71st ICoMST
GIRONA - CATALONIA - SPAIN
3rd-8th August - 2025

GEOGRAPHICAL ORIGIN AUTHENTICATION OF LAMB: ADVANCING LABEL CONTROL THROUGH 4D-DIA QUANTITATIVE PROTEOMICS AND INTERPRETABLE MACHINE LEARNING METHODS

Chongxin Liu [1,2], Simona Grasso [2], Nigel Patrick Brunton [2], Minghui Gu [1], Mengyu Shang [1], Shaobo Li [1], Li Chen [1,*], Dequan Zhang [1]

[1] Institute of Food Science and Technology, Chinese Academy of Agriculture Sciences, Key Laboratory of Agro-Products Quality and Safety Control in Storage and Transport Process, Ministry of Agriculture and Rural Affairs, Beijing, China.
[2] School of Agriculture and Food Science, University College Dublin, Belfield, Dublin, Ireland
*Corresponding author email: chenliwork@126.com

I. INTRODUCTION

This study focuses on the authentication of lamb meat's geographical origin, addressing growing consumer concerns about transparency, quality, and safety amid increasing market demand and complex supply chains [1]. Fraudulent labelling not only threatens consumer health and rights but also undermines fair trade and brand integrity [2]. Traditional techniques like stable isotope and multi-element analysis have been effective but face limitations, such as susceptibility to dietary influences [3], necessitating faster and more reliable methods. Recent advances in food omics, particularly proteomics, offer promising solutions. Proteins, influenced by both genetic and environmental factors, serve as potential biomarkers for origin identification [4]. Among proteomic technologies, four-dimensional data-independent acquisition (4D-DIA) proteomics stands out for its enhanced data resolution and quantitative accuracy [5]. However, the high-dimensional datasets generated pose analytical challenges [6]. Interpretable machine learning emerges as a robust solution, improving analysis accuracy, model transparency, and confidence in biomarker validation by revealing key features and decision processes [7]. This study aims to develop a reliable framework combining 4D-DIA proteomics and interpretable machine learning for lamb origin authentication. It uses 4D-DIA analysis on lamb samples, applies LASSO regression for biomarker identification, and compares machine learning models. The most accurate model is analysed using SHAP to reveal biomarker-origin relationships. The hypothesis suggests this approach ensures transparent, scientifically validated origin verification, benefiting broader food labelling and authenticity control.

II. MATERIALS AND METHODS

A total of 24 lamb samples (8 per region) from three major Tan sheep production areas were analyzed using 4D-DIA proteomics, following the protocol by Wang et al. [8] with specific adjustments. After database search and quantification, the proteomic profiles were analyzed with a LASSO regression to identify biomarkers. Two groups of potential biomarkers were used to develop origin discrimination models with six machine learning algorithms: Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree Classifier (DTC), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and LightGBM. Model performance was compared to select the best algorithm for SHAP analysis interpretation.

III. RESULTS AND DISCUSSION

The 4D-DIA quantitative proteomics analysis identified 26,442 peptides and 3,790 proteins in Tan sheep meat from Ningxia (NTS), Gansu (GTS), and Inner Mongolia (ITS), with 3,511 proteins (92.64%) common across regions and 279 region-specific proteins. Variations in protein composition, influenced by environmental and dietary factors, suggest potential links to meat quality differences. LASSO regression identified 14 and 6 potential protein biomarkers under λmin and λ1se penalty values, respectively, with four proteins (W5PF65, W5PQE5, W5Q501, W5QGW4) shared by both sets (**Figure 1**). Machine learning models—Logistic Regression (LR), Random Forest (RF), and XGBoost—were constructed to classify lamb origin (**Table 1**). LR achieved the highest accuracy (1.0000) on training and 0.7500 on test datasets, outperforming RF and XGBoost, which showed overfitting. SHapley Additive exPlanations (SHAP) analysis provided insights into model interpretability, identifying W5PF65, W5PQE5, and W5Q501 as critical contributors to geographical classification (**Figure 2**). SHAP demonstrated superior stability and interpretability compared to traditional methods, revealing complex protein interactions affecting model predictions. These findings highlight the effectiveness of integrating 4D-DIA proteomics with interpretable machine learning for reliable geographical traceability of lamb, offering a robust framework for food authenticity and labelling.

Table 1. Comparative performance evaluation of machine learning models on test sets.

Group	**Lasso_min**				**Lasso_1se**			
Method	Accuracy	Recall	Precision	F1 Score	Accuracy	Recall	Precision	F1 Score
Logistic Regression	1.0000	1.0000	1.0000	1.0000	0.7500	0.7500	0.8750	0.7500
Support Vector Machine	0.6250	0.6250	0.7917	0.5667	0.6250	0.6250	0.6667	0.6357
Decision Tree Classifier	0.5000	0.5000	0.4167	0.4500	0.3750	0.3750	0.2083	0.2500
Random Forest	0.6250	0.6250	0.6250	0.6107	0.7500	0.7500	0.8750	0.7500
XGBoost	0.6250	0.6250	0.5500	0.5833	0.7500	0.7500	0.8333	0.7333
LightGBM	0.2500	0.2500	0.0625	0.1000	0.2500	0.2500	0.0625	0.1000

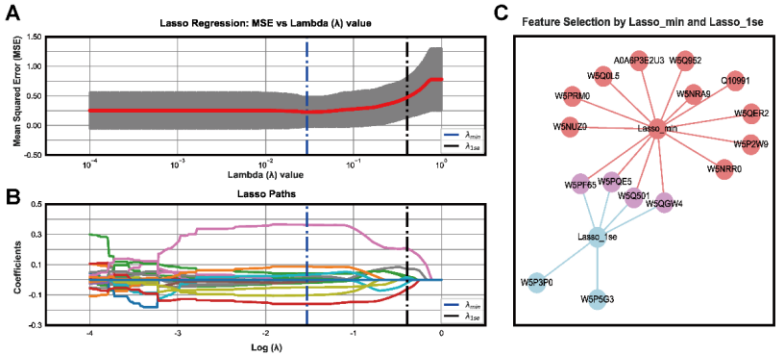


Figure 1. Results from Lasso regression-based protein biomarker screening: (A) mean squared error (MSE), (B) coefficient path trajectories, and (C) biomarker identification at λ_{min} (best λ value with minimal MSE) and λ_{1se} (λ value where MSE remains within one standard error of the best value).

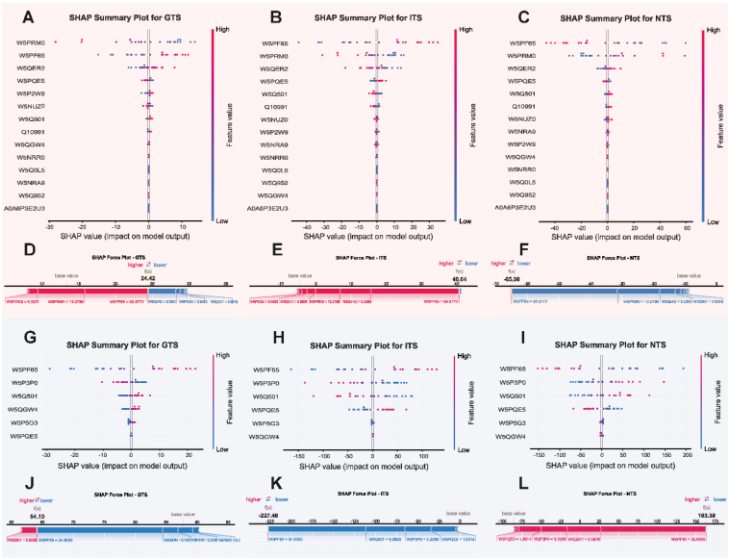


Fig. 2. SHAP-based interpretability analysis of Logistic Regression models: (A-C) biomarker summary plots and (D-F) force plots for an independent sample in the λ_{min} group; (G-I) biomarker summary plots and (J-L) force plots for an independent sample in the λ_{1se} group.

IV. CONCLUSION

Authenticating lamb origin is vital for detecting high-risk samples and preventing origin fraud, protecting consumers. This study pioneers the use of 4D-DIA proteomics combined with machine learning for lamb origin authentication, offering high accuracy, efficiency, and transparency. The approach also shows potential for other meat products. However, further research is needed to validate key protein biomarkers, focusing on their stability across regions, time, and storage conditions to improve the method's precision and practicality.

ACKNOWLEDGEMENTS

This work was supported by the National Key R&D Program of China (2022YFD2100500) and the National Key R&D Program of Ningxia Hui Autonomous Region, China (2024BBF02022). Chongxin Liu gratefully acknowledges the financial support from the China Scholarship Council (CSC) under grant number 202403250010.

REFERENCES

1. Kantono, K., Hamid, N., Ma, Q., Chadha, D., Oey, I. (2021). Consumers' perception and purchase behaviour of meat in China. Meat Science, 179, 108548.
2. Berry, C., Romero, M. (2021). The Fair Trade Food Labeling Health halo: Effects of fair trade labeling on consumption and perceived healthfulness. Food Quality and Preference, 94, 104321.
3. Liu, H., Nie, J., Liu, Y., Wadood, S. A., Rogers, K. M., Yuan, Y., Gan, R.-Y. (2023). A review of recent compound-specific isotope analysis studies applied to food authentication. Food Chemistry, 415, 135791.
4. Chen, G., Qi, L., Zhang, S., Peng, H., Lin, Z., Zhang, X., Nie, Q., Luo, W. (2024). Metabolomic, lipidomic, and proteomic profiles provide insights on meat quality differences between Shitou and Wuzong Geese. Food Chemistry, 438, 137967.
5. Sun, X., He, Z., Yang, L., Wu, H., Li, H. (2024). Quantitative proteomic analysis to identify potential biomarkers linked to quality traits of beef tripe from different sources. Food Chemistry, 449, 139224.
6. Chen, R.-C., Dewi, C., Huang, S.-W., Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. Journal of Big Data, 7(1).
7. Bifarin, O. O. (2023). Interpretable machine learning with tree-based Shapley additive explanations: Application to metabolomics datasets for binary classification. PLOS ONE, 18(5).
8. Wang, J., Lin, H., Wang, Q., Wu, Y., Zhang, C. (2025). 4D DIA proteomic analysis of gender and age influences on meat quality and flavor in Hetian white sheep. Food Chemistry, 464, 141851.