Precision phenotyping using FTIR for expensive-to-measure traits

F. Tiezzi[1], H. Toledo-Alvarado[2], G. Bittante[3], A. Cecchinato[3]

[1]North Carolina State University, Department of Animal Science, Raleigh NC, USA,

[2]Universidad Nacional Autónoma de México, Departamento de Genética y Bioestadística, Ciudad Universitaria, Ciudad de México, Mexico,

[3]University of Padova, Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE)

70th Annual Meeting of the European Federation of Animal Science
City of Ghent (Belgium), 26 - 30 Aug 2019

Aims

To assess the effectiveness of large-scale FTIR phenotyping using two approaches:

1. estimating the additive genetic correlations between LAB-measures and FIELD-FTIR-predictions

2. evaluating the predictive ability of FTIR-derived phenotyping using different genotyping strategies

Data structure

LAB measures

Phenotypes
MCP
CY and RECs
Protein Fractions
Fatty Acids
Minerals
etc. ~ 90 novel phenotypes

~ 1,300 cows

$$y_i = \beta_0 + \sum_{j=1}^{1,060} x_{ij}\beta_j + \varepsilon_i$$

245 sires in common (LAB and FIELD)

FIELD predictions

Predictions based on calibration equation previously defined (Ferragina et al., 2015; Cecchinato et al., 2015)

~720,000 FTIR prediction records
~51,000 cows

Genetic analysis

LAB and FIELD as 2 distinct traits

$$y_{LAB} = PARITY_i + DIM_j + HERD_k + addgen_l + e_{ijklm}$$

$$y_{FIELD} = PARITYDIM_i + HTD_j + cowpe_k + addgen_l + e_{ijklm}$$

Assumptions:

- $\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \sim N(0, \mathbf{GA})$
- Residuals uncorrelated

REMLF90 and AIREMLF90

Results

Traits	h^2 LAB (n = 1,200)	h^2 FIELD (n = 700,000)	r_g	R^2_{CV}
MCP (4 'standard' MCP, 4 Optigraph, 6 new modeling parameters)	~0.24	~0.32	~0.77	~0.54
CY - RECs (3 % of CYs and 4 nutrients recovery in curd)	~0.25	~0.30	~0.85	~0.69
Protein profile (6 fractions: αS1-CN, αS2-CN, β-CN, k-CN, β-lactoglobulin (LG), and α-lactalbumin (LA) + total casein, protein and whey)	~0.44	~0.22	~0.64	~0.46

30 traits in total
r_g additive genetic correlation between LAB and FIELD
R^2_{CV} estimated on LAB dataset using Bayes-B with the BGLR library on R

Genomic analysis

LAB and FIELD as 2 different traits (same model)

Single-Step GBLUP method

The pool of genotyped individuals consisted of:

- 1,011 LAB cows (genotyped at 50k)
- 1,463 FIELD cows (genotyped mostly at LD and imputed to 50k)
- 181 sires with both LAB and FIELD daughters (genotyped at 50k or HD and imputed to 50k)
- 540 sires with FIELD daughters only genotyped at 50k or HD and imputed to 50k.
- The software fImpute was used for imputation (Sargolzaei and Schenkel, 2013)

Mat&Met / 2

FIELD.t 75% of records

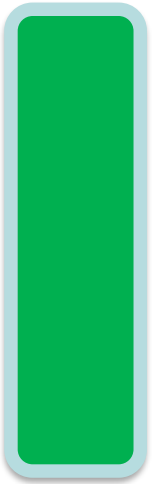
LAB.t 75% of records

FIELD.v 25% of records

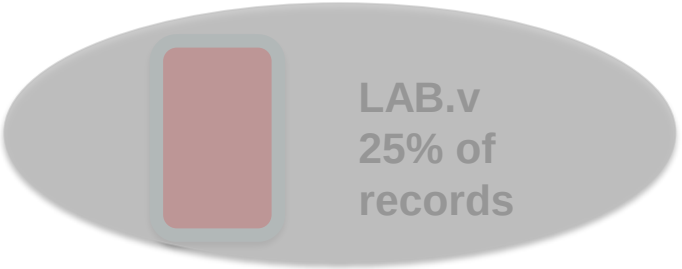
LAB.v 25% of records

Mat&Met / 2

FIELD data is not recorded

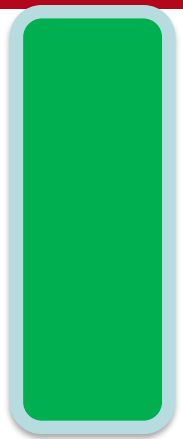


LAB.t
75% of records



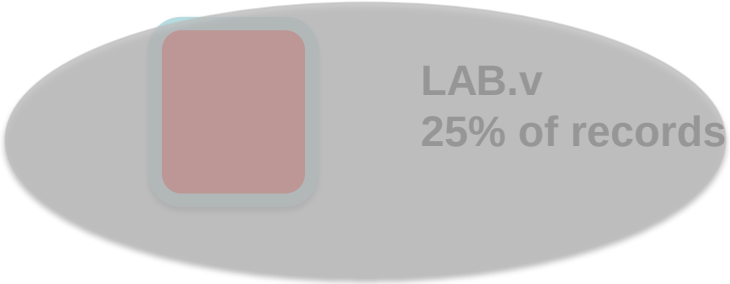
Mat&Met / 2

FIELD.t
75% of records



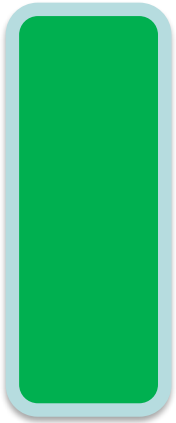
LAB.t
75% of records

FIELD is recorded, but only for proven bulls

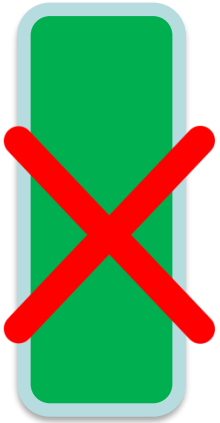


Mat&Met / 2

FIELD.t
75% of records

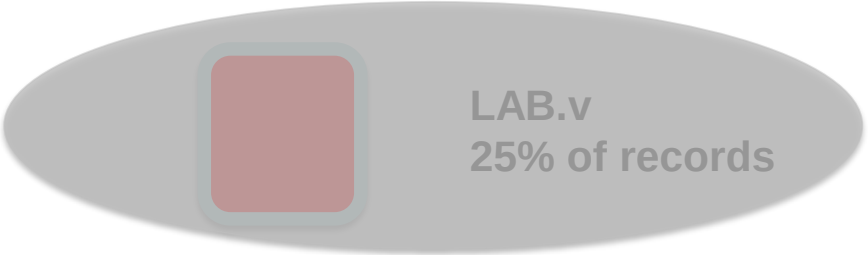


FIELD is recorded for proven and progeny testing bulls



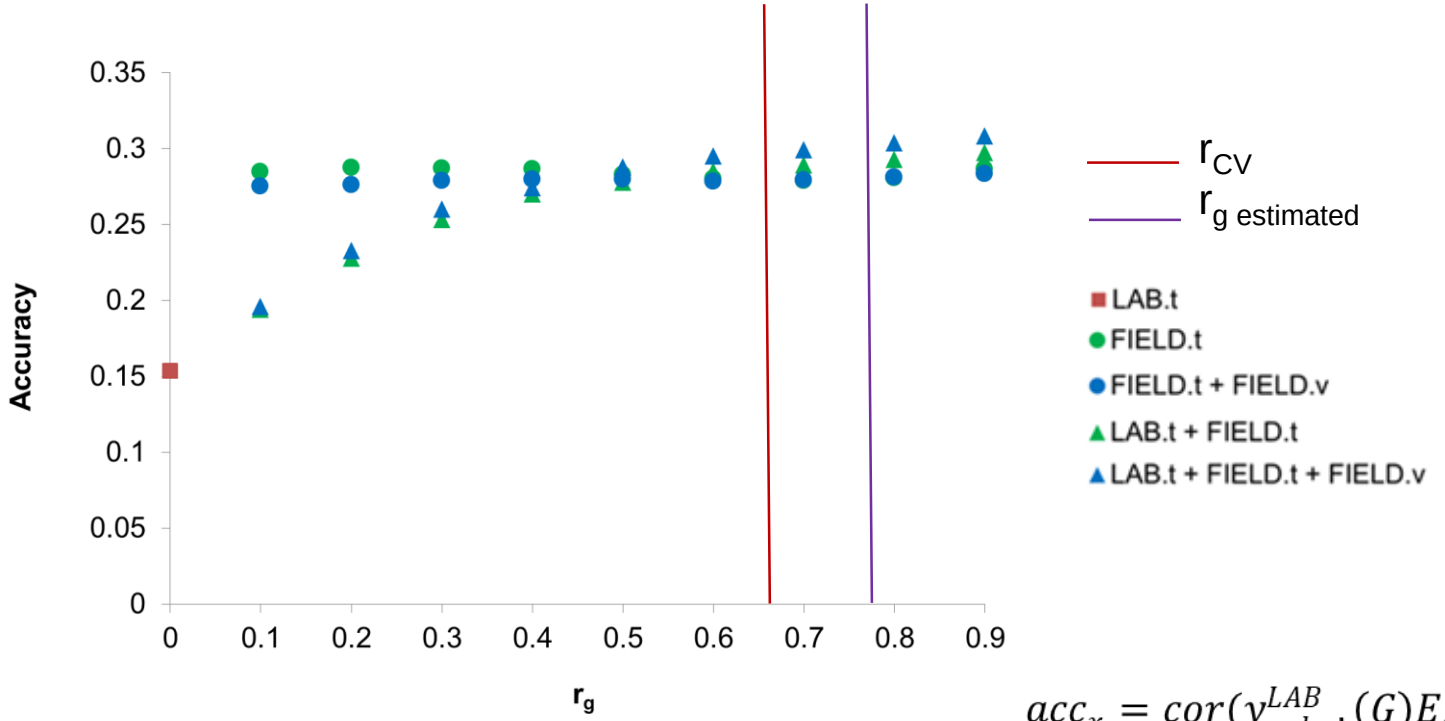
LAB.t
75% of records

FIELD.v
25% of records



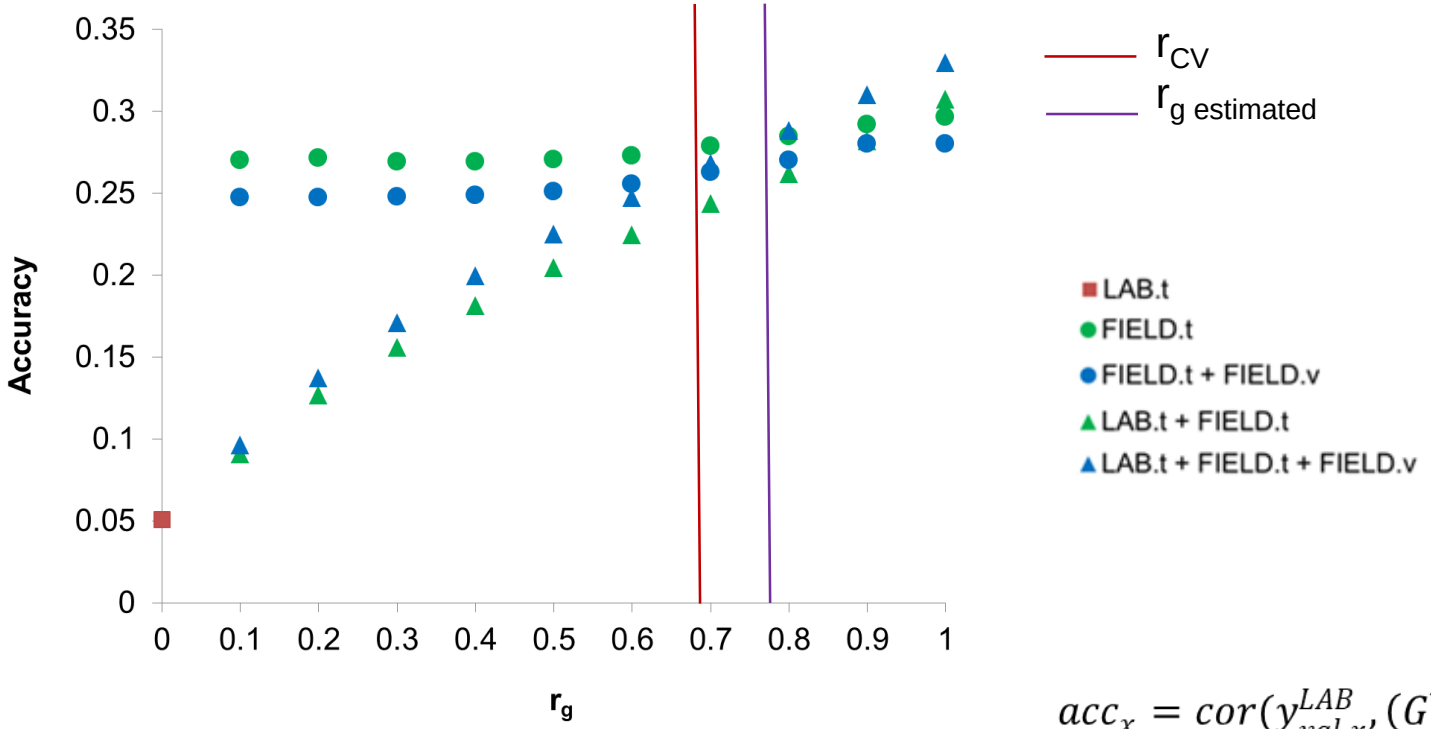
LAB.v
25% of records

Results: β-casein



$$acc_x = cor(y^{LAB}_{val.x}, (G)EBV^{LAB}_{val.x})$$

Results: K-casein



$$acc_x = cor(y_{val.x}^{LAB}, (G)EBV_{val.x}^{LAB})$$

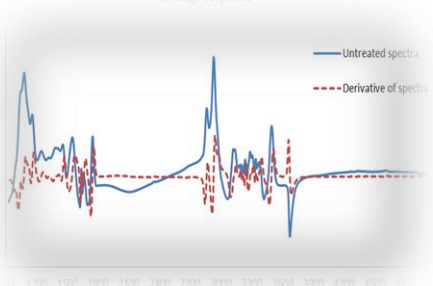
Conclusions

- For genetic purposes it is important having relatively large FTIR (LAB) calibration sets
- The r_g between LAB and FIELD could be an important criterion to assess the robustness of calibrations equations
- In terms of prediction, results demonstrated the relevance of FIELD measures, especially when coming from progeny of proven bulls
- Collection of FIELD measures of progeny testing bulls appears to be unnecessary for the traits studied
- As there is no evidence of an advantage of including FIELD information for progeny testing bulls, progeny testing schemes could be replaced by a more vast and robust collection of FIELD measures on daughters of proven bulls

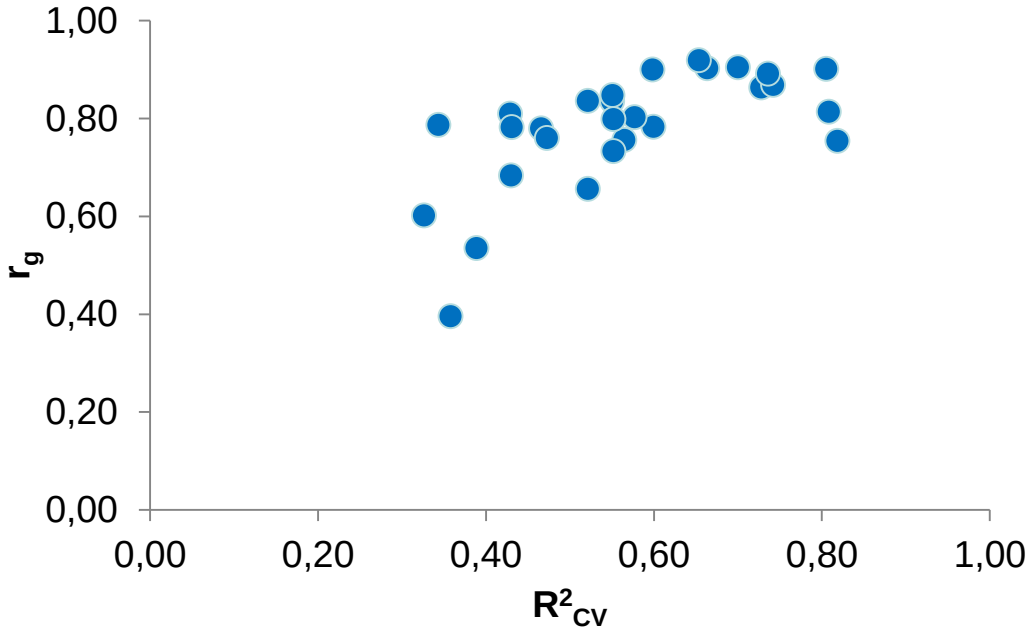
Thank you for your attention!

Rationale

- FTIR-based phenotyping is a compelling choice for genetic improvement of milk quality traits
- Genomic selection (GS) has revolutionized animal breeding, especially in dairy cattle
- Validation strategies??



Results - r_g e R^2_{CV}



Models / 1

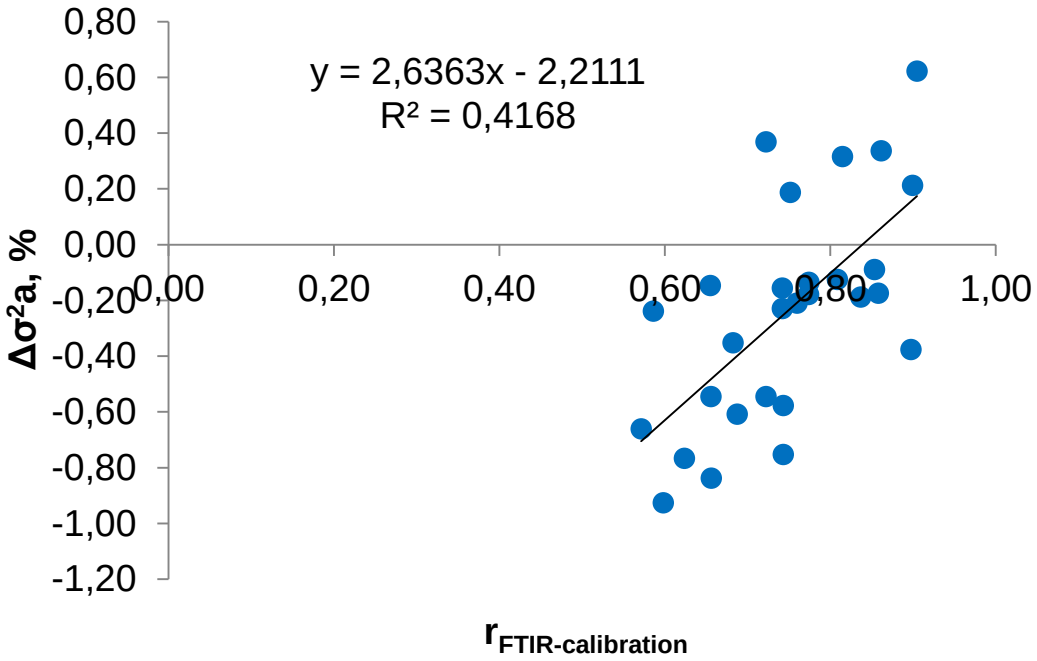
Spectra editing of the population

•The spectrum T transformed to A = log(1/T)

•PCA on the FTIR spectra with Mahalanobis distances

•Samples with a probability level <0.01 were considered to be outliers and removed from the data set

•Absorbance values for every wave centered to a null mean and standardized to a unit sample variance within year periods

•For each trait a BayesB regression was applied, with 100,000 iterations and 20,000 chains discarded as burn in

Results - Δσ²a e r FTIR -calibration