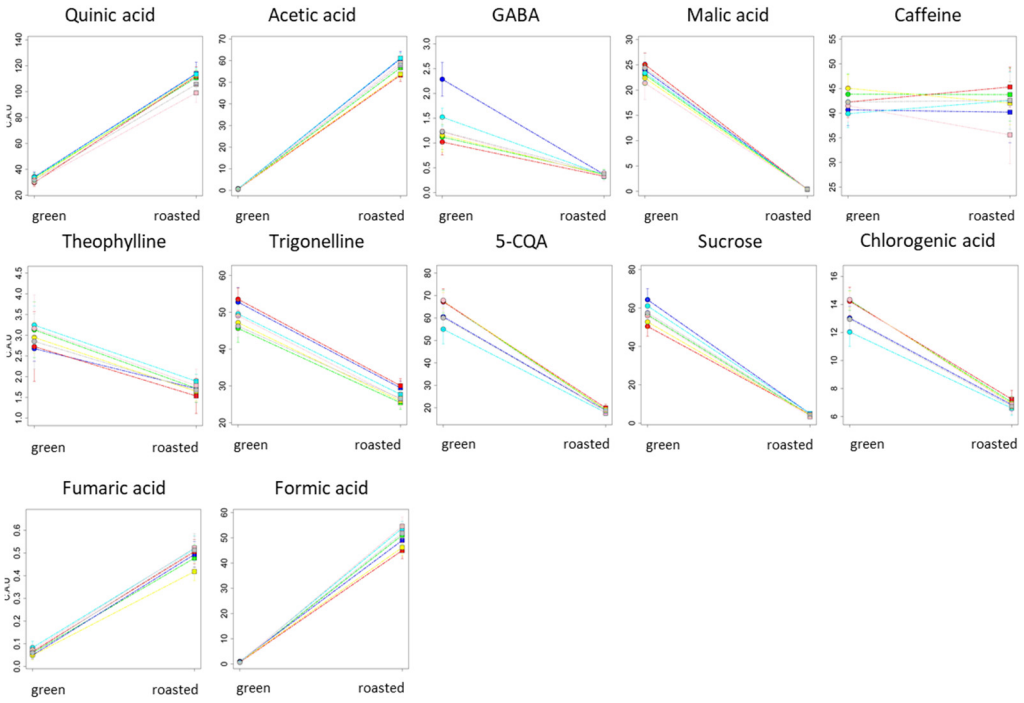

Figure S6. Roasting effect on coffee beans' metabolites. Concentrations of metabolites are expressed in arbitrary units, referring to the area under the identified peaks (C.A.U).

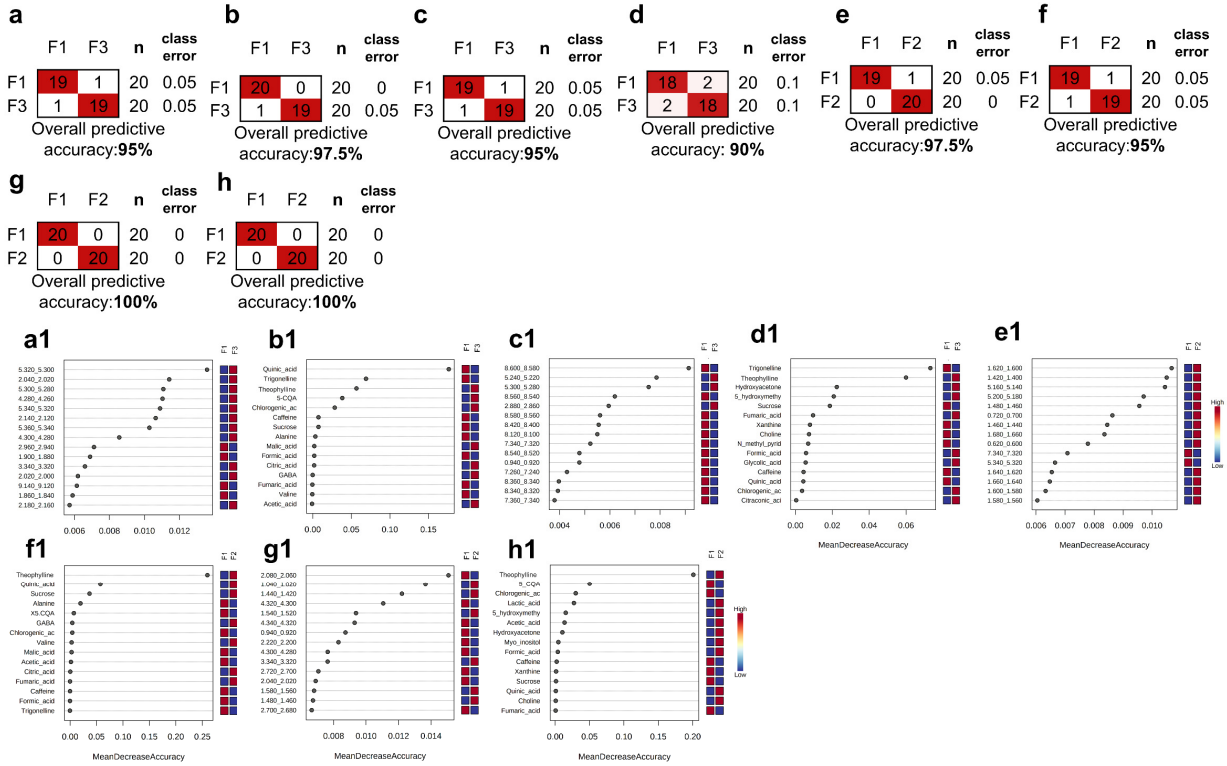

Figure S7. Farm's fingerprint and profiling assessment through RF. a-h) Confusion matrices of RF algorithm of green (a, b, e, f) and roasted (c, d, g, h) coffee beans' spectra. Catuai rojo batches are used to compare the products of farm1 and farm 3 (a-d) and bourbon batches are used to compare farm 1 and farm 2 (e-h). Summary of the variable importance measures for the buckets (a1, c1, e1, h1) and metabolites (b1, d1, e1, h1,) of coffee NMR bucketed spectra (a, c, e, g)/metabolites (b, d, f ,h) with farm as the response variable, are reported and labeled as 1 with the same letter of the corresponding RF model (*es.* the summary of the variable importance for model "a" is the "a1" picture, *etc.*)

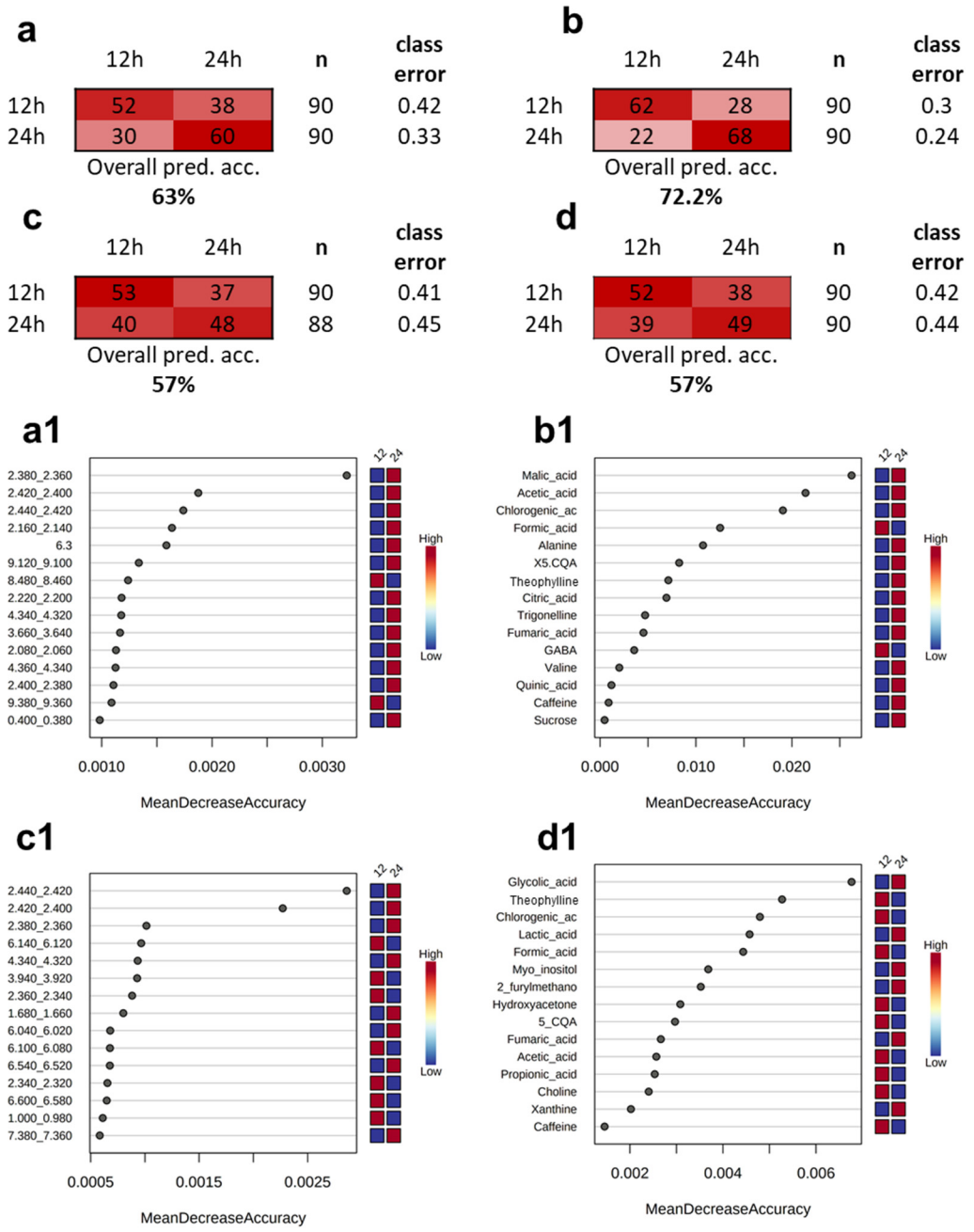

Figure S8. RF models built on green and roasted NMR bucketed spectra and metabolites: beans fermented 12 hours vs beans fermented 24 hours. Models discriminating coffee batches according to the fermentation time are built using bucketed spectra (panel "a", for green coffee, and panel "c" for roasted coffee) and identified metabolites (panel "b", for green coffee, and panel "d" for roasted coffee). A summary of the variable importance measures for the buckets (panels "a1" and "c1") and metabolites ("b1" and "d1") of coffee NMR spectra with fermentation time as the response variable are

reported and labeled as 1 with the same letter of the corresponding RF model.

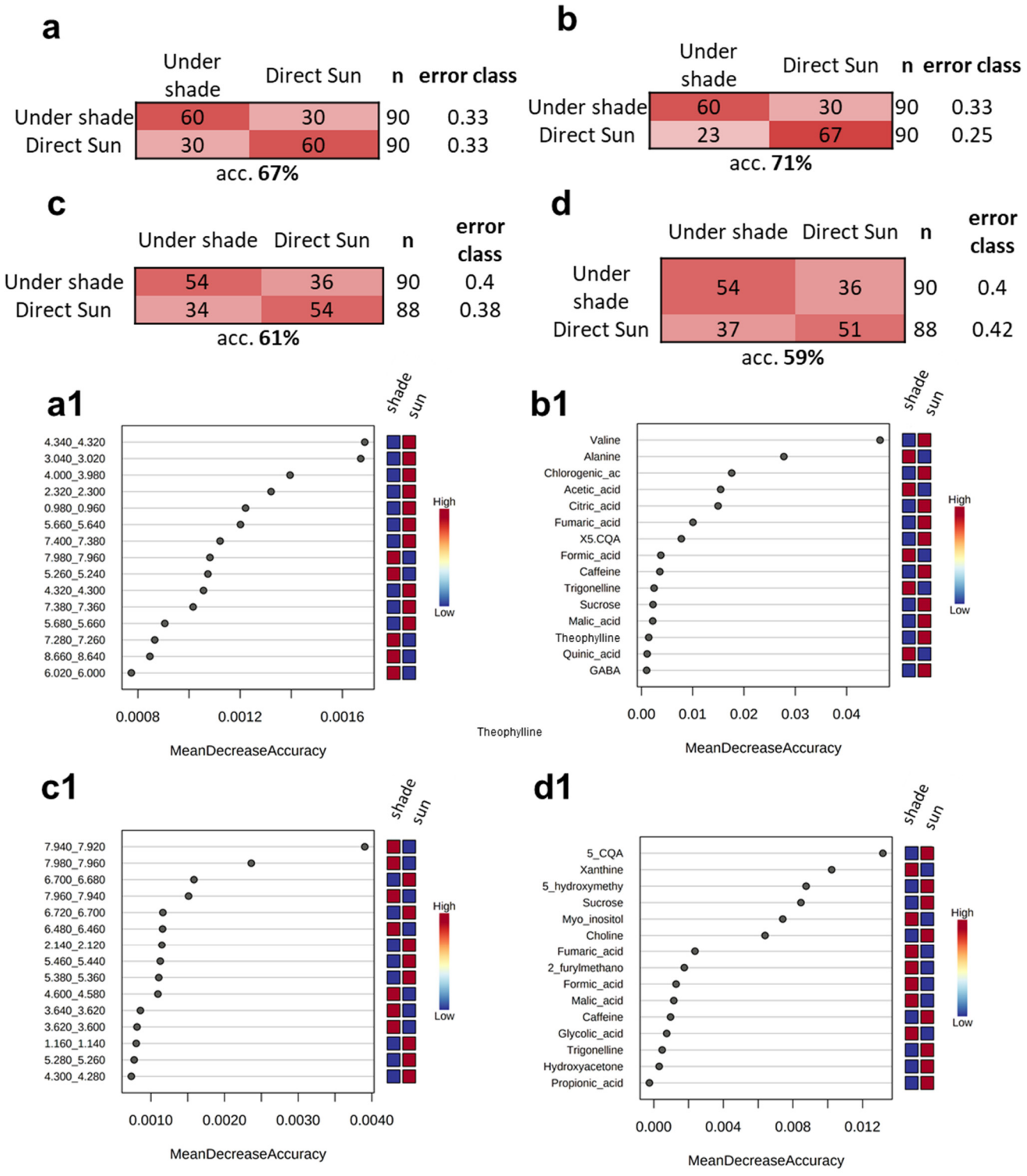

Figure S9. RF models built on green and roasted NMR bucketed spectra and metabolites: beans dried under shade vs beans dried at direct sun Models discriminating coffee batches according to the different drying procedure are built using bucketed spectra (panel "a", for green coffee, and panel "c" for roasted coffee) and identified metabolites (panel "b", for green coffee, and panel "d" for roasted coffee). A summary of the variable importance measures for the buckets (panels "a1" and "c1") and metabolites ("b1" and "d1") of coffee NMR spectra with the drying procedure as the response variable are reported and labeled as 1 with the same letter of the corresponding RF model.