

Untargeted Metabolomic Profiling by LC–MS Reveals Discriminant Features Between Vanilla Species

Giovanni Bonatti Bevilaqua¹, Gabriel Franco dos Santos¹, Gesiane Silva Lima¹, Hugo Gontijo Machado¹, Roberto Fontes Vieira², Rosa de Belem das Neves Alves², Luciano de Bem Bianchetti², Lanaia Ítala Louzeiro Maciel¹, Bianca Micaela Macario Gonçalves Acioli¹, Nerilson Marques Lima³, Boniek Gontijo Vaz¹

¹. UFG, Universidade Federal de Goiás; Avenida Esperança s/n, Câmpus Samambaia, CEP 74690-900 Goiânia - Goiás - Brasil.

². EMBRAPA, Embrapa Recursos Genéticos e Biotecnologia; Parque Estação Biológica, Brasília, 70770-901, Distrito Federal, Brasil

³. UNIFAL, Universidade Federal de Alfenas; Alfenas, 37130-001 Minas Gerais, Brasil.

Introduction

The genus *Vanilla* holds significant economic and ecological relevance, with the species *Vanilla planifolia* being the primary global source of natural vanilla flavoring. However, the biochemical diversity within the genus, especially among native Neotropical species, remains largely unexplored. In this context, untargeted metabolomics emerges as a powerful approach to characterize and differentiate closely related species, such as *V. pompona*, *V. phaeantha*, and *V. columbiana*, based on their chemical profiles. This analysis enables not only taxonomic distinction but also the prospecting of biomarkers with biotechnological potential, revealing the influence of genetic and environmental factors (biomes) on metabolic composition. Analysis by liquid chromatography coupled to high-resolution mass spectrometry (LC-MS) is the tool of choice for these studies, as it generates complex chemical "fingerprints" of each sample. The subsequent application of chemometric tools is fundamental to extracting relevant information from this data, allowing for the construction of classification models and the identification of molecules responsible for the differentiation between groups. The objective of this work is to establish a discriminant chemical profile for the studied *Vanilla* species and to identify the most significant ions for this separation.

Methodology

113 samples (3 *V.* species + QCs) were analyzed by LC-MS (Orbitrap). Raw data was extracted from .RAW files via a custom Python script. Data processing included filtering low-intensity signals, base peak normalization, and mass alignment. This created a final 113x1081 data matrix (samples vs *m/z* features). All statistical processing was done with custom Python routines.

Results

A LC-MS investigation analyzed 102 samples from 3 vanilla species (*V. pompona*, *V. phaeantha*, *V. columbiana*) across 4 biomes to differentiate their chemical profiles. The initial unsupervised analysis (PCA) did not reveal any clustering by species or biome. The high natural chemical variability within each species (intraspecific) masked the differences between the groups.

To focus on discrimination, a supervised method (PLS-DA) was used. The models created for the species were excellent, with high predictive power (Q^2 from 0.74 to 0.90) and accuracy (97-100%), proving that each species possesses a robust and distinct chemical fingerprint. In contrast, the models for the biomes failed completely, with low or negative Q^2 values, indicating the lack of a consistent chemical profile that could differentiate them. With the species separated, the next step was to identify

the responsible chemical biomarkers. A dual approach was used: Volcano Plots (for statistical significance) and VIP scores from PLS-DA (for predictive importance). The intersection of these two lists provided the most reliable biomarkers, which were then compared to identify markers unique to each species and those shared among them.

Conclusion

This study successfully used LC-MS and PLS-DA to differentiate three Vanilla species with high accuracy (97-100%), confirming each possesses a unique chemical fingerprint. The method failed to classify samples by biome, showing genetics are more influential than environment in this dataset. A dual statistical approach identified key biomarkers for this separation. Future work should focus on identifying these compounds for species authentication and exploring their biotechnological potential.

Agradecimentos: The authors acknowledge CEMEP for the institutional support and facilities, and thank CAPES, CNPq, FAPEG, and FUNAPE for the financial assistance.