



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Machine learning approaches for classifying the Iberian pig strains based on microbiome

L. Azougagh¹, C. Casto-Rebollo¹, L. Varona², J. Casellas³, S. Negro⁴,
M. Martínez-Álvaro¹ y N. Ibáñez-Escriche¹

¹ Instituto de Ciencia y Tecnología Animal , Universitat Politècnica de Valencia, Valencia, España,

² Instituto Agroalimentario de Aragón (IA2), Universidad de Zaragoza, Zaragoza, España,

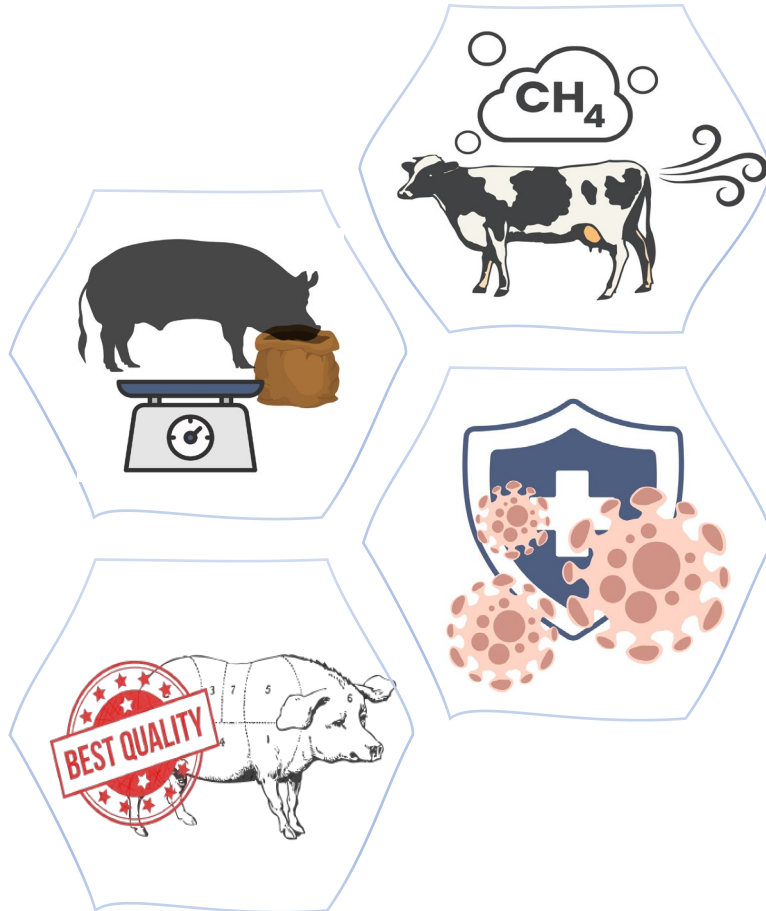
³ Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, España, ⁴ Inga Food, 06200, Almendralejo, España.



ICTA

Introduction

- Growing interest in the last decade in studying the Microbiome
- Microbiota composition regulates important phenotypic traits ..



- **Methane emission**
- **Feed efficiency**
- **Meat quality**
- **Health status**

Introduction : Microbiota drivers

- Main drivers of microbiota composition : Environment

- **Diet**

- **Housing and farm management**

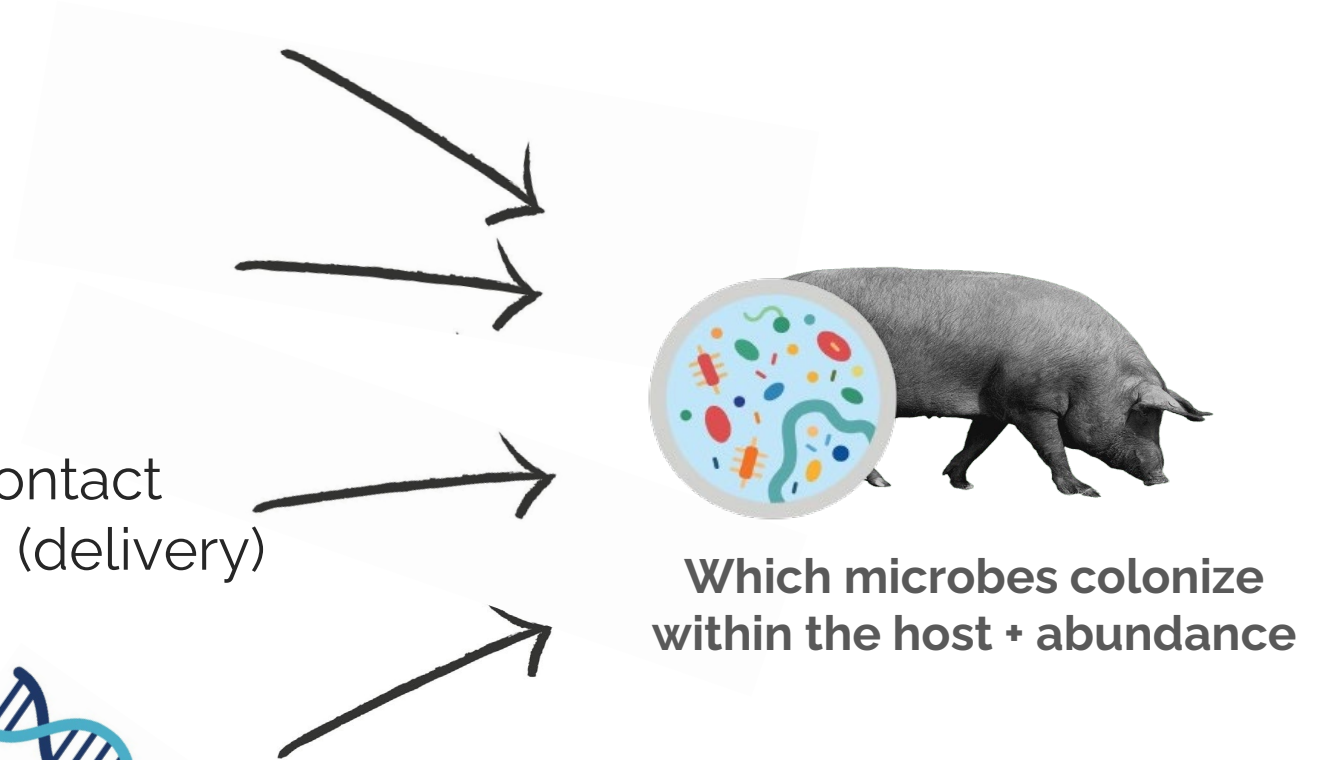
- **Age**

- **Treatment/Antibiotics**

- **Stressors**

- Vertical transmission : physical contact between newborns and mothers (delivery)

- ... And also **host genetics**, an emerging factor of interest



Which microbes colonize within the host + abundance

Introduction : The Iberian pig



- Breed of high economical interest
- High-quality meat : oriented to the production of cured ham (some of the most prized in the world)
- ↑ Intra-muscular fat, ↑ Oleic acid content
- Five officially recognized Iberian pig strains: **Entrepelado**, **Retinto**, Torbiscal, Lampiño, and Manchado de Jabugo
- All of them exhibit substantial variability in productive features (meat quality, carcass traits, etc.).

Introduction : Iberian strains diversity

Phenotypic



Genomic

Article | [Open access](#) | Published: 14 February 2019

Five genomic regions have a major impact on fat composition in Iberian pigs

[Open Access](#) [Article](#)

Additive and Dominance Genomic Analysis for Litter Size in Purebred and Crossbred Iberian Pigs

by Houssemeddine Srihi¹, José Luis Noguera², Victoria Topayan³, Melani Martín de Hijas⁴, Noelia Ibañez-Escriche³, Joaquim Casellas⁴, Marta Vázquez-Gómez⁴, María Martínez-Castillero¹, Juan Pablo Rosas⁵ and Luis Varona^{1,*}

Genetic

Genetic structure of the Iberian pig breed using microsatellites

A M M

Intra-breed genetic diversity characterization of the Iberian pig

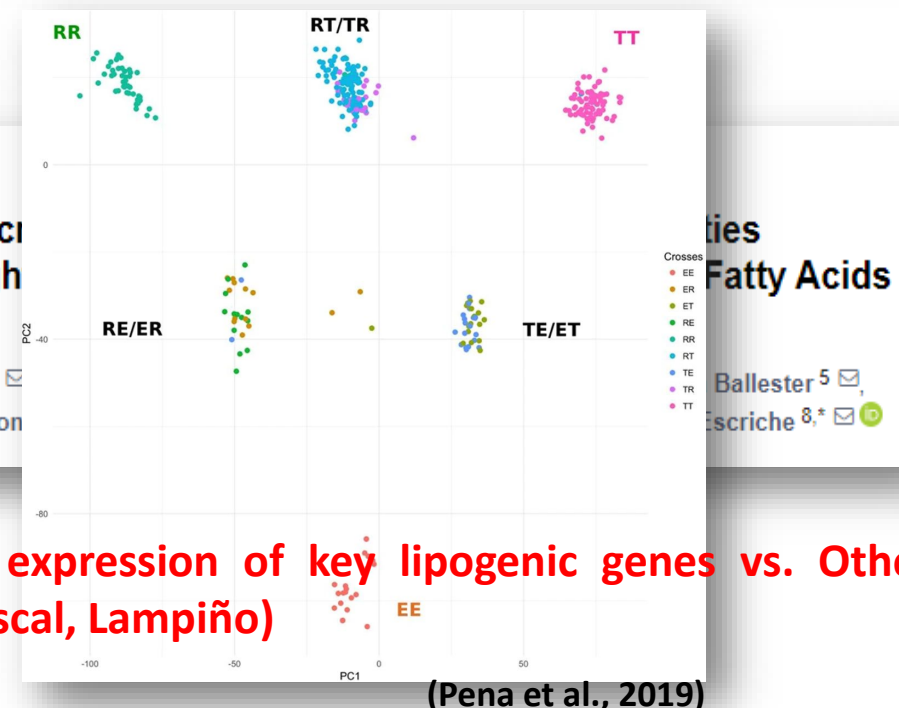
I. Clemente*, A. Membrillo*, P.J. Azor*, O. Polvillo**, M. Juárez**, E. Santos*, A.M. Jiménez*, E. Diéguez***, A. Molina*

Transcriptomic

[Open Access](#) [Article](#)

Comparative Transcriptomic Analysis Provides New Insights into Lipid Metabolism in Iberian Pigs

by Ana Villaplana-Velasco^{1,2}, Lourdes Muñoz⁶, Elena González



Introduction : Iberian strains diversity

- **Can we see this diversity at the microbiome level of the Iberian pigs ?**
 - Genetic diversity plays an important role in phenotypes of interest in Iberian pigs (meat quality)
 - Association between microbiota and those phenotypes in pigs

➤ **Becomes crucial to explore this link further in the context of Iberian pigs..**



Introduction : Iberian strains diversity

- **Maternal effect** : Litter size ($\uparrow$ in Entrepelado), palmitoleic acid (C16:1)
- **Heterosis effect** : Intramuscular fat ($\uparrow$ in Entrepelado $\times$ Retinto cross), litter size
- **Imprinting Effect** : Reproductive traits (Entrepelado $\times$ Retinto cross)

Genetic parameters and direct, maternal and heterosis effects on litter size in a diallel cross among three commercial varieties of Iberian pig

J. L. Noguera¹, N. Ibáñez-Escriche², J. Cas

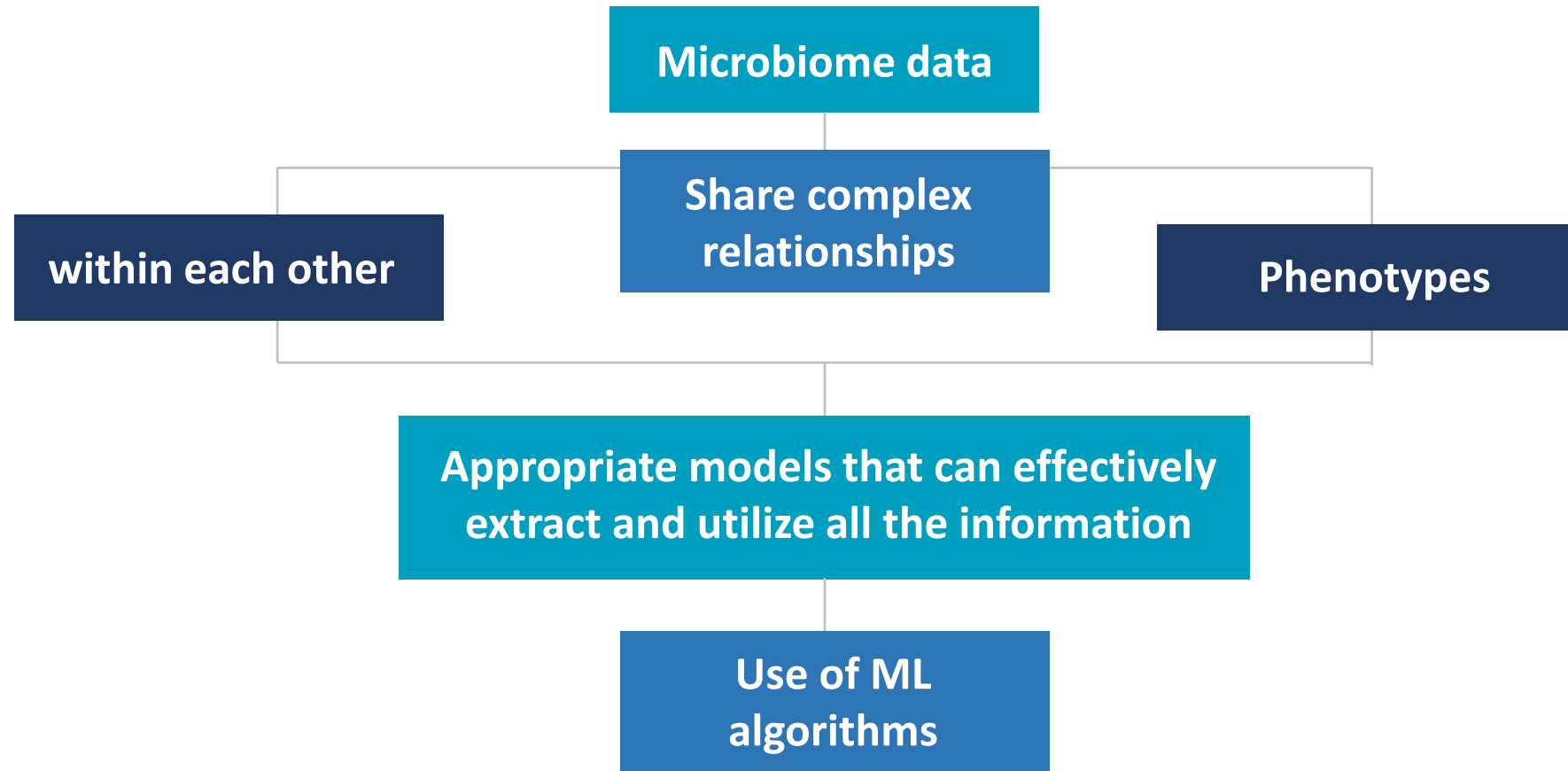
Genetic parameters and crossbreeding effects of fat deposition and fatty acid profiles in Iberian pig lines¹

N. Ibáñez-Escriche,^{*}2 E. Magallón,[†] E. Go

A Bayesian Multivariate Gametic Model in a Reciprocal Cross with Genomic Information: An Example with Two Iberian Varieties

by Houssemeddine Srihi ¹ , David López-Carbonell ¹ , Noelia Ibáñez-Escriche ² , Joaquim Casellas ³ , Pilar Hernández ² , Sara Negro ⁴ and Luis Varona ^{1,*} 

Introduction : Microbiome data peculiarities



- Choice of the most effective one ? Depends on the specific use case
 - Testing multiple models is advisable

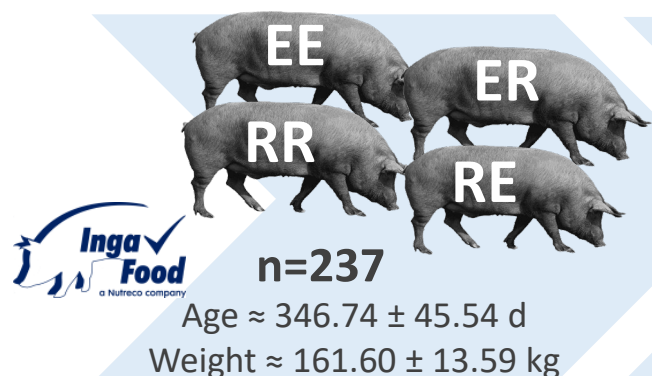
Objetives



- Explore the gut microbiota of Iberian pigs of two different strains (EE;Entrepelado and RR;Retinto) and their reciprocal crosses:
 - Evaluating the classification performance of a set of widely used ML models based on microbial abundances
- Identify the key taxa relevant for distinguishing the genetic backgrounds of these Iberian pigs.

Material and methods

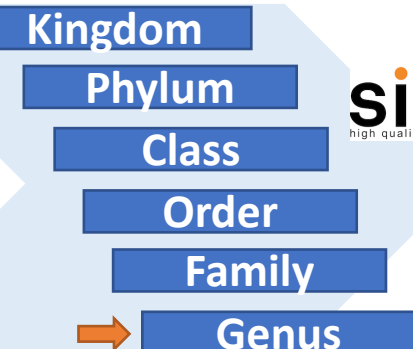
Sample collection & DNA extraction



Sequencing & 16S analysis

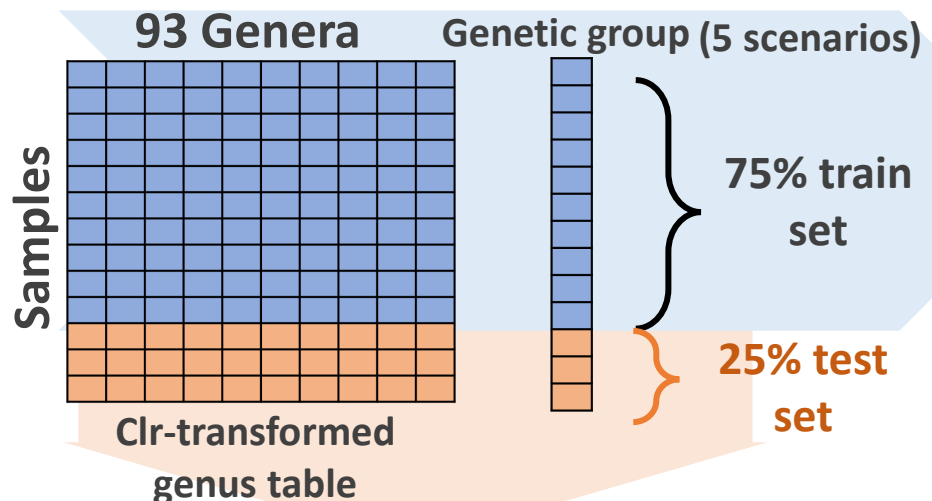


Taxonomic annotation

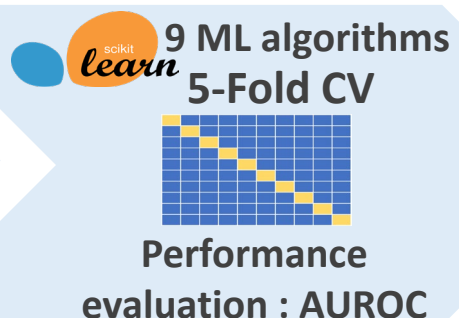


silva
high quality ribosomal RNA databases

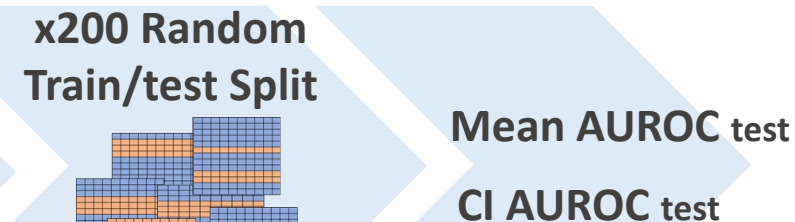
Data pre-processing : filtering, transform.



Model building and optimization



Final evaluation



Important features
(RandomForest Feature selection)

Material and methods

- Five scenarios were explored:

- Four genetic groups



- Purebred



- Maternal



- Paternal



- Heterosis



Material and methods



- ML algorithms used
 - **PLS-Discriminant Analysis (PLS-DA)**
 - **Support vector Machine (SVM)**
 - **Tree-based algorithms** : Decision tree (DT), Random forest (RF), Adaboost (AB), XGBoost (XGB), Catboost (CB)
 - **Probabilistic algorithms** : Gaussian Naïve Bayes (GNB) and Logistic regression (LR)

Material and methods



- Identification of taxa with relevant differential abundance (DA) per scenario, using a Bayesian linear model :

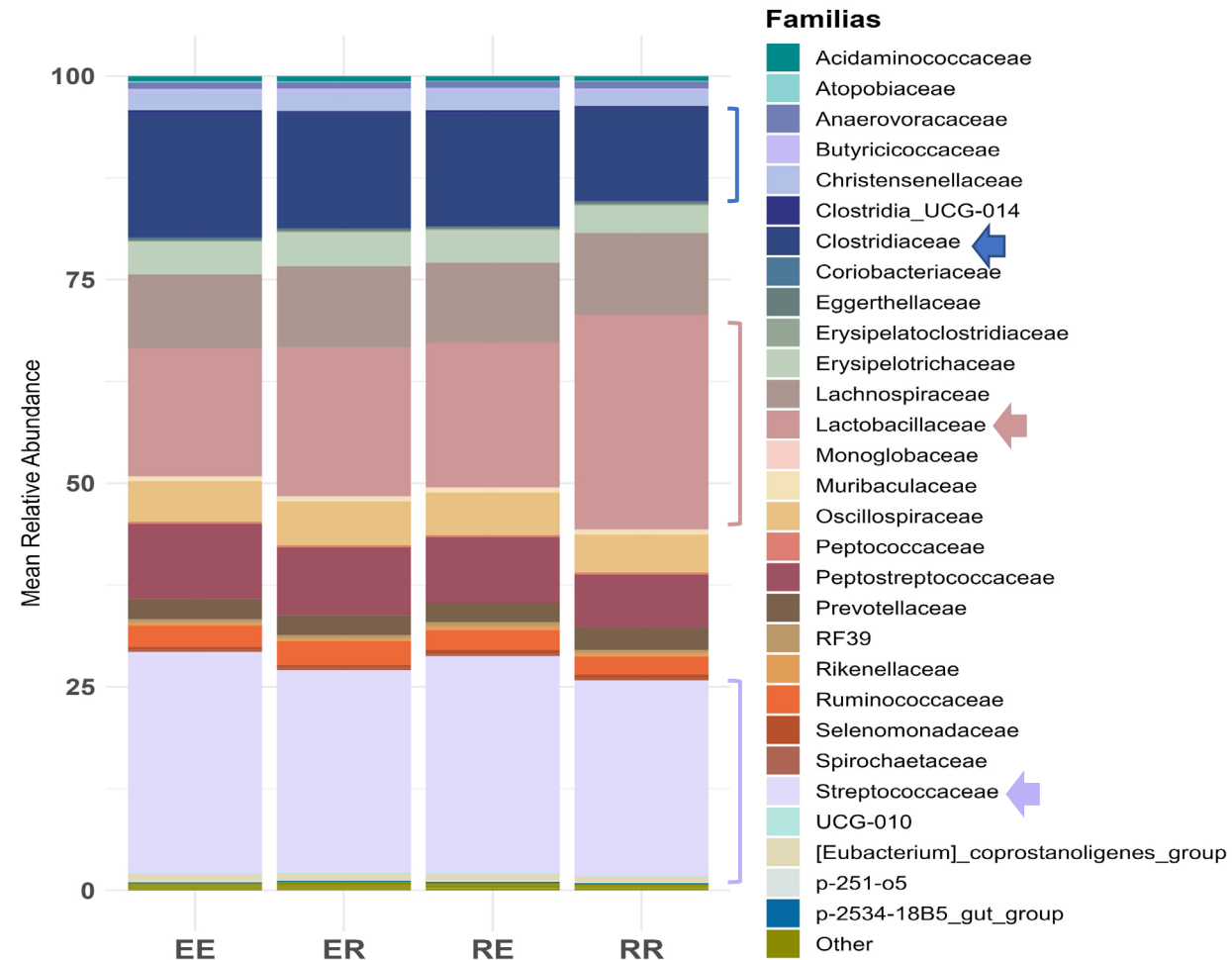
$$Abundance(clr) = \mu + \textit{Effect of the scenario related group} + e$$

- Genera with relevant DA : a minimum mean difference between groups 0.50 SD and the probability of the (|difference| > 0) > 0.95

Taxonomic results

• Relative abundances

- 11,960 ASVs, 12 Phyla, 51 families, 130 genera (prevalence filter 75%: 93 genera)
- Predominance of :
 - 2 Phyla (98.7%):
 - Firmicutes (94.6%)
 - Bacteroidetes (4.1%)
 - 3 genera (58.9%)
 - *Streptococcus* (25.7%)
 - *Lactobacillus* (19.5%)
 - *Clostridium sensu stricto 1* (13.8%)



Microbial diversity



Alpha diversity

- Shannon, Pielou & Chao1
- Pairwise Kruskal-Wallis test:
 - $\nexists$ differences between the Four genetic groups
 - $\exists$ differences between Batches and Age groups



Beta diversity

- Aitchison distance matrix
- PERMANOVA ; $\exists$ relevant compositional divergence among:
 - Four genetic groups
 - Batches
 - Age groups

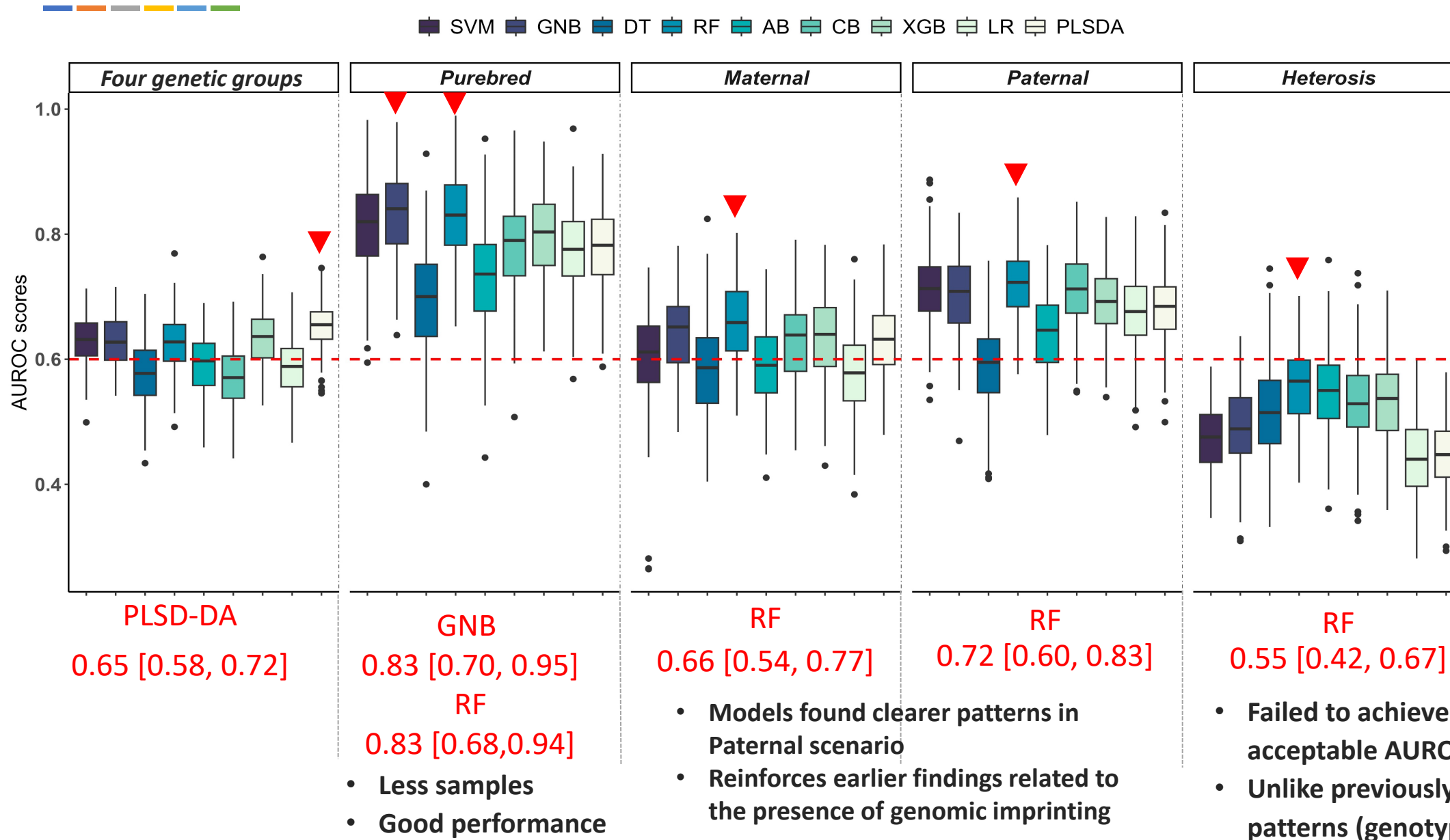
- Similar diversity of the microbial community within the four genetic groups, but overall bacterial abundance profile was different among them
- Adjusting the genera abundances for the confounding effects of age and batch

Classification results

- Beta-diversity analysis → differences were also evident at the microbiota level
- First classification task : Four genetic groups
 - Best performing model : **PLS-DA** AUROC of 0.65 (IC del 95% [0.58, 0.72])
 - Confusion matrix of the test sets over 200 iterations

Predicted classification						
Actual classification	classes	EE	ER	RE	RR	Total
	EE	15	0	0	4	19
	ER	6	1	0	5	12
	RE	5	1	0	7	13
	RR	3	1	0	12	16

Classification results



Classification results post FS

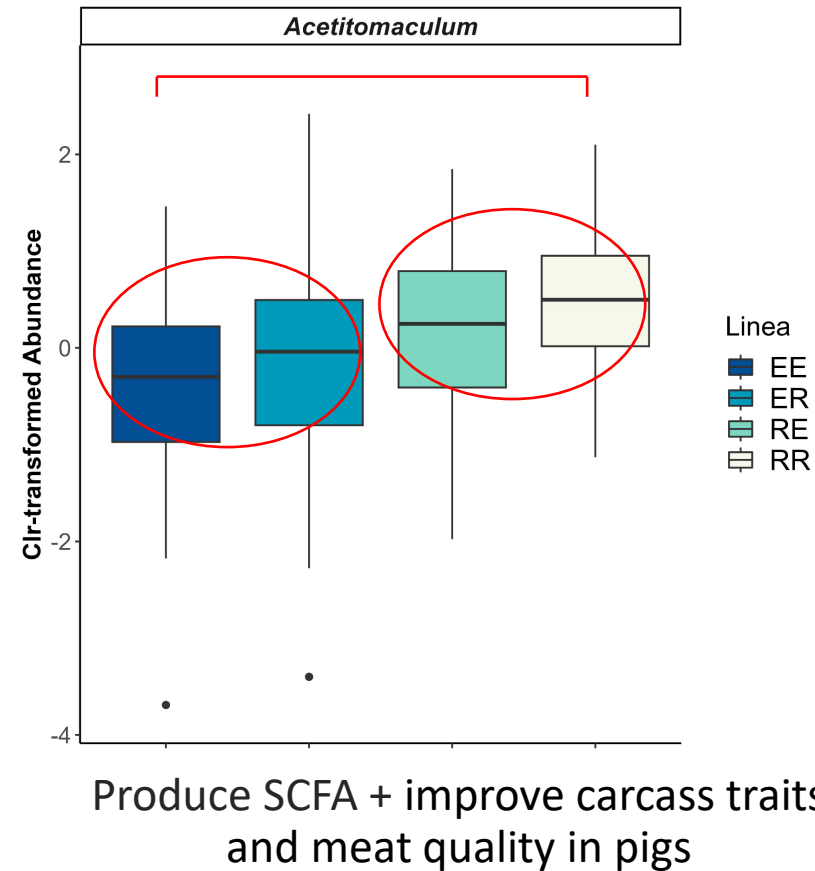
- RF allows feature selection (FS) based on the importance scores of the genera
- Classification using newly selected genera → increase in performance

Scenario	Mean increase FS (%)	Best performance no FS	Best model no FS	Best performance FS	Best model FS
Four gen. groups	6.30%	0.65 ± 0.04	PLS-DA	0.70 ± 0.04	RF
Purebred	3.30%	0.83 ± 0.07	GNB-RF	0.88 ± 0.05	GNB
Maternal	6%	0.66 ± 0.06	RF	0.71 ± 0.06	RF
Paternal	4.90%	0.72 ± 0.05	RF	0.76 ± 0.05	RF
Heterosis	13.90%	0.55 ± 0.06	RF	0.63 ± 0.07	RF

➤ ↗ initial classification performance → ↘ performance gain observed post-FS

Feature selection vs. Differential abundance

- FS → address classification aspect + Identify most important genera
- Several were consistent between scenarios
- Associated to SCFA production and lipid metabolism
- Found to be DA between groups using Bayesian linear model



Higher abundance in RR vs. EE animals

→ Aligns with RR higher quality meat

Summary

- The genetic background is an important factor influencing the microbiota profile of Iberian pigs
- RF and GNB reached the best classification performance, especially in the Purebred and Paternal scenarios and the FS improved the results.
- The genera that exhibited DA between different genetic groups were also identified by the RF-based FS method as important features
- Several important genera were linked to SCFA production and lipid metabolism → the differences in the microbiota composition could be contributing to differences in meat quality traits between Iberian pigs



Thanks for your attention!



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ICTA

