



Article

The Compositional HJ-Biplot—A New Approach to Identifying the Links among Bioactive Compounds of Tomatoes

Marcos Hernández Suárez ¹, Daniel Molina Pérez ², Elena M. Rodríguez-Rodríguez ², Carlos Díaz Romero ², Francisco Espinosa Borreguero ^{3,*} and Purificación Galindo-Villardón ^{4,5}

¹ Aula Dei Scientific Technological Park Foundation, Av. Montañana 930, 50059 Zaragoza, Spain; marcoshsuarez@gmail.com

² Department of Chemical Engineering and Pharmaceutical Technology, Food Science and Nutrition Area, University of La Laguna, 38206 La Laguna, Santa Cruz de Tenerife, Spain; dmolina.perez@gmail.com (D.M.P.); emrguez@ull.es (E.M.R.-R.); cdiaz@ull.es (C.D.R.)

³ Plant Biology, Ecology and Earth Sciences Department, Extremadura University, Av. Elvas s/n, 06071 Badajoz, Spain

⁴ Statistics Department, University of Salamanca, c/Alfonso X El Sabio s/n, 37008 Salamanca, Spain; pgalindo@usal.es

⁵ Higher Polytechnic School of the Litoral, Campus Gustavo Galindo, 09-01-5863 Guayaquil, Ecuador

* Correspondence: espinosa@unex.es; Tel.: +34-924-316502

Academic Editor: Maurizio Battino

Received: 28 July 2016; Accepted: 28 October 2016; Published: 2 November 2016

Abstract: Tomatoes have been described as a functional food because of their particular composition of different bioactive compounds. In this study, the proximate composition, minerals and trace elements, and antioxidant compounds were determined in two tomato cultivars (Mariana and Duncan) that were grown in Gran Canaria (Spain) either conventionally or hydroponically. Although compositional data of this type require being subjected to the specific statistical techniques of compositional analysis, this approach has not usually been considered in this context. In the present case, a compositional Mann–Whitney U test of the data showed significant differences for each factor (cultivar and cultivation system) in several of the compositional variables studied. For the differences between cultivars, these parameters were the protein, Mg, lycopene, ascorbic acid, citric acid, and fumaric acid contents. For the differences between cultivation systems, they were mainly those of the mineral and trace elements group. Although one-year data are insufficient to make clear relationship among compounds because more repetitions in several localities and years are necessary, the compositional HJ-biplot (in which the links provide estimates of the linear relationship among variables) results agreed with other scientific results about linear relationship among some compounds analyzed.

Keywords: tomato; link; compositional Mann–Whitney U test; compositional HJ-biplot

1. Introduction

The tomato is one of the world's most consumed vegetables and fruits. It is regarded as a functional food since the chemical composition of tomato samples has been found to include such bioactive constituents as fiber, minerals, vitamins, and phenolic and carotenoid compounds [1]. The Canary Islands are a major producer of tomatoes, with many of the cultivars used having particular and interesting characteristics [2]. Various factors are known to affect the nutrient content, including cultivar, climate, geochemistry, and agricultural practices. Hence the numerous studies aimed at evaluating and improving the quality of the tomato fruit [1]. However, the conclusions of those studies

have been based on the application of standard statistical tools, while it has not been usual to consider this kind of data from the mathematical viewpoint of compositional analysis.

Compositional data are vectors of positive values that sum to unity (or, in general, to some constant that is fixed for all the vectors) [3]. In the present study, each tomato sample was defined by 25 chemical content variables whose sum accounts for almost 100% of the sample's chemical composition. In the case of compositional data, the information that standard descriptive statistics can contribute is limited. The problem is that arithmetic means and variances are clearly defined in the framework of a Euclidean geometry, but this geometry is not appropriate for compositional data because the central tendency of a compositional sample is best represented by the geometric mean instead of the arithmetic mean, and this has consequences if comparisons are made using standard statistical tools based on this latter class of mean. The overall dispersion of a compositional sample is not the standard deviation but the total variance, given by the mean squared Aitchison distance [4].

The measure of correlation can also be quite misleading with compositional data. This phenomenon is known as “spurious correlation”, i.e., a false presumption that two variables are correlated when in reality they are not. This has consequences when a standard multivariate analysis based on a correlation matrix is applied to this kind of data [3]. Compositional biplots are graphical techniques that allow the visual identification of real associations between variables in this type of data. They were introduced by Aitchison and Greenacre [5] who proposed the terms “form biplot” and “covariance biplot” as equivalents for Gabriel's “JK-biplot” and “GH-biplot”, respectively [6]. Such compositional biplots consist of vertices, rays, links, and markers related to each individual observation (Figure 1). A ray provides an estimate of the standard deviation of a variable, while a link is a measure of the strength of the linear relationship between two components of a composition. Links are not represented as arrows in the biplot. Unlike other biplots, in a compositional biplot, the links are the most relevant elements.

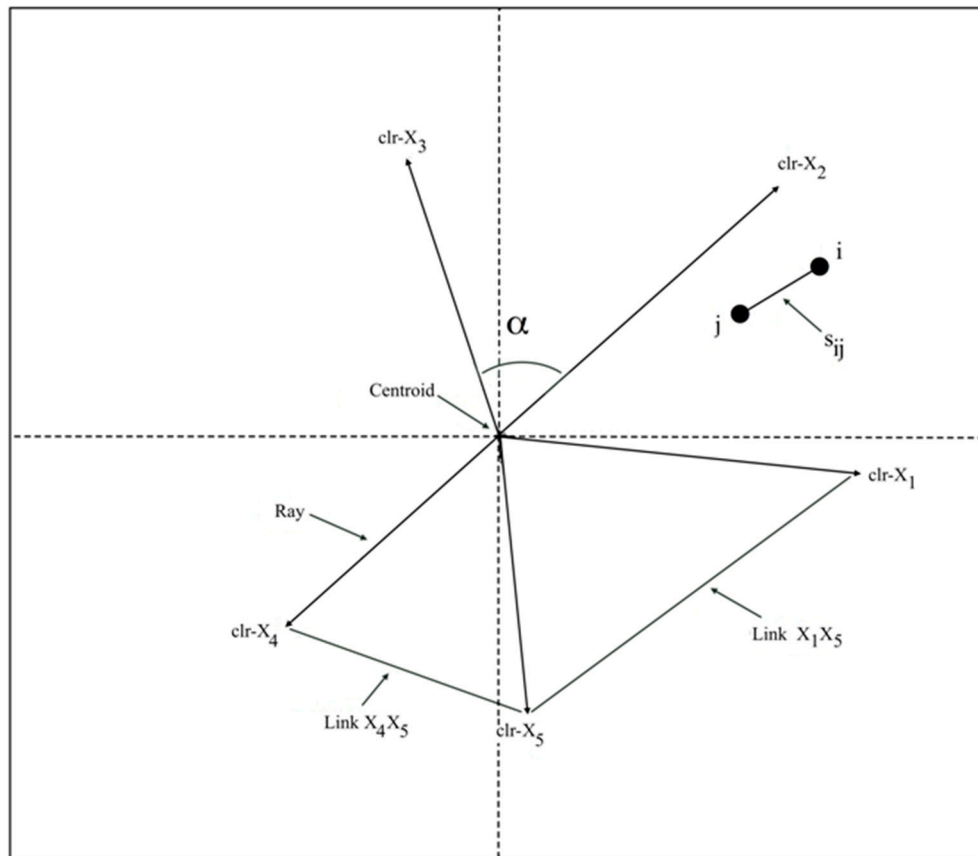


Figure 1. Elements of a compositional biplot.

In plant physiology and plant nutrition, it is common to study relationships between variables. Examples are the Ca/Mg ratio for the preparation of balanced nutrient solutions, the Ca/oxalic acid ratio in order to analyze the bioavailability of Ca [7], and the Deviation from Optimum Percentage (DOP) index, expressed as $DOP = C/C_{ref}$ where C is the nutrient concentration in the sample to assess and C_{ref} is the optimal nutrient concentration used as reference value, for information on nutritional status [8]. The link concept can be regarded as a statistical tool with which to study those and other relationships.

The purpose of this paper is to identify the link (and rays) among bioactive compounds of tomatoes through compositional HJ-biplot as alternative to the GH and JK compositional biplots. HJ-biplot was selected because it provides a higher goodness-of-fit than the Gabriel's [9–11] for rows and columns of the matrix. Besides, the HJ-biplot is line with the prerequisite of singular value decomposition (SVD) of the original log-transformed data matrix indicated by Aitchison and Greenacre [5].

It would be interesting to extend this idea to other Biplots (Canonical biplot for example) but the impact on links should be carefully studied since working with the matrix centroids and not the original data matrix.

Prior to constructing the compositional HJ-biplot, we applied a compositional Mann–Whitney U-test to the data to assess the effect of cultivar and cultivation system on the overall chemical composition of the tomatoes. Both of these procedures require a log transformation to be performed [3,5].

2. Results and Discussion

2.1. General Description of the Tomato Samples

Tables 1 and 2 give the results of the influence of cultivar and cultivation system on the chemical composition of the tomatoes (proximate composition, mineral elements, organic acids, and antioxidant compounds) together with the compositional Mann–Whitney U test results after centered log-ratio (clr)-transformation.

Moisture and ash showed no significant differences between cultivar or cultivation system. However, protein content, while similar to that reported by Valdivia-Mares et al. [12], was higher in cv. Mariana than in cv. Duncan, and in the no-soil than in the soil condition. There were no differences in glucose or fructose content whose values were similar to the maximum values reported by Figàs et al. [13] and Lahoz et al. [14]. While fruit and vegetable mineral element concentrations are strongly influenced by the content of those elements in the soil or nutritional solution [15], no significant differences were found in total Fe, Cu, or Zn. However, P, Na, K, Ca, and Mn differed significantly by cultivation system, and Mg by both cultivar and cultivation system. Compared with the results of Erba et al. [16], our P, Ca, and Mg values were greater, our Cu and Zn lower, and the rest of the mineral elements similar. The lycopene, ascorbic acid, and phenol contents were similar to the low values reported by Figàs et al. [13]. For citric acid, the contents were higher than those reported by Lahoz et al. [14], while for malic acid they were lower.

Considering just the effect of the tomato cultivar, the cv. Mariana's content was higher in protein, Mn, and ascorbic acid, and lower in Mg, lycopene, and citric and fumaric acids than cv. Duncan's.

Considering the cultivation system, the main differences were in the group of mineral and trace elements. For the soil-grown tomatoes, the Ca, Mn, and phenol contents were higher, and the protein, P, Na, K, and Mg lower than in the no-soil tomatoes.

Table 1. Proximate and mineral elements composition (mean \pm standard deviation, data expressed in fresh weight, FW) of tomato samples grouped according to cultivar and cultivation system.

Compound	Cultivar	Cultivation System		<i>p</i> -Value
		Soil	No-Soil	
Proximate Composition (% FW)				
Moisture	Dunkan	93.78 \pm 0.55	94.11 \pm 0.51	0.061 ¹
	Mariana	94.28 \pm 0.47	93.88 \pm 0.58	
	<i>p</i> -Value	0.129 ²		
Ash	Dunkan	0.75 \pm 0.09	0.73 \pm 0.04	0.159
	Mariana	0.69 \pm 0.07	0.74 \pm 0.07	
	<i>p</i> -Value	0.952		
Protein	Dunkan	0.68 \pm 0.05	0.75 \pm 0.13	0.049
	Mariana	0.70 \pm 0.07	0.77 \pm 0.07	
	<i>p</i> -Value	0.015		
Fructose	Dunkan	1.41 \pm 0.09	1.38 \pm 0.24	0.765
	Mariana	1.24 \pm 0.22	1.39 \pm 0.17	
	<i>p</i> -Value	0.145		
Glucose	Dunkan	1.40 \pm 0.09	1.37 \pm 0.24	0.798
	Mariana	1.23 \pm 0.22	1.38 \pm 0.17	
	<i>p</i> -Value	0.137		
Mineral Elements (mg/kg FW)				
P	Dunkan	255 \pm 26.0	280 \pm 36.3	0.455
	Mariana	242 \pm 20.8	276 \pm 32.3	
	<i>p</i> -Value	0.014		
Na	Dunkan	42.29 \pm 5.31	46.23 \pm 2.67	0.405
	Mariana	40.66 \pm 9.63	53.00 \pm 12.7	
	<i>p</i> -Value	0.001		
K	Dunkan	2654 \pm 340	2844 \pm 332	0.165
	Mariana	2570 \pm 275	2859 \pm 214	
	<i>p</i> -Value	0.093		
Ca	Dunkan	88.76 \pm 8.53	83.00 \pm 5.74	0.295
	Mariana	85.65 \pm 14.4	82.00 \pm 7.67	
	<i>p</i> -Value	0.046		
Mg	Dunkan	159 \pm 13	167 \pm 18	3.05 \times 10⁻⁴
	Mariana	115 \pm 16	133 \pm 13	
	<i>p</i> -Value	0.009		
Fe	Dunkan	3.36 \pm 0.65	3.14 \pm 0.48	0.109
	Mariana	2.66 \pm 0.32	2.98 \pm 0.48	
	<i>p</i> -Value	0.176		
Cu	Dunkan	0.35 \pm 0.18	0.47 \pm 0.20	0.623
	Mariana	0.32 \pm 0.08	0.33 \pm 0.04	
	<i>p</i> -Value	0.260		
Zn	Dunkan	0.80 \pm 0.18	0.99 \pm 0.29	0.146
	Mariana	0.98 \pm 0.31	0.96 \pm 0.12	
	<i>p</i> -Value	0.718		
Mn	Dunkan	1.04 \pm 0.12	0.99 \pm 0.12	0.053
	Mariana	1.04 \pm 0.11	1.02 \pm 0.13	
	<i>p</i> -Value	0.009		

¹ *p*-Value by tomato cultivar; ² *p*-Value by cultivation system. *p*-Values in bold mean significant differences after clr-transformation according to Mann–Whitney U test.

Table 2. Organic acids and antioxidant compounds composition (mean \pm standard deviation, data expressed in fresh weight, FW) of tomato samples grouped according to cultivar and cultivation system.

Compound	Cultivar	Cultivation System		p-Value
		Soil	No-Soil	
Organic Acids (mg 100 g⁻¹ FW)				
Citric acid	Dunkan	422 \pm 64.7	443 \pm 93.4	0.001 ¹
	Mariana	541 \pm 138	561 \pm 131	
	p-Value	0.999 ²		
Malic acid	Dunkan	25.7 \pm 9.87	22.1 \pm 4.16	0.732
	Mariana	25.5 \pm 9.14	19.9 \pm 5.95	
	p-Value	0.001		
Ascorbic acid	Dunkan	14.8 \pm 0.61	14.4 \pm 1.38	0.009
	Mariana	15.2 \pm 2.04	15.0 \pm 1.67	
	p-Value	0.087		
Oxalic acid	Dunkan	16.9 \pm 6.13	17.3 \pm 4.06	0.104
	Mariana	13.2 \pm 2.93	14.1 \pm 4.43	
	p-Value	0.798		
Fumaric acid	Dunkan	5.69 \pm 1.437	5.62 \pm 1.37	0.026
	Mariana	4.33 \pm 1.37	4.08 \pm 1.09	
	p-Value	0.499		
Pyruvic acid	Dunkan	1.40 \pm 0.44	1.87 \pm 0.71	0.608
	Mariana	1.80 \pm 0.89	1.40 \pm 0.74	
	p-Value	0.079		
Antioxidant Compounds (mg 100 g⁻¹ FW)				
Total phenols	Dunkan	24.3 \pm 3.00	23.5 \pm 1.07	0.091
	Mariana	26.3 \pm 6.77	23.6 \pm 6.81	
	p-Value	0.016		
Lycopene	Dunkan	2.34 \pm 0.12	2.63 \pm 0.36	0.001
	Mariana	1.73 \pm 0.36	1.76 \pm 0.40	
	p-Value	0.741		
Chlorogenic acid	Dunkan	1.08 \pm 0.24	1.09 \pm 0.13	0.814
	Mariana	0.996 \pm 0.26	1.15 \pm 0.42	
	p-Value	0.344		
Ferulic acid	Dunkan	0.124 \pm 0.03	0.125 \pm 0.02	0.685
	Mariana	0.121 \pm 0.03	0.130 \pm 0.02	
	p-Value	0.548		
Caffeic acid	Dunkan	0.115 \pm 0.02	0.115 \pm 0.05	0.010
	Mariana	0.083 \pm 0.03	0.093 \pm 0.02	
	p-Value	0.674		

¹ p-Value by tomato cultivar; ² p-Value by cultivation system. p-Values in bold mean significant differences after clr-transformation according to Mann–Whitney U test.

Tomato consumption is associated with healthy diets for its contribution to the daily intake of carotenoids (providing an estimated 80% of daily intake of lycopene) and of folate, ascorbic acid, flavonoids, α -tocopherol, and potassium [17,18]. However, other nutritionally important variables such as the oxalic acid content must also be considered. This dicarboxylic acid forms insoluble salts with calcium and other divalent cations, leading to a decrease in their bioavailability. It is therefore considered to be an anti-nutrient compound inhibiting the correct absorption of Ca [7]. According to Mitjavila [19], 2.5 g of oxalic acid precipitates with 1 g of Ca, so that its anti-nutrient effect is determined by the oxalic acid/Ca ratio. According to the values given in Tables 1 and 2, for cv. Dunkan, this ratio

was 1.90 and 2.08 for the soil and no-soil cultivation systems, respectively, while for cv. Mariana, the ratio was lower, 1.54 and 1.72, respectively. These differences were not found to be significant, however.

2.2. Differences in Bioactive Compounds by Cultivar and Cultivation System

The antioxidant profile of the diet has recently been recognized as being an important variables for predicting the impact of food on human health [20]. The main antioxidant of tomato, lycopene, has been widely reported to play a major role in the prevention of cardiovascular disease and various types of cancer [21]. One must also consider the content and relationships of phenolics (including chlorogenic, caffeic, and ferulic acids) and other bioactive compounds such as ascorbic acid, K, and Mg.

In the context of compositional statistics, the analysis of such relationships requires the use of logarithms of the ratios between variables, i.e., between the parts of the compositional vector. When these log-ratios are taken to be those between each component and the geometric mean, they are known as centred log-ratios [5,22].

It is important to emphasize how the problem of spurious correlations is resolved with compositional statistics. By way of illustration, a standard HJ-biplot (Figure 2A) and a compositional HJ-biplot (Figure 2B) were constructed for lycopene, total phenolics, chlorogenic acid, caffeic acid, ferulic acid, ascorbic acid, K, and Mg. The HJ-biplot was constructed with data expressed as percentages, and the compositional HJ-biplot with centred log-ratios, in both case with a double-centring transformation of the matrix. The angles between the vectors approximate the correlations between variables in such a way that small acute angles are associated with variables that are strongly positively correlated, obtuse angles close to 180° with variables that are strongly negatively correlated, and right angles with uncorrelated variables. One observes different correlations between the variables, and different percentages of the variance explained in each biplot, with an overall fit of 54.00% in the HJ-biplot and 60.72% in the compositional HJ-biplot.

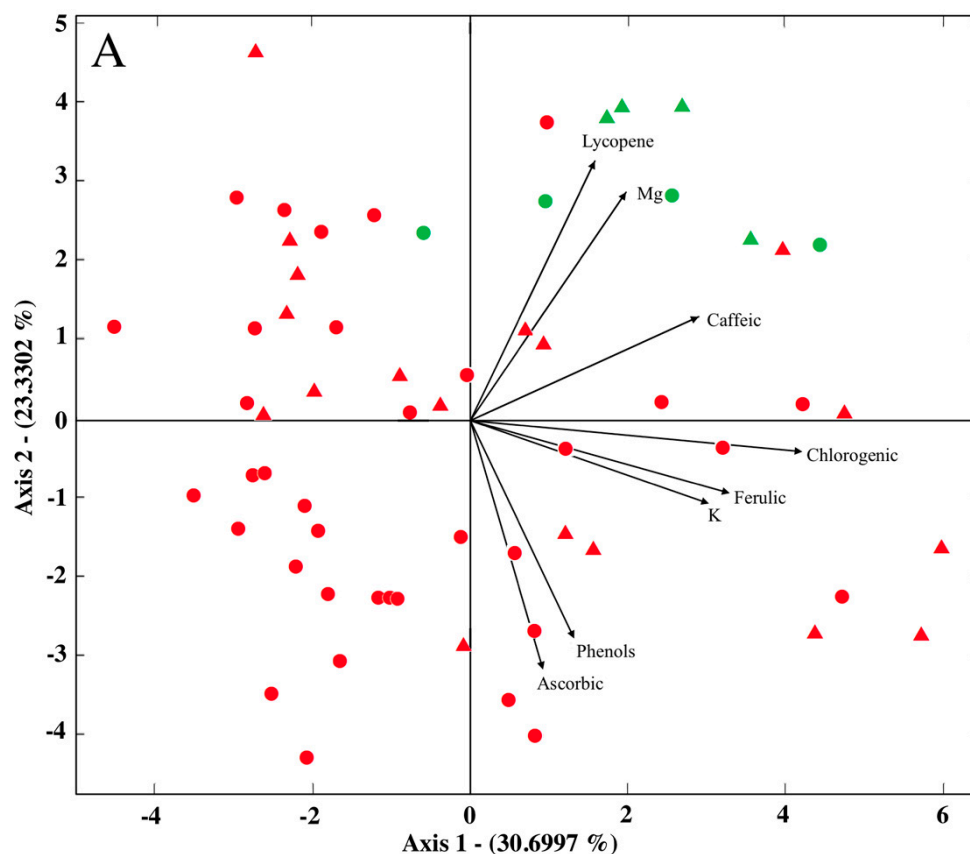


Figure 2. Cont.

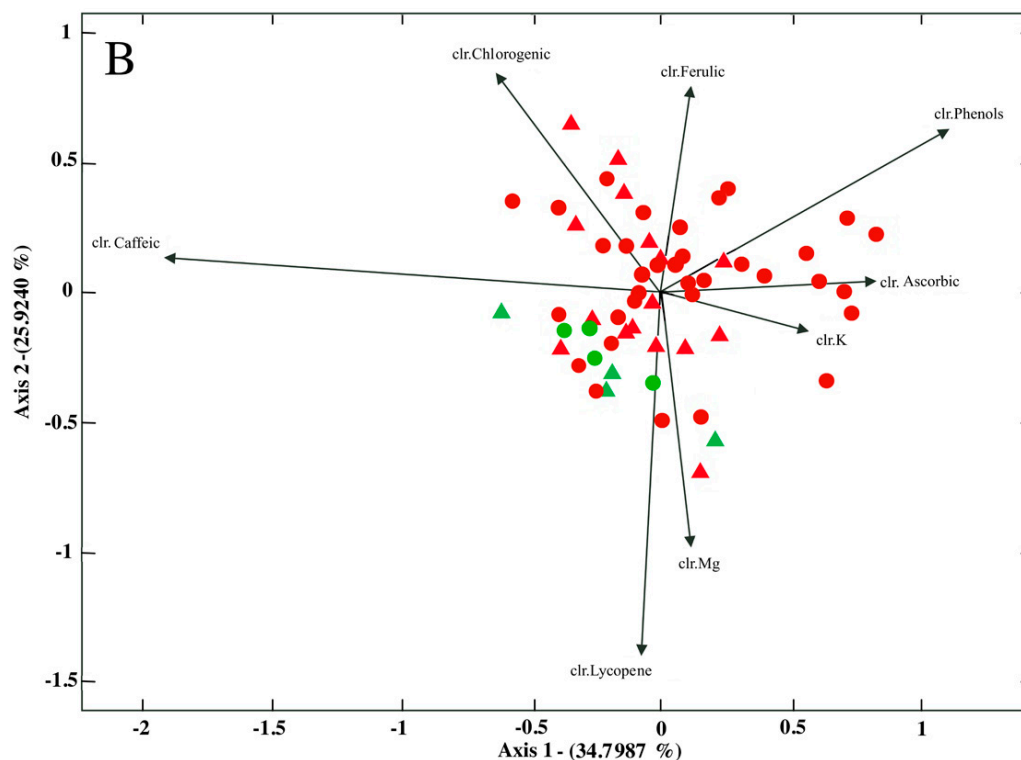


Figure 2. (A) Standard HJ-biplot; and (B) Compositional HJ-biplot of the chemical composition data. The Mariana tomato cultivar is represented in red; the Duncan in green. The soil cultivation method is represented with the marker ●; the no-soil with the marker ▲.

Regarding compositional HJ-biplot (Figure 2B), samples near the centroid have values similar to the geometric mean for each component, while those far from the center can be diagnosed as having a large relative variation with respect to the geometric mean for the clr-transformed chemical variables that the points are near. In the case of the rays, caffeic acid has the greatest variation with respect to its mean, while the smallest variation corresponds to K. One observes a tendency that distinguishes between the two cultivars. The cv. Mariana samples are mainly located near the rays of chlorogenic, ferulic, and ascorbic acids, phenolics, and K, while the cv. Duncan samples are near the rays of lycopene, caffeic acid, and Mg. Indeed, the highest values for each of those chemical variables correspond to the respective cultivars (Tables 1 and 2). Considering just the no-soil samples in particular, the main effect of the cultivar is in the greater content of chlorogenic and ferulic acids for cv. Mariana, and of lycopene, caffeic acid, and Mg for cv. Duncan.

All these relationships required more detailed consideration, especially that of lycopene with the rest of the variables. The acute angles for Mg/lycopene and K/ascorbic acid indicate positive correlations, a correlation that is especially strong in the case of Mg. This is coherent with the shortness of the Mg/lycopene link (see the rules for interpretation described in the Statistical Methods Section). In order to quantify the relationship, the values of the variance given in the upper triangle of Table 3 must be considered. For the Mg/lycopene log-ratio, the variance corresponds to a standard deviation of 0.207. Since this value is expressed on a logarithmic scale, it must be back-converted to a Euclidean scale, i.e., $e^{0.207} \approx 1.23$. Thus, an almost linear increase in the lycopene content is due to the Mg content of the tomato fruit. However, K is more strongly correlated with the ascorbic acid than the lycopene content, with an estimated standard deviation of 1.14 (K/ascorbic acid). Fanasca et al. [23] showed that the cations K, Ca, and Mg affect the antioxidant content of tomatoes, especially under hydroponic cultivation. In our study too, the highest values of Mg and K were in the samples grown under no-soil conditions. In the case of lycopene/ferulic acid, the obtuse, close to 180°, angle between their rays indicates a possible strong negative correlation. This could be because the antioxidant content

then mixed and homogenized to homogeneous purée. A fraction of this purée was desiccated at 105 °C, homogenized again, and stored in a polyethylene tube at room temperature until assay for metals and protein. The rest was immediately stored in a polyethylene tube at –80 °C in order to avoid enzymatic changes and for the measurement of the other chemical variables.

3.3. Analytical Methods

All assays were performed in triplicate, and the results expressed per fresh weight (FW). Moisture was determined using the oven-drying method [26]. The residue was heated at 550 °C for 24 h for the ash determination [26]. Organic nitrogen was determined by the Kjeldahl method [26], and the protein concentration estimated using 6.25 as the conversion factor. The method of Kujala et al. [27] was used to determine the phenolic compound content: after the colorimetric reaction with the Folin-Ciocalteu reagent (Sigma-Aldrich Chemical Co., St. Louis, MO, USA), absorbance was measured at 750 nm using gallic acid (Sigma-Aldrich) standard solutions for the external standard calibration curve. Ascorbic acid was determined in the individual tomatoes using the 2,6-dichlorophenolindophenol titration procedure [26]. The lycopene concentration was determined spectrophotometrically at 503 nm following extraction in 20 mL of acetone:ethanol:hexane (5:5:10 *v/v*) in the dark, in accordance with the method described by Fish et al. [28].

Mineral and trace elements were determined by atomic absorption spectrometry using a Varian SpectrAA-10Plus (Varian Iberica SL, Madrid, Spain) atomic absorption spectrometer equipped with a deuterium-lamp background correction system using an air-acetylene flame. The dried tomato samples were previously acid-digested in nitric acid, in accordance with the procedure described by Hernández Suárez et al. [9].

Sugars were determined as described by Rodríguez Galdón et al. [29]. About 1 g of the frozen homogenized tomato was weighed directly in polypropylene tubes and mixed with ethanol:water (80:20 *v/v*). One milliliter of the solution was filtered (0.45 µm filter GHP, Waters, Milford, MA, USA), and the sugars were assayed by high-performance liquid chromatography (HPLC) using a Waters 2690 chromatograph equipped with a Waters model 2414 differential refractive index (DRI) detector (Waters, Milford, MA, USA). The separation was performed using a Waters carbohydrate analysis column (3.9 mm × 300 mm i.d.) with a particle-size diameter of 10 µm, equipped with a Waters CarboTM4-mm carbohydrate guard column. The eluent was acetonitrile:water (80:20 *v/v*), and the peaks were identified by comparison with the retention times of standard compounds.

Organic acids were determined by HPLC as described by Hernández Suárez et al. [30]. About 1 g of the frozen homogenized tomato was weighed directly in polypropylene tubes and mixed with ethanol:water (80:20 *v/v*). After filtering, the liquid phase was concentrated with a nitrogen stream, and the residue was redissolved in ultrapure water (Milli-Q water system, Waters, Milford, MA, USA), followed by filtration through a 0.45 µm pore size membrane and a Sep-Pak Accell Plus QMA cartridge (Waters, Milford, MA, USA) which had been pre-conditioned with 3 mL of ultrapure water. The eluent used was 2 mL of 20 mM sodium dihydrogen phosphate (adjusted to pH 1.0 with phosphoric acid buffer). The analytical system was a Waters 2690 high-performance liquid chromatograph equipped with a Waters 996 photodiode array detector (Waters, Milford, MA, USA). Separation was performed using a Waters Atlantis dC18 steel column (150 mm × 4.6 mm i.d.) with a particle diameter of 3 µm equipped with a Waters Atlantis dC18 guard column (20 mm × 4.6 mm). The eluent was 20 mM sodium dihydrogen phosphate (adjusted to pH 2.7 with phosphoric acid buffer), and the 210-nm wavelength peaks were identified by comparison with the retention times of standard compounds.

Free hydroxycinnamic acids were determined by HPLC as described by Hernández Suárez et al. [31]. About 1 g of the frozen homogenized tomato was mixed with 75% methanol (pH 2.5 adjusted with trifluoroacetic acid (TFA)), heated to 40 °C for 30 min, and then centrifuged. The supernatant was filtered (0.45 µm pore size) prior to HPLC analysis. The analytical system was the same as that used for the organic acid assays. The separation was performed using a Nova-Pak C18 steel cartridge (150 mm × 3.9 mm i.d.) with a particle diameter of 4 µm equipped with a Waters C18 guard column.

The eluent was 0.05% TFA (pH 2.0) and methanol. For the detection, three wavelengths were used (280, 290, and 329 nm).

3.4. Statistical Methods

In order to apply standard statistical multivariate analysis, the raw compositional data were subjected to a clr-transformation. This transformation is symmetric with respect to the compositional parts, and maintains the same number of components as the number of parts in the composition. Its advantage is that it is an isometric transformation of the simplex with Aitchison metric onto a subspace of real space with the ordinary Euclidean metric [3,5,22]. The clr-transformation is given by the expression:

$$clr(x) = \left[\ln \frac{x_1}{g_m(x)}, \ln \frac{x_2}{g_m(x)}, \dots, \ln \frac{x_D}{g_m(x)} \right]$$

where $x = (x_1, x_2, \dots, x_D)$ is a compositional data vector, the x_i are the chemical variables all of which must be expressed in the same units, and $g_m(x)$ is the geometric mean of that compositional data vector. Calculations were performed using the CoDaPack 2.01 software package [32].

The sample sizes in the present study were unbalanced 8 tomato Dunkan cultivar tomato samples and 53 Mariana tomato cultivar samples from Gran Canaria (Spain) grown either conventionally or hydroponically. To compare the means of the variables, a Mann–Whitney U test was applied on the clr-transformed data (i.e., a compositional Mann–Whitney U test), with the significance level for differences among them being taken as $p < 0.05$. This calculation was performed using the SPSS v. 21.0 program package (SPSS Inc., Chicago, IL, USA). The HJ-biplot is a joint representation in a low-dimensional vector space (usually a plane) of the rows and columns of a data matrix X , using markers (points/vectors) j_1, j_2, \dots, j_n for its rows and h_1, h_2, \dots, h_p for its columns. The markers are obtained from the usual singular value decomposition (SVD) of the data matrix $X = U\Sigma V^T$, where U is formed by the eigenvectors of the matrix XX^T , V by the eigenvectors of the matrix X^TX , and Σ is a diagonal matrix containing the singular values (i.e., the square roots of the non-zero eigenvalues of both XX^T and X^TX), taking as rows the marker rows of $J = U\Sigma$ and as columns the marker rows of $H = V\Sigma$, in the appropriate dimensions. Thus, the matrix X is formed by clr-transformed data, and then double-centered through its SVD to ensure that the components are analyzed on a ratio scale [5] in the appropriate dimensions.

A biplot of compositional data consists of the elements shown in Figure 1 [4,5]:

The similarity (S_{ij}) between two samples or individuals is taken to be an inverse function of their distance, in such a way that closer samples are more similar.

The centroid represents the center-of-gravity formed by the geometric mean of the compositional parts used in the clr-transformation.

ray provides information on the variance of the corresponding log-ratio with respect to the geometric mean (gm):

$$var \left(\ln \frac{X_i}{g_m(X)} \right)$$

The square root of this expression is the standard deviation of the clr-transformed variable X_i , and is represented by the length of the ray. The cosine of the angle (α) between two rays represents the approximate correlation coefficient between the corresponding variables.

The length of a link between the vertices of two rays represents the standard deviation of the log-ratio between the associated compositional parts, with the corresponding variance being defined as:

$$var \left(\ln \frac{X_i}{X_j} \right)$$

Angles between links provide information on the relationships between pairs of variables. If two links intersect at a right angle then this is an indication that the pairs of variables are possibly uncorrelated (in Figure 1, links X_1X_5 and X_4X_5), while if they are parallel (or the angle is obtuse) then the pairs of parameters may be strongly correlated (in Figure 1, links X_1X_5 and X_4X_3). Coincident vertices or short links mean that the two variables are linearly proportional, so that the two parts involved can be assumed to be redundant. If a subset of links is collinear, this might indicate a possible one-dimensional variability.

The compositional HJ-biplot analyses were done using the MULTBILOT program package (MULTivariate analysis using BILOT) developed by Vicente-Villardón [33].

4. Conclusions

Compositional Data Analysis (CoDA) refers to the analysis of data, which have been defined as random vectors with strictly positive components whose sum is constant. In this paper we propose a new tool to analyze and represent compositional data, “Compositional HJ-biplot”, based on HJ-biplot and Aitchison Geometry to capture the algebraic-geometric structure of the data. Besides, this is the first time that an HJ-biplot multivariate analysis has been applied to chemical compositional data. Taking into account that HJ-biplot provides a better quality of representation than Gabriel biplots (GH and JK biplot), we can state that Compositional HJ-biplot get the best representation of the compositional data, better than “form biplot” and “covariance biplot”, proposed by Aitchison and Grenacre.

To extend to other Biplots, such as MANOVA biplot, the impact on links should be carefully studied. Using this novel statistical technique, linear relationships or links between the variables were identified that are different from those provided by the usual correlation coefficient studies. The value of links allows new ways to study the relationship among variables. In order to validate this statistical methodology, the results were contrasted with other scientific results, however more repetitions in several localities and year are necessary to test the results.

Acknowledgments: We wish to thank the COAGRISAN company (Sociedad Cooperativa Agrícola de San Nicolás de Tolentino de Gran Canaria, Spain) for their collaboration in providing the tomatoes. We also gratefully acknowledge the help of Vanesa Raya Ramallo (Instituto Canario de Investigaciones Agrarias, Tenerife, Spain) for her collaboration in checking the agronomical data.

Author Contributions: Marcos Hernández Suárez and Francisco Espinosa Borreguero conceived and designed the experiments. Marcos Hernández Suárez, Daniel Molina Pérez, Francisco Espinosa Borreguero, Elena M. Rodríguez-Rodríguez and Carlos Díaz Romero collected the materials and carried out the experiments and analytical determinations. Marcos Hernández Suárez and Purificación Galindo-Villardón analyzed statistical data. Elena M. Rodríguez-Rodríguez and Carlos Díaz Romero interpreted the results of compositional MANOVA. Marcos Hernández Suárez, Purificación Galindo-Villardón and Francisco Espinosa Borreguero interpreted the results of compositional HJ-biplot. Marcos Hernández Suárez, Daniel Molina Pérez, Francisco Espinosa Borreguero, Elena M. Rodríguez-Rodríguez, Purificación Galindo-Villardón and Carlos Díaz Romero wrote the manuscript. All authors read and approved the finalized manuscript.

Conflicts of Interest: The authors declare no conflict of interests. The COAGRISAN company had no role in the design of the study, in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Hernández Suárez, M.; Rodríguez Rodríguez, E.M.; Díaz Romero, C. Chemical composition and nutritional value of tomatoes. In *Tomatoes: Cultivation, Varieties and Nutrition*; Higashide, T., Ed.; Nova Science Publishers: New York, NY, USA, 2013; pp. 191–222.
2. ISTAC. Instituto Canario de Estadística. Available online: <http://www.gobiernodecanarias.org/istac> (accessed on 13 January 2016).
3. Aitchison, J. *The Statistical Analysis of Compositional Data*; Chapman and Hall: London, UK, 1986; p. 416.
4. Pawlowsky-Glahn, V.; Egozcue, J.J.; Tolosana-Delgado, R. *Modeling and Analysis of Compositional Data*; John Wiley and Sons, Ltd.: Chichester, UK, 2015; p. 272.

5. Aitchison, J.; Greenacre, M. Biplots for compositional data. *Appl. Stat.* **2002**, *51*, 375–382. [[CrossRef](#)]
6. Gabriel, K.R. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* **1971**, *58*, 453–467. [[CrossRef](#)]
7. Guil, J.L.; Torija, M.E.; Giménez, J.J.; Rodríguez-García, I.; Giménez, A. Oxalic acid and calcium determination in wild edible plants. *J. Agric. Food Chem.* **1996**, *44*, 1821–1823. [[CrossRef](#)]
8. Montañés, L.; Heras, L.; Abadía, J.; Sanz, M. Plant analysis interpretation based on a new index: Deviation from optimum percentage (DOP). *J. Plant Nutr.* **1993**, *16*, 1289–1308. [[CrossRef](#)]
9. Galindo-Villardón, M. Una alternativa de representación simultánea: HJ-Biplot. *Qüestió* **1986**, *10*, 13–23.
10. Frutos, E.; Galindo, P.; Leiva, V. An interactive biplot implementation in R for modeling genotype-by-environment interaction. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 1629–1641. [[CrossRef](#)]
11. Hernández, M.; Espinosa, F.; Galindo, P. Tomato fruit quality as influenced by interactions in agricultural techniques and harvest period. *J. Plant Nutr. Soil Sci.* **2014**, *177*, 443–448. [[CrossRef](#)]
12. Valdivia-Mares, L.E.; Rodríguez, F.A.; Sánchez, J.J.; Vargas-Ponce, O. Phenology, agronomic and nutritional potential of three wild husk tomato species (*Physalis*, *Solanaceae*) from Mexico. *Sci. Hortic. (Amsterdam)* **2016**, *200*, 83–94. [[CrossRef](#)]
13. Figàs, M.R.; Prohens, J.; Raigón, M.D.; Fita, A.; García-Martínez, M.D.; Casanova, C.; Borràs, D.; Plazas, M.; Andújar, I.; Soler, S. Characterization of composition traits related to organoleptic and functional quality for the differentiation, selection and enhancement of local varieties of tomato from different cultivar groups. *Food Chem.* **2015**, *187*, 517–524. [[CrossRef](#)] [[PubMed](#)]
14. Lahoz, I.; Pérez-de-Castro, A.; Valcárcel, M.; Macua, J.I.; Beltrán, J.; Roselló, S.; Cebolla-Cornejo, J. Effect of water deficit on the agronomical performance and quality of processing tomato. *Sci. Hortic.* **2016**, *200*, 55–65. [[CrossRef](#)]
15. Hernández Suárez, M.; Rodríguez Rodríguez, E.M.; Díaz Romero, C. Mineral and trace element concentrations in cultivars of tomatoes. *Food Chem.* **2007**, *104*, 489–499. [[CrossRef](#)]
16. Erba, D.; Casiraghi, M.C.; Ribas-Agustí, A.; Cáceres, R.; Marfà, O.; Castellari, M. Nutritional value of tomatoes (*Solanum lycopersicum* L.) grown in greenhouse by different agronomic techniques. *J. Food Compos. Anal.* **2013**, *31*, 245–251. [[CrossRef](#)]
17. Bramley, P.M. Is lycopene beneficial to human health? *Phytochemistry* **2000**, *54*, 233–236. [[CrossRef](#)]
18. Willcox, J.K.; Catignani, G.L.; Lazarus, S. Tomatoes and cardiovascular health. *Crit. Rev. Food Sci. Nutr.* **2003**, *43*, 1–18. [[CrossRef](#)] [[PubMed](#)]
19. Mitjavila, S. Sustancias naturales nocivas en los alimentos. In *Toxicología y Seguridad de los Alimentos*; Derache, J., Ed.; Omega: Barcelona, Spain, 1990.
20. Danesi, F.; Bordoni, A. Effect of home freezing and Italian style of cooking on antioxidant activity of edible vegetables. *J. Food Sci.* **2008**, *73*, H109–H112. [[CrossRef](#)] [[PubMed](#)]
21. Omoni, A.O.; Aluko, R.E. The anti-carcinogenic and anti-atherogenic effects of lycopene: A review. *Trends Food Sci. Technol.* **2005**, *16*, 344–350. [[CrossRef](#)]
22. Egozcue, J.J.; Pawlowsky-Glahn, V.; Mateu-Figueras, G.; Barceló-Vidal, C. Isometric logratio transformations for compositional data analysis. *Math. Geol.* **2003**, *35*, 279–300. [[CrossRef](#)]
23. Fanasca, S.; Colla, G.; Maiani, G.; Venneria, E.; Roupheal, Y.; Azzini, E.; Saccardo, F. Changes in antioxidant content of tomato fruits in response to cultivar and nutrient solution composition. *J. Agric. Food Chem.* **2006**, *54*, 4319–4325. [[CrossRef](#)] [[PubMed](#)]
24. Flueret, A.; Macheix, J.J. Quinyl esters and glucose derivatives of hydroxycinnamics during growth and ripening of tomato fruit. *Phytochemistry* **1981**, *20*, 667–671. [[CrossRef](#)]
25. Clé, C.; Hill, L.M.; Niggeweg, R.; Martin, C.R.; Guisez, Y.; Prinsen, E.; Jansen, M.A. Modulation of chlorogenic acid biosynthesis in *Solanum lycopersicum*; consequences for phenolic accumulation and UV-tolerance. *Phytochemistry* **2008**, *69*, 2149–2156. [[CrossRef](#)] [[PubMed](#)]
26. AOAC (Association Official of Analytical Chemists). *Official Methods of Analysis of AOAC International*, 18th ed.; Horwitz, W., Latimer, G.W., Jr., Eds.; AOAC International: Gaithersburg, MD, USA, 2006.
27. Kujala, T.S.; Loponen, J.M.; Klika, K.D.; Pihlaja, K. Phenolic and betacyanins in red beetroot (*Beta vulgaris*) root: Distribution and effect of cold storage on the content of total phenolic and three individual compounds. *J. Agric. Food Chem.* **2000**, *48*, 5338–5342. [[CrossRef](#)] [[PubMed](#)]
28. Fish, W.; Perkins-Veazie, P.; Collins, J. A quantitative assay for lycopene that utilizes reduced volumes of organic solvents. *J. Food Comp. Anal.* **2002**, *15*, 309–317. [[CrossRef](#)]

29. Rodríguez Galdón, B.; Tascón Rodríguez, C.; Rodríguez Rodríguez, E.M.; Díaz Romero, C. Fructans and major compounds in onion cultivars (*Allium cepa*). *J. Food Compos. Anal.* **2009**, *22*, 25–32. [[CrossRef](#)]
30. Hernández Suárez, M.; Rodríguez Rodríguez, E.; Díaz Romero, C. Analysis of organic acid content in cultivars of tomato harvested in Tenerife. *Eur. Food Res. Technol.* **2008**, *226*, 423–435. [[CrossRef](#)]
31. Hernández, M.; Rodríguez, E.; Díaz, C. Free hydroxycinnamic acids, lycopene, and color parameters in tomato cultivars. *J. Agric. Food Chem.* **2007**, *55*, 8604–8615.
32. Comas-Cufí, M.; Thió-Henestrosa, S. CoDaPack 2.0. A stand-alone, multi-platform compositional software. In Proceedings of the 4th International Workshop on Compositional Data Analysis, CoDaWork'11, Sant Feliu de Guíxols, Spain, 9–13 May 2011.
33. Vicente Villardón, J.L. *MULTBILOT: A Package for Multivariate Analysis Using Biplots*; Departamento de Estadística, Universidad de Salamanca: Salamanca, Spain, 2015.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).