



HAL
open science

Application of a path-modelling approach for deciphering causality relationships between microbiota, volatile organic compounds and off-odour profiles during meat spoilage

Ngoc-Du Martin Luong, Jeanne-Marie Membré, Louis Coroller, Monique Zagorec, Simon Poirier, Stéphane Chaillou, Marie-Hélène Desmots, Dalal Werner, Véronique Cariou, Sandrine Guillou

► To cite this version:

Ngoc-Du Martin Luong, Jeanne-Marie Membré, Louis Coroller, Monique Zagorec, Simon Poirier, et al.. Application of a path-modelling approach for deciphering causality relationships between microbiota, volatile organic compounds and off-odour profiles during meat spoilage. *International Journal of Food Microbiology*, 2021, 348, pp.109208. 10.1016/j.ijfoodmicro.2021.109208 . hal-03210645

HAL Id: hal-03210645

<https://hal.science/hal-03210645v1>

Submitted on 28 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journal Pre-proof

Application of a path-modelling approach for deciphering causality relationships between microbiota, volatile organic compounds and off-odour profiles during meat spoilage

Ngoc-Du Martin Luong, Jeanne-Marie Membré, Louis Coroller, Monique Zagorec, Simon Poirier, Stéphane Chaillou, Marie-Hélène Desmonts, Dalal Werner, Véronique Cariou, Sandrine Guillou



PII: S0168-1605(21)00167-7

DOI: <https://doi.org/10.1016/j.ijfoodmicro.2021.109208>

Reference: FOOD 109208

To appear in: *International Journal of Food Microbiology*

Received date: 18 December 2020

Revised date: 26 March 2021

Accepted date: 18 April 2021

Please cite this article as: N.-D.M. Luong, J.-M. Membré, L. Coroller, et al., Application of a path-modelling approach for deciphering causality relationships between microbiota, volatile organic compounds and off-odour profiles during meat spoilage, *International Journal of Food Microbiology* (2021), <https://doi.org/10.1016/j.ijfoodmicro.2021.109208>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Application of a path-modelling approach for deciphering causality relationships between microbiota, volatile organic compounds and off-odour profiles during meat spoilage.

Ngoc-Du Martin Luong¹ (ngocdu.luong@gmail.com); Jeanne-Marie Membré¹ (jeanne-marie.membre@oniris-nantes.fr); Louis Coroller² (louis.coroller@univ-brest.fr); Monique Zagorec¹ (monique.zagorec@oniris-nantes.fr); Simon Poirier³ (simon.poirier@ifpen.fr); Stéphane Chaillou³ (stephane.chaillou@inrae.fr); Marie-Hélène Desmots⁴ (mh.Desmots@Aerial-Crt.Com); Dalal Werner⁴ (da.werner@aerial-crt.com); Véronique Cariou⁵ (veronique.cariou@oniris-nantes.fr); Sandrine Guillou¹ (sandrine.guillou@inrae.fr).

1. Secalim, INRAE, Oniris, Nantes, France

2. Univ Brest, Laboratoire Universitaire de Biodiversité et Ecologie Microbienne (LUBEM), UMT Alter'ix, Quimper, France

3. Université Paris-Saclay, INRAE, AgroParisTech Micalis Institute, F78352 Jouy-en-Josas, France

4. Aerial, Illkirch, France

5. StatSC, Oniris, INRAE, Nantes, France

Corresponding author

Sandrine Guillou: sandrine.guillou@inrae.fr

Abstract

Microbiological spoilage of meat is considered as a process which involves mainly bacterial metabolism leading to degradation of meat sensory qualities. Studying spoilage requires the collection of different types of experimental data encompassing microbiological, physicochemical and sensorial measurements. Within this framework, the objective herein was to carry out a multiblock path modelling workflow to decipher causality relationships between different types of spoilage-related responses: composition of microbiota, volatilome and off-odour profiles. Analyses

were performed with the Path-ComDim approach on a large-scale dataset collected on fresh turkey sausages. This approach enabled to quantify the importance of causality relationships determined *a priori* between each type of responses as well as to identify important responses involved in spoilage, then to validate causality assumptions. Results were very promising: the data integration confirmed and quantified the causality between data blocks, exhibiting the dynamical nature of spoilage, mainly characterised by the evolution of off-odour profiles caused by the production of volatile organic compounds such as ethanol or ethyl acetate. This production was possibly associated with several bacterial species like *Lactococcus piscium*, *Leuconostoc gelidum*, *Psychrobacter sp.* or *Latilactobacillus fuchuensis*. Likewise, the production of acetoin and diacetyl in meat spoilage was highlighted. The Path-ComDim approach illustrated here with meat spoilage can be applied to other large-scale and heterogeneous datasets associated with pathway scenarios and represents a promising key tool for deciphering causality in complex biological phenomena.

Keywords (6 keywords)

Data integration, Sensory analyses, Modified Atmosphere Packaging, Next-Generation Sequencing, Heterogeneous dataset, Path-ComDim.

Abbreviations

- HTS High-throughput sequencing
- HS-GC-MS Headspace Gas Chromatography Mass Spectrometry
- LAB Lactic Acid Bacteria
- LRM-LV Linear Regression Model on Latent Variable
- MAP Modified Atmosphere Packaging
- OTU Operational Taxonomic Unit
- PLS-LV Partial Least Squares regression on Latent Variable

- VOC Volatile organic compound

Colour should be used for all figures in print.

1 Introduction

Meat spoilage can be defined as any change in the product organoleptic quality during storage, such as off-odours or discolouration, that renders it unsuitable for human consumption (Huis in't Veld, 1996; Pellissery et al., 2019). Meat spoilage is mainly due to the contamination by microorganisms hosted by animals as well as spread into the factory environment or transmitted through human handling (Hultman et al., 2015). During storage, the growth of these microorganisms may lead to defects such as olfactory side-effects associated with undesirable volatile compounds produced from mainly bacterial metabolisms (Casaburi et al., 2015). In particular, off-odours are often a cause of rejection by consumers of spoiled products (Ammer et al., 2009). Therefore, studying meat spoilage process by bacteria requires generally different types of analyses encompassing microbiological, physicochemical and sensorial measurements. Hence, the collected experimental data may represent high-dimension heterogeneous datasets. These datasets can be divided into several "data blocks" according to the associated experimental analyses. Within the meat study framework, several types of analyses can be nowadays investigated, providing then different spoilage responses. Firstly, the characterisation of bacterial microbiota has been widely investigated thanks to recent advances in high throughput sequencing (HTS) methods of 16S rRNA amplicons, giving relative abundances of Operational Taxonomic Units (OTUs) (Cao et al., 2017; Cardinali et al., 2017). Volatilome profiles of spoiled foods defined as the complete set of volatile organic compounds (VOC) generated during storage is quantified using gas chromatography – mass spectrometry (GC-MS) analyses (Holm et al., 2013; Jääskeläinen et al., 2016). Finally, the characterization of off-odours is commonly done by sensory evaluation carried out by trained assessors on the basis of several appropriate attributes (Pereira et al., 2015).

Since the above responses can be nowadays easily measured, a novel challenge consists in integrating these large, heterogeneous, datasets and analysing the relationships between these responses. In general, each data block (microbiota, volatilome, sensory) is analysed separately or by relating one block to another one, using unsupervised or supervised approaches. Unsupervised approaches, which correspond to situations where no dependency relationship has been assumed between blocks (Tchandao Mangamana et al., 2019), have already been applied in the context of microbiology (Ferrocino et al., 2018; Mazerolles et al., 2006; Parlapani et al., 2020). Supervised approaches correspond to predictive modeling where blocks do not play a symmetrical role in the data integration, aiming especially to explain a response block by the others (Cocchi, 2019). Such approaches have become increasingly popular in food science as they can shed light on relationships between responses from different sources and exhibit common dimensions from related datasets (Biancolillo et al., 2019; Borràs et al., 2015).

Integrating multivariate data in meat spoilage can generally serve completely different purposes. One can cite, for example, analyses seeking for data structuration (similarities/differences between meat samples) with Principal Component's Analysis (PCA)-based methods (Argyri et al., 2011; Holm et al., 2013; Vestergaard et al., 2007) or Principal Coordinates Analysis (PCoA) (Raimondi et al., 2018; Säde et al., 2017), discrimination purposes depending on experimental design with discriminant analyses such as Linear Discriminant Analysis (LDA), Partial Least Squares – Discriminant Analysis (PLS-DA) or Multiblock (MB)-PLS-DA (Argyri et al., 2011; Biancolillo et al., 2015; Chen et al., 2014; Girón et al., 2014), or classification purposes (Fontana et al., 2016; Rahkila et al., 2012), etc. Herein, we are interested in evaluating the importance of relationships between data blocks corresponding to different types of instrumental measurements. Relationships in meat spoilage can express supervised situations where the off-odour sensory profiles characterizing the various products under study are assumed to result from the volatilome and/or underlying bacterial metabolic pathways. Such assumptions on causality relationships can be included in the analysis by specifying *a priori* directed (oriented) paths between data blocks. This directional modelling specifically refers to path

modelling approach. Herein, not only the sensory profile of the sausages has to be explained but also intermediate links between microbiota and volatilome may be evaluated (Casaburi et al., 2015; Koutsoumanis and Nychas, 1999). This leads to set up a generalized path scenario where a block can be simultaneously assumed to be explanatory and to be explained by the other blocks. In the case of meat spoilage, such a path structure reflects the extent to which the formation of volatile compounds is the consequence of the microbiota dynamics and the cause of off-odour attributes detected by panellists (Casaburi et al., 2015; Pellissery et al., 2019). Several path fitting methods have been developed in the literature within the framework of Structural Equation Modelling (Bollen, 1989; Kline, 2010; Wardrop and Loehlin, 1987), for example PLS-PM (Lohmöller, 1989), rGCCA (Tenenhaus and Tenenhaus, 2014; Tenenhaus et al., 2017), Path-ComDim (Cariou et al., 2018) or SO-PLS (Menichelli et al., 2014; Naes et al., 2011). Herein, we focus on the application of Path-ComDim which has been already applied in sensometrics and environment (Breitwieser et al., 2018; Cariou et al., 2018) but not in food microbiology yet. The originality of such an approach lies in the determination of saliences associated with each dependency relationship which makes it possible to evaluate assumptions on generalised causal scenario as well as to evaluate the importance of the causality links between data blocks associated with spoilage responses.

The objective of this paper was to present a multiblock path modelling workflow as a data analysis tool within the meat spoilage area for evaluating causality relationships between microbiota composition, volatilome and off-odour sensory profiles. To this end, a large-scale dataset on spoilage of fresh turkey sausages, previously collected (Poirier et al., 2020), was analysed. The data analysis was carried out with Path-ComDim, an extension of the ComDim approach for analysing several data blocks having oriented relations between them (Cariou et al., 2018).

2 Materials and methods

2.1 The raw dataset

The dataset used in this work was extracted from data collected in a collaborative project in which we took part (Redlosses Project, Grant ANR-16-CE21-0006, funded by the French National Research Agency) and published in a data paper (Poirier et al., 2020). The dataset corresponded to several measurements of spoilage responses of fresh turkey sausages during chilled storage. Turkey sausages were industrially produced in ten independent batches with different formulations in each batch, packed under three gas mixtures for packaging, and monitored at four time points (day 2, 8, 15, 22 after production). For simplification purpose, we denoted: “Full dose”, “Half-dose” and “Zero lactate” for the two formulations of lactate (2%, 1% and 0% w/w in turkey sausages, respectively); “Air”, “MAP1” and “MAP2” for the three initial atmospheres used for packaging (Air packaging, MAP1: 70%O₂-30%CO₂ and MAP2: 50%CO₂-50%N₂, respectively). The overall experimental design and all protocols are provided in the data paper.

The extracted dataset used in the present article included three blocks of measurements: (i) microbiota composition monitored by *gyrB* amplicon sequencing, (ii) quantification of VOCs by Headspace Gas Chromatography and Mass Spectrometry (HS-GC-MS) and (iii) sensory profile of the products assessed by trained panellists. Since the experimentations were conducted independently by different laboratories, raw data were structured differently from block to block with more or less samples considered due to differences in techniques, equipment and resources. The dimensions of each block, the data pre-processing workflow as well as the pattern of the directed path between blocks are shown in Figure 1.

Microbiota composition

OTU-based (Operational taxonomic units) microbial diversity analyses were carried out using high-throughput *gyrB* amplicon sequencing in turkey sausages from ten batches, formulated with three initial lactate contents and conditioned under three MAPs at four time points. The total number of samples considered across all batches and all conditions was equal to 360. The extracted raw data included the sequencing read counts of 150 OTUs and the counts had been merged by the authors of the data paper at the 'Species' taxonomy level. These OTUs were assigned according to a recent taxonomic classification proposed by Zheng et al. (2020).

Volatilome

Identification and quantification of VOCs were conducted by comparing gas chromatography retention times and mass spectra with those of standard compounds. Peak area in 1-g samples was used as quantitative data. Raw data in the block 'Volatilome' were obtained from the quantification of twenty-seven detected volatile organic compounds across all measurements in turkey sausages collected from the ten production batches, formulated with two initial concentrations of lactate ("Full dose" and "Zero-lactate"), packed under the three packaging atmospheres, monitored at the four time points and measured in triplicate (technical replicates).

Sensory profile

Sensory profiles were assessed for turkey sausages from six independent batches chosen among the 10 batches after a sensory pre-screening step (Luong et al., 2020). Fresh sausages, with two lactate concentrations ("Full dose" and "Zero lactate"), were packed at day 2 under three different MAPs and assessed afterward at days 2, 8 and 15. The total number of samples considered in sensory analyses across all the above conditions was 80 products ($N = 80$). Off-odours were assessed by a panel of 15 experts who had developed during an 8-month training period a consensus vocabulary to define relevant descriptors for spoiled sausages (see details in Appendix A-2). The off-odours in each sample were then evaluated by at least 12 panellists on the basis of seven descriptors. These descriptors included the global odour of rotten meat, rancid, pungent/sour, ethereal/fermented

fruit, fermented meat/old dry sausage-like, old cheese-like and sulphurous odour. Each panellist rated each descriptor intensity of each sample on a continuous scale from 0 to 10 (from “absent” to “very intense”, respectively).

2.2 Statistical pre-treatments of raw data

Microbiota

Two criteria were applied for filtering OTUs with very low counts as previously done by Bouju-Albert et al. (2017). A normalisation by Centered Log Ratio transformation was performed on the retained OTUs as suggested by Bodein et al. (2019) (see Appendix A.1).

Volatilome

Quantification of each VOC (expressed in peak area) was firstly averaged across the three technical measurements from the same sample. Afterwards, averaged data were \log_{10} -transformed for all VOCs to reduce right-skewness in distributions due to low frequencies of very high counts, since their distributions were right-skewed.

Sensory profile

To take into account the degree of agreement of each assessor with the panel sensitivity to different off-odour descriptors, a weighted average across the panellists of their product ratings was determined by applying the STATIS method (Schlich, 1996). This led to identifying a product configuration compromise where each product is described for each descriptor by a weighted average of the various scores of panellists (see Appendix A.2).

2.3 Multiblock Path-ComDim approach

For the investigation of the relationships between microbiota composition, volatilome and sensory profile, the path-ComDim approach was applied (Cariou et al., 2018). Path-ComDim is an extension of

the unsupervised multi-block method ComDim which aims at simultaneously exhibiting common dimensions shared by the various blocks and, associated with each common dimension, a vector of saliences depicting the degree of agreement of each block with this common dimension (Cariou et al., 2018; Mazerolles et al., 2006; Qannari et al., 2001). Herein, a path modelling strategy was considered with the *a priori* oriented links retrieved from the literature: 1) from “Microbiota” and “Volatilome” to the block “Sensory profile” and 2) from the “Microbiota” to “Volatilome” block (Figure 1). These links were chosen since the off-odours were *a priori* assumed to result from the bacterial development leading to the formation of the VOCs. As all blocks must share the same observations (only variables differ from one block to another), only measurements corresponding to samples evaluated by sensory analyses ($N = 80$) were kept in the path modeling. All blocks were assumed to be column-wise centred.

Path-ComDim criterion is given in Appendix B. For a detailed presentation of Path-ComDim we referred to studies of Cariou et al. (2018, 2017). Synthetically, the rationale behind Path-ComDim analysis is to seek a common latent variable t associated with the explanatory blocks together with a common latent variable u associated with the blocks to be explained, so that they maximize the association between the block components linked by a causal path. Moreover, a salience reflecting the strength of each dependency link is determined jointly with the two common components t and u . This is achieved by a sequential procedure whereby, at each stage, the couple of common components is determined for the corresponding dimension. As output, the relative importance of each dimension in retrieving the structuration of data was computed as well as the explained inertia of each block (variance accounted for each dimension). The strength of the links between blocks in each common dimension was evaluated using its relative salience (expressed in %). In the present study, the significance of each link was discussed using the p-value associated with a Linear Regression model (LRM-LV) as well as R² coefficient from Partial Least Squares between latent variables (PLS-LV) (see Appendix B). To get insights into the contributions of each pair of variables to each common dimension retrieved, indices have been defined which makes it possible to decompose

each salience into pairwise covariance between the variables involved. Considering the dimension d , for each link oriented from the block X_k to X_l , a contribution matrix, denoted $A_{kl(d)}$ was computed in order to evaluate how each pair of variables from X_k and X_l contributes to the determination of the common latent variables of Path-ComDim.

In summary, the Path-ComDim approach determines common dimensions across data blocks “Microbiota”, “Volatilome” and “Sensory profile” in such a way that these blocks are highly associated while assuming causality links between them. The importance of these links was evaluated (tested and quantified) using the relative importance of each dimension as well as the inertia of each block. In each dimension, the importance of each link between blocks was discussed based on its relative salience as well as p-value from LRM-LV model or R2 coefficient from PLS-LV. Within the food microbiology context, for simplification and comprehensive purposes, in this work: (i) the notion of “data structuration” can be interpreted as the different ways that the food samples could be similar or different from each other while taking into account all measured responses, (ii) these above ways were observed by the determination of common “dimensions” which possibly reflect, after exploring the dataset, some factors in the experimental design, (iii) the “salience” of causality links show how important one type of spoilage responses is related to others, and (iv) the “inertia” quantities of each data block correspond to the amount of variance accounted for each dimension, that is the proportion of the block retrieved by each dimension. Finally, specific relationships between spoilage-related responses from different blocks were identified by representing graphically the contribution matrices between blocks. The R software (R Core Team, 2017) was used for the pre-treatments of raw data, executing the Path-ComDim approach as well as for generating different graphs illustrated in this paper. The R scripts are available upon request to the authors.

3 Results

3.1 Spoilage associated responses in the pre-treated dataset

After the pre-treatment step, the number of OTUs in microbial community across all 80 samples retained for the Path-ComDim was reduced from 150 to 86 species, belonging to 37 different genera, according to the recent taxonomic classification proposed by (Zheng et al., 2020). Among these 86 species, the data showed mostly the presence of several Lactic Acid Bacteria such as genera from the family of *Lactobacillaceae* (union from *Lactobacillaceae* and *Leucosporocaceae*), *Lactococcus* as well as different species of *Psychrobacter* and *Pseudomonas*. Besides, some species of *Serratia*, *Enterobacter* and *Aeromonas* were also observed. The abundance observed in the 80 samples under study varied to a large extent from one species to another (the species list and the variance plot for all species is provided in Supplementary data Figure S1). Concerning the volatilome block, the 27 detected VOCs belonged to five groups of chemical compounds including alcohol, aldehyde, ketone, sulphurous and some compounds associated with spices.

3.2 Structuration of the data by Path-ComDim common dimensions

The importance of the determined common dimensions is given by absolute and relative values as in Table 1. The three blocks, Microbiota, Volatilome and Sensory, were mainly structured by the three first dimensions, with relative importance of 74%, 7.2% and 6.4%, for dimension 1, 2 and 3, respectively (Table 1). An important drop in inertia of the subsequent dimensions below 3% was then observed. In the three first dimensions, the variability of the block "Sensory profiles" was the best explained between the three blocks, with a cumulative inertia of 58.5%. This was mainly associated with the first and the third dimensions with 42.4% and 15.2% of explained inertia, respectively. The volatilome profile was also well explained by the three first dimensions with a cumulative inertia of

44.4%. The microbiota composition was the least well explained, with a cumulative inertia of 20.8% mainly associated with the second dimension (9.8% explained).

An example of visualisation of data structuration in Path-ComDim dimensions is given in Figure 2. In Figure 2, projected data on the first and second dimension of Path-ComDim were plotted and coloured according to storage time (2, 8 or 15 days). Such graphs could also be generated for projection in other dimensions and according to other factors such as lactate formulation and modified atmosphere packaging: some graph examples are provided in Supplementary data, Figure S2, since the analysis of these factors was out of scope for the present study. As shown in Figure 2, spoilage-related data were strongly structured according to the storage time in the first dimension (74%), confirming the dynamical nature of spoilage in turkey sausages. Table 1 shows that this dimension explained important proportions of variability in the blocks "Sensory profile" (42.4%) and "Volatilome" (26.1%), suggesting that the spoilage dynamics was mainly highlighted by the production of VOCs over time and the evolution of off-odours. In dimension one, the causality link from volatilome (26.1% explained) to sensory profiles (42.4% explained) was also shown as the most important among the three links: the relative salience associated with this link equalled to 78.3% associated with a significant LRM-LV coefficient ($p < 0.05$) (Table 1). In contrast, dimension one only explained 5.2% of the variability in the microbiota block. This low value of inertia might be associated with the high number of species in the microbiota compared to the number of samples. Moreover, all species might not contribute to the spoilage dynamics or might exhibit nonlinear evolutions. On the other hand, the link oriented from the microbiota block to volatilome (26.1% explained) was significant (LRM-LV p -value < 0.05), suggesting then that 5.2% of the variability in the microbiota could explain the observed change in the volatilome structure.

The second dimension explained the best the microbiota block with inertia of 9.8%, *i.e.* almost two-fold higher than for the other two dimensions. This dimension also explained 14.1% of the volatilome. Nevertheless, unlike dimension one, dimension two (relative importance estimated at

7.2%) did not allow the data to be clearly separated according to storage time, atmosphere or lactate formulation (Supplementary data, Figure S2 a,b,c). Finally, the third dimension showed respectively 5.8%, 4.2% and 15.2% of inertia in each block “Microbiota”, “Volatilome” and “Sensory profile”.

3.3 Causality relationships between specific spoilage-associated responses

With regard to dimension one, the contribution matrices corresponding to the path directed links are plotted in the heatmap shown in Figure 3. Figure 3 highlights the pairwise variables which contributed the most to the determination of each dimension and then causality relationships between specific responses from the different blocks. Data can be structured differently in different dimensions, these matrices appeared thus to be different from one to another dimension, thus giving specific contribution profiles for each dimension.

As already mentioned, the explained inertia was the highest for the blocks sensory profile (42.4%) and volatilome (26.1%) in the first dimension of Path-ComDim, which highlighted the link oriented from volatilome to off-odours (relative salience 78.3% and LRM-LV $p < 0.05$). The overall odour of rotten meat appeared to contribute the most to this dimension compared to its contribution value as shown in Figure 3. This off-odour descriptor was related to different volatile molecules, such as ethanol or ethyl acetate (Figure 3, right graph). Since the data were well structured by storage time in dimension one, these results suggested that the spoilage dynamics was characterized by the evolution of volatilome leading to off-odour profiles, in particular by the production of ethanol and ethyl acetate probably causing the overall odour of rotten meat. The dimension one also highlighted the causality link from microbiota to volatilome (relative salience 4.9%, LRM-LV $p < 0.05$, Figure 3). In spite of the low inertia in microbiota (5.2% explained), plotting contribution matrices pairing microbiota and volatilome can represent a useful tool for understanding microbial mechanisms involved in spoilage process. In this respect, as illustrated in Figure 3 (left graph), the production of ethanol and ethyl acetate were strongly linked to different LAB species such as *Lactococcus piscium*, *Leuconostoc gelidum*, *Latilactobacillus fuchuensis* as well as several *Serratia* and *Psychrobacter*

species. The species *Psychrobacter*, *Lc. piscium* and *L. gelidum* were also shown to be linked to other detected molecules such as 1-propanol, 3-methyl-1-butanol as well as 2-butanone. The above observed links could likely correspond to different fermentation products of LAB over time. In dimension one, since the data were strongly structured depending on the sampling time, the fermentation process of LAB may play an important role in the spoilage dynamics. Besides the LAB species mentioned above, contributions of several *Serratia* species were also highlighted, and the production of ethyl acetate or ethanol could correspond to fermentation products by these species.

In the same way as described above, one can visualise contribution matrices for other dimensions of Path-ComDim as graphical tools for understanding causal relationships between the blocks or the responses. For instance, matrices contribution associated with the second and third dimension are shown in Figure 4 and Figure 5 respectively. In Figure 4 (dimension 2), the contribution matrix pairing the blocks 'Microbiota' (9.8%) and 'Volatilome' (14.1%) showed strong contribution of the quantity of acetoin, diacetyl and some LAB species, for example, *Dellagليا algida*, *Latilactobacillus curvatus*, *Latilactobacillus sakei* as well as *Prostobacterium iliopiscarium*, *Pseudomonas versuta* and *Psychrobacter*. In the same way, in Figure 5, one can observe important contributions of *Aeromonas salmonicida*, *Dellagليا algida* as well as the production of 2-methylbutyraldehyde. Finally, the molecules α -thujene, α -pinene, sabinene, α -phellandrene and α -terpinene seemed to be neither related to the microbiota composition nor to the off-odours descriptors, since their contribution values were low whatever the dimension. This result was expected since these molecules originated from the spices mix of the sausages formulation. These spices-related molecules appeared with small variabilities across different conditions and across different time points as well, because concentration of spices mixes in sausages did not vary over time.

4 Discussion

In this paper, a multiblock path modelling approach was carried out in order to evaluate causality relationships between different types of responses within a meat spoilage framework. The major novelties in our data analysis approach with Path-ComDim consisted in including assumptions of oriented connections between data blocks while taking into account the particularities in each block by appropriate statistical pre-treatment. These causality assumptions were tested and quantified for a better biological interpretation of relationships between blocks. Model outputs made it possible to also assess the contribution of a variable to the strength of each causal relationship. To the best of our knowledge, these tests and quantifications have never been carried out in the meat spoilage area. Such assumptions help identify significant patterns of specific spoilage-related relationships complementary to other unsupervised approaches such as correlation analyses, using a graphical representation of contributions matrices. In the present study, correlations were also pre-assessed between responses from different blocks of pre-treated data using Spearman rank sum correlation, independently from Path-ComDim (Supplementary data, Figure S3, S4). The correlations profiles between blocks did not clearly show any particular link between one or several specific responses. Spearman correlation based on the rank sum correlation is generally useful in cases where relationships between variables are non-linear (as for microbiota composition, production of metabolites, etc.). Path-ComDim enables to take into account additional differences in variabilities between responses, unconsidered in Spearman correlation, since this approach is multidimensional with a global optimisation criterion based on squared covariance. Therefore, among a large number of responses highly correlated together, Path-ComDim may shed light to more specific relationships from one response to another by including variabilities in the model.

It is important to note that, in the present study, we focused particularly on the evaluation of causality between different types of spoilage responses, rather than looking for a way to discriminate / separate meat samples depending on initial factors of experimental design such as lactate

formulation or atmosphere, hence the choice of Path-ComDim. Unlike other commonly used data integration approaches such as PLS-DA, sPLS-DA or MB-PLS-DA (Biancolillo et al., 2015; Cao et al., 2016; Panagou et al., 2014) for which these factors should be included as model input, Path-ComDim can be run without requiring these inputs since its algorithm was not intended to separate samples. It still gives a possibility to visualise the structuring of the data from its outputs according to factors as we have illustrated here and could provide complementary results compared to other data integration approaches.

Among the initial large number of spoilage responses (more than 100 variables from three different blocks) included in model input, our results enabled to identify specific microbial and spoilage-related responses globally consistent with the literature. For instance, Path-ComDim identified many alcohol molecules such as ethanol, isopropanol, 1-propanol or 3-methyl-1-butanol, which have been found as some of the most common alcohols in naturally spoiled meat and in model meat systems inoculated with different spoilage microorganisms (Casaburi et al., 2015). Alcohol molecules are likely to be associated with fermentation processes by LAB as suggested in a previous work of Nordvi et al. (2007). For instance, ethanol or 3-methyl-1-butanol have been associated with LAB species such as *L. carnosum* or *C. divergens* since they have been detected in monocultures of these species in a study of Vasilopoulos et al. (2010). Ethanol is one of the end-products of LAB heterofermentation such as by *L. gelidum* or *L. fischeri* (Matamoros et al., 2009) and is a precursor of ethyl esters such as ethyl acetate, also identified in the present study. *Lc. piscium* and *L. gelidum* were shown to contribute mainly to the link between the microbiota and the volatilome blocks, which is consistent with their fermentative capabilities (Jung et al., 2012, Marché et al., 2017). 1-propanol has been detected in spoiled beef samples in the presence of one or several LAB species such as *Lactobacillus sakei*, *Lc. piscium*, or *Carnobacterium divergens* (Hernández-Macedo et al., 2012), also identified in our study. As expected, our results mainly highlighted the importance of the fermentation process in spoilage dynamics of fresh turkey sausages. Another example of deciphering the spoilage mechanisms of meat using the approach carried out here was the production of acetoin

and diacetyl. Acetoin, the ketone the most found in meat no matter the storage conditions, may also originate from the glucose catabolism of *B. thermosphacta* under aerobic atmospheres (Casaburi et al., 2015; Pin et al., 2002). Acetoin could also come from the microbial degradation of aspartate (Ardö, 2006), as well as a consequence of heme-mediated respiration metabolism by *L. gelidum* (Pothakos et al., 2015). Acetoin and diacetyl have also been associated with natural spoiled meat under air packaging (Dainty and Mackey, 1992; La Stora et al., 2012). Finally, the analysis with Path-ComDim also identified causality relationships between some volatile compounds and several bacterial species/groups even if these species were generally not abundant in microbiota of spoiled meat such as for *Serratia*, *Aeromonas* or *Psychrobacter*. *Serratia* spp. has been reported as belonging to *Enterobacteriaceae* species commonly present on work surfaces in the processing of meat or meat taken from slaughterhouses (Casaburi et al., 2015; Douberaki et al., 2011). Contamination and growth of these species of *Enterobacteriaceae* during storage must correspond to a possible fermentation of glucose (Illeghems et al., 2011; Rafii, 2014). For instance, Illeghems et al. (2015) suggested that several *Enterobacteriaceae* species could be associated with mixed-acid fermentation by methylglyoxal pathway converting glucose into acid lactic, as well as the production of acetate, lactate, succinate and ethanol. A recent study by Wang et al., (2017) also figured out that *S. liquefaciens* and *A. salmonicida* showed important activities on chicken meat, leading to the production of alcohol molecules such as 1-propanol or 3-methyl-1-propanol. Regarding *Psychrobacter* spp., to our best knowledge, there is little information concerning their possible spoilage-associated metabolisms. These results, consistent and complementary to the literature, showed the promising capacities of Path-ComDim and, in general, path modelling methods, as a data integration tool for evaluating causal relationships in biological phenomena such as meat spoilage.

In this work considering three data blocks, the sensory and volatilome blocks were globally well explained by Path-ComDim. The limited inertia in the microbiota could probably be due to the important number of variables in this block compared with the others. Another possible limitation for exploring the composition of the microbiota was indeed the complexity of the intra-block

dynamics of the bacterial community, in particular with regard to interspecies interactions, some species possibly using substrates produced by other species. Complexity in the integration of microbiota and volatilome data could also be translated by the fact that different bacterial species could have similar metabolic pathways, leading to similar end-products, and should probably then contribute to the formation of the same volatilome. Finally, although primarily associated with microbial growth, quality changes resulting from meat spoilage and production of volatile compounds may also be associated optionally with other chemical mechanisms independent of microbial metabolisms such as lipid and protein oxidation due to lipases and proteases. However, as our results point out, once the volatilome is structured, the off-odour outcome profiles could be well explained.

Another potential limitation could lie in the possible influence of some VOCs on the microbiota development over time. In the present study, the unidirectional influence oriented from the microbiota towards VOCs was assumed since VOCs are mainly produced from the microbiota development. The reverse influence from VOCs towards microbiota attesting to the potential effect of certain VOCs on the microbiota development, was not considered because only some VOCs and not the entire block of VOCs would be likely to influence the microbiota development. Through the present paper, we aimed to focus on the presentation of the statistical approach to explore a multiblock dataset especially when prior information is limited. This approach could then be adapted to further studies by experts in microbiology field in order to decipher more specific causality relationships, for example by splitting data into smaller sub-blocks and/or by adapting the corresponding path diagram in the model.

In this paper, we aimed at suggesting a generic data integration workflow for evaluating causality for complex path scenarios occurring in meat spoilage using Path-ComDim. At first glance, the notion of “causality” used herein could appear as a strong wording for correlations. However, directed (oriented) relationships between the different blocks were carefully designed and included in the

statistical model by the means of a path diagram. Thus, relationships between blocks can then be encoded from one block to another one, without considering the reverse direction (orientation). The proposed path diagram mainly aimed to evaluate and assess these directed relationships designed from initial global knowledge about microbiological spoilage (microbiota, volatiles, etc.). Such path diagram, which is asymmetrical by design, represented a key step in the model, since oriented relationships between blocks could be interpreted further than correlations which do not consider any orientation. Finally, the sequential procedure of Path-ComDim for determining common dimensions, while taking into account the above path diagram, also considered distinctly the explanatory blocks and the blocks to be explained as well.

Herein, only three blocks were considered, but this workflow could be easily extended for other scenarios with more blocks and/or more connections to test. For example, to overcome the limited inertia observed in microbiota, one could subdivide microbiota into several blocks corresponding to bacterial genera. The reduced number of responses in each block as well as new causality connections may thus provide interesting additional results. Another possible subdivision of blocks could be done depending on factors of experimental design (such as lactate or atmosphere) in order to take into account these factors, which could be an alternative approach to other existing discriminant ones. The main goal of this work was to explore a large-scale dataset for better understanding causality relationships, and the approach carried out enabled interpreting several causality relationships. The relationships observed from the results were globally consistent with the literature. It is worth noting that these relationships were not “discovered” from a biological or chemical point of view. Notwithstanding, one novelty in our approach consisted herein in the way that the model was set-up: we arbitrarily included little expert information *a priori* from the beginning in order to conduct a data-driven approach, providing consistent results eventually. Regarding these results in such an approach, we are convinced that the approach illustrated here with meat spoilage may be applied to other path scenarios or other datasets as a key tool for

deciphering causality in complex biological phenomena through the analysis of a large-scale heterogeneous dataset.

In a context where numerous multivariate tools can be mainly used in recent food microbiology studies, we think that the success key for exploring spoilage, especially for non-modeller researchers, is to determine the main purposes from the very beginning of the analysis. These purposes (classification, discrimination, seeking for similarities, causality, etc.) should be defined before manipulating multivariate data because the research questions will undeniably condition the choice of statistical techniques (and not the reverse). Afterwards, the dataset can be investigated thoroughly (type of responses, scale, variability, number of samples, etc.) and pre-treatments (pre-processing or data transformation), should be undertaken carefully depending on the data type. These pre-treatments could represent a key step in the overall workflow as they can greatly improve the reliability of data integration. Depending on the considered purposes, the choice of one integration method can only be made afterwards, for example investigating data configuration with PCA or PCoA, discriminant purposes with LDA or PLS-DA, etc. Based on these purposes and the chosen methods, other biological assumptions can be included and tested in models with expert knowledge, for example the orientation of different pathways between blocks in our study. To our knowledge, a “universal” statistical workflow for exploring spoilage or other biological phenomena with multivariate data from pre-processing to dedicated integration steps unfortunately does not yet exist. However, several generic frameworks have been proposed for processing different kinds of data, for example the integration of microbiome data (Bodein et al., 2019), data fusion strategies in sensory evaluation in food (Biancolillo et al., 2019, 2015; Cariou et al., 2019). Thus, in view of possible difficulties due to different processing steps, we focused merely on exploratory aspect in our illustrated workflow. In respect of the above, Path-ComDim seems to be a promising workflow since it provides not only the comprehensive structuring of spoilage profiles but also the most important relationships between specific responses. For further studies, validating the model proposed herein on external data and building generic predictive models could undoubtedly be of primary interest to

microbiologists, even if they must be handled with great care, especially for data pre-processing steps. Although there were no predictive insights developed in our study, several perspectives could be investigated. Indeed, on the basis of the obtained common components, approaches such as regression model on latent variables involving cross-validation strategies, could be undertaken for a predictive purpose, for example in estimating off-odour intensities from microbiota and VOCs compositions.

Funding sources

This work was funded by the French Research National Agency (ANR, REDLOSSES project (ANR-16-CE21-0006)).

Acknowledgments

The authors are grateful to collaborators of the REDLOSSES consortium for their helpful comments and expertise for the experiment design. Émilie Cauchie, Marie-Christine Champomier-Vergès, Tatiana Defosse, Carole Feurer, Rodérick Gohier, Emmanuel Jaffrès and Sabine Jeuge.

Conflicts of interest: none

Authors contribution

- Conceptualization: S. Guillou; N.-D.M. Luong; V. Cariou; J.-M. Membré; L. Coroller
- Data curation: M. M. Desmonts, S. Guillou, N.-D.M. Luong, S. Poirier, D. Werner, M. Zagorec;
- Formal analysis: V. Cariou, S. Guillou, N.-D.M. Luong;
- Funding acquisition: M. Zagorec;
- Methodology: V. Cariou, S. Guillou, N.-D.M. Luong;
- Project administration: S. Guillou, M. Zagorec;
- Software: V. Cariou, N.-D.M. Luong;
- Supervision: V. Cariou, L. Coroller, S. Guillou, J.-M. Membré;
- Visualisation: N.-D.M. Luong;

- Writing – original draft: N.-D.M. Luong;
- Writing – review and editing: V. Cariou, S. Chaillou, L. Coroller, M.-H. Desmonts, S. Guillou, J.-M. Membré, D. Werner, M. Zagorec.

Appendices

A – Data pre-treatment of the block

A.1 Microbiota

Only raw data associated with microbiota in turkey sausages were considered from the data paper published by Poirier et al., (2020). These data could present a high number of zero or very low counts in several OTUs possibly when (i) the latter were observed exclusively in pork samples and/or (ii) because of the very low counts of non-dominant bacterial species as well as their low occurrence frequency across all samples. Two criteria were then considered for low count data removal. Firstly, only OTUs with non-zero counts in at least 2% of the total number of samples were kept, as previously done by Bouju-Albert, Pilet, & Guillou (2017). Afterwards, only OTUs with a total count higher than 0.05% of the total count from all OTUs across all samples were kept (arbitrary threshold). This resulted in a decrease of the total number of OTUs from 150 to 86. For each considered sample, a normalization by Centered Log Ratio transformation (CLR) was performed as suggested by Bodein et al. (2019).

A.2 Sensory profiles

The entire procedure for sensory analysis was carried out according to the Quantitative Descriptive Analysis (QDA) by a panel of 15 persons ($n_{panel} = 15$). The descriptors used for the analyses were proposed and preliminarily selected by the panel in several steps. Firstly, each panellist proposed potential off-odour descriptors after sausage samples sniffing, and 59 descriptors in total were gathered. These descriptors were afterwards scored by the panellists during separated sessions. Analyses of Variance (ANOVA) were then performed on these scores in order to select discriminating

descriptors against products and led to retain 19 descriptors. A second evaluation using these 19 descriptors was carried out and 7 descriptors were finally retained on the basis of analysis of variance as previously done as well as expert elicitation. Panellists were eventually trained on the 7 retained descriptors: tainted meat odour (global spoilage odour), rancid, sour/pungent, ethereal/ fermented fruit, fermented meat/ old dry sausage-like, old cheese-like and sulphurous odour. References used in the training of the panellists are described below.

Descriptor	Natural food or chemical substance used for the training of the panellists	Corresponding scale
Tainted meat odour	Turkey meat spoiled at day-22	Very intense
Rancid	Pork fat with rancid meat	Medium
	Rancid pork fat	Very intense
Sour/pungent	8% white alcohol vinegar	Very intense
Ethereal/ fermented fruit	1% (v/v) ethyl acetate solution	Very intense
	CAS 141-78-6 (ref: W241415,-Sigma-aldrich)	
Fermented meat/ old dry sausage-like	Dry sausage (saucisson sec)	Medium
	Dry pork sausage (rosette)	Very intense
Old cheese-like	1% (v/v) 3-methylbutyraldehyde	Very intense
	CAS 590-86-3 (ref: W269204, sigma-aldrich)	
Sulphurous odour	Cooked egg yolk aged at room temperature for 10 days	Very intense

For these last six attributes, the subjects validated for each of them a chemical reference (nature and concentration) or food reference among an initial choice of 15 references with 1 to 5 possible concentrations. Panellists were trained to recognize the different odours and to rate their intensity.

After training sessions, the sensory profile was carried out for evaluating the 7 retained descriptors on 80 products ($N = 80$) across all two conditions of lactate, 3 MAP and 3 time points. Each product (randomly chosen during different analysis sessions) was evaluated by at least 12 panellists. All panellists did not evaluate the same number and type of products. Firstly, data with off-odour score evaluated by the three 'less participating' panellists mentioned above were removed from the raw dataset. The remaining data were then divided into 12 subsets corresponding to 12 remaining 'relevant' panellists ($N_{panellists} = 12$), the aim was then to determine the consensus profile associated with the panel by weighting each remaining panellist according to his degree of agreement with the panel. The STATIS method was carried out (Llobell et al., 2018; Pizarro et al., 2013) to seek the best consensus by weighting each assessor scores with a weight, denoted $\alpha_k (k = 1, \dots, N_{panellists})$. These weights were obtained by computing the first eigenvector of the matrix of the RV coefficients between the 12 data subsets (RV coefficient between two subsets can be interpreted as a generalized correlation index between the associated panellist matrices). The STATIS method required that the scores need to be measured on exactly the same products (Llobell et al., 2018). In our case, only 40 products were commonly evaluated by the 12 relevant panellists ($N'_{prod} = 40$). Our strategy was then to firstly perform the STATIS method on data associated with these 40 products in order to determine the weights $\alpha_k (k = 1, \dots, N_{panellists})$ corresponding to each panellist, and afterwards to take into account these weights in the calculation of the global score for each product among the initial 80 products (which were not necessarily evaluated by the same panellists).

The STATIS method applied in the 12 data subsets concerning 40 products, denoted $V_{k,k=1,\dots,N_{panellists}}$ was performed using the R package *ClustBlock* (Llobell et al., 2019) to obtain the computed weights $\alpha_k (k = 1, \dots, N_{panellists})$ associated with the 12 panellists. STATIS also computed the inertia of each panellist' matrix by calculating the corresponding Frobenius norm, denoted $F_k (k = 1, \dots, N_{panellists})$, which should be taken into account to deal with the between-panellists

scale factor since the score range effectively used might be different from one panellist to another.

Considering the panelist k , the corresponding Frobenius norm is given by:

$$F_k = \sqrt{\sum_q \sum_r w_{qr(k)}^2}, \#(Eq. VI. 1)$$

where $w_{qr(k)}$ is the $(q, r)^{th}$ entry of the matrix $W_k = V_k V_k^T$.

The calculation of a global score of each product among the initial 80 products should consider the fact that all products were not necessarily evaluated by $N_{panellists} = 12$. Indeed, the product i was rather evaluated by $N_{panellists(i)} \leq 12$, and the panelists could be different from one product to another. In the raw data, the score for the off-odour descriptors $j = 1, \dots, p_s$ ($p_s = 7$) of the product $i = 1, \dots, N$ ($N = 80$) evaluated by the panelist $k = 1, \dots, N_{panellists(i)}$ was denoted $s_{jk(i)}$. Hence, for the product i , the weighted average global score for the off-odour descriptor j ($j = 1, \dots, p_s, p_s = 7$) evaluated by $N_{panellists(i)}$ persons with the corresponding weights $(\alpha_{1(i)}, \dots, \alpha_{N_{panellists(i)}})$ and inertia, denoted $gs_{j(i)}$ is given by:

$$gs_{j(i)} = \frac{\sum_{k=1}^{N_{panellists(i)}} \frac{\alpha_{k(i)}}{F_k(i)} * s_{jk(i)}}{\sum_k \alpha_{k(i)}}. \#(Eq. VI. 2)$$

The global score for all descriptors across all the 80 products were then structured in the block “Sensory profile” ($\dim N * t, N = 80, P_s = 7$).

B – Details for Path-ComDim approach

Let us denote X_1, X_2, X_3 the data blocks “Microbiota”, “Volatilome” and “Sensory profile” measured on the same products, respectively. The given directed links in the path diagram between K blocks ($K = 3$) were encoded using the path matrix $\Delta_{K,K} = (\delta_{kl})$ where $\delta_{kl} = 1$ if we considered a directed link from X_k to X_l and $\delta_{kl} = 0$ otherwise. The Path-ComDim approach determines, in subsequent dimensions $d = 1, 2, \dots$, common components $t_{(d)}$ and $u_{(d)}$ together with the saliences $\lambda_{kl(d)}$ associated with the links going from the block X_k to X_l , by maximising the quantity:

$$C'_{(d)} = \sum_{k=1}^K \sum_{l=1}^K \delta_{kl} cov^2(t_{k(d)}, u_{l(d)}), \#(Eq. VI. 3)$$

where cov stands for the covariance, $t_{k(d)} = X_k X_k^T t_{(d)}$ and $u_{l(d)} = X_l X_l^T u_{(d)}$ are denoted as latent variables (LV) associated respectively with the blocks X_k and X_l connected by a directed link from X_k to X_l . (Cariou et al., 2019, 2018)

The strength of the relationships between blocks in the dimension d of Path-ComDim was evaluated by computing the absolute value of the obtained saliences, $\lambda_{kl(d)}$. The importance of the subsequent dimensions $d = 1, 2, \dots$, denoted $I_{a(d)}$ was expressed by the quantities $I_{a(d)} = \sum_{k=1}^K \sum_{l=1}^K \delta_{kl} |\lambda_{kl(d)}|$. The relative importance of each dimension (expressed in %), was given by $I_{r(d)} = I_{a(d)} / \sum_{d=1,2,\dots} I_{a(d)}$. In the present study, we fixed a total number of 10 subsequent dimensions for the Path-ComDim algorithm. For each dimension, the relative contribution of each link (oriented from X_k to X_l) to the determination of the common components was given by the relative saliences (expressed in %). The significance of each link was evaluated by Linear Regression Models between latent variables determined by Path-ComDim (LRM-LV) (model without intercept since latent variables are centred) as well as by Partial Least Square regression between Latent Variables (PLS-LV) $u_{l(d)} \sim t_{k(d)}$. The significance was then discussed based on the p-value associated with LRM-LV coefficients and the coefficient of determination R² of the PLS model.

Considering the dimension d , for each link oriented from the block X_k to X_l , the contribution matrix, denoted $A_{kl(d)}$ for evaluating the contribution by each pair of variables from X_k and X_l to the determination of the common components of Path-ComDim (with dimensions $p_k * p_l$ corresponding to the number of variables of X_k and X_l) was determined by computing the following quantities:

$$L_{kl(d)} = \lambda_{kl(d)} t_{(d)} u_{(d)}^T, \#(Eq. VI. 4)$$

$$V_{kl(d)} = \begin{pmatrix} v_{11} & \cdots & v_{1p_l} \\ \vdots & v_{ij} & \vdots \\ v_{p_k 1} & \cdots & v_{p_k p_l} \end{pmatrix} = X_k^T L_{kl(d)} X_l, \#(Eq. VI. 5)$$

where $i = 1 \dots p_k, j = 1 \dots p_l$. Finally, the entries a_{ij} at the i -th row and j -th column of the contribution matrix $A_{kl(d)}$ were computed by $a_{ij} = v_{ij}^2 / (\sum_i \sum_j v_{ij}^2)$.

Journal Pre-proof

References

- Ammor, M.S., Argyri, A.A., Nychas, G.-J.E., 2009. Rapid monitoring of the spoilage of minced beef stored under conventionally and active packaging conditions using Fourier transform infrared spectroscopy in tandem with chemometrics. *Meat Sci.* 81, 507–514. <https://doi.org/10.1016/j.meatsci.2008.10.015>
- Ardö, Y., 2006. Flavour formation by amino acid catabolism. *Biotechnol. Adv.* <https://doi.org/10.1016/j.biotechadv.2005.11.005>
- Argyri, A.A., Doulgeraki, A.I., Blana, V.A., Panagou, E.Z., Nychas, G.-J.E., 2011. Potential of a simple HPLC-based approach for the identification of the spoilage status of minced beef stored at various temperatures and packaging systems. *Int. J. Food Microbiol.* 150, 25–33. <https://doi.org/10.1016/j.ijfoodmicro.2011.07.011>
- Biancolillo, A., Boqué, R., Cocchi, M., Marini, F., 2019. Data fusion strategies in food analysis, in: *Data Handling in Science and Technology*. Elsevier Ltd, pp. 271–310. <https://doi.org/10.1016/B978-0-444-63984-4.00010-7>
- Biancolillo, A., Måge, I., Næs, T., 2015. Combining SO-PLS and linear discriminant analysis for multi-block classification. *Chemom. Intell. Lab. Syst.* 141, 58–67. <https://doi.org/10.1016/j.chemolab.2014.12.001>
- Bodein, A., Chapleur, O., Droit, A., Lê Cao, K.-A., 2019. A generic multivariate framework for the integration of microbiome longitudinal studies with other data types. *Front. Genet.* 10. <https://doi.org/10.3389/fgene.2019.00963>
- Bollen, K.A., 1989. *Structural equations with latent variables*. Wiley.
- Borràs, E., Ferré, J., Boqué, R., Mestres, M., Aceña, L., Busto, O., 2015. Data fusion methodologies for food and beverage authentication and quality assessment - A review. *Anal. Chim. Acta.* <https://doi.org/10.1016/j.aca.2015.04.042>

- Bouju-Albert, A., Pilet, M.-F., Guillou, S., 2018. Influence of lactate and acetate removal on the microbiota of french fresh pork sausages. *Int. J. Food Microbiol.* 76, 328–336. <https://doi.org/10.1016/j.fm.2018.06.011>
- Breitwieser, M., Vigneau, E., Viricel, A., Becquet, V., Lacroix, C., Erb, M., Huet, V., Churlaud, C., Le Floch, S., Guillot, B., Graber, M., Thomas, H., 2018. What is the relationship between the bioaccumulation of chemical contaminants in the variegated scallop *Mimachlamys varia* and its health status? A study carried out on the French Atlantic coast using the Path ComDim model. *Sci. Total Environ.* 640–641, 662–670. <https://doi.org/10.1016/j.scitotenv.2018.05.317>
- Cao, K.A.L., Costello, M.E., Lakis, V.A., Bartolo, F., Chua, X.Y., Brozeilles, R., Rondeau, P., 2016. MixMC: A multivariate statistical framework to gain insight into microbial communities. *PLoS One* 11, e0160169. <https://doi.org/10.1371/journal.pone.0160169>
- Cao, Y., Fanning, S., Proos, S., Jordan, K., Srikanth, S., 2017. A review on the applications of next generation sequencing technologies as applied to food-related microbiome studies. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2017.01829>
- Cardinali, G., Corte, L., Robert, V., 2017. Next Generation Sequencing: problems and opportunities for next generation studies of microbial communities in food and food industry. *Curr. Opin. Food Sci.* <https://doi.org/10.1016/j.cofs.2017.09.009>
- Cariou, V., Jouan-Rimbaud Bouveresse, D., Qannari, E.M., Rutledge, D.N., 2019. ComDim methods for the analysis of multiblock data in a data fusion perspective, in: *Data Handling in Science and Technology*. Elsevier Ltd, pp. 179–204. <https://doi.org/10.1016/B978-0-444-63984-4.00007-7>
- Cariou, V., Qannari, E.M., Rutledge, D.N., Vigneau, E., 2018. ComDim: From multiblock data analysis to path modeling. *Food Qual. Prefer.* 67, 27–34. <https://doi.org/10.1016/J.FOODQUAL.2017.02.012>
- Casaburi, A., Piombino, P., Nychas, G.-J.J.E., Villani, F., Ercolini, D., 2015. Bacterial populations and

- the volatilome associated to meat spoilage. *Food Microbiol.* 45, 83–102.
<https://doi.org/10.1016/j.fm.2014.02.002>
- Chen, Q., Li, H., Ouyang, Q., Zhao, J., 2014. Identification of spoilage bacteria using a simple colorimetric sensor array. *Sensors Actuators, B Chem.* 205, 1–8.
<https://doi.org/10.1016/j.snb.2014.08.025>
- Cocchi, M., 2019. Introduction: Ways and means to deal with data from multiple sources. *Data Handl. Sci. Technol.* 31, 1–26.
- Dainty, R.H., Mackey, B.M., 1992. The relationship between the phenotypic properties of bacteria from chill-stored meat and spoilage processes. *J. Appl. Bacteriol.* 73, 103s-114s.
<https://doi.org/10.1111/j.1365-2672.1992.tb03630..>
- Doulgeraki, A.I., Paramithiotis, S., Nychas, G.-J.E., 2011. Characterization of the *Enterobacteriaceae* community that developed during storage of minced beef under aerobic or modified atmosphere packaging conditions. *Int. J. Food Microbiol.* 145, 77–83.
<https://doi.org/10.1016/j.ijfoodmicro.2010.11.030>
- Ferrocino, I., Bellio, A., Giordano, M., Macori, G., Romano, A., Rantsiou, K., Decastelli, L., Cocolin, L., 2018. Shotgun metagenomics and volatilome profile of the microbiota of fermented sausages. *Am. Soc. Microbiol.* 24, 1–15. <https://doi.org/10.1128/AEM.02120-17CE>
- Fontana, C., Bassi, D., López, C., Pisacane, V., Otero, M.C., Puglisi, E., Rebecchi, A., Coconcelli, P.S., Vignolo, G., 2016. Microbial ecology involved in the ripening of naturally fermented llama meat sausages. A focus on lactobacilli diversity. *Int. J. Food Microbiol.* 236, 17–25.
<https://doi.org/10.1016/j.ijfoodmicro.2016.07.002>
- Girón, J., Ivorra, I., SánchezAJ, Fernández-Segovia, I., Barat, J.M., Grau, R., 2014. Preliminary study using visible and SW-NIR analysis for evaluating the loss of freshness in commercially packaged cooked ham and Turkey ham Czech J. Food Sci. 32, 376–383.

<https://doi.org/10.17221/370/2013-CJFS>

Hernández-Macedo, M.L., Contreras-Castillo, C.J., Tsai, S.M., Da Cruz, S.H., Sarantopoulos, C.I.G.L., Padula, M., Dias, C.T.S., 2012. Gases and volatile compounds associated with micro-organisms in blown pack spoilage of Brazilian vacuum-packed beef. *Lett. Appl. Microbiol.* 55, 467–475. <https://doi.org/10.1111/lam.12004>

Holm, E.S., Schäfer, A., Koch, A.G., Petersen, M.A., 2013. Investigation of spoilage in saveloy samples inoculated with four potential spoilage bacteria. *Meat Sci.* 93, 687–695. <https://doi.org/10.1016/j.meatsci.2012.11.016>

Huis in't Veld, J.H.J., 1996. Microbial and biochemical spoilage of foods: an overview. *Int. J. Food Microbiol.* 33, 1–18. [https://doi.org/10.1016/0168-1605\(96\)01139-7](https://doi.org/10.1016/0168-1605(96)01139-7)

Hultman, J., Rahkila, R., Ali, J., Rousu, J., Björkroth J., 2015. Meat processing plant microbiome and contamination patterns of cold-tolerant bacteria causing food safety and spoilage risks in the manufacture of vacuum-packaged cooked sausages. *Appl. Environ. Microbiol.* 81, 7088–7097. <https://doi.org/10.1128/AEM.02225-15>

Illegghems, K., Weckx, S., De Vuyst, L., 2015. Applying meta-pathway analyses through metagenomics to identify the functional properties of the major bacterial communities of a single spontaneous cocoa bean fermentation process sample. *Food Microbiol.* 50, 54–63. <https://doi.org/10.1016/j.fm.2015.03.005>

Jääskeläinen, E., Hultman, J., Parshintsev, J., Riekkola, M.-L., Björkroth, J., 2016. Development of spoilage bacterial community and volatile compounds in chilled beef under vacuum or high oxygen atmospheres. *Int. J. Food Microbiol.* 223, 25–32. <https://doi.org/10.1016/j.ijfoodmicro.2016.01.022>

Kline, R.B., 2010. Principles and practice of structural equation modeling, Third. ed. Guilford Press.

Koutsoumanis, K., Nychas, G.J.E., 1999. Chemical and sensory changes associated with microbial flora

- of mediterranean boque (Boops boops) stored aerobically at 0, 3, 7, and 10°C. *Appl. Environ. Microbiol.* 65, 698–706. <https://doi.org/10.1128/aem.65.2.698-706.1999>
- La Stora, A., Ferrocino, I., Torrieri, E., Di Monaco, R., Mauriello, G., Villani, F., Ercolini, D., 2012. A combination of modified atmosphere and antimicrobial packaging to extend the shelf-life of beefsteaks stored at chill temperature. *Int. J. Food Microbiol.* 158, 186–194. <https://doi.org/10.1016/j.ijfoodmicro.2012.07.011>
- Llobell, F., Cariou, V., Vigneau, E., Labenne, A., Qannari, E.M., 2018. Analysis and clustering of multiblock datasets by means of the STATIS and CLUSTALIS methods. Application to sensometrics. *Food Qual. Prefer.* <https://doi.org/10.1016/j.foodqual.2018.05.013>
- Llobell, F., Vigneau, E., Cariou, V., Qannari, E.M., 2019. *ClustBlock*: Clustering of datasets. R packaging version 1.0.0.
- Lohmöller, J.-B., 1989. Predictive vs. Structural Modeling: PLS vs. ML, in: *Latent Variable Path Modeling with Partial Least Squares*. Springer, pp. 199–226. https://doi.org/10.1007/978-3-642-52512-4_5
- Luong, N.-D.M., Jeuge, S., Coroller, L., Feurer, C., Desmonts, M.-H., Moriceau, N., Anthoine, V., Gagniet, S., Rapin, A., Frémaux, B., Robieue, E., Zagorec, M., Membré, J.-M., Guillou, S., 2020. Spoilage of fresh turkey and pork sausages: Influence of potassium lactate and modified atmosphere packaging. *Food Res. Int.* 137, 109501. <https://doi.org/10.1016/j.foodres.2020.109501>
- Matamoros, S., Pilet, M.F., Gigout, F., Prévost, H., Leroi, F., 2009. Selection and evaluation of seafood-borne psychrotrophic lactic acid bacteria as inhibitors of pathogenic and spoilage bacteria. *Food Microbiol.* 26, 638–644. <https://doi.org/10.1016/j.fm.2009.04.011>
- Mazerolles, G., Hanafi, M., Dufour, E., Bertrand, D., Qannari, E.M., 2006. Common components and specific weights analysis: A chemometric method for dealing with complexity of food products.

- Chemom. Intell. Lab. Syst. 81, 41–49. <https://doi.org/10.1016/j.chemolab.2005.09.004>
- Menichelli, E., Almøy, T., Tomic, O., Olsen, N.V., Næs, T., 2014. SO-PLS as an exploratory tool for path modelling. Food Qual. Prefer. 36, 122–134. <https://doi.org/10.1016/j.foodqual.2014.03.008>
- Naes, T., Tomic, O., Mevik, B.-H., Martens, H., 2011. Path modelling by sequential PLS regression. J. Chemom. 25, 28–40. <https://doi.org/10.1002/cem.1357>
- Nordvi, B., Langsrud, Ø., Egelandsdal, B., Slinde, E., Vogt, G., Gutierrez, M., Olsen, E., 2007. Characterization of volatile compounds in a fermented and dried fish product during cold storage. J. Food Sci. 72, S373–S380. <https://doi.org/10.1111/j.1750-3841.2007.00421.x>
- Panagou, E.Z., Papadopoulou, O., Carstensen, J.M., Nychas, G.-J.E., 2014. Potential of multispectral imaging technology for rapid and non-destructive determination of the microbiological quality of beef filets during aerobic storage. J. Food Microbiol. 174, 1–11. <https://doi.org/10.1016/j.ijfoodmicro.2013.12.026>
- Parlapani, F.F., Ferrocino, I., Michailidou, S., Argiriou, A., Haroutounian, S.A., Kokokiris, L., Rantsiou, K., Boziaris, I.S., 2020. Microbiota and volatiline profile of fresh and chill-stored deepwater rose shrimp (*Parapenaeus longirostris*). Food Res. Int. 132, 109057. <https://doi.org/10.1016/j.foodres.2020.109057>
- Pellissery, A.J., Vinayamohan, P.G., Amalaradjou, M.A.R., Venkitanarayanan, K., 2019. Spoilage bacteria and meat quality. Meat Qual. Anal. Adv. Eval. Methods, Tech. Technol. 307–334. <https://doi.org/10.1016/B978-0-12-819233-7.00017-3>
- Pereira, J.A., Dionísio, L., Patarata, L., Matos, T.J.S., 2015. Effect of packaging technology on microbiological and sensory quality of a cooked blood sausage, *Morcele de Arroz*, from Monchique region of Portugal. Meat Sci. 101, 33–41. <https://doi.org/10.1016/j.meatsci.2014.09.147>
- Pin, C., García de Fernando, G.D., Ordóñez, J.A., 2002. Effect of modified atmosphere composition on

- the metabolism of glucose by *Brochothrix thermosphacta*. *Appl. Environ. Microbiol.* 68, 4441–7. <https://doi.org/10.1128/AEM.68.9.4441-4447.2002>
- Pizarro, C., Esteban-Díez, I., Rodríguez-Tecedor, S., González-Sáiz, J.M., 2013. A sensory approach for the monitoring of accelerated red wine aging processes using multi-block methods. *Food Qual. Prefer.* 28, 519–530. <https://doi.org/10.1016/j.foodqual.2013.01.002>
- Poirier, S., Luong, N.-D.M., Anthoine, V., Guillou, S., Membré, J.-M., Moriceau, N., Rezé, S., Zagorec, M., Feurer, C., Frémaux, B., Jeuge, S., Robieu, E., Champomier Vergès, M., Coeuret, G., Cauchie, E., Daube, G., Korsak, N., Coroller, L., Desriac, N., Desmonts, M.-H., Gohier, R., Werner, D., Loux, V., Rué, O., Dohollou, M.-H., Defosse, T., Chaillou, S., 2020. Large-scale multivariate dataset on the characterization of microbiota diversity, microbial growth dynamics, metabolic spoilage volatiles and sensorial profiles of two industrially produced meat products subjected to changes in lactate concentration and pH. *Data Br.* 105453. <https://doi.org/10.1016/j.dib.2020.105453>
- Pothakos, V., Devlieghere, F., Villani, F., Björkroth, J., Ercolini, D., 2015. Lactic acid bacteria and their controversial role in fresh meat spoilage. *Meat Sci.* 109, 66–74. <https://doi.org/10.1016/j.meatsci.2015.04.014>
- Qannari, E.M., Courcoux, P., Vigneau, E., 2001. Common components and specific weights analysis performed on preference data. *Food Qual. Prefer.* 12, 365–368. [https://doi.org/10.1016/S0950-3293\(01\)00026-X](https://doi.org/10.1016/S0950-3293(01)00026-X)
- R Core Team, D.T., 2017. R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna, Austria.
- Rafii, F., 2014. *Serratia*, in: *Encyclopedia of Food Microbiology: Second Edition*. Elsevier Inc., pp. 371–375. <https://doi.org/10.1016/B978-0-12-384730-0.00304-9>
- Rahkila, R., Nieminen, T.T., Johansson, P., Säde, E., Björkroth, J., 2012. Characterization and

- evaluation of the spoilage potential of *Lactococcus piscium* isolates from modified atmosphere packaged meat. *Int. J. Food Microbiol.* 156, 50–59.
<https://doi.org/10.1016/j.ijfoodmicro.2012.02.022>
- Raimondi, S., Nappi, M.R., Sirangelo, T.M., Leonardi, A., Amaretti, A., Ulrici, A., Magnani, R., Montanari, C., Tabanelli, G., Gardini, F., Rossi, M., 2018. Bacterial community of industrial raw sausage packaged in modified atmosphere throughout the shelf life. *Int. J. Food Microbiol.* 280, 78–86. <https://doi.org/10.1016/j.ijfoodmicro.2018.04.041>
- Säde, E., Penttinen, K., Björkroth, J., Hultman, J., 2017. Exploring lot-to-lot variation in spoilage bacterial communities on commercial modified atmosphere packaged beef. *Food Microbiol.* 62, 147–152. <https://doi.org/10.1016/j.fm.2016.10.004>
- Schlich, P., 1996. Defining and validating assessment compromises about product distances and attribute correlations, in: *Data Handling in Science and Technology*. Elsevier Ltd, pp. 259–306.
[https://doi.org/10.1016/S0922-3487\(96\)80333-5](https://doi.org/10.1016/S0922-3487(96)80333-5)
- Tchandao Mangamana, E., Cariou, V., Vigneau, E., Glèlè Kakai, R.L., Qannari, E.M., 2019. Unsupervised multiblock data analysis: A unified approach and extensions. *Chemom. Intell. Lab. Syst.* 194. <https://doi.org/10.1016/j.chemolab.2019.103856>
- Tenenhaus, A., Tenenhaus, M., 2014. Regularized generalized canonical correlation analysis for multiblock or multigroup data analysis. *Eur. J. Oper. Res.* 238, 391–403.
<https://doi.org/10.1016/j.ejor.2014.01.008>
- Tenenhaus, M., Tenenhaus, A., Groenen, P.J.F., 2017. Regularized generalized canonical correlation analysis: A framework for sequential multiblock component methods. *Psychometrika* 82, 737–777. <https://doi.org/10.1007/s11336-017-9573-x>
- Vasilopoulos, C., De Mey, E., Dewulf, L., Paelinck, H., De Smedt, A., Vandendriessche, F., De Vuyst, L., Leroy, F., 2010. Interactions between bacterial isolates from modified-atmosphere-packaged

- artisan-type cooked ham in view of the development of a bioprotective culture. *Food Microbiol.* 27, 1086–1094. <https://doi.org/10.1016/j.fm.2010.07.013>
- Vestergaard, J.S., Martens, M., Turkki, P., 2007. Analysis of sensory quality changes during storage of a modified atmosphere packaged meat product (pizza topping) by an electronic nose system. *LWT - Food Sci. Technol.* 40, 1083–1094. <https://doi.org/10.1016/j.lwt.2006.06.009>
- Wang, G., Wang, H., Han, Y., Xing, T., Ye, K., Xu, X., Zhou, G.-H., 2017. Evaluation of the spoilage potential of bacteria isolated from chilled chicken *in vitro* and *in situ*. *Food Microbiol.* 63, 139–146. <https://doi.org/10.1016/j.fm.2016.11.015>
- Wardrop, J.L., Loehlin, J.C., 1987. Latent variable models: An introduction to factor, path, and structural analysis. *J. Educ. Stat.* 12, 410. <https://doi.org/10.2307/1165058>
- Zheng, J., Wittouck, S., Salvetti, E., Franz, C.M.A.P., Garras, H.M.B., Mattarelli, P., O'Toole, P.W., Pot, B., Vandamme, P., Walter, J., Watanabe, K., Wuyts, S., Felis, G.E., Gänzle, M.G., Lebeer, S., 2020. A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int. J. Syst. Evol. Microbiol.* 70, 2782–2858. <https://doi.org/10.1099/ijsem.0.004107>

Table 1 Importance of the first common components of Path-ComDim (absolute value of saliences) and contribution of each link to the determination of these components. For LRMLV (Linear Regression Model on Latent Variables) coefficients: p-value significant codes: '*' < 0.05, '**' < 0.01, '***' < 0.001.

Dimension	Importance of the common components		Inertia in each block (proportion explained)			Contribution of each link								
	Absolute I_a	Relative I_r	Microbiota	Volatilome	Sensory profile	Microbiota → Volatilome			Microbiota → Sensory profile			Volatilome → Sensory profile		
d						Relative salience (%)	LRM-LV coefficient	R2 (PLS-LV)	Relative salience (%)	LRM-LV coefficient	R2 (PLS-LV)	Relative salience (%)	LRM-LV coefficient	R2 (PLS-LV)
1	0.1831	74.0%	5.2%	26.1%	42.4%	1.9%	1.75 *	0.48	16.8%	1.56	0.48	78.3%	1.79 *	0.48
2	0.0177	7.2%	9.8%	14.1%	0.4%	39.9%	-0.03	0.01	1.2%	-1.59 *	0.18	8.9%	-1.14	0.19
3	0.0158	6.4%	5.8%	4.2%	15.2%	3.5%	-1.42 *	0.24	57.4%	-0.05	0.06	39.1%	-0.04	0.06
Total			20.8%	44.4%	58.5%									

CAPTIONS**TABLE**

Table 1 - Importance of the first common components of Path-ComDim (absolute value of saliences) and contribution of each link to the determination of these components. For LRMLV (Linear Regression Model on Latent Variables) coefficients: p-value significant codes: '*' < 0.05, '**' < 0.01, '***' < 0.001.

FIGURES

Figure 1 – Data pre-treatment workflow. n_m and p_m (resp. n_v and p_v , n_s and p_s) correspond to the dimension (number of rows and column) of the raw data block: Microbiota (resp. Volatilome, Sensory profile). N_m and P_m (resp. N_v and P_v , N_s and P_s) correspond to the dimension (number of rows and columns) of the pre-treated data block Microbiota (resp. Volatilome, Sensory profile). N correspond to the total number of samples for which sensory analyses were assessed. The dotted arrows correspond to the pre-treatment in each block. The solid arrows correspond to the *a priori* oriented relationships patterns between blocks assumed in the model.

Figure 2 – Example of representation of data structuration (including microbiota, volatilome and sensory profiles) in the two first common dimensions of Path-ComDim (with their relative importance) depending on storage time. The horizontal X-axis (resp. vertical Y-axis) represents the Path-ComDim dimension 1 (resp. 2). The given percentage indices for these shows how each one is important among all determined common dimensions during the overall procedure. The light blue (resp. brown, red) points correspond the storage time of day 2 (resp. day 8 and 15).

Figure 3 – Path-ComDim – first dimension. The upper right graph represents inertia (% exp.) of each block, relative salience (%) and significance of each link between blocks (Solid arrow: LRM-LV p-value < 0.05. Dotted arrow: LRM-LV p-value > 0.05). The lower graph contribution matrices corresponding to the links oriented from microbiota to volatilome and from volatilome to sensory profiles in dim 1.

The size of the diamond-shaped points associated with a pair of responses from two blocks is proportional to the contribution values this pair to the links.

Figure 4 – Path-ComDim – second dimension. The upper right graph represents inertia (% exp.) of each block, relative salience (%) and significance of each link between blocks (Solid arrow: LRM-LV p-value < 0.05. Dotted arrow: LRM-LV p-value > 0.05). The lower graph contribution matrices corresponding to the links oriented from microbiota to volatilome and from volatilome to sensory profiles in dim 1. The size of the diamond-shaped points associated with a pair of responses from two blocks is proportional to the contribution values this pair to the links.

Figure 5 – Path-ComDim – third dimension. The upper right graph represents inertia (% exp.) of each block, relative salience (%) and significance of each link between blocks (Solid arrow: LRM-LV p-value < 0.05. Dotted arrow: LRM-LV p-value > 0.05). The lower graph contribution matrices corresponding to the links oriented from microbiota to volatilome and from volatilome to sensory profiles in dim 1. The size of the diamond-shaped points associated with a pair of responses from two blocks is proportional to the contribution values this pair to the links.

Conflicts of interest: none

The authors declare no conflict of interest.

Journal Pre-proof

Highlight

- Volatilome and off-odour profiles were well explained by the Path-ComDim approach
- The integrated dataset was strongly structured according to storage time
- The causality link oriented from volatilome to off-odour descriptors was significant

Journal Pre-proof

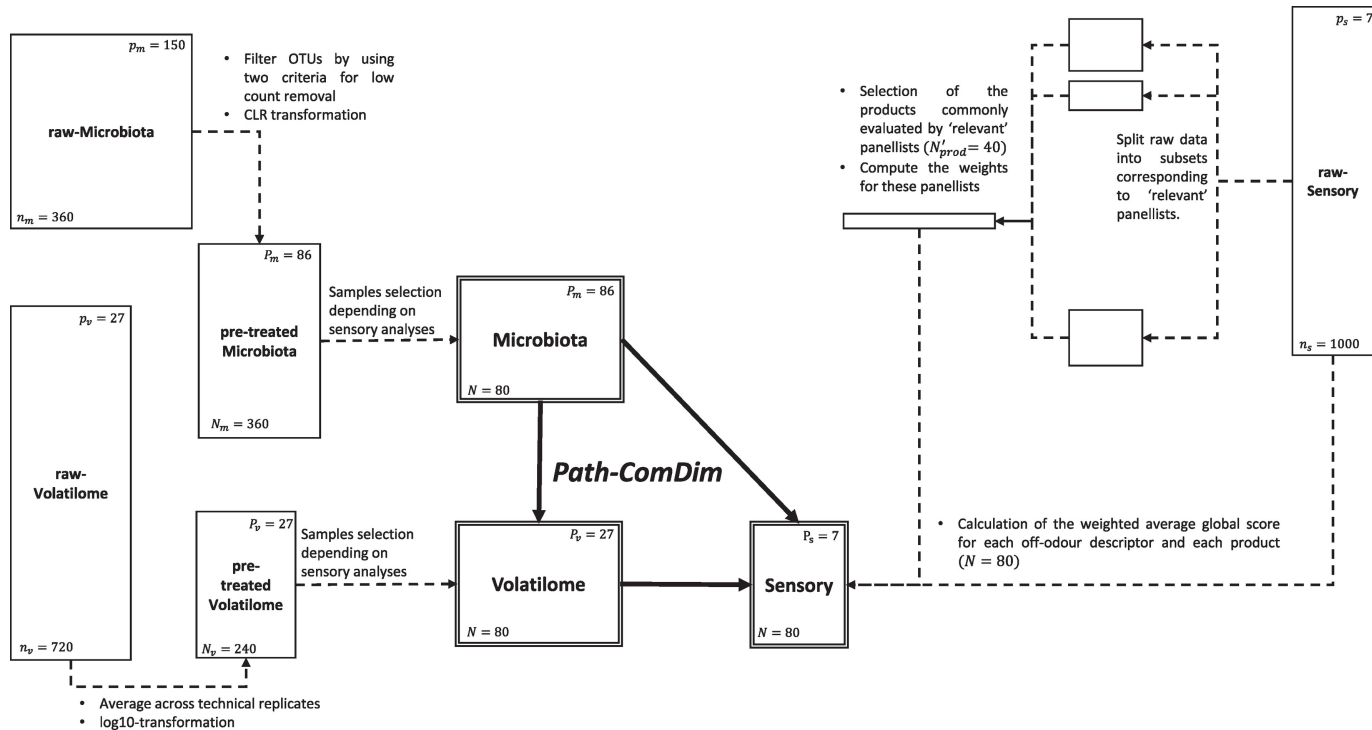


Figure 1

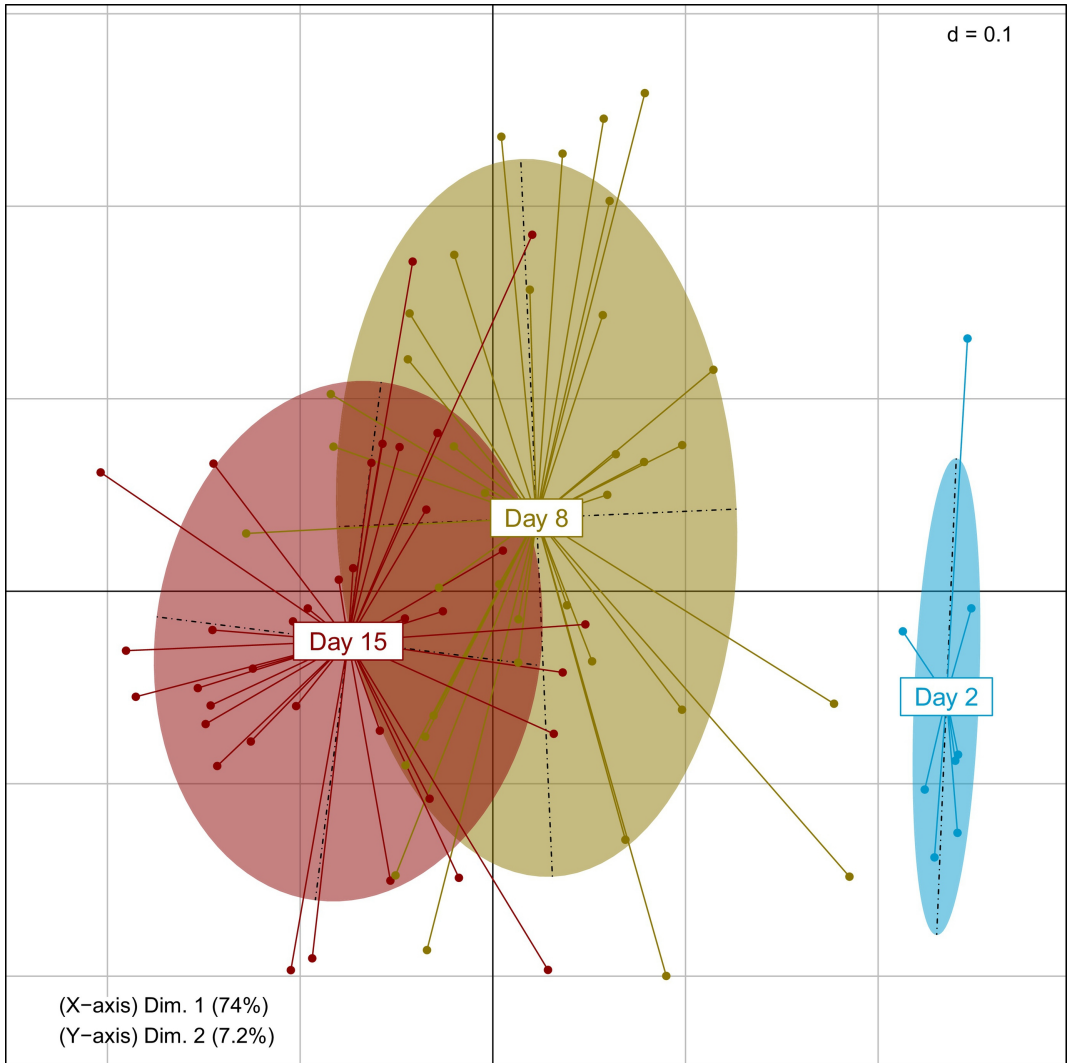


Figure 2

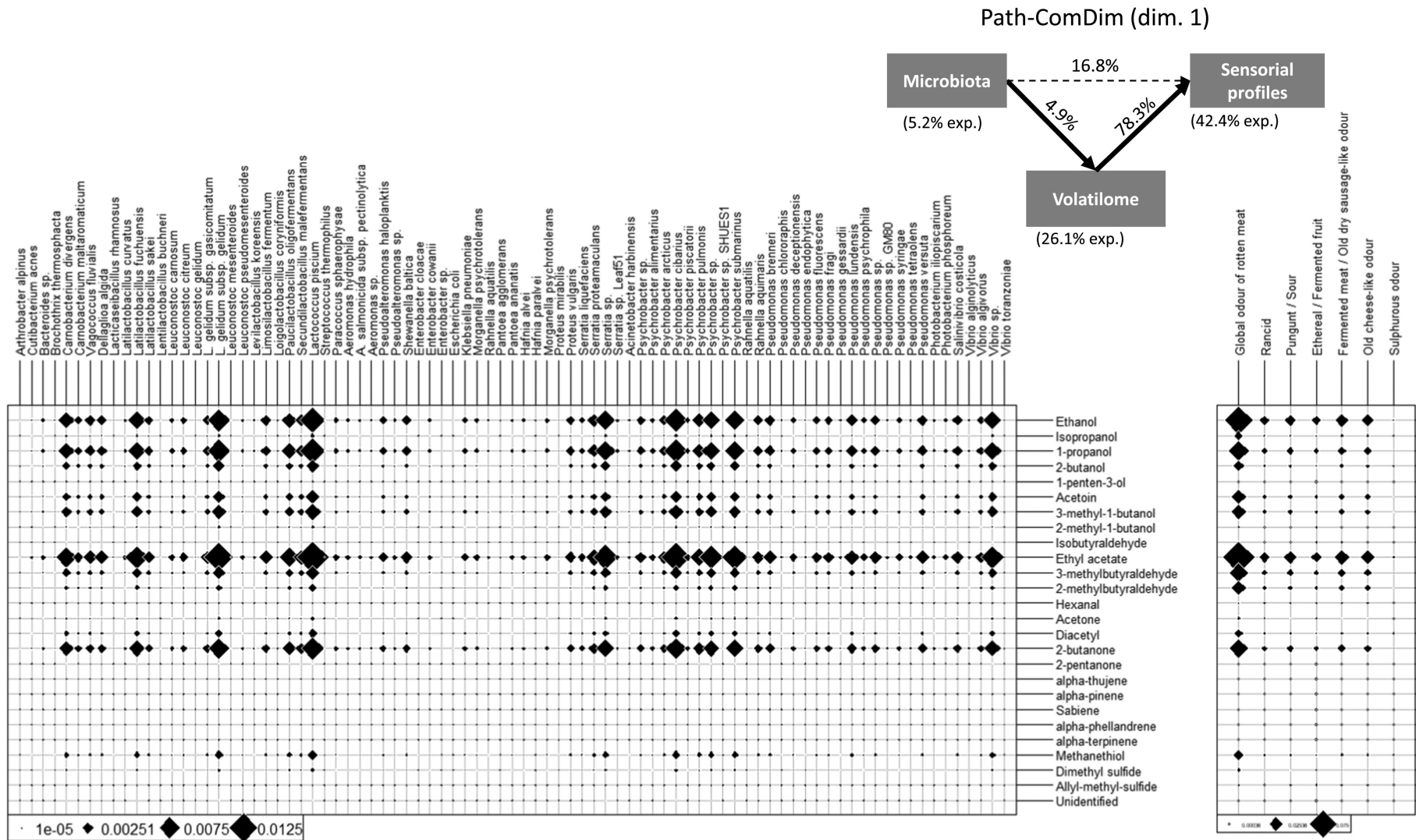


Figure 3

Path-ComDim (dim. 2)

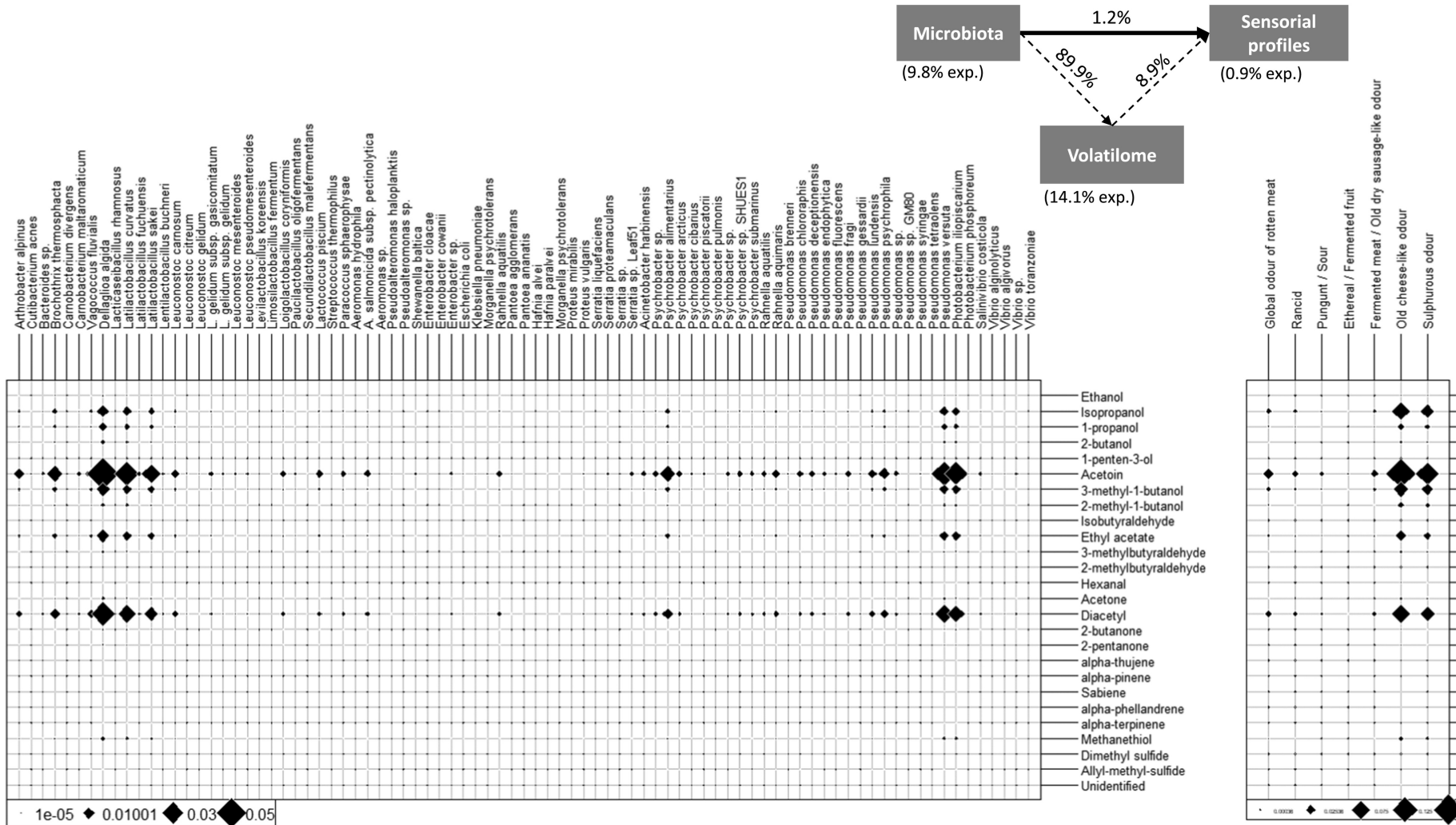


Figure 4

Path-ComDim (dim. 3)

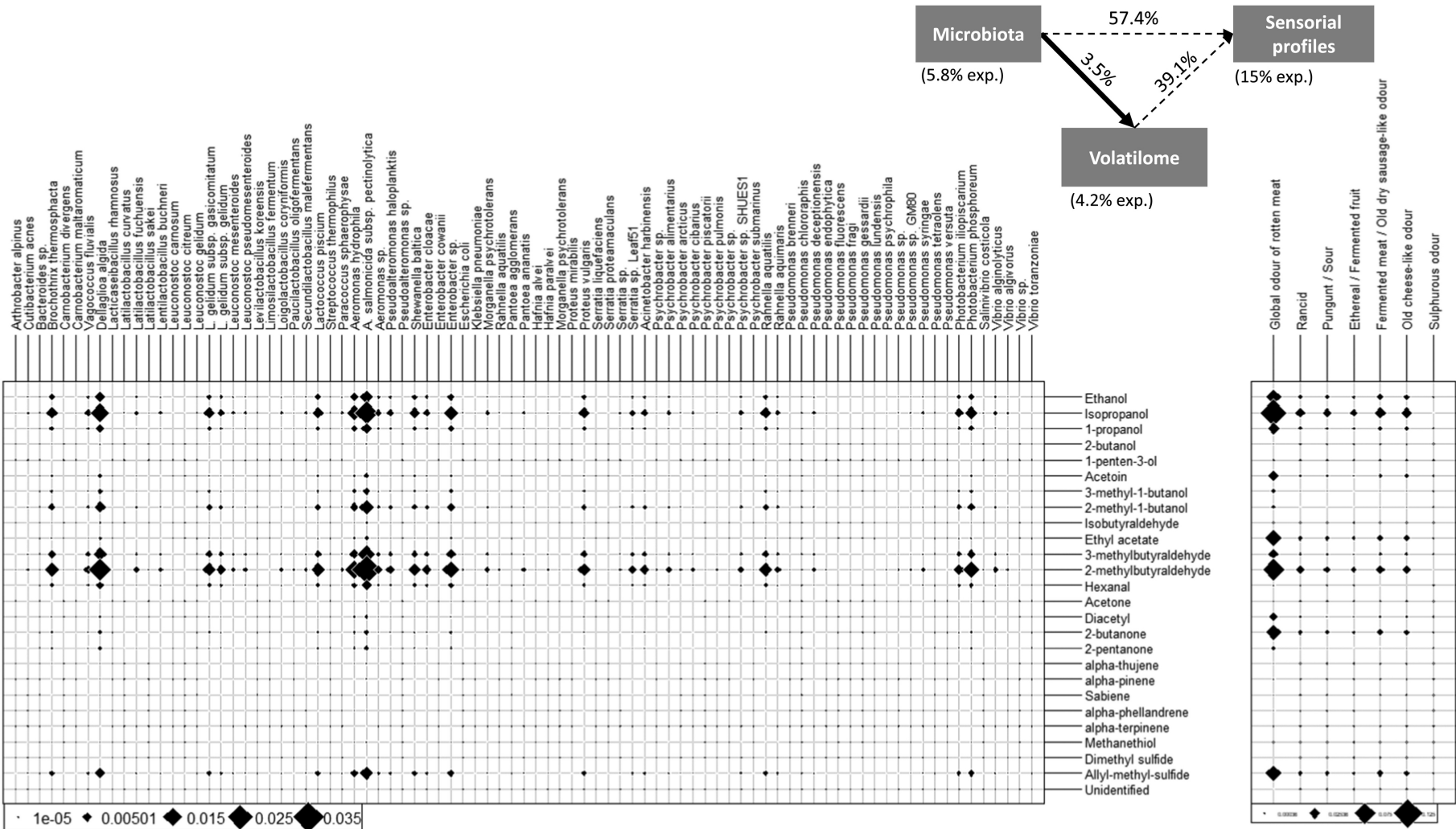


Figure 5