Using machine learning to predict feed intakes of meat sheep from animal traits and ruminal microbiota

Q. Le Graverand, C. Marie-Etancelin, J.L. Weisbecker, A. Meynadier, D. Marcon, F. Tortereau

Context

Why predict feed intake ?

Necessity for complex traits (e.g. feed efficiency)

Environmental/societal/economic stakes

Only two breeding sheep compagnies record feed intakes in France

Study objectives:

Check:

- The accuracy of feed intake predictions from microbiota data
  / host traits

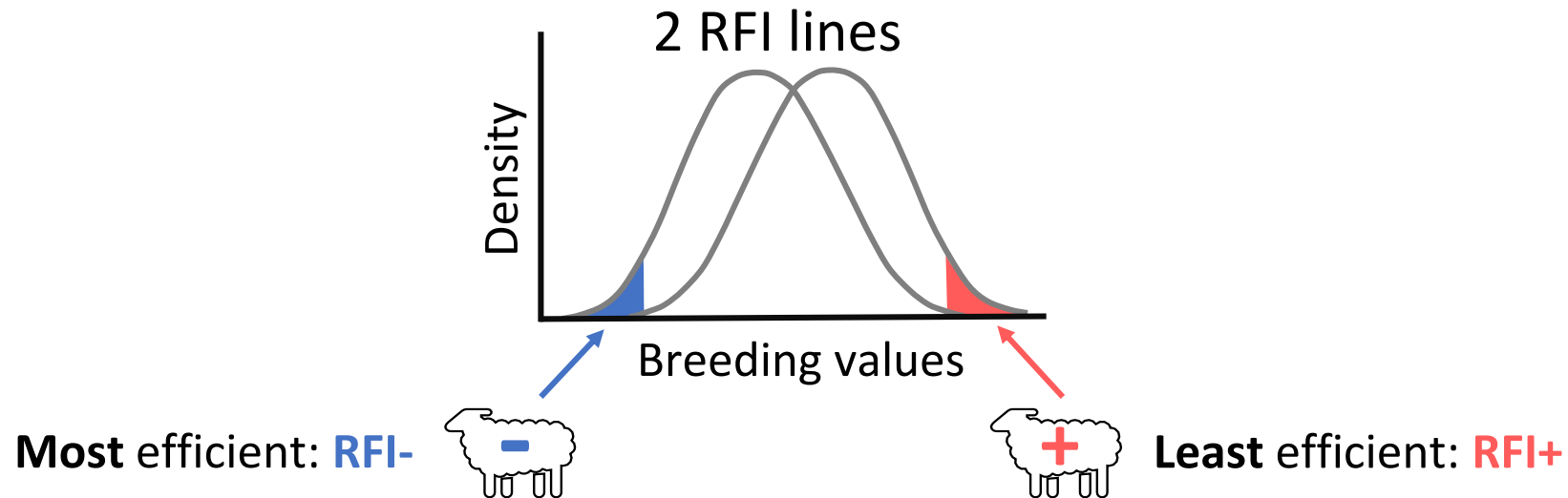- The relevance of predictions for the genetic evaluation

Automatic concentrate feeder

Study population – Residual Feed Intake lines

277 Romane ♂ lambs of two RFI divergent lines (G2 & G3)

Regression: $\text{Feed intake} = \mu + \text{Fixed effects} + \beta_1\,\text{ADG} + \beta_2\,\text{fBW}^{0.75} + \beta_3\,\text{MD} + \beta_4\,\text{BFT} + \underbrace{\varepsilon}_{\textbf{RFI}}$

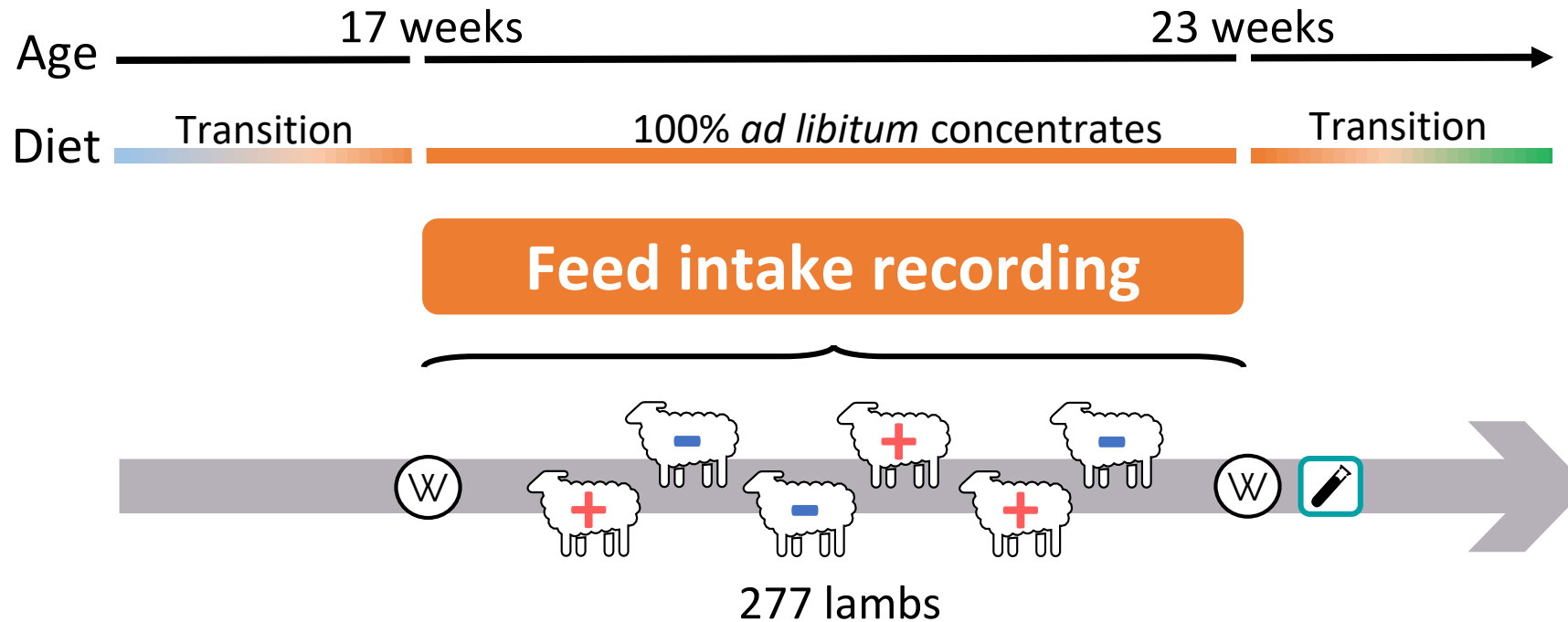Divergent selection to get the study population:

Most efficient: **RFI-**   **Least** efficient: **RFI+**

For the G2 & G3:   divergence of 1.9 $\sigma_{genetic\ RFI}$

$\mu_{\text{RFI-, feed intake}} = 2{,}043\ \text{g/d}$ $\mu_{\text{RFI+, feed intake}} = 2{,}155\ \text{g/d}$

Experimental protocol

I - Accuracy of feed intake predictions from microbiota data and host traits

Statistical approach to predict feed intake

Three sets of predictors

Gold standard

Host traits
Weights (at 145d, start, end)
Average Daily Gain
Muscle Depth
Fat Thickness

(6 variables)

Rumen microbiota

16S (bacteria + archaea)

496 OTUs

(496 variables)

Host traits
&
Rumen microbiota

(592 variables)

Three machine learning approaches

sPLSR: sparse Partial Least Squares Regression *(R, mixOmics)*

SVR: Support Vector Regression *(R, e1071)*

RFR: Random Forest Regression *(R, caret + randomForest)*

K-fold cross-validations nested in leave-one-group out cross-validations

With 277 lambs raised over 3 years

Predicting feed intake of an independent cohort:

Two different years   Third year

1 **Training set**   **2** **Testing set**

Tuning through repeated 5-fold cross-validations   Prediction of feed intake

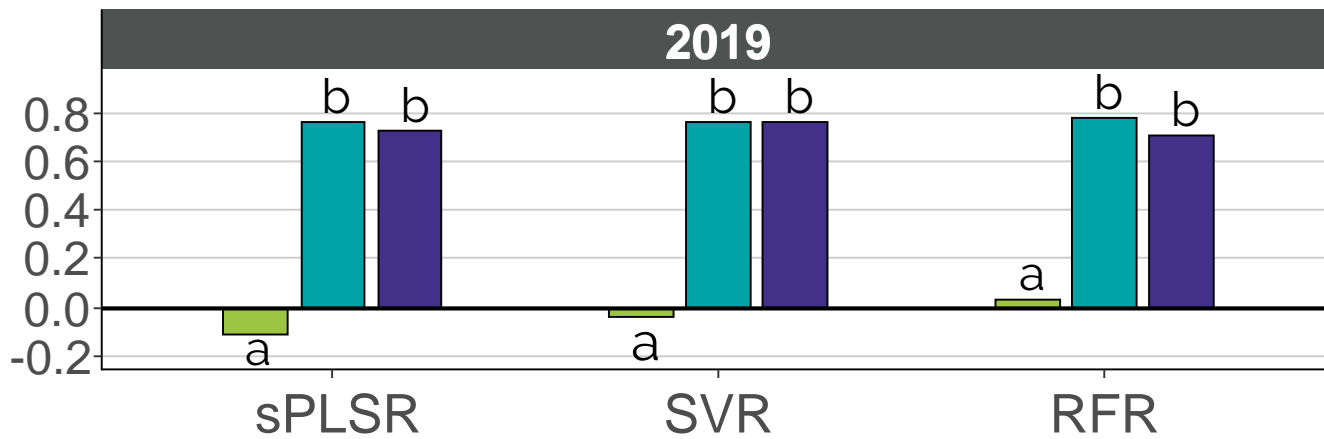Model fitting   Correlation between predictions & real

                                                       phenotypes

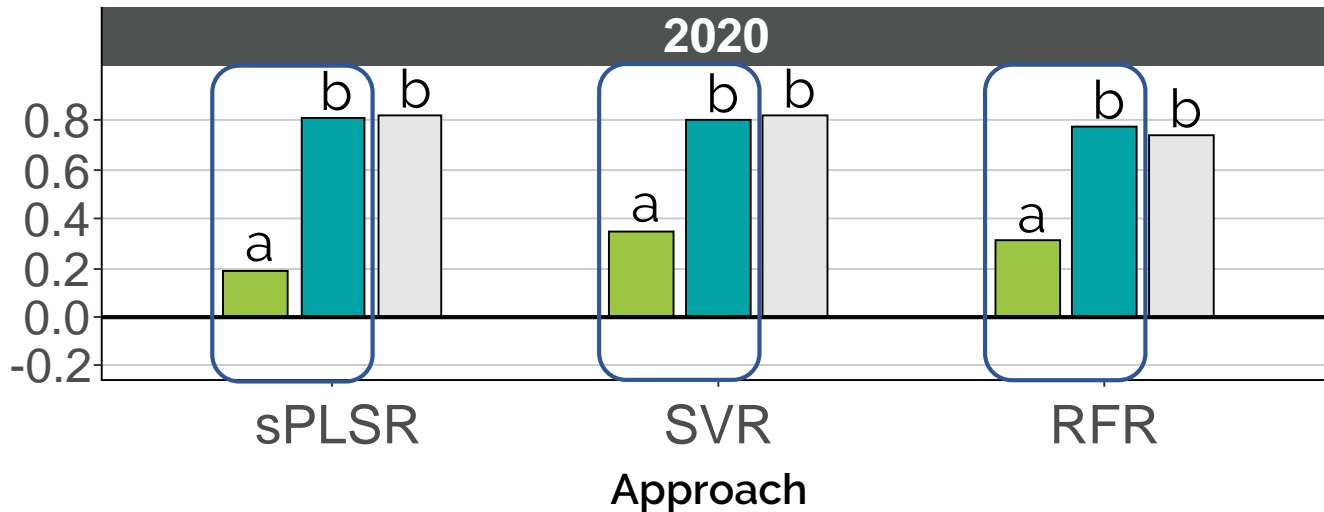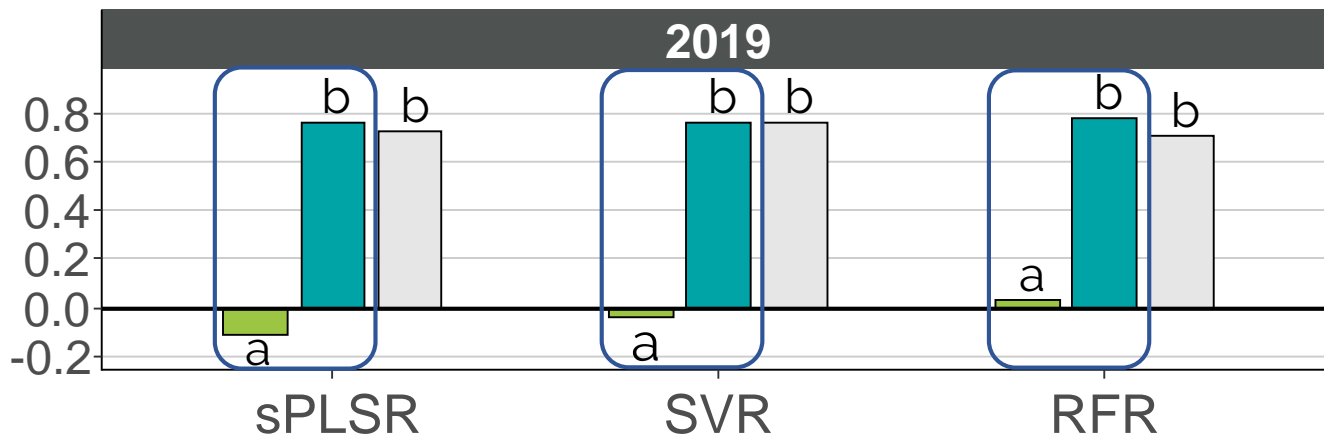# Pearson correlation between real feed intake and predictions

Predictor set: ■ 16S ■ Traits ■ Traits+16S

- Note: 2020 lambs' sires are part of 2018
  → 2018 is not used as a testing set

a,b Comparisons with Dunn & Clark's z test (adjusted p-value <0,05)

Pearson correlation between real feed intake and predictions

Pearson correlation between real feed intake and predictions

Predictor set: ☐ 16S ■ Traits ■ Traits+16S

- Note: 2019 lambs' sires are part of 2018 → 2018 is not used as a testing set

- Higher correlations with host traits predictors than 16S data

- Combining 16S data and traits does not improve correlations

a,b Comparisons with Dunn & Clark's z test (adjusted p-value <0,05)

Pearson correlation between real feed intake and predictions

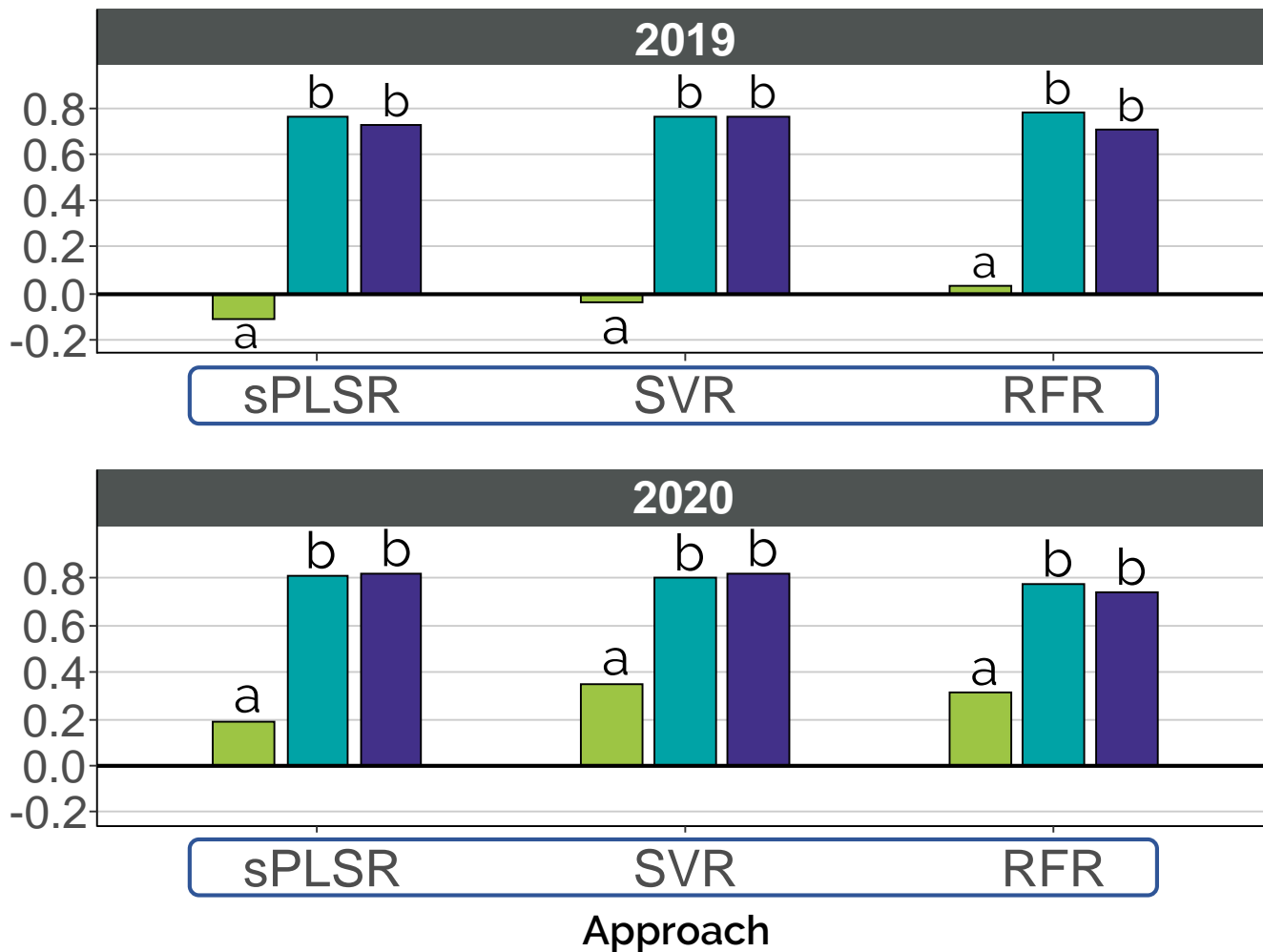Predictor set: ■ 16S ■ Traits ■ Traits+16S

- Note: 2020 lambs' sires are part of 2018 → 2018 is not used as a testing set

- Higher correlations with host traits predictors than 16S data

- Combining 16S data and traits does not improve correlations

- No difference between approaches

[a,b] Comparisons with Dunn & Clark's z test (adjusted p-value <0,05)

Pearson correlation between real feed intake and predictions
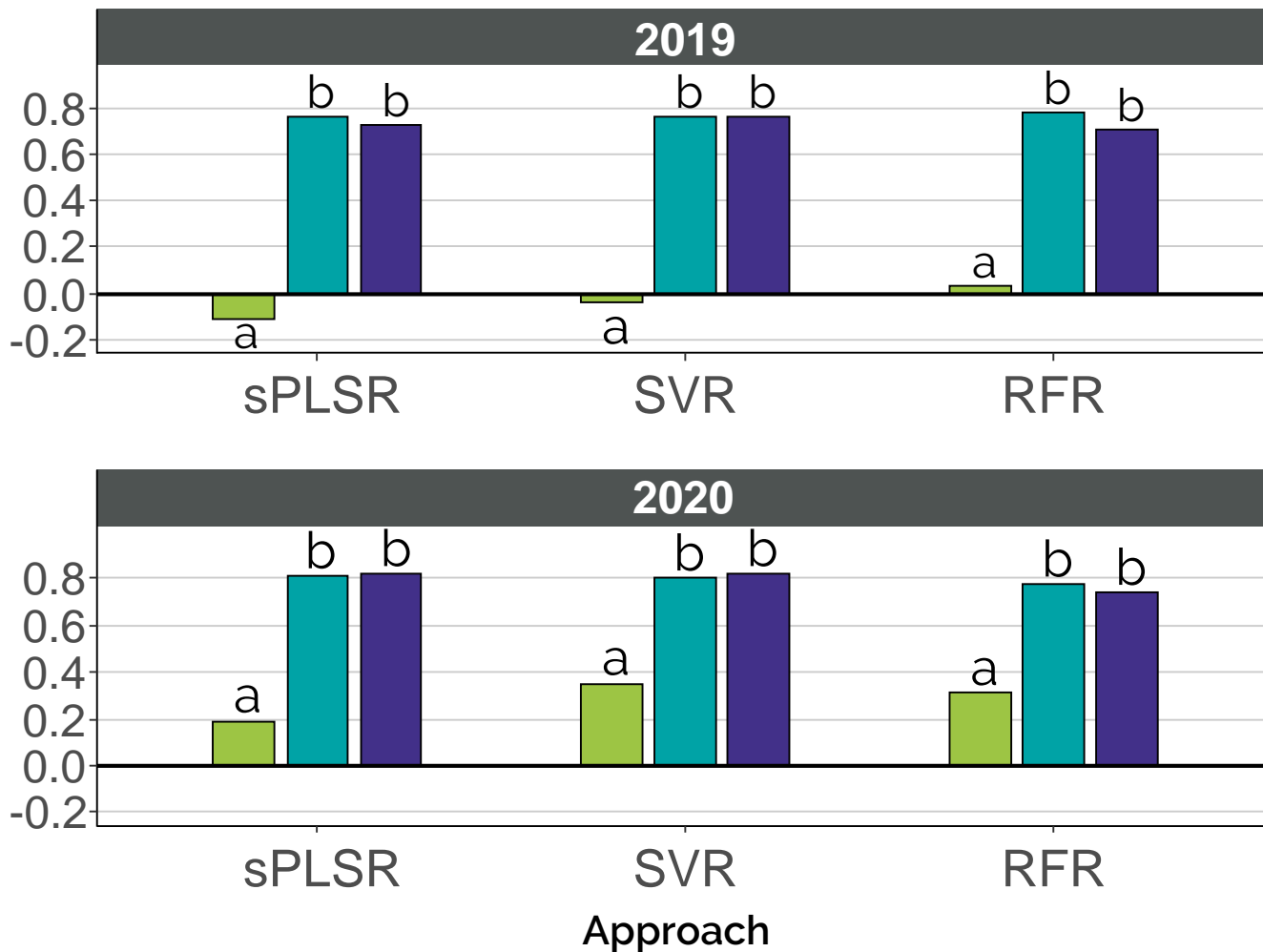

Predictor set: ■ 16S ■ Traits ■ Traits+16S

- Note: 2020 lambs' sires are part of 2018 → 2018 is not used as a testing set

- Higher correlations with host traits predictors than 16S data

- Combining 16S data and traits does not improve correlations

- No difference between approaches

💡 **Microbiota alone or combined with traits is not an advisable predictor**

a,b Comparisons with Dunn & Clark's z test (adjusted p-value <0,05)

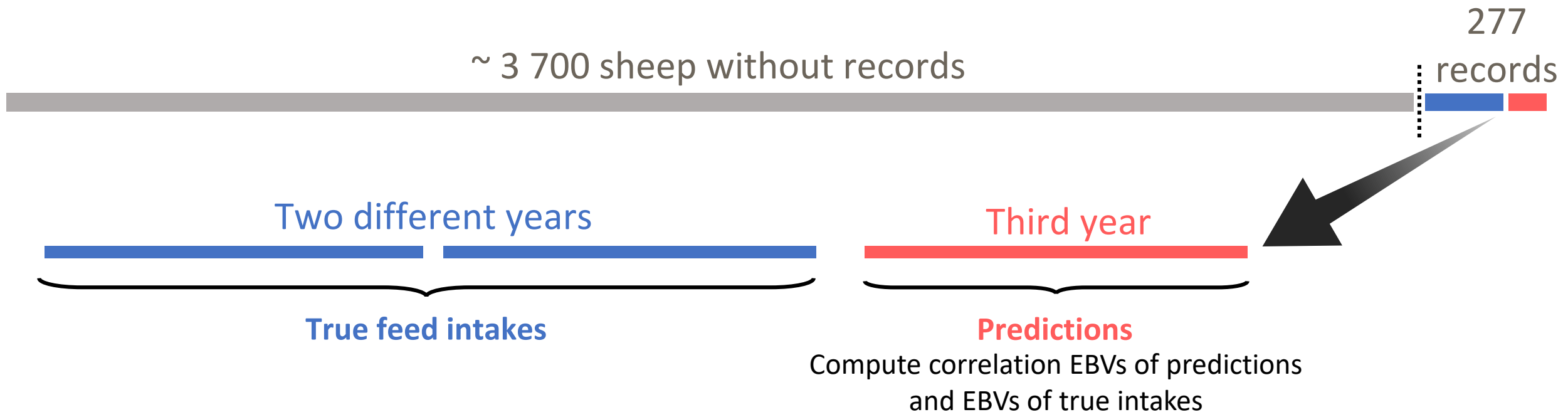# II - Relevance of predictions for the genetic evaluation

Estimation of breeding values for feed intake

With:

- PEST software

- $h^2 = 0.28$ *(Tortereau et al., 2020)*

- a pedigree of ~ 4 000 animals

Population for the genetic evaluation (subset from 2018 to 2020):

277 records

~ 3 700 sheep without records

Two different years

Third year

True feed intakes

Predictions
Compute correlation EBVs of predictions
and EBVs of true intakes

Pearson correlation between EBVs of real intake and EBVs of predictions
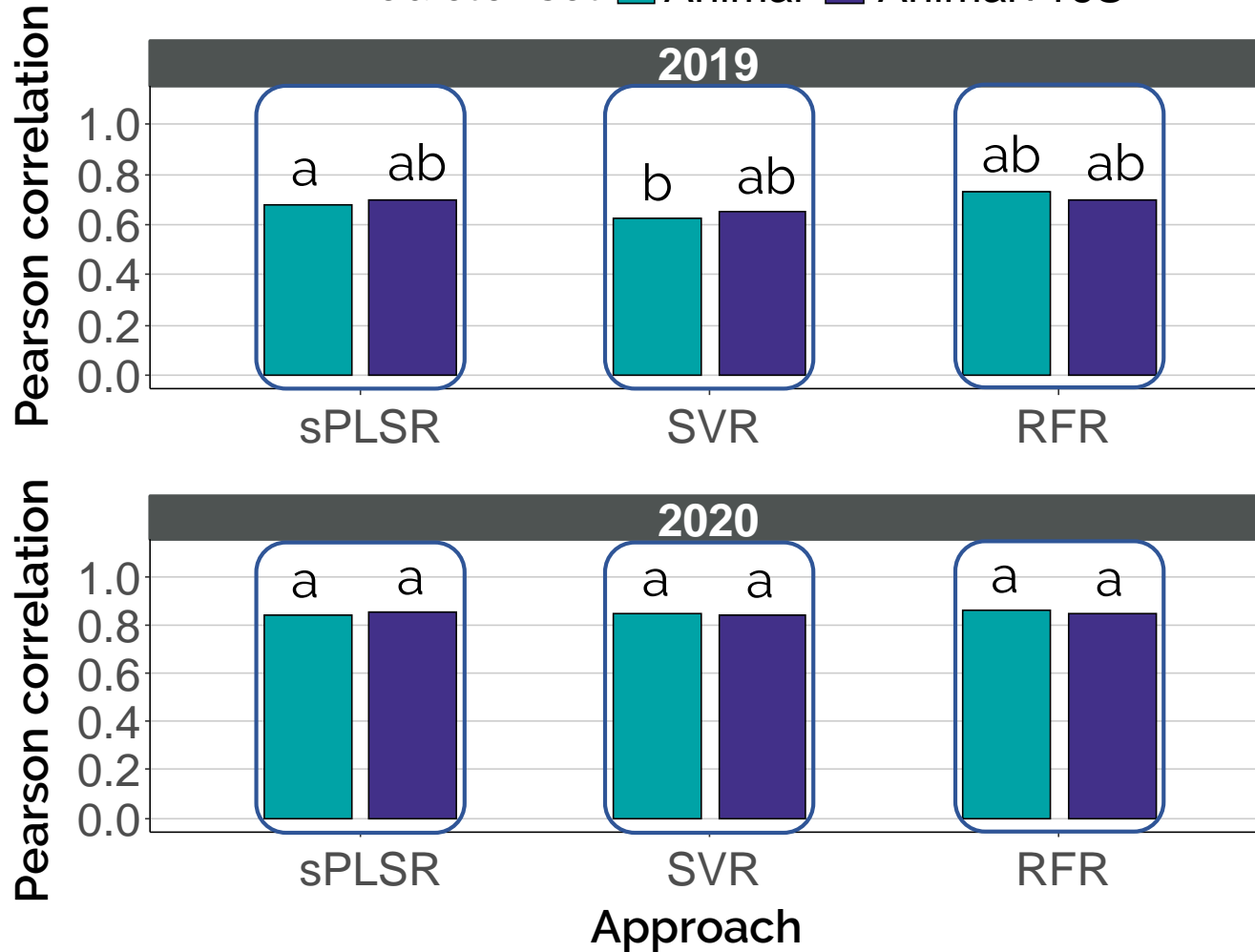

Predictor set: Animal / Animal+16S

2019
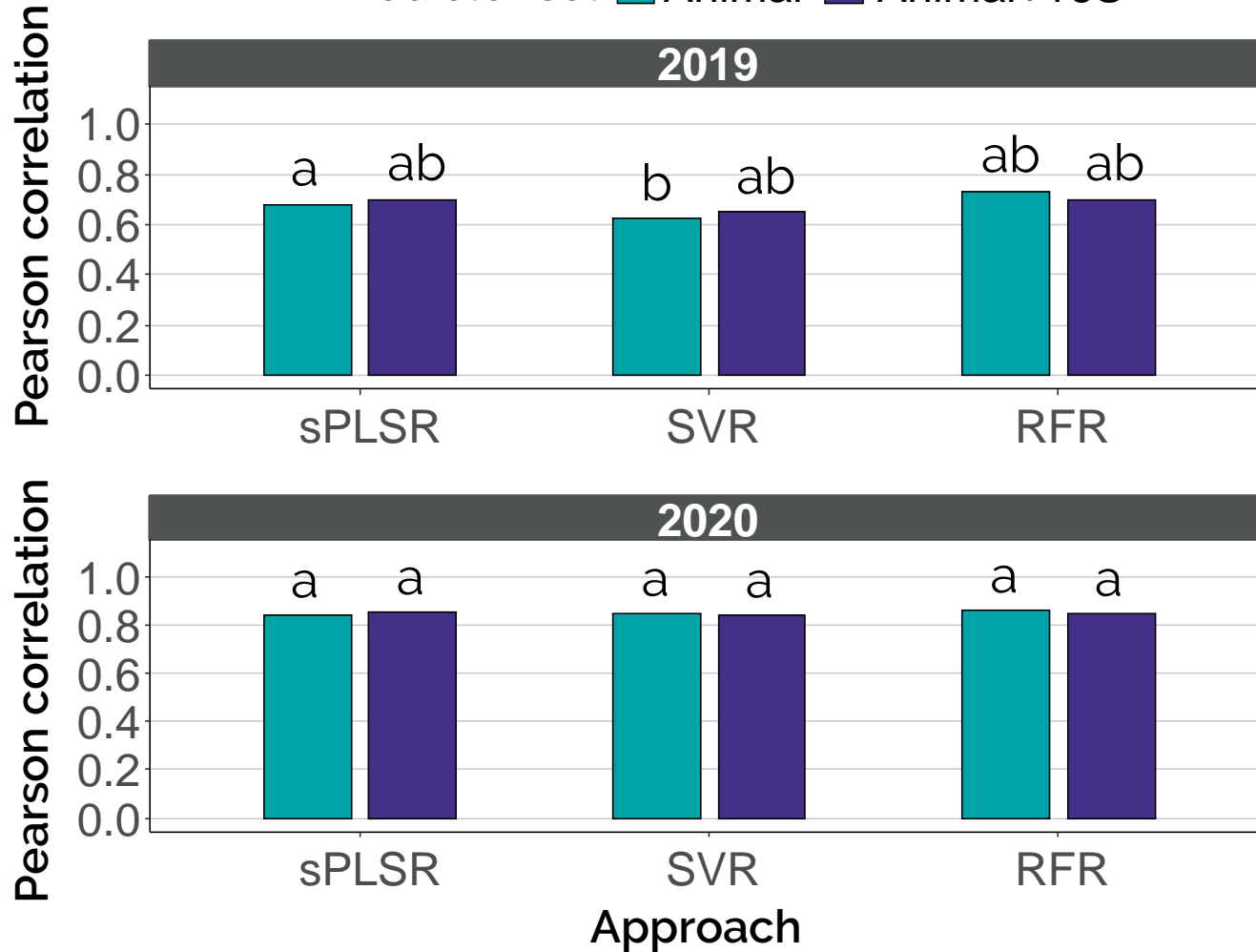
2020

- Note: 2020 lambs' sires are part of 2018

a,b Comparisons with Dunn & Clark's z test (adjusted p-value <0,05)

Pearson correlation between EBVs of real intake and EBVs of predictions

- Note: 2020 lambs' sires are part of 2018

- Combining 16S data and traits for predictions does not improve correlations between EBVs

a,b Comparisons with Dunn & Clark's z test (adjusted p-value <0,05)

Pearson correlation between EBVs of real intake and EBVs of predictions

- Note: 2020 lambs' sires are part of 2018

- Combining 16S data and traits for predictions does not improve correlations between EBVs

- One difference between machine learning approaches

[a,b] Comparisons with Dunn & Clark's z test (adjusted p-value <0,05)

Pearson correlation between EBVs of real intake and EBVs of predictions
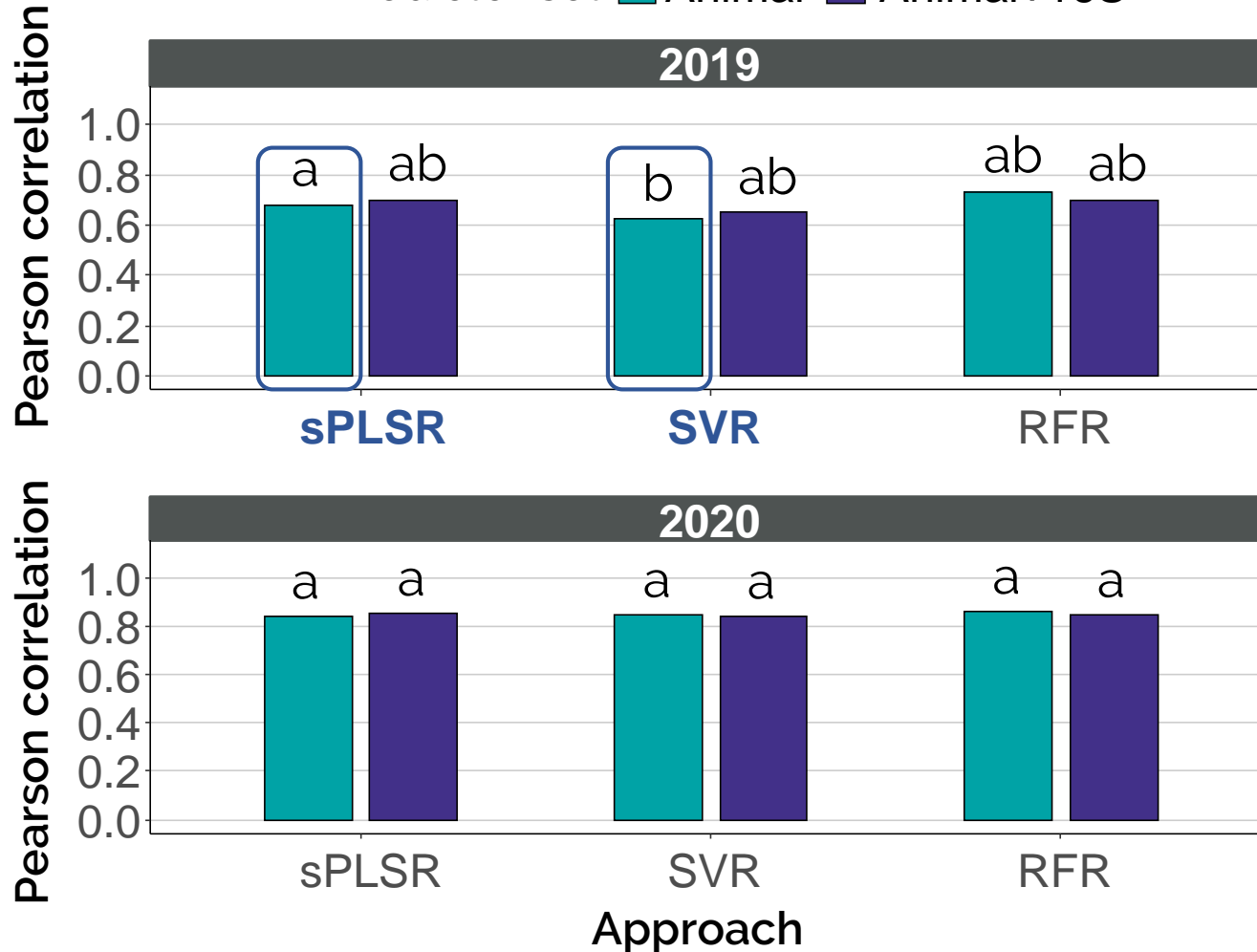

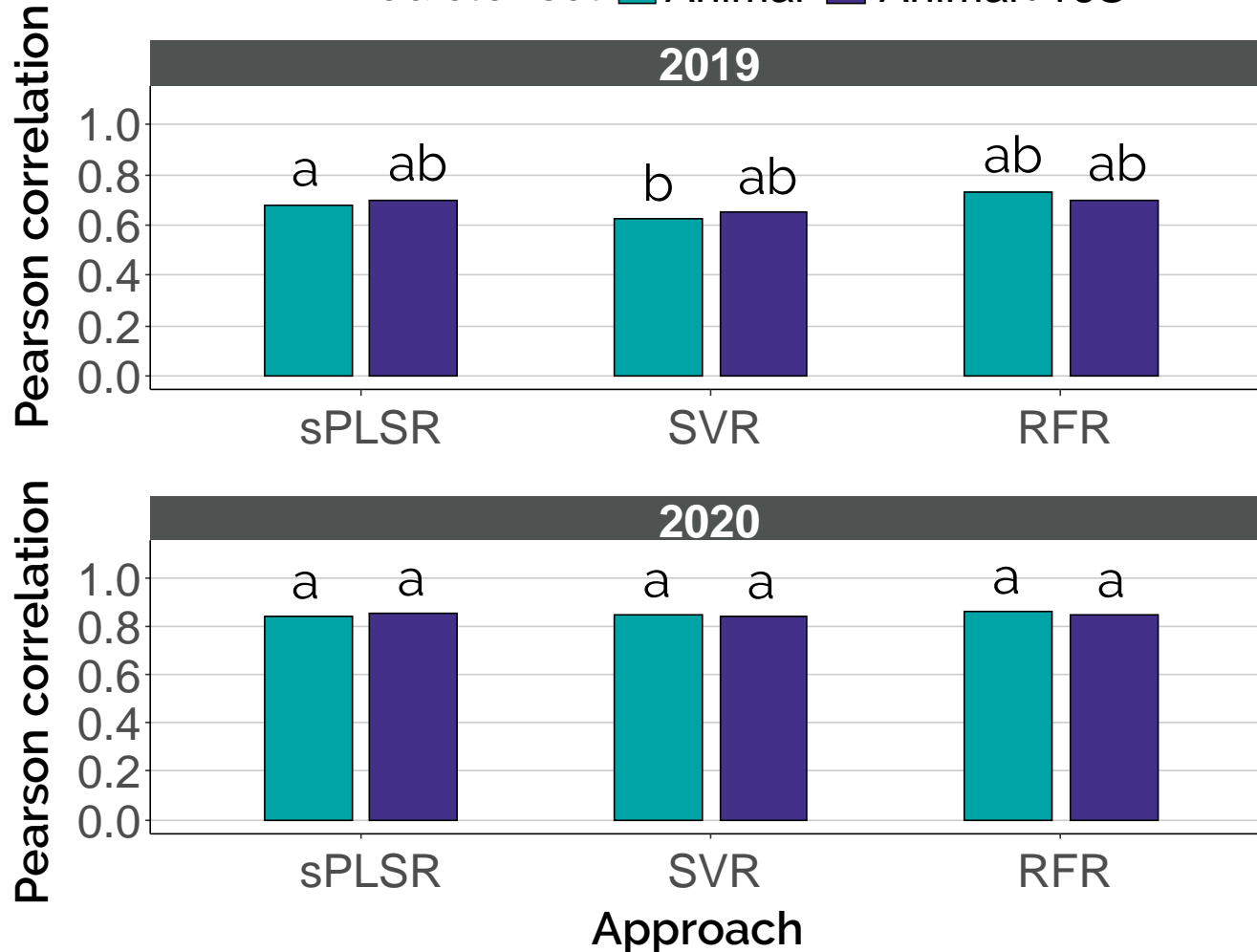Predictor set: ■ Animal ■ Animal+16S

2019

2020

- Note: 2020 lambs' sires are part of 2018

- Combining 16S data and traits for predictions does not improve correlations between EBVs

- One difference between machine learning approaches

- **16S data: no improvement for the genetic evaluation of _predicted_ feed intake**

a,b Comparisons with Dunn & Clark's z test (adjusted p-value <0,05)

Conclusions

✓ Rumen microbiota data: **no improvement of feed intake predictions in sheep**

Similar to sPLSR results in rabbits (Velasco-Galilea *et al.*, 2021)

✓ Correlations between actual RFI and predictions from 16S: from -0.15 to 0.19

New results

✓ Microbiota data: no improvement for the genetic evaluation of **predicted** feed intake

Perspectives

✓ Predict from different omics: metabolomics, genomics, phenomics

✓ Predict additional traits: greenhouse gases emissions

✓ Need for research into the phenotyping strategy

- Training/testing sets partitioning: contemporaneous animals, genetic connections

- Number of samples/records: machine learning predictions, genetic evaluation (see abstract)

Any question ?

quentin.le-graverand@inrae.fr