Bolner, M., Bovo, S., Holl, J., Valente, B., Lewis, C., Schiavo, G., Bertolini, F., Fontanesi, L. (2026) Metabolomic profiles of divergent pigs for feed intake: towards the genetic dissection of a complex trait using molecular phenotypes, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 23748: https://doi.org/10.31274/wcgalp.23748

Metabolomic profiles of divergent pigs for feed intake: towards the genetic dissection of a complex trait using molecular phenotypes

M. Bolner[1], S. Bovo[1], J. Holl[2], B. Valente[2], C. Lewis[3], G. Schiavo[1], F. Bertolini[1], and L. Fontanesi[1*]

[1]University of Bologna, Bologna, Italy; [2]PIC North America - Genus, Hendersonville, TN, USA; [3]PIC Europe - Genus, Barcelona, Spain; [*]Corresponding author: luca.fontanesi@unibo.it

Introduction

The pig production sector is increasingly directed toward the identification of novel and informative phenotypic traits to extend the conventional phenotypic framework, with the objective of improving the accuracy of selection programs and fostering the development of advanced breeding strategies. Traditional phenotypes, including production and performance traits, are regulated by intricate and multilayered molecular networks, thereby complicating the dissection of their genetic architecture. In this regard, the metabolome – defined as the comprehensive repertoire of low-molecular-weight metabolites circulating within biological systems – constitutes a relevant source of intermediate phenotypes. Given the strong association between metabolite abundance and the underlying genetic background of the animal, metabolomic features may act as robust proxies for complex traits and can be strategically integrated into genetic evaluation and breeding programs (Fontanesi, 2016). Feed intake is a key trait in pig production, as it directly influences growth performance, feed efficiency, production costs, and environmental sustainability. In this study, we applied an untargeted metabolomic approach to identify metabolites that could describe feed intake in pigs. The investigated population was constituted by two groups of Duroc pigs with extreme and divergent values of feed intake whose plasma was analysed with more than 700 metabolites.

Materials & Methods

Animal ethics, pigs included in the study and blood samples. All pigs used in this study were kept in compliance with European and US legislation regarding pig husbandry. The pigs were not raised or treated in any way for the purpose of this study, so no additional ethical statement is necessary. The study analyzed 290 male Duroc pigs sampled in two rounds conducted in October and November 2023. Animals were selected from the high (H) and low (L) extremes of voluntary individual feed intake within a population of approx. 1,000 performance tested pigs. Feed intake measurements were taken between 62 and 71 days of age using individual feed intake recording stations, while blood samples were collected between 91 and 133 days of age. All pigs were reared under uniform management and environmental conditions. Blood samples were drawn into EDTA tubes and subsequently processed for plasma separation and metabolomic analyses as previously described (Bovo *et al.*, 2025). Genotyping data were available for these animals and were used to assess population stratification via principal component analysis (PCA). No clustering according to feed intake group was observed.

Metabolomics of plasma samples. Untargeted metabolomics analysis was conducted on plasma samples from animals representing two extreme feed intake groups using the HD4 platform at Metabolon, Inc. (Morrisville, NC, USA). After a methanol-based extraction, the

samples underwent ultra-performance liquid chromatography coupled with tandem mass spectrometry (UPLC-MS/MS). Raw peak areas were subsequently extracted and normalized.

Data cleaning and imputation. Data preprocessing included the following steps, applied within each of the two differential feed intake groups: (i) identification of extreme outlier values for each metabolite, defined as those exceeding five interquartile ranges beyond the median value; (ii) replacement of identified outliers with a missing value; (iii) elimination of metabolites with missing data in over 25% of the samples; and (iv) exclusion of individual samples with missing values for more than 30% of the metabolomic profile. The remaining missing values were imputed following the approach described by Faquih *et al.* (2020), using the Multivariate Imputation by Chained Equations (MICE) framework (Azur *et al.*, 2011; Buuren and Groothuis-Oudshoorn, 2011). To account for the stochastic nature of the imputation procedure, we repeated the imputation process five times using distinct random seed values, each time generating five different imputed datasets, yielding a total of 25 imputed datasets for each extreme group. Then, the two extreme groups were combined into a single dataset for each combination of random seed and imputation cycle. Confounding effects attributable to sex and sampling date were removed from the imputed datasets with an Ordinary Least Squares model, with the resulting residuals being used for downstream analyses.

Identification of differentially abundant metabolites. To identify metabolites that could discriminate between the two extreme groups, we utilized the Boruta algorithm (Kursa *et al.*, 2010), an all-relevant feature selection approach that leverages Random Forest classification models. This method has been demonstrated to be robust in handling stochastic components compared to other feature selection algorithms in untargeted metabolomics applications (Bovo *et al*., 2025). Boruta was applied to each of the 25 residual datasets that were generated during the imputation phase. To address for the random component of the feature selection process, each dataset underwent five independent Boruta runs using unique random seed values, resulting in a total of 125 feature selection iterations. To ensure the consistency of results in relation to the sample composition of the dataset, we incorporated a 10-fold cross-validation (10CV) procedure. Each of the 25 imputed datasets was divided into ten equal subsets, and by iteratively excluding each subset, we created 250 reduced datasets, each containing 90% of the original samples. The Boruta feature selection procedure, along with five random seed iterations, was then applied to each reduced dataset, resulting in an additional 1,250 feature selection iterations. Each metabolite was evaluated based on the results from all feature selection runs. Two metabolite sets were defined: those labeled as "Confirmed" in all 125 primary Boruta runs and all 1,250 cross-validation runs, were assigned to the "Robust" set, while those labeled as "Confirmed" in all 125 primary Boruta runs and in at least 1,000 of the 1,250 cross-validation runs (80%) were assigned to the "Permissive" set.

Evaluation of discriminatory capacity of the selected metabolites. The selected metabolites underwent the following univariate analyses, to evaluate their discriminatory power: i) Mean difference in abundance, calculated as

$$\Delta\%_{\mathrm{i}} = \frac{\bar{x}_{\mathrm{i}}^{\mathrm{H}} - \bar{x}_{\mathrm{i}}^{\mathrm{L}}}{\bar{x}_{\mathrm{i}}^{\mathrm{H}}} * 100 \quad (1)$$

where $\bar{x}_{\mathrm{i}}^{\mathrm{H}}$ and $\bar{\mathrm{x}}_{\mathrm{i}}^{\mathrm{L}}$ are the average metabolite abundance of the *i*th metabolite in the high (H) and low (L) feed intake groups, respectively (Bovo *et al.*, 2016); ii) Mann-Whitney U test (MWU) with results corrected for multiple testing using the Bonferroni correction; iii) Area Under the Curve (AUC) derived from the Receiver Operating Characteristic (ROC) curve analysis to quantify binary classification performance across all possible decision thresholds. Higher AUC values indicate superior discriminatory capacity; iv) Correlation with raw feed

intake. To assess the relationship between metabolite abundance and feed intake, we calculated the Pearson correlation between the two values. Principal Component Analysis (PCA) was conducted on the entire sample dataset using the complete metabolite profile and then on the two selected metabolite subsets, to identify clustering and other sample-level effects before and after feature selection.

Results & Discussion

Description of the plasma metabolomic profile in the population. A total of 756 metabolites were analysed, along with their relative abundance data. Of these, 616 metabolites were annotated as endogenous and covered eight super-pathways: lipid (39% of the dataset), amino acid (24.9%), nucleotide (5.6%), peptide (3.4%), carbohydrate (3.2%), cofactors and vitamins (3%), energy (1%), and partially characterized molecules (0.8%). The datas*et al*so included a total of 104 defined sub-pathways. Figure 1 illustrates the distribution of super and sub-pathways in the dataset, both before and after the filtering process, which excluded all xenobiotic metabolites from the analysis. After filtering, 104 metabolites were removed, due to having over 25% missing values, while none of the samples were discarded. The relative distribution of super pathways remained largely unchanged, as shown in Figure 1. Consequently, the final dataset used for all subsequent analyses consisted of 290 samples and 587 metabolites.

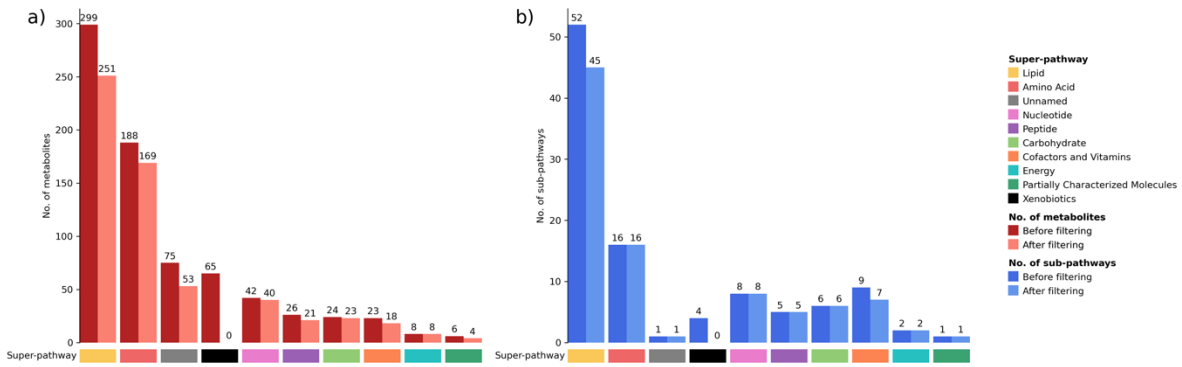

Figure 1. Metabolomic datasets before and after filtering. (a) Number of metabolites across super-pathways. (b) Number of sub-pathways characterizing each super-pathway.

Identification of metabolites differentiating feed intake extreme groups. An initial exploratory analysis was conducted using PCA on the entire metabolomic profile, to assess potential sample clustering and differences between the two extreme groups. The first two principal components (PC) explained approximately 26% of the total variance (PC1 = 19.6% and PC2 = 6.7%), with no sample clustering and the two extreme groups overlapping with each other (Figure 2A). The feature selection process identified a total of 29 metabolites selected across all 125 Boruta runs and in at least 1,000 Boruta CV runs. Nineteen of them were selected in all 1250 CV runs (Robust-CV threshold) and 10 were selected between 1,000 and 1,250 (Permissive-CV threshold). Among the selected metabolites, eight (28%) had higher abundance in high feed intake pigs, with mean difference across samples ranging from 8% to 59% and an average of 27% ± 19%. The other 19 metabolites which were more abundant in low feed intake pigs had mean differences ranging from -8% to -87%, with an average of -28% ± 18%. The correlation between metabolite abundance and raw feed intake for the selected metabolites ranged from -0.4 to +0.34. The mean correlation was 0.28 ± 0.07. The overall Pearson correlation coefficient between mean difference in abundance and feed intake was 0.86. AUC values demonstrated good discriminatory power for the selected metabolites, ranging from 0.61 to 0.75, with an average of 0.69 ± 0.03. The results of the Mann-Whitney U test were significant ($P < 0.05$) for all but two of the selected metabolites, indicating a

significant relationship with AUC (Spearman's ρ = -0.99, $P = 8.0\times10^{-209}$) as in our previous findings (Bovo *et al.*, 2025) and in Mason and Graham (2002).
PCA for the Permissive-CV subset of 29 metabolites showed a marked improvement in separation between the two extreme groups, with an increased amount of explained variance in PC1 (25.3%) and PC2 (16.2%) (Figure 2B). Similarly, the Robust-CV subset of 19 metabolites showed comparable separation, with increased amounts of variance explained by PC1 (33.9%), while PC2 explained 10.5% (Figure 2C).

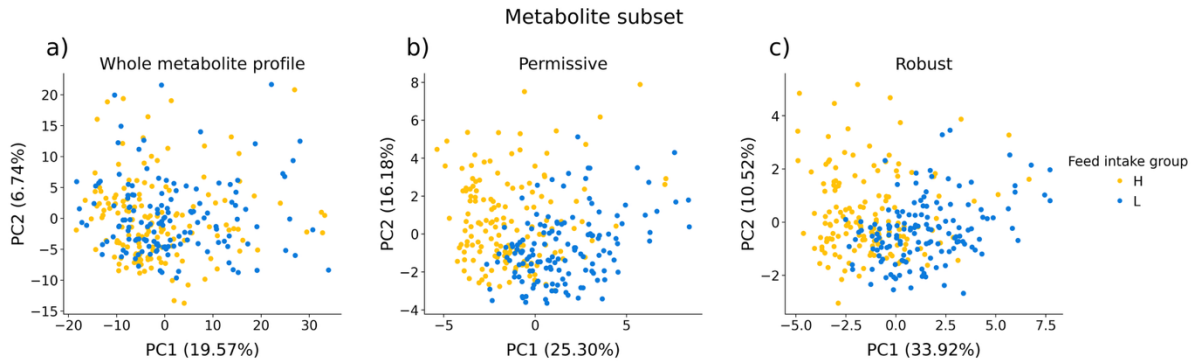

Figure 2. Principal component analysis (PCA) based on: (a) whole metabolite profile (n. 587 metabolites), (b) subset-specific metabolites based on the permissive-CV criterion, and (c) subset-specific metabolites based on the robust-CV criterion. For each subfigure, the X axis shows principal component 1 (PC1) and the Y axis shows principal component 2 (PC2), along with each PC's respective explained variance.

Conclusions

These findings demonstrated that proxies for feed intake can be identified in the pig metabolome, which can aid in dissecting it, as well as other complex production traits. With further investigation and more in-depth studies, such as the addition of samples, as well as association studies with genomic information, these results can be utilized to dissect complex traits at the molecular level and more clearly define the genetic complexity that underlies their variability.

Acknowledgements

This study received funding from the European Union's Horizon Europe research and innovation program under grant agreement No. 101059609 (Re-Livestock project).

References

Azur M.J., Stuart E.A., Frangakis C., *et al.* (2011) Int. J. Methods Psychiatr. Res. 20:40-49. doi.org/10.1002/mpr.329
Bertolini F. *et al.* (2025) Animal 20:101725. doi.org/10.1016/j.animal.2025.101725
Bovo S. *et al.* (2016) Animal 10:1741-1748. doi.org/10.1017/S1751731116000483
Bovo S. *et al.* (2025) Animal 19:101393. doi.org/10.1016/j.animal.2024.101393
Buuren S. van, and Groothuis-Oudshoorn K. (2011) J. Stat. Softw. 45:1-67. doi.org/10.18637/jss.v045.i03
Cantalapiedra-Hijar G., *et al.* (2024) Sci. Rep. 14:8540. doi.org/10.1038/s41598-024-59253-7
Faquih T., *et al.* (2020) Metabolites 10:486. doi.org/10.3390/metabo10120486
Fontanesi L. (2016) Anim. Front. 6:73-79. doi.org/10.2527/af.2016-0011
Kursa M.B., Jankowski A., and Rudnicki W.R. (2010) Fundam. Inform. 101:271-285. doi.org/10.3233/FI-2010-288
Mason S.J., and Graham N.E. (2002) Q. J. R. Meteorol. Soc. 128:2145-2166. doi.org/10.1256/003590002320603584
Stanley B.G., Urstadt K.R., Charles J.R., *et al.* (2011) Physiol. Behav. 104:40-46. doi.org/10.1016/j.physbeh.2011.04.046