



OPEN

# Exploring bacteria diversity in commercialized table olive biofilms by metataxonomic and compositional data analysis

Antonio Benítez-Cabello<sup>1</sup>, Verónica Romero-Gil<sup>2</sup>, Eduardo Medina-Pradas<sup>1</sup>, Antonio Garrido-Fernández<sup>1</sup> & Francisco Noé Arroyo-López<sup>1</sup>✉

In this work, a total of 72 samples of non-thermally treated commercial table olives were obtained from different markets of the world. Then, prokaryotic diversity in olive biofilms was investigated by metataxonomic analysis. A total of 660 different OTUs were obtained, belonging to Archaea (2.12%) and Bacteria domains (97.88%). From these, 41 OTUs with a proportion of sequences  $\geq 0.01\%$  were studied by compositional data analysis. Only two genera were found in all samples, *Lactobacillus*, which was the predominant bacteria in the biofilm consortium (median 54.99%), and *Pediococcus* (26.09%). *Celerinatantimonas*, *Leuconostoc*, *Alkalibacterium*, *Pseudomonas*, *Marinilactibacillus*, *Weissella*, and the family *Enterobacteriaceae* were also present in at least 80% of samples. Regarding foodborne pathogens, only *Enterobacteriaceae*, *Vibrio*, and *Staphylococcus* were detected in at least 91.66%, 75.00%, and 54.10% of samples, respectively, but their median values were always below 0.15%. Compositional data analysis allowed discriminating between lye treated and natural olive samples, as well as between olives packaged in glass, PET and plastic bags. *Leuconostoc*, *Celerinatantimonas*, and *Alkalibacterium* were the bacteria genera with a higher discriminant power among samples. These results expand our knowledge of the bacteria diversity in olive biofilms, providing information about the sanitary and hygienic status of this ready-to-eat fermented vegetable.

The world's olive grove consists of more than 10 million hectares, of which over 1 million are destined to table olives, which constitute the most important fermented vegetable in the Mediterranean countries, with also noticeable productions in South America, USA and Australia. The last consolidated worldwide table olive balance shows that its consumption in 2018/2019 season was above 2.7 million tons<sup>1</sup>.

Among the various processing methods, alkali-treated olives (Spanish style), ripe olives by alkaline oxidation (Californian style) and directly brined olives (natural black or green olives) are the most common<sup>2</sup>. Besides, the industry makes diverse commercial presentations (mainly whole, pitted, and sliced) and use different packaging material (glass, PET, or bags, sometimes under vacuum) which lead to a great variety of products. Packaged table olives can be stabilized by pasteurization or sterilization, guaranteeing the absence of viable microbial cells. However, many times, they are preserved by their physico-chemical characteristics (pH, free acidity and salt) with or without use of preservatives. In these cases, the presence of microorganisms is usual<sup>2</sup>.

Because of the natural antimicrobial compounds of these fruits and the low pH and high salt levels reached during fermentation, table olives have a long history of microbial safety. However, in the absence of correct and hygienic handling practices, diverse microbial hazards may appear such as botulism, mycotoxins, biogenic amines or foodborne pathogens<sup>3</sup>. Thus, the study and control of the bacteria biodiversity during table olive processing can be useful for the evaluation and management of the safety risk associated with this ready-to-eat fermented vegetable, especially microorganisms with the ability to form biofilms on olive epidermis. Even so, most of the microbial studies on table olives have focused exclusively on the fermentative process, with scarce

<sup>1</sup>Food Technology Department, Instituto de La Grasa (CSIC), Crta Utrera km 1, Campus Universitario Pablo de Olavide. Building 46, 41013 Seville, Spain. <sup>2</sup>Technological Applications for Improvement of Quality and Safety in Foods, Carretera de Marbella nº22. Planta-1, 29108 Guaro, Málaga, Spain. ✉email: fnoe@ig.csic.es

attention to the packaging and distribution phases<sup>4–6</sup>. Furthermore, the most-used approach has been based on culture-dependent techniques, which only allow counting and identifying viable cells<sup>7</sup>. However, the emergence of omic technologies and metataxonomic analysis represented a basic step for improving the knowledge of the microbial presence in foods, as they provide a broader vision of the existing biodiversity which could not be reached through the classical culture-dependent techniques. In table olives, various omic studies have been carried out to determine the bacterial and fungal evolution during the fermentation process<sup>8–14</sup>, but not in the finished products.

Metataxonomic analysis generates a considerable amount of information. However, standard multivariate statistical analyses may lead to, at least, formal incongruences when data are expressed in frequencies<sup>15</sup>. A first approach to analyse metataxonomic data could be Correspondence Analysis (CA)<sup>16</sup>. Metataxonomic results also represent a clear example of Compositional Data (CoDa), which are typically defined as vectors of positive components representing parts of a whole which carry relative information, usually with a constant sum (100 or 1). These conditions render most classical statistical techniques incoherent when applied to compositions, as such tools were devised for random variables in the Euclidean sampling space. The log-ratio approach was introduced to analyze CoDa data<sup>17</sup>. This solution was based on some log-ratio transformations of the original data, followed by the application of the standard techniques. The CoDa concept has already been applied in several fields, from genetic<sup>18,19</sup> to the Spatial exploration<sup>20</sup> among others. Its implementation in metagenomics could lead to more realistic segregation between samples than using standard multivariate methods. Recently, this approach was used to relate different inocula with the bacterial diversity found at the end of the Spanish-style green table olive fermentation<sup>8</sup>.

This work aimed to elucidate the bacteria composition in the biofilms of different commercial table olive samples through a metagenomic approach. Its relationship with the type of elaboration, presentation and packaging material/system, by applying CA and CoDa analysis, is also studied. The information could be useful to assess the hygienic and safety status of this important ready-to-eat fermented vegetable.

## Material and methods

**Sampling of olive packaging.** A total of 72 commercial table olive samples were obtained from different supermarkets of Spain (n = 50), Greece (n = 6), France (n = 4), Chile (n = 4), Peru (n = 4), Portugal (n = 2), and Argentina (n = 2), between 2015 and 2017 years. None of them had undergone heat treatment such as pasteurization or sterilization for preservation, which was exclusively based on their physicochemical characteristics (pH, titratable acidity, and salt) or the use of authorized preservatives. The samples were analysed in the Food Biotechnology Department of Instituto de la Grasa (CSIC) within the first month of shelf life. They were classified according to the type of elaboration (SS: Spanish style n = 36, GN: green natural n = 26, or BN: black natural n = 10), presentation (P: pitted n = 12, W: whole n = 34, or S: sliced n = 26), and packaging material/system (P: PET = 28, B: bag n = 20, G: glass n = 18, or V: vacuum n = 6). They belonged to 16 different olive cultivars (*Manzanilla* n = 16, *Aloreña* n = 14, *Hojiblanca* n = 6, *Gordal* n = 2, *Morona* n = 2, *Verdial* n = 2, *Empeltre* n = 2, *Empeltre Mallorquina* n = 4, *Kalamata* n = 2, *Conservolea* n = 4, *Galega* n = 2, *Picholine* n = 2, *Criolla* n = 4, *Lucques du Languedoc* n = 3, *Azapa* n = 4, and *Arauco* n = 2). Table S1 (supplementary material) shows their references and origins.

**Physicochemical and microbiological analyses.** The brine pH was measured using a Titroprocessor model 670 (METROHM, Switzerland). For determination of NaCl concentration, 0.5 mL of brine was mixed with 100 mL of distilled water. Then, titration of Cl<sup>-</sup> was carried out with silver nitrate (AgNO<sub>3</sub>), using as the indicator a solution of potassium chromate (K<sub>2</sub>CrO<sub>4</sub>). The results were expressed as percentage (w/v) of NaCl.

For determining the viable bacteria populations on the olive surface, 25 g of fruits were washed with sterile saline solution (0.85% v/v) to remove no-adhered cells and pitted at sterile conditions. Then, fruit flesh was homogenised in a Stomacher (SEWARD LABORATORY SYSTEMS, Inc. Bohemia, NY, USA) for 2 min with 100 ml of sterile saline solution. Afterwards, direct or decimal dilutions were plated on selective culture media. Lactic acid bacteria (LAB) were plated on the Man, Rogosa and Sharpe (MRS) agar (OXOID, Basingstoke, Hampshire, England) supplemented with 0.02% sodium azide (SIGMA, St. Luis, USA), while *Enterobacteriaceae* was counted on Crystal-violet Neutral-Red bile glucose (VRBD) agar (MERCK, Darmstadt, Germany). Results were expressed as log<sub>10</sub> CFU/g.

**DNA extraction from olive samples and sequencing.** Twenty-five grams of olives from every sample were homogenized with 100 mL of sterile saline solution (0.9% NaCl) in a Stomacher homogenizer for 5 min, and the aqueous phase centrifuged at 9,000 × g for 15 min. In all cases, the supernatant was withdrawn, and the pellets were washed twice with sterile saline solution before stored at -80 °C until use. Total genomic DNA from fruit samples was extracted and purified using the Power Food Microbial DNA Isolation Kit (MOBIO, Carlsbad, Calif) according to the manufacturer instructions, and sent for sequencing to FISABIO (Valencia, Spain). Before sequencing, the content of purified DNA was measured using a Qubit fluorometer (THERMO FISHER SCIEN-TIFIC, Waltham, USA), always obtaining values above 0.2 ng/μL.

The V3 and V4 region of the 16S ribosomal RNA gene was amplified<sup>21</sup> following the 16S rDNA Gene Metagenomic Sequencing Library Preparation Illumina protocol (Part #1504423 Rev. A). Libraries were sequenced using a 2 × 300 pb paired-end run (MiSeq Reagent kit v3 (MS-102-3001) on a MiSeq Illumina platform, according to the manufacturer's instructions. The quality evaluation of the sequencing was developed by the prinseq-lite program<sup>22</sup> using a minimum sequence length of 50 bp, trim\_qual\_right of 30, triam\_qual\_type of mean and trim\_qual\_window of 20 bp. R1 and R2 read from sequencing were joined using the FLASH program<sup>23</sup> by applying default parameters.

**Metataxonomic analysis.** Metagenomic sequences in \*.fna formats were labelled, concatenated and then analysed using the Quantitative Insights into Microbial Ecology (QIIME) pipeline (version 1.9.1), running Ubuntu v16.04. Sequences were first filtered by quality, excluding from downstream analyses those with a length out from 220 to 300 pb and a mean sequence Phred quality score < 35.

Chimeric sequences were identified and removed using Chimera slayer, and Operational Taxonomic Unit (OTU) tables were generated by using the UPGMA clustering method (Unweighted Pair Group Method with Arithmetic mean) and uclust to assign the OTUs<sup>24</sup>, through the script “Pick\_de\_novo\_otus.py”. SILVA108 was used as a reference database<sup>25</sup> defining OTUs at a level of  $\geq 97\%$  sequence homology. The Pick\_de\_novo\_otus.py script generates an OTU mapping file, a representative set of sequences, a sequence alignment file, taxonomy assignment file, a filtered sequence alignment a phylogenetic tree, and a biom-formatted OTU table.

Cyanobacteria (which are confused with vegetable chloroplasts from olive cells), and mitochondria were excluded from the OTU tables using the script “filter\_taxa\_from\_otu\_table.py”. OTUs with a presence below 0.01% were also removed from the multivariate statistical analysis, but their frequencies are also shown in Table S2 (supplementary material).

The OTU tables were rarefied to 2,600 sequences (lowest number of reads obtained in one sample), and alpha diversity indexes (Chao, Observed Species, Shannon, Simpson, Good’s coverage) were calculated, using the Alpha\_diversity.py script.

**Multivariate analysis.** For each sample, the sum of all assigned OTUs was 100%. Therefore, for comparing associations containing counts of taxa, CA is an appropriate technique<sup>16</sup>. The objective of CA was to determine the similarities between the categories of samples and their populations of bacteria. This multivariate technique is more suitable when expecting that species have unimodal responses to the underlying parameters and become rare for lower parameters<sup>16</sup>. The analysis was performed using Past4 (Hammer et al., 2001) and the R package version<sup>26</sup>.

The microbiome data present a clear compositional structure and could be studied by CoDa<sup>15,27</sup>. Most of the CoDa analyses are based on log-ratio methodologies. Both ratios and logarithms are operations that require non-zero elements in the data matrix. As a consequence, any analysis of a vector of components should be preceded by a treatment of the zeros. For this purpose, CoDa considers three main types: rounded, count and essential<sup>28</sup>. The zeros in our data set match the first type, which applied when the value in the data matrix is not a true zero but rather represents an observed value below a particular rounding-off error or a very low value that cannot be recorded. The strategy to prevent the so-called zero- problem consists essentially of their replacement (imputation) by a small quantity. One of the most common is the multiplicative strategy. It simply consists of replacing each rounded zero in the data set by a small value (habitually 50–65% of the detection limit or the smaller value per variable) and, then, modifying the non-zero values in a multiplicative way<sup>28</sup>. In this work, the zeros were replaced by 65% of the lowest detected level for each variable. Then, diverse CoDa multivariate analyses could be applied. The biplot<sup>29,30</sup> is a method which has been regularly applied to visualize the rows and columns of many different kinds of data matrices. The CoDa biplot was developed by Aitchison and Greenacre (2002)<sup>31</sup>. Two types of biplot are considered: covariance (rows in standard coordinates and columns in principal coordinates), which favour the display of the variables; and form (a reversed type of coordinates), which favour the display of individual. In both biplots, the row and column points are centred at the origin because of the double-centring transformation of the initial matrix. Other exploratory tools used were variation array and sequential binary partition (SBP)<sup>28</sup>. For the ANOVA performance and Discriminant Analysis (DA), the original data (in the Simplex) were subjected to an isometric log-ratio transformation (SBP or pivot coordinates) which transformed them into the Euclidean space, where the traditional statistical techniques can be applied. The statistical software used were: (i) CoDaPack, which is a stand-alone software developed by the University of Girona, Spain<sup>32</sup>, (ii) Package *compositions*<sup>33</sup>, and (iii) *robCompositions*<sup>34</sup>.

## Results

**Physicochemical and microbiological analysis.** All table olive samples were analysed by microbiological and physicochemical methods (Table S1). The average pH value obtained for all of them was  $3.94 \pm 0.63$ , ranging from a minimum of 2.47 (S146 sample) to a maximum of 5.40 (S168). Depending of the type of elaboration, the average pH value was  $4.05 \pm 0.33$  in GN,  $3.96 \pm 0.28$  in BN, and  $3.78 \pm 0.81$  in SS table olives. According to olive legislation<sup>35</sup>, the maximum pH value allowed for SS olives is 4.0 when fruits are preserved by refrigeration, addition of preservatives, modified atmospheres or by their own chemical characteristics. In the case of GN and BN olives, the pH value must be less than 4.3. On the contrary, the average salt concentration value was  $5.97 \pm 1.39\%$ , ranging from a minimum of 3.28% (S139) to a maximum of 10.59% (S178). According to the type of elaboration, the average salt levels in brine were  $5.50 \pm 1.13\%$  in SS,  $6.17 \pm 1.40\%$  in GN, and  $7.21 \pm 1.53\%$  in BN. In this case, the olive legislation<sup>35</sup> establish a lower limit of 4.0% for SS olives preserved by refrigeration or by the addition of preservatives, while this limit must exceed 5.0% when SS olives are preserved by modified atmospheres or by their own chemical characteristics. In the case of GN and BN olives, the salt concentration must exceed 6%.

Because of the lack of thermal treatment, viable bacteria were present during shelf life. The average LAB population obtained for all treatments after the first month was  $3.98 \pm 1.97 \log_{10}$  CFU/g, with certain samples (S104, S105, S140, S141, S146, S147, S148, S172, S173, S179) below the detection limit of the technique (< 1.60  $\log_{10}$  CFU/g) (see Table S1 supplementary material). The maximum LAB counts were obtained in SS samples, with an average value of  $4.32 \pm 1.82 \log_{10}$  CFU/g, followed by GN with an average count of  $3.77 \pm 2.17 \log_{10}$  CFU/g, and finally BN table olives with  $3.38 \pm 2.05 \log_{10}$  CFU/g. On the contrary, *Enterobacteriaceae* population was always below the detection limit.

**Phylogenetic analysis.** The massive sequencing of the 72 samples generated a total of 4,493,422 raw sequences using as a target the V3 and V4 domains of the 16S ribosomal RNA gene. After screening the data for poor-quality sequences, a total of 4,463,697 high-quality sequences were recovered and 3,731,003 assigned into OTUs. After removing chloroplasts and mitochondria, a total of 2,901,384 sequences (64.99% from the total of the raw sequences) were finally used for phylogenetic assignment, making a mean of 40,297 sequences per sample. S104 was the sample with the minimum number of sequences (2,623), while S159 had the maximum number (127,416 sequences). A total of 660 different prokaryotic OTUs were obtained, taking into consideration all table olive samples analysed. They belong to Archaea (phyla Euryarchaeota 2.12%) and Bacteria domains (main phyla reported: Proteobacteria 47.27%, Firmicutes 18.33%, Bacteroidetes 13.03%, and Actinobacteria 10.76%) (see Table S2 in supplementary material). An average of 198 OTUs per sample was obtained (Table S3, supplementary material), with GN and BN olives showing similar average values (172 and 178, respectively), and slightly higher for SS (average 221 OTUs per sample).

The diversity of the bacterial community was also analysed using rarefaction curves and richness estimator (Chao1 index). The Chao1 index varied from 100.28 (S126 sample) to 1,280.00 (S173). Prokaryotic biodiversity richness did not have a clear relation with any type of elaboration, presentation or packaging material/system. The average Chao1 indexes for GN, BN, and SS elaborations were very similar (613.78, 623.86, and 606.07, respectively). Overall, despite the diversity of sequencing depth between samples, the rarefaction analysis indicated that a number of bacterial reads above 2,600 per sample was enough. Thus, there was a satisfactory coverage of the bacteria diversity for all the samples analysed with Good's coverage values in many cases above 95% (see Table S3).

From the 660 prokaryotic OTUs detected, only 41 bacteria were found in at least one sample with a proportion of sequences  $\geq 0.01\%$  (Table 1). Two of them were present in all samples. *Lactobacillus* (B1) was the predominant genera ranging from a minimum of 4.32 (S105) to a maximum of 98.11% (S116) (median value 54.99%), followed by *Pediococcus* (B2) with a proportion of sequences ranging from 0.07 (S169) to 87.01% (S137) (median value 26.09%). The rests of OTUs were absent in at least one sample, but *Celerinatantimonas* (B9), *Leuconostoc* (B18), *Alkalibacterium* (B6), *Pseudomonas* (B12), *Marinilactibacillus* (B3), *Weissella* (B16), and the family *Enterobacteriaceae* (B14), were also present in at least 80% of table olive packages, albeit with median values in all cases below 0.5% (Table 1). As mentioned previously, Table S2 in supplementary material shows the frequency of sequences obtained for the 660 assigned OTUs (even when the frequency was lower 0.01%) in all samples.

The distribution of the leading 41 OTUs in the 72 olive samples (Fig. 1) shows clearly that *Lactobacillus* (B1) was the most abundant genera in many packages (but not in S173, S170, S105 or S178) followed by *Pediococcus* (B2) which was, nevertheless, the most important in some others (S114, S137, S132, S124, S110, and S178). Also, *Celerinatantimonas* (B9) was abundant in several samples (S173, S171, S170, S169, S168, S167, and S189), while *Pseudomonas* (B12) and *Enterobacteriaceae* (B14) had an outstanding presence in S105. The other bacteria were found mostly in low proportion, except some samples (22.22% of the total) in which showed a complex mixture of genera (S115, S114, S137, S133, S132, S130, S126, S151, S150, S106, S105, S104, S180, S179, S121, and S120). Among the main 41 OTUs identified, the presence of microorganisms' sequences with the potential to produce illness was reduced. *Enterobacteriaceae* (B14) was detected at least in 91.66% of samples, but with a median value of only 0.05%; *Vibrio* (B4) 75% of the samples, median value of 0.11%; and *Staphylococcus* (B17) 54% of samples, median value of 0.001%. The abundance of sequences of other foodborne pathogens was ever lower, detecting just in sporadic cases the presence of *Listeria*, *Legionella*, *Campylobacter*, *Clostridium*, *Yersinia*, *Escherichia* or *Salmonella* (Table S2, supplementary material). Among spoilage microorganisms, only *Pseudomonas* (B12) (found in 87.50% of samples with a median value of 0.03%) and *Propionibacterium* (B13) (found in 59.72% of samples with a median value of 0.002%), were detected (Table 1).

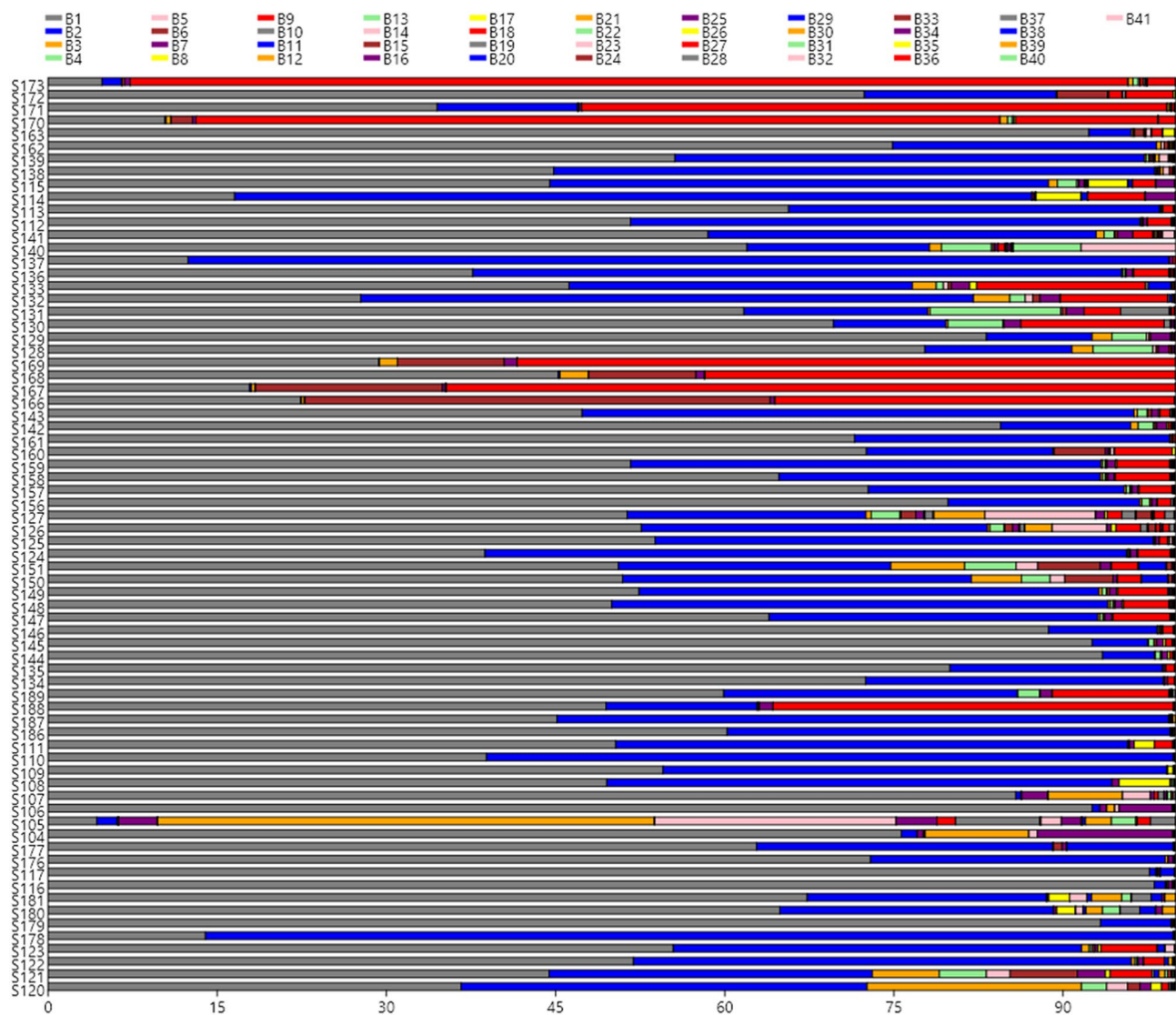
**Correspondence analysis.** A test of independence between rows and columns yielded a Chi-square of 148.268 ( $p < 0.0001$ ), concluding that samples and bacteria were significantly associated. The percentage of inertia (variability) accounted by the first three axes was 23.21, 21.79, and 19.03, respectively, with a total of 64.08% explained inertia (Fig. 2). According to Fig. 2A, several table olive samples showed long distances from the mean. They were S166, S167, S168, and S169 along Axis 1 (all of them from *Picholine* and *Lucques du Languedoc* cultivars processed as SS), S105 along Axis 2 (*Aloreña* fruits processed as GN), and S170, S171, S173, and S188 samples along Axis 3 (*Empeltre* and *Criolla* cultivars processed as GN and BN, respectively). A few other samples were also relatively different from the majority group (S126, S127, S104, and S107, corresponding the two first to *Hojiblanca* olives processed as SS and the two later to *Aloreña* as GN olives).

According to the CA analysis, a total of 18 OTUs had discriminant power among samples. The most different OTUs profiles across samples (Fig. 2B) were *Chryseobacterium* (B23), *Paenibacillus* (B25), *Novosphingobium* (B30), *Pantoea* (B19), *Sphingomonas* (B31), *Pseudomonas* (B12), *Serratia* (B37), *Enterobacteriaceae* (B14), *Pseudoalteromonas* (B35), *Weissella* (B16), *Uncultured* (B29), and *Myroides* (B24) along Axis 2, *Alkalibacterium* (B6) and *Leuconostoc* (B18) along Axis 1, and *Celerinatantimonas* (B9), *Enterobacter* (B36), *Propionibacterium* (B13), and *Alteromonas* (B33) along Axis 3. Thus, comparing Fig. 2A,B, it is observed that the bacteria along Axis 1, *Alkalibacterium* and *Leuconostoc*, were more abundant in S166, S167, S168, and S169 samples. Similarly, *Chryseobacterium* (B23), *Paenibacillus* (B25), *Novosphingobium* (B30), *Pantoea* (B19), *Sphingomonas* (B31), *Pseudomonas* (B12), *Serratia* (B37), *Enterobacteriaceae* (B14), *Pseudoalteromonas* (B35), *Weissella* (B16), *Uncultured* (B29), and *Myroides* (B24) were mainly detected in the sample segregated along the same axis (S105). Moreover, *Celerinatantimonas* (B9), *Enterobacter* (B36), *Propionibacterium* (B13), and *Alteromonas* (B33) along Axis 3 would be characteristic of samples S170, S171, S173, and S188. The rest of 23 OTUs, including *Lactobacillus* (B1), *Pediococcus* (B2), *Marinilactibacillus* (B3), *Vibrio* (B4), and *Halolactibacillus* (B5), which had a substantial presence in many samples, did not have discriminant power.

OTU reference	OTU assignment	Frequency of samples with presence (%)	Median (%)	Maximum value (%)	Minimum value (%)
B1	<i>g. Lactobacillus</i>	100.00	54.99 (27.77)	98.11 (S116)	4.32 (S105)
B2	<i>g. Pediococcus</i>	100.00	26.09 (33.71)	87.01 (S137)	0.07(S169)
B3	<i>g. Marinilactibacillus</i>	84.72	0.11 (0.36)	19.01 (S120)	0.00
B4	<i>g. Vibrio</i>	75.00	0.11 (0.67)	11.58 (S131)	0.00
B5	<i>g. Halolactibacillus</i>	72.22	0.03 (0.08)	2.10 (S121)	0.00
B6	<i>g. Alkalibacterium</i>	86.11	0.13 (0.32)	41.27 (S166)	0.00
B7	Unassigned	100.00	0.44 (0.52)	3.48 (S105)	0.01 (S162)
B8	<i>F. Cardiobacteriaceae</i>	73.61	0.03 (0.12)	4.54 (S108)	0.00
B9	<i>g. Celerinantimonas</i>	87.50	0.41 (2.89)	88.49 (S173)	0.00
B10	<i>g. Salinivibrio</i>	62.50	0.02 (0.06)	4.32 (S131)	0.00
B11	<i>g. Amphibacillus</i>	62.50	0.01 (0.08)	9.43 (S177)	0.00
B12	<i>g. Pseudomonas</i>	87.50	0.03 (0.07)	44.02 (S105)	0.00
B13	<i>g. Propionibacterium</i>	59.72	0.00 (0.01)	0.47 (S173)	0.00
B14	<i>F. Enterobacteriaceae</i>	91.66	0.05 (0.12)	21.39 (S105)	0.00
B15	<i>F. Bacillaceae</i>	51.38	0.00 (0.01)	0.44 (S151)	0.00
B16	<i>g. Weissella</i>	81.94	0.02 (0.06)	12.04 (S104)	0.00
B17	<i>g. Staphylococcus</i>	54.10	0.00 (0.01)	3.99 (S114)	0.00
B18	<i>g. Leuconostoc</i>	83.30	0.06 (0.17)	64.67 (S167)	0.00
B19	<i>g. Pantoea</i>	56.90	0.00 (0.01)	7.47 (S105)	0.00
B20	<i>g. Marinobacterium</i>	65.20	0.00 (0.02)	1.27 (S117)	0.00
B21	<i>g. Arthrobacter</i>	25.00	0.00 (0.00)	2.67 (S181)	0.00
B22	<i>g. Aequorivita</i>	6.94	0.00 (0.00)	1.51 (S180)	0.00
B23	<i>g. Chryseobacterium</i>	23.61	0.00 (0.00)	1.81 (S105)	0.00
B24	<i>g. Myroides</i>	20.83	0.00 (0.00)	1.36 (S127)	0.00
B25	<i>g. Paenibacillus</i>	16.67	0.00 (0.00)	1.76 (S105)	0.00
B26	<i>g. Oenococcus</i>	12.50	0.00 (0.00)	0.02 (S163)	0.00
B27	<i>F. Rhodobacteraceae</i>	41.66	0.00 (0.01)	5.08 (S114)	0.00
B28	<i>g. Rhodobacter</i>	15.27	0.00 (0.00)	1.74 (S181)	0.00
B29	Uncultured	33.33	0.00 (0.00)	1.39 (S180)	0.00
B30	<i>g. Novosphingobium</i>	34.72	0.00 (0.00)	2.29 (S105)	0.00
B31	<i>g. Sphingomonas</i>	40.27	0.00 (0.00)	2.18 (S105)	0.00
B32	<i>g. Helicobacter</i>	41.66	0.00 (0.00)	1.06 (S141)	0.00
B33	<i>g. Alteromonas</i>	59.72	0.00 (0.03)	0.15 (S188)	0.00
B34	<i>g. Marinobacter</i>	55.55	0.00 (0.02)	2.68 (S114)	0.00
B35	<i>g. Pseudoalteromonas</i>	38.89	0.00 (0.00)	0.11 (S127)	0.00
B36	<i>g. Enterobacter</i>	62.50	0.00 (0.02)	2.49 (S173)	0.00
B37	<i>g. Serratia</i>	55.55	0.00 (0.02)	2.24 (S105)	0.00
B38	<i>g. Cobetia</i>	34.72	0.00 (0.00)	0.02 (S149)	0.00
B39	<i>g. Halomonas</i>	64.11	0.00 (0.01)	1.17 (S180)	0.00
B40	<i>g. Marinomonas</i>	22.22	0.00 (0.00)	6.01 (S140)	0.00
B41	<i>g. Photobacterium</i>	16.67	0.00 (0.00)	8.39 (S140)	0.00

**Table 1.** References of OTUs assigned at genera or family level (with a minimum percentage of sequences of 0.01% obtained in at least one sample) reported in this study. Percentage of samples with each OTU, median (interquartile range 25–75% in parenthesis), maximum, and minimum frequencies are also included.

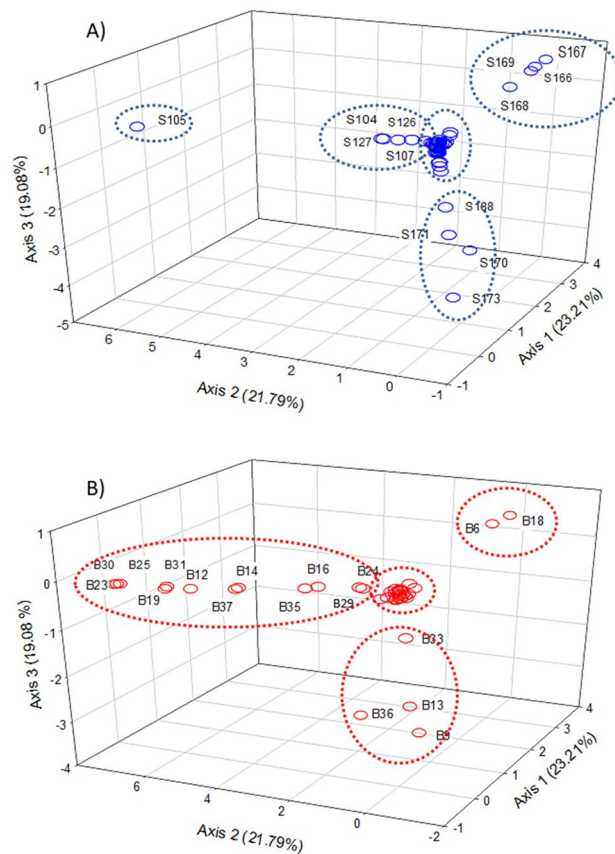
**CoDa analysis.** The two principal components (PCs) of the CoDa biplot accounted for 39.36% of the total variance, which reached 49.62% when including the third PC. The characteristics and distribution of bacterial OTUs (regardless of grouping factor) were similar in covariance and form biplot (Fig. 3). Variables (bacteria) were mainly grouped along PC1, leading to large standard deviations (distance between two ends of rays) of their log-ratios. Besides, these log-ratios between the variables on both sides form small angles between them and, therefore, are strongly correlated. On the contrary, only a few variables were distributed along PC2: up (only *Leuconostoc*, B18) and down (*Pediococcus*, B2; *Cardiobacteriaceae*, B8; *Marinobacterium*, B20; *Marinobacter*, B34; and *Halomonas*, B39) and their log-ratios are also high but strongly correlated since their links are almost parallel. Also, as the angles between the links of the variables from the first and second groups are around 90°, their log-ratios are hardly related. Together, they contribute to the segregation of samples (PC1



**Figure 1.** Abundance of the main OTUs. 41 OTUs (>0.01% relative frequency) at genera and family level were found in the different table olive samples analysed in this work.

along the horizontal axis and PC2 along the vertical). Table S4 (supplementary material) shows the variances of their log-ratios, which sum corresponds to the common variance of the *clr* transformed data. The largest values corresponded to *Celerinatantimonas* (clr B9) (10.89), *Leuconostoc* (clr B18) (8.54), and *Alkalibacterium* (clr B6) (7.39). The log-ratios of these variables eventually lead to the best segregation among samples. As the interest of this study was mainly focused on the difference between samples, the form biplot, according to the three factors involved in the design, is particularly appropriate: the distances in the plot approximate those between the samples (rows). Regarding elaboration, there is a trend for the segregation of SS from GN, although with some overlapping (Fig. 3A). A quite similar situation is observed for the type of presentation (Fig. 3B), with S olives situated on the right zone of the biplot and P on the left, but W distributed around the plot. In the case of the packaging material/system (Fig. 3C), those using P containers are mainly located on the right, B (together with V) on the left, and G spread from first to the third quadrant.

When clustering analysis was applied (Fig. 4), considering the height of 60, two very different major groups were observed. Cluster A included 35 samples, many of them (71.42%) belonging to SS, while 28.58% were natural (GN or BN). Cluster B was composed of 37 samples, but, on the contrary, many of them (70.28%) were natural (GN or BN), and only 29.72% lye treated (SS). Hence, such dissimilarity allowed making gross discrimination between lye treated and natural olives. Considering a height of 20, a total of 14 different sub-clusters (C1-C14) were obtained. Notably, they included most of the samples from a similar origin in the same sub-cluster like S140 and S141 (cluster C4); S180 and S181 (C12); S170, S171, and 173 (C5). Nevertheless, some samples were singular and were not included in their analogous characteristic sub-clusters but in other groups of samples, like S105, S172, and S161 (C13) assigned to clusters C8, C7 and C13, respectively. Therefore, clustering was appropriate for grouping most of the samples from a similar type of elaboration and origins, but also displayed the presence of



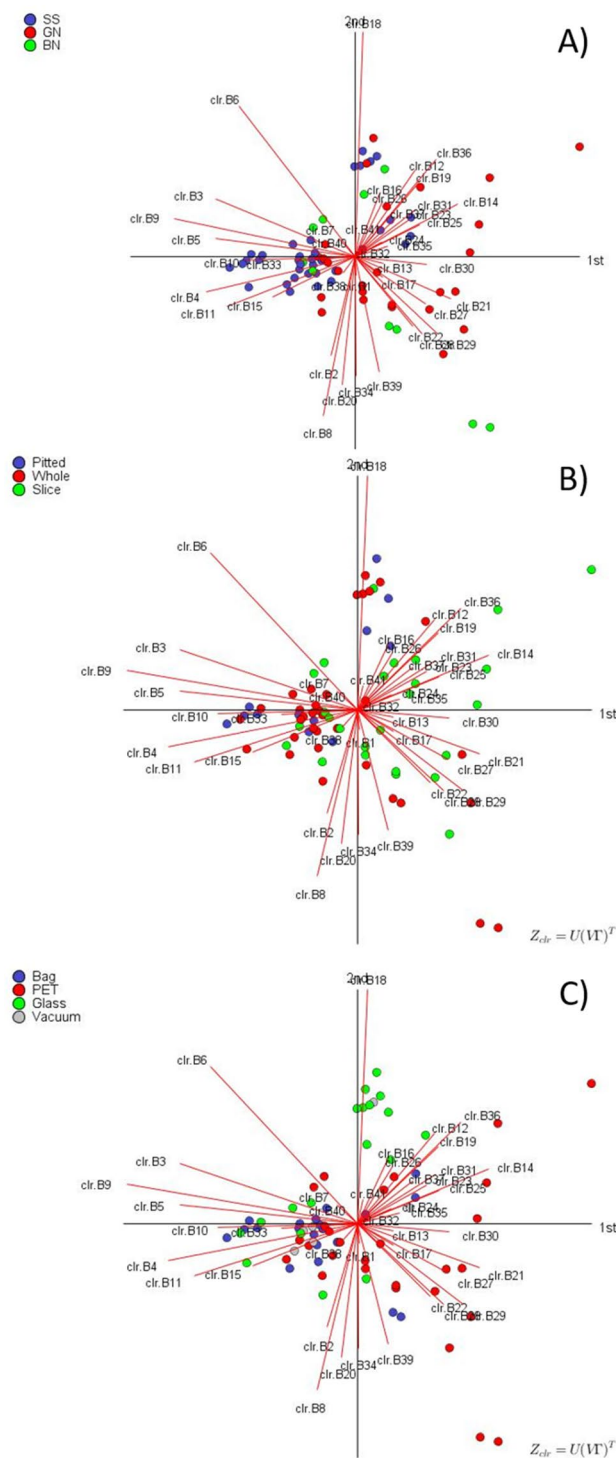
**Figure 2.** Study of similarity. Correspondence analysis among samples (A) and bacteria (B) was developed to study their similarity, and subsequent segregation of those different from the average.

some others which bacteria consortium could not be related to the packaging conditions but other uncontrolled factors. These segregations are in agreement to the significant differences already observed for some groups in the CA (among others, S166, S167, S168, S169 (C6), S105 (C8), or S170, S171, S173 (C5)).

CoDa ANOVA was also applied to evaluate the effect of elaboration, presentation and packaging material/system on bacteria diversity, using *ilr* transformed *coordinates*. Individually considered, the effect of type of elaboration was significant (at  $p=0.0005165$ ) while the effect of type of presentation ( $p=0.1191$ ) and packaging material ( $p=0.0522$ ) were not significant. However, the contribution of a new variable can be assessed by the additional variance explained by the model (already including the other two) when it is included. For this purpose, the multi-factor ANOVA, obtained by successively placing a different variable in the last position, was tested. In this case, the presentation form was not significant neither in the intermedia position or when it was added as the last term (see Table S5 in supplementary material). Therefore, the only two significant variables (factors) in the final ANOVA model were the type of elaboration and packaging material/system. Even so, despite the significant influence of both variables on the model, it only explained a reduced variance ( $\approx 15\%$ ), indicating that the bacterial diversity in the samples was, at least partially, randomly induced (or due to uncontrolled factors). As the model coefficients are in *ilr coordinates*, it was necessary the back-transformation of the ANOVA model predictions to present them in the original scale (Table S6, supplementary material). Regardless of elaboration, *Lactobacillus* (B1) would be the most abundant, with the highest level in the G containers, followed by its growth in P, V and B material/system, possibly associated with the permeability to oxygen of the different materials.

On the contrary, *Pediococcus* (B2), which is always expected, descends in the order  $B > V > P > G$  containers; that is, according to their decrease oxygen permeability. The presence of the remaining bacteria was not so clearly related to the packaging material. However, *Marinilactobacillus* (B3) was expected to be absent in olives packaged under vacuum (V), *Vibrio* (B4) and *Halolactibacillus* (B5) should be absent in P (GN) and V (GN) and P (GN), respectively. *Alkalibacterium* (B6) could have a higher proportion in G (SS) and lower in the other materials. *Cardiobacteriaceae* (B8) was not present in G (SS) and B (BN), *Celerinatantimonas* (B9) would be absent from V (GN), *Leuconostoc* (B18) was not found in B (GN) or P (BN), *Salinivibrio* (B10) and *Amphibacillus* (B11) could mainly be related to SS while *Cardiobacteriaceae* with GN olives. As observed, overall, GN was more restrictive regarding biodiversity than the other elaborations.

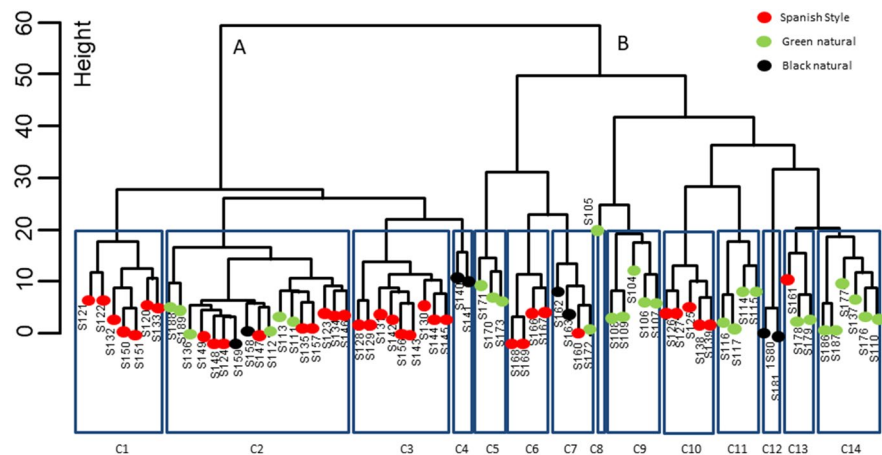
LDA analysis was also performed on *ilr coordinates* (Euclidean space). There was an overall 95.83%, 94.44%, and 95.83% success in the classification according to the type of elaboration, presentation and packaging material,



**Figure 3.** CoDa form biplot. Analysis done according to the type of elaboration (A), presentation (B), and packaging material (C), based on the proportion of each bacterium in the commercial samples analysed in this work.

respectively (Table 2). Therefore, this supervised classical method provided fairly good classification. LDA was also applied using the package *compositions* (together with MASS for *lda*). This technique also provided the means of the different bacteria according to factors (elaboration, presentation, and packaging material/system). First in *ilr* coordinates, where mean and standard deviations are common parameters, and then as original values by back-transformation. The information (Table S7 in supplementary material) allows comparing the bacteria average according to levels of the factor co-variables. E.g., in elaboration, *Lactobacillus* (B1), as an average, is present similarly in SS and GN in a proportion of  $\approx 78\%$  but in a slowly lower frequency in BN ( $\approx 72\%$ ). On





**Figure 4.** Cluster dendrogram of the samples. Dendrogram was done using the Ward.D2 on the pivot coordinates from the original data set. Heights of 60 and 20 were used for assessing dissimilarity.

the contrary, *Pediococcus* (B2) was more expected in BN ( $\approx 27\%$ ) than in GN ( $\approx 21\%$ ). Regarding the type of presentation, the differences were more limited for *Lactobacillus* (from approx. 75 to 79%) and approximately similar for *Pediococcus* (from 17 to 24%). On the contrary, in packaging material, the differences in the main two bacteria were sensible for both *Lactobacillus* (from 63 to 88%) and *Pediococcus* (from 8 to 35%). Situation that is in agreement with the significant differences observed among the levels of this factor in the ANOVA. The package compositions also works in *ilr* coordinates (Euclidean space) when performing LDA. The discriminant functions, for original values, is reproduced in Table S8 (supplementary material). This package showed similar segregation than *robComposition* but provided a visualization of the results. Two BN and one GN samples were wrongly classified as SS in case of the type of elaboration (Fig. 5a). S presentations were wholly differentiated from the rest of presentations, but several P samples (only two with *robCompositions*) were included in the W group (Fig. 5b), in agreement with the non-significant difference deduced in the ANOVA analysis. Regarding packaging material, based on the first two LDA functions, those using G and P containers were well segregated from that using B and V, but a few of V samples were wrongly classified as B (Fig. 5c), a situation compatible with the use plastic films in both samples. The discrimination according to packaging material using any other combination of LD functions did not improve the results since led to a similar or high degree of overlap between levels.

## Discussion

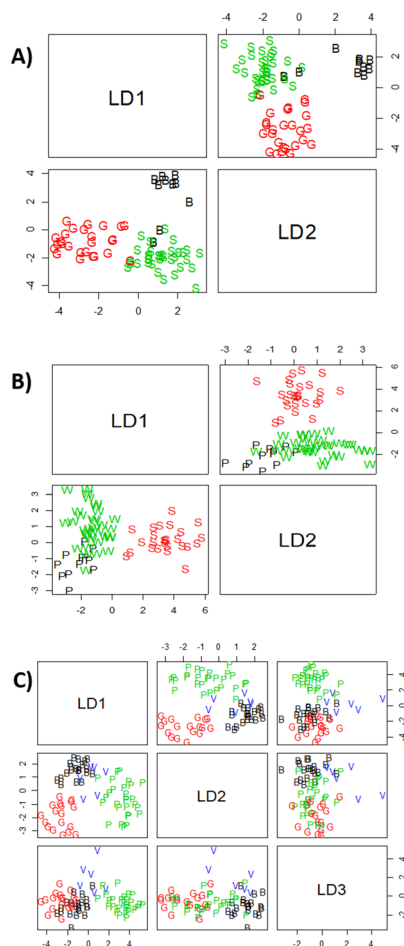
Consumers increasingly demand healthier and safer foods. Table olives are an important component of the Mediterranean diet and culture, and as fermented vegetables, they are considered safe in a health perspective, provided that certain levels of pH and salt are obtained. The international legislation specifies the salt content and maximum pH limit for table olive packaging as a function of the type of elaboration<sup>35</sup>. In this study, albeit average values of pH (3.94) and salt (5.97%) were within the levels established by international legislation, 22% of the total samples were above the pH limit of 4.3 (close limit to germination of *Clostridium botulinum* spores). Such deviations might represent an obvious safety risk. However, *Clostridium* spp. were not detected in any table olive brine analysed in this work above 0.01% of frequency. Some authors<sup>36</sup>, recently reported that a cocktail of diverse strains of *C. sporogenes* needed a pH value above 5.0 to germinate in both laboratory medium and table olive brine, which could explain why *Clostridium* was neither detected albeit pH value in certain samples was above 4.3. Our data are very similar to other studies on commercial table olives performed in markets of Spain<sup>5</sup> and Portugal<sup>6</sup>, although with more than 12 years of difference. This way, while the first authors<sup>5</sup> reported the average pH value of 3.9 and salt content of 5.3%, the second ones<sup>6</sup> found average pH values of 4.0.

Regarding the microbial population in the products, the legislation only establishes that they must be free of any microorganisms, or their toxins, which could be a risk for consumer health<sup>35</sup>. A considerable number of the packages showed the presence of viable LAB populations (85% of samples) but, on the contrary, absence of *Enterobacteriaceae*. The average LAB population was 3.98  $\log_{10}$  CFU/g, with no statistical differences as a function of the type of elaboration. The presence of LAB populations in approximately 30% of Spanish products, reaching levels up to 7  $\log_{10}$  cfu/mL was also reported<sup>5</sup>. Others<sup>4</sup> found that practically all the packages of lye treated olives showed microbial presence (ranging from 2 to 7  $\log_{10}$  cfu/ml). In natural olives, the percentage of samples with microorganisms was also high (78%). Moreover, some authors<sup>6</sup> detected microbial presence in 64% of samples from table olives commercialized in Portugal. The presence of microorganisms habitually involved in the fermentation (LAB) in non-thermally treated olive packages could be considered usual, provided they do not produce changes that might compromise the safety of the product<sup>2,35</sup>.

As far as we know, this is the first study evaluating by molecular methods the prokaryotic diversity associated with olive biofilms from diverse international table olive markets. As above-commented, previous works have investigated, using culture-dependent methods, the microbial populations in the cover brines of commercial table olive packages from Italy<sup>4</sup>, Spain<sup>5</sup>, and Portugal<sup>6</sup>. However, microorganisms forming biofilms on fruit's

Elaboration (BN-Black natural, GN-Green natural, SS-Spanish style) (95.83%)				Presentation (P-Pitted, S-Sliced, W-Whole) (94.44%)				Packaging material (B-Bag, G-Glass, P-PET, V-Vacuum) (95.83%)				
	Predicted				Predicted				Predicted			
Actual	BN	GN	SS	Actual	P	S	W	Actual	B	G	P	V
BN	8	0	2	P	10	0	2	B	19	0	0	1
GN	0	25	1	S	0	26	0	G	0	18	0	0
SS	0	0	36	W	2	0	32	P	1	0	27	0
								V	1	0	0	5

**Table 2.** Classification table of the LDA according to elaboration, presentation, and packaging material, using pivot *coordinates* and the classical option in the R package *robCompositions* (apparent success rate in parenthesis).



**Figure 5.** Segregation of the samples by CoDa linear discriminant analysis. Analysis was done according to elaboration (S-Spanish style, G-Green Natural, B-Black natural), presentation (S-Sliced, P-Pitted, W-Whole), and packaging material (G-Glass, P-PET, B-Bags, V-Vacuum).

epidermis could be ingested by the final consumers and depending on the bacteria consortium, had positive or negative connotations. Despite fermented table olives having a long history of microbial safety, different biological hazards (botulism, mycotoxins, biogenic amines and presence of foodborne pathogens) may be present in the finished products, usually due to improper olive handling practices<sup>3</sup>.

Metataxonomic analysis for the study of prokaryotic and eukaryotic microorganisms in table olive fermentation has been introduced recently<sup>8–10,12–14,37</sup>, but not in finished commercial products. The application of massive sequencing techniques to this fermented vegetable could help to assess the safety status of final products. This independent culture technique is based on the massive sequencing of total bacteria DNA obtained from olive biofilms, belonging to both alive and died cells. Results showed that the proportion of sequences of foodborne

pathogens was in general very low (median value always obtained below 0.15%), being family *Enterobacteriaceae* and the genera *Vibrio* and *Staphylococcus* the undesirable microorganisms found in higher frequency. These data are in agreement with previous works<sup>6</sup>, which detected the presence of viable *Staphylococcus* in table olives commercialized in Portugal, and others, which did the same in seasoning material of natural green olives<sup>38</sup>. Furthermore, Benítez-Cabello et al. (2016)<sup>39</sup> and Lucena-Padrós et al. (2014)<sup>40</sup> detected, using molecular methods, *Vibrio* and *Staphylococcus* genera during industrial Spanish-style green table olive fermentation, while Abriouel et al. (2011)<sup>41</sup> reported the presence of *Vibrio* spp. during the processing of natural green olives. *Vibrio* is a genus of halophilic Proteobacteria, which includes several species associated with human gastroenteritis diseases. *Staphylococcus* is a genus of bacteria very resistant to high salt, low pH (below 5.0) and fluctuating temperatures. Both genera include pathogenic and non-pathogenic species but, in any case, their relative presence should be taken into account as an indicator of the safety state of the final product. Finally, *Enterobacteriaceae* was always below the detection limit in all table olive samples analysed in this work, albeit DNA sequences for this bacteria family were detected in 91.66% of samples at low proportions (except in S105 sample with 21.39%). The presence of *Enterobacteriaceae* during table olive processing is frequent, but they are inhibited at the usual low pH levels that are reached during table olive packaging<sup>2</sup>. Our data are also in agreement with other authors<sup>5</sup>, who did not report viable *Enterobacteriaceae* presence in commercial Spanish cultivar table olives.

Bacteria and Archaea are the most ancient and most widespread forms of life on Earth. In this work, a total of 660 different prokaryotic OTUs (97.9% bacteria) were obtained when analysing the 72 olive packages, but only 41 OTUs were a majority and had frequencies >0.01% in at least one sample. However, our biodiversity indexes were higher than those obtained in other metatranscriptomics studies performed during table olive fermentations<sup>12,13,37</sup>. These authors reported the presence of OTUs at family/genera level in many cases below 100. On the contrary, in the study carried out on the seasoning material (salt, aromatic herbs, fermented garlic, or pepper) used during table olive packaging, a higher number of OTUs (> 3.000) were found<sup>42</sup>. This abundance could explain the intermediate values between fermented fruits and seasoning material found in this survey because, in olive packages, diverse types of seasoning material are added. It was reported using several statistical approaches and analysing > 700.000 prokaryotic 16S-V4 sequences, that could exist about 2.2–4.3 million bacterial OTUs worldwide, refuting recent predictions of trillions of prokaryotic OTUs in the biosphere<sup>43</sup>. According to their data, the bacterial biodiversity in table olive packaging could stand for 0.015% of the worldwide prokaryotic diversity.

CoDa analysis identified *Lactobacillus* and *Pediococcus* as the ubiquitous bacteria in the packages of this fermented vegetable, widely distributed around the world. Both genera are usually isolated from diverse types of table olive fermentations and be responsible for its lactic acid process<sup>2,44</sup>. This information reinforces the healthy and nutritional aspects of this fermented food, proving that table olives in a splendid carrier of potential probiotic *Lactobacillus* strains to final consumers. In this sense, *L. pentosus* LPG1 isolated from table olive biofilms recently proved its high probiotic potential in mammals<sup>45</sup>. Furthermore, the work has demonstrated that following methodologies developed for CoDa, as the proportion of OTUs, the statistical analysis can lead to results that allow testing the different hypothesis of the experiment while working in the suitable sampling space. Therefore, the work might also be considered as a query for the appropriate study of metagenomic information, as was previously reported<sup>8</sup>.

## Conclusion

This work has proven the effectiveness of the application of metagenomics and multivariate analysis based in CoDa to expand our knowledge of the prokaryotic diversity in olive biofilms. Results have shown the absence of food-borne pathogens in many samples, but suggesting to *Vibrio*, *Staphylococcus*, and *Enterobacteriaceae* as potential biomarkers of the safety and hygienic status of this ready-to-eat fermented vegetable. Moreover, the CoDa exploratory techniques presented the distribution map of samples and discrimination as a function of elaboration (lye treated vs natural olives) and packaging material (among glass, PET, and bags). *Lactobacillus* and *Pediococcus* were the most widely distributed and ubiquitous bacteria in all the samples. At the same time, the presence of other genera like *Propionibacterium*, *Pantoea*, *Marinobacterium*, *Alteromonas* or *Marinobacter* could be considered as sporadic and randomly distributed. However, further studies are necessary to corroborate the microbial activity of all these microorganisms after packaging, because microbial DNA could belong to viable or died cells.

Received: 28 April 2020; Accepted: 23 June 2020

Published online: 09 July 2020

## References

1. International Olive Oil Council. *World Table Olive Figures* (2020). Available online: <https://www.internationaloliveoil.org>.
2. Garrido-Fernández, A. et al. *Table Olives Production and Processing* (Chapman & Hall, London, 1997).
3. Medina-Pradas, E. & Arroyo-López, F. N. Presence of toxic microbial metabolites in table olives. *Front. Microbiol.* **6**, 873 (2015).
4. Marsilio, V. & De Angelis, M. Indagine su alcuni parametri di qualità di campioni di olive da tavola del commercio. *Industria Alimentari* **32**, 1093–1095 (1993).
5. López-López, A. et al. Physicochemical and microbiological profile of packed table olives. *J. Food Prot.* **67**, 2320–2325 (2004).
6. Pereira, A. P. et al. Microbiological characterization of table olives commercialized in Portugal in respect to safety aspects. *Food Chem. Toxicol.* **46**, 2895–2902 (2008).
7. Botta, C. & Cocolin, L. Microbial dynamics and biodiversity in table olive fermentation: culture-dependent and -independent approaches. *Front. Microbiol.* **6**, 3 (2012).
8. Benítez-Cabello, A. et al. The use of multifunctional yeast-lactobacilli starter cultures improves fermentation performance of Spanish-style green table olives. *Food Microbiol.* **91**, 103497 (2020).
9. De Castro, A. et al. Microbiota and metabolite profiling of spoiled Spanish-style green table olives. *Metabolites*. **8**(4), 73 (2018).

10. Medina, E. *et al.* Microbial ecology along the processing of Spanish olives darkened by oxidation. *Food Control* **86**, 35–41 (2018).
11. Rodríguez-Gómez, F. *et al.* Microbiological and physicochemical changes in natural green heat-shocked Aloreña de Málaga table olives. *Front. Microbiol.* **8**, 2209 (2017).
12. Cocolin, L. *et al.* NaOH-debittering induces changes in bacterial ecology during table olives fermentation. *PLoS ONE* **8**(7), e69074 (2013).
13. De Angelis, M. *et al.* Microbiota and metabolome of un-started and started Greek-type fermentation of Bella di Cerignola table olives. *Food Microbiol.* **52**, 18 (2015).
14. Arroyo-López, F. N. *et al.* Enhancement of the knowledge on fungal communities in directly brined Aloreña de Málaga Green olive fermentations by metabarcoding analysis. *PLoS ONE* **11**, 9 (2016).
15. Gloor, G. B. *et al.* Microbioma data sets are compositional: and this is not optional. *Front. Microbiol.* **8**, 2224 (2017).
16. Greenacre, M. *La Práctica del Analisis de Correspondencias*. Fundación BBVA, Barcelona, Spain, 310 pp (2008).
17. Aitchison, J. The statistical analysis of compositional data. *J. R. Stat. Soc. B.* **44**(2), 139–160 (1982).
18. Bowen, B. W. & Karl, S. A. Population genetics and phylogeography of sea turtles. *Mol. Ecol.* **16**(23), 4886–4907 (2007).
19. Pierotti, M. E. R. & Martín-Fernández, J. A. Compositional analysis in behavioural and evolutionary ecology. *CoDA* **218**, 234 (2011).
20. Lammer, H. *et al.* Compositional data analysis in planetology: the surfaces of Mars and Mercury. In *Compositional Data Analysis: Theory and Applications* (eds Pawłowsky, V. & Buccianti, A.) 267–281 (John Wiley & Sons, Chichester, 2011).
21. Klindworth, A. *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**(1), e1–e1 (2013).
22. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**(6), 863–864 (2011).
23. Magoč, T. & Salzberg, S. L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**(21), 2957–2963 (2011).
24. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**(19), 2460–2461 (2010).
25. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**(21), 7188–7196 (2007).
26. Nemadić, O. & Greenacre, M. Correspondence analysis in R, with two- and three dimensional graphics: the ca package. *J. Stat. Softw.* **20**(3), 1 (2007).
27. Tsilimigras, M. C. B. & Fodor, A. A. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Ann. Epidemiol.* **26**, 330–335 (2016).
28. Pawłowsky-Glahn, V. *et al.* Modeling and analysis of compositional data. JWS, <http://www.whiley.com/go/glahn/practical> (2015).
29. Gabriel, K. R. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* **58**, 453–467 (1971).
30. Greenacre, M. *Biplots in practice*. Fundación BBVA, Barcelona, Spain, 237 pp (2010).
31. Aitchison, J. & Greenacre, M. Biplots of compositional data. *Appl. Statist.* **51**(4), 375–392 (2002).
32. Comas-Cufí, M. & Thió-Henestrosa, S. *CoDaPack Department of Computer Science and Applied Mathematics* (University of Girona, Girona (Spain), 2018).
33. Van den Boogaart, K. G. & Tolosana-Delgado, R. *Analyzing compositional data with R* (Springer-Verlag, Berlin, 2013).
34. Filzmoser, P. *et al.* *Applied Compositional Data Analysis, with worked examples* (Springer Nature Switzerland AG, Cham, 2018).
35. International Olive Oil Council. Trade standard applying to table olives. Madrid, Spain (2004).
36. Valero, A. *et al.* Influence of acid adaptation on the probability of germination of *Clostridium sporogenes* spores against pH NaCl and time. *Food* **9**, 127 (2020).
37. Medina, E. *et al.* Assessment of the bacterial community in directly brined Aloreña de Málaga table olive fermentations by meta-genetic analysis. *Int. J. Food Microbiol.* **236**, 47–55 (2016).
38. Ruiz-Bellido, M. A. *et al.* A Probabilistic decision-making scoring system for quality and safety management in Aloreña de Málaga table olive processing. *Front. Microbiol.* **8**, 2326 (2017).
39. Benítez-Cabello, A. *et al.* RT-PCR–DGGE analysis to elucidate the dominant bacterial species of industrial spanish-style green table olive fermentations. *Front. Microbiol.* **7**, 1291 (2016).
40. Lucena-Padrós, H. *et al.* Microbial diversity and dynamics of Spanish-style green table-olive techniques. *Int. J. Food Microbiol.* **42**, 154–165 (2014).
41. Abriouel, H. *et al.* Culture-independent study of the diversity of microbial populations in brines during fermentation of naturally-fermented Aloreña green table olives. *Int. J. Food Microbiol.* **144**, 487–496 (2011).
42. Benítez-Cabello, A. *et al.* Metataxonomic analysis of the bacterial diversity in table olive dressing components. *Food Control* **105**, 190–197 (2019).
43. Louca, S. *et al.* A census-based estimate of earth's bacterial and archeal diversity. *PLoS Biol.* **7**, e00999–e1016 (2019).
44. Hurtado, A. *et al.* Lactic acid bacteria from fermented table olives. *Food. Microbiol.* **31**, 1–8 (2012).
45. Benítez-Cabello, A. *et al.* Probiotic properties of *Lactobacillus* strains isolated from table olive biofilms. *Probiotic Antimicrobial Prot.* In press. (2020).

## Acknowledgements

The research has received funding from the Spanish Government (Project TOBE RTI2018-100883-B-I00: [www.tobeproject.csic.es](http://www.tobeproject.csic.es)).

## Author contributions

F.N.A.-L. conceived and designed the project, writing the paper with input from the rest of co-authors. A.B.-C. prepared and analysed the metataxonomic data, and helped to write the paper. A.G.-F. realized the CoDA analysis, and helped to write the paper. V.R.-G. and E.M.-P. performed the laboratory work (microbiological, physicochemical analysis, DNA extraction and purification).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-68305-7>.

**Correspondence** and requests for materials should be addressed to F.N.A.-L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020