

# Assessment of the genetic structure of two different tomato breeding populations by Principal Component Analysis

## *Evaluación de la estructura genética de dos poblaciones de mejoramiento diferentes de tomate mediante Análisis de Componentes Principales*



Vitelleschi, María Susana; Pratta, Guillermo Raúl

 **María Susana Vitelleschi**

[mvitelle@fcecon.unr.edu.ar](mailto:mvitelle@fcecon.unr.edu.ar)

Instituto de Investigaciones Teóricas y Aplicadas en Estadística (IITAE), Argentina

Consejo de Investigaciones de la Universidad Nacional de Rosario (CIUNR), Argentina

Universidad Nacional de Rosario, Argentina

 **Guillermo Raúl Pratta**

[gpratta@unr.edu.ar](mailto:gpratta@unr.edu.ar)

Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR), Argentina

Universidad Nacional de Rosario, Argentina

CONICET, Argentina

**Revista FAVE Sección Ciencias Agrarias**

Universidad Nacional del Litoral, Argentina

ISSN: 2346-9129

ISSN-e: 2346-9129

Periodicity: Semestral

no. 23, e0024, 2024

[revistafave@fca.unl.edu.ar](mailto:revistafave@fca.unl.edu.ar)

Received: 01 August 2023

Accepted: 11 March 2024

URL: <http://portal.amelica.org/ameli/journal/586/5864885008/>

DOI: <https://doi.org/10.14409/fa.2024.23.e0024>

Corresponding author: [gpratta@unr.edu.ar](mailto:gpratta@unr.edu.ar)



This work is licensed under Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International.

**Abstract:** In Plant Breeding, different populations are generated, which frequently represent different gene arrangement from a common selected gene pool. The technique of Principal Component Analysis (PCA) has been widely applied to evaluate the genetic structure of different populations. The objective of this research was to assess PCA for evaluating the genetic structure of two breeding tomato populations, one representing a final step of a breeding program (RILs population) and the other, an initial step (a six basic generations' population, composed by two homozygous parent, their heterozygous F1 and the segregating F2 and two backcrosses). Both populations were evaluated for phenotypic quantitative traits and population structure was assessed in terms of variances and covariances. PCA was adequate for evaluating differences in genetic structure for evaluated fruit quality traits in both populations.

**Keywords:** fruit quality, Classical Quantitative Genetics, Plant Genetic Resources, Multivariate Statistics Techniques.

**Resumen:** En el fitomejoramiento, se generan diferentes poblaciones que con frecuencia representan diferentes arreglos genéticos de un acervo genético común seleccionado. La técnica de Análisis de Componentes Principales (PCA) ha sido ampliamente aplicada para evaluar la estructura genética de diferentes poblaciones. El objetivo de esta investigación fue aplicar PCA para evaluar la estructura genética de dos poblaciones de tomate de mejoramiento, una que representa un paso final de un programa de mejoramiento (población de RILs) y otra, un paso inicial (una población de seis generaciones básicas, compuesta por dos progenitores homocigotos y su F1 heterocigota, como materiales genéticamente uniformes, y las generaciones segregantes F2 y las dos retrocruzas). Ambas poblaciones fueron evaluadas para rasgos cuantitativos fenotípicos y la estructura de la población fue evaluada en términos de varianzas y covarianzas. La técnica de PCA fue adecuada para evaluar las diferencias en la estructura genética de los rasgos de calidad de la fruta evaluados en ambas poblaciones.

**Palabras clave:** calidad del fruto, Genética Cuantitativa Clásica, Recursos Fitogenéticos, Técnicas Estadísticas Multivariadas.

## INTRODUCTION

Plant breeding is the art and science of modifying the genetic structure of plant populations to better satisfy human requirements (Acquaah, 2015). In fact, plant breeding implies the development of different populations which are voluntarily created by crossing selected genotypes followed by selfing, backcrossing, or open pollinations, among other strategies for obtaining the desired genetic structure. Most traits of agronomic interest are quantitative, hence the genetic structure of populations are measured in terms of variance and covariance. In consequence, the strategy to be implemented mainly depends on reproductive biology of the crop and the genetic variance and covariance composition underlying the target traits (Kearsey and Pooni 1996).

Frequently, a considerable amount of data is generated in this evaluation of plant breeding populations (Acquaah, 2015). Principal Component Analysis (PCA) is a tool used not only to reduce the dimension of the data keeping as much variability as possible but also for pre-processing data which will be then analyzed by unsupervised multivariate statistical techniques. Since PCA is performed on covariance matrices of the dataset, it has been widely used for studying different population genetic structures (McVean, 2010; Lu and Xu, 2013; Nachimuthu et al., 2014) according to phenotypic and/or genotypic data.

Tomato (*Solanum lycopersicum* L.) is one of the most important horticultural crops worldwide (FAOSTAT, 2017). Also, it is a model species for plant genetics and breeding in both classical and new strategies (Gerszberg et al., 2015). Phenotypic evaluation is essential in different steps of a breeding program, especially when variability for quantitative agronomic traits was increased by crosses to wild germplasm (Dempewolf et al., 2017). Rodríguez et al. (2006) obtained eighteen recombinant inbred lines (RILs) by crossing the Argentinean cultivar Caimanta to *S. pimpinellifolium* L. LA0722 followed by selection for fruit weight and shelf life from the F2 segregating population. Second cycle hybrids (SCH, *i.e.*, F1 obtained from crossing selected RILs) and their corresponding segregating generations were developed to continue with the breeding program. Phenotypic and molecular variation and covariation was assessed in all populations which have different linkage disequilibrium (Pratta et al. 2011a, Pereira da Costa et al. 2014, Cabodevila et al. 2021), and PCA was widely used in these studies (Pratta et al., 2010; Pratta et al., 2011b).

The objective of this research was to assess the genetic structure for phenotypic quantitative fruit traits of two sets of tomato populations, which were derived from the same interspecific cross but represent different steps of a breeding program, by PCA application.

## MATERIAL AND METHODS

### Plant populations and traits under study

Two sets of populations were evaluated with the aim of considering two different plant breeding activities. Both of them represent genomic recombination among genes contributed by the same parental genotypes (cv. Caimanta of *S. lycopersicum* and LA0722 of *S. pimpinellifolium*) in discrepant conditions of linkage disequilibrium, genotypic composition and inbreeding level.

The first set, the RILs population, comprised eight of the 18 different genotypes obtained by Rodríguez et al. (2006), hereafter named as L1, L5, L6, L8, L9, L15, L17, and L18 (total N = 396 plants because

---

## AUTHOR NOTES

gpratta@unr.edu.ar

some individuals were lost during the transplant; final number of individual per RIL is given in Table 1 and selected for being representative of total variability. In this set, linkage disequilibrium is low, inbreeding level is high and genotypes are homozygous, representing potential new tomato commercial cultivars derived after several cycles of artificial selection from a cross among cultivated and exotic germplasms. Data analyzed in this research are the mean values over 6 years of agronomic evaluation and this population represents a final step in plant breeding programs: the development of new genotypes according to society's requirements.

The second set, the six basic generations (SBG, according to Kearsey and Pooni, 1986) population, comprised two selected RILs (L1 and L18), the SCH F1(L18 x L1) and its segregating generations F2(L18 x L1), obtained by selfing, and both backcrosses F1(L18 x L1) x L18 and F1(L18 x L1) x L1, hereafter named as F1, F2, BC1 and BC2, respectively (N = 218 plants). In this set, linkage disequilibrium is high, inbreeding level is low and genotypes are both homozygous and heterozygous, representing early generations from a cross among elite genotypes that gives new opportunity of recombining and selecting over the genotypes resulting from the previous breeding actions that allowed deriving the parental RILs. Data analyzed in this research were measured just in one year of agronomic evaluation. This population represents an initial step in plant breeding programs; the creation of new variability by crossing discrepant parents and recombining their alleles in early segregating generations with the aim of obtaining new genotypes to satisfy human demands.

Both sets were assayed under greenhouse conditions at the experimental field station "J.F. Villarino", Universidad Nacional de Rosario, Argentina (Latitude: -33.016667°, Longitude: -60.883333°, Altitude: 50 masl) according to a completely randomized design. Following Mahuad *et al.* (2013), 11 quantitative traits were evaluated, five of them in fruits harvested at breaker stage (when carotenoids accumulation becomes visible) and the other five in fruits harvested at red ripe stage (with approximately 90% of red surface). In 10 fruits per plant at breaker stage, Weight (W, in g), Diameter (D, in cm), Height (H, in cm), Shape Index (SI, the ratio between Height and Diameter), and Shelf Life (SL, number of days from harvest until the fruit stored at  $25 \pm 3$  °C loses commercial value due to, for instance, excessive softening), were measured. In fruits at red ripe stage, the following traits were evaluated: Soluble Solids content (SS, in Brix degrees) as the percentage of fructose and glucose in the fruit juice, pH and Titratable acidity (TA, in g of citric acid per 100 g of homogenate) of the fruit juice, Firmness (F, measured on two opposite equatorial sides with a digital firmness type Shore A tester Durofel, DFT 100, with a 0.10 cm. cap), ratio a/b or Chroma index (parameter related to color tone, being "a" the absorbance at 540 nm wavelengths and "b" the absorbance at 675 nm wavelengths), and L value or Reflectance Percentage (L, parameter related to color intensity, presenting values that range from +100 for white to 0 for black). Values "a", "b" and L were determined with a Chroma Meter CR 400. The color parameters and Firmness were determined in five intact fruits per plant, whereas the Soluble Solids content and the pH were measured in the juice obtained by homogenizing a variable number of three to eight fruits per plant, which depended on the fruit size. In the first set, the mean locule number per fruit (LN) was also measured in 5 fruits per plant.

## Statistical Analyses

Classical Quantitative Genetics proposes that phenotypic quantitative traits are determined by many *loci* dispersed among the genome, with small individual effects, additive action and influenced in their expression by the environment. Hence, and in contrary to qualitative traits, genetic structure of populations is not assessed by allele and genotype frequencies for quantitative traits but by phenotypic means, variances and covariances. Also, decomposition of these phenotypic estimators in genotype, environment and their interaction components are generally achieved, but in the multivariate approach proposed in this research, the phenotype will be considered as enough representative of the genotype because assays were accomplished in the same environment and with a unique agronomic management. Following these considerations, the genetic structure was first approached with univariate Statistics by calculating the mean values and standard

deviations for all traits in each group of both populations. The adjustment to normal distribution was assessed by Shapiro-Wilk test, and Pearson's correlation coefficient was calculated between pairs of traits with mean values of individual plants.

Then, a multivariate approach was done by applying PCA to both datasets. Two results from PCA were observed: the first was the proportion of total variance explained by the first PC, which was related to the genetic structure of population in terms of covariance, i.e., of correlation among traits. The second result was related to the coefficient of each trait on a same PC: when these coefficients were different in sign or in value among populations, a different genetic structure relative to mean value or variance of these traits was assumed. Finally, the valorization of plants of each population in the PC1 and PC2 were plotted in two-dimensional graphs. SAS software was used in all analyses.

## RESULTS

Mean values and standard deviations for each evaluated traits are presented by group of genotypes in both populations in Table 1. Phenotypic correlations among pairs of traits, estimated by the respective Pearson's coefficients because traits showed a normal distribution ( $W > 0.95$ , ns in all cases) in both populations are presented in Tables 2 and 3, respectively.

PCA results are presented in Table 4 for RILs population. The first and the second Principal Components (PC1 and PC2, respectively) account the 64% of the total variance. The variables which highly contribute to the PC1 are weight, diameter, height and firmness, and in a moderate way, locule number, shelf life, titratable acidity, chroma index and pH. Instead, the variables which highly contribute to the PC2 are reflectance percentage and soluble solids, and moderately to a/b and SL. Also, PCA results are presented in Table 5 for SBG population. PC1 and PC2 account the 39% of the total variance. The variables which highly contribute to the PC1 are weight, diameter and height. Otherwise, the variables which highly contribute to the PC2 are chroma index, shelf life and reflectance percentage.

Plants of each population valorized in the corresponding PC1 and PC2 are presented in Figures 1 and 2, respectively. Plants are differentially colored according to the genetic group they represent in both populations. The Figure 1 shows that plants are rather differentially positioned according to RIL genotype while in Figure 2, F2 plants are distributed across all the area of the graphic, both backcrosses plants had a narrow distribution and F1, L1 and L18 plant are more grouped.

TABLE 1/TABLA 1

Mean values  $\pm$  standard deviation of the fruit traits weight (W, in g), diameter (D, in cm), height (H, in cm), shape index (SI, ratio H/D), shelf life (SL, in days), reflectance percentage (L, in %), chroma index (ratio a/b, being “a” the absorbance at 540 nm wavelengths and “b” the absorbance at 675 nm wavelengths), locule number (LN), soluble solids content (SS, in °Brix), pH, titratable acidity (TA, in g of citric acid per 100 g of homogenate juice), and firmness (F, in %) in a two different populations (Pop) composed by different genotypes (G): a set of Recombinant Inbred lines (RILs, N: number of plants per RIL) and a set of six basic generations (SBG, L1: parental RIL 1, L18: parental RIL 18, F1; second cycle hybrid L18 x L1, F2: selfed F2 generation from F1, R5: backcross F1 x L18, R6: backcross F1 x L1, N: number of plants per generation). / *Valores medios  $\pm$  desviación estándar de las caracteres del fruto peso (W, en g), diámetro (D, en cm), altura (H, en cm), índice de forma (SI, relación H/D), vida poscosecha (SL, en días), porcentaje de reflectancia (L, en %), índice de croma (relación a/b, siendo “a” la absorbancia en longitudes de onda de 540 nm y “b” la absorbancia en longitudes de onda de 675 nm), número de lóculos (LN), contenido de sólidos solubles (SS, en °Brix), pH, acidez titulable (TA, en g de ácido cítrico por 100 g de jugo homogeneizado) y firmeza (F, en %) en dos poblaciones diferentes (Pop) compuestas por diferentes genotipos (G): un conjunto de líneas endocriadas recombinantes (RILs, N: número de plantas por RIL) y un conjunto de seis generaciones básicas (SBG, L1: RIL 1 parental, L18: RIL 18 parental, F1; híbrido de segundo ciclo L18 x L1, F2: generación F2 autofecundada a partir de F1, R5: retrocruza F1 x L18, R6: retrocruza F1 x L1, N: número de plantas por generación).*

Pop	G	N	Fruit Traits												
			W	D	H	SI	SL	L	a/b	LN	SS	pH	TA	F	
RILs	L1	54	26.3 $\pm$ 12.8	3.6 $\pm$ 0.8	3.1 $\pm$ 0.6	0.9 $\pm$ 0.1	25.2 $\pm$ 8.8	45.9 $\pm$ 1.5	0.87 $\pm$ 0.01	3.8 $\pm$ 0.7	6.9 $\pm$ 0.1	4.7 $\pm$ 0.1	0.3 $\pm$ 0.1	54.1 $\pm$ 1.2	
			10.6 $\pm$ 5.8	2.6 $\pm$ 0.5	2.2 $\pm$ 0.5	0.9 $\pm$ 0.1	22.2 $\pm$ 7.6	44.6 $\pm$ 1.4	0.82 $\pm$ 0.01	4.4 $\pm$ 0.8	8.5 $\pm$ 0.2	4.9 $\pm$ 0.3	0.4 $\pm$ 0.5	36.5 $\pm$ 0.5	
	L17	30	15.1 $\pm$ 5.0	3.1 $\pm$ 0.3	2.8 $\pm$ 0.4	0.9 $\pm$ 0.1	68.7 $\pm$ 15.6	36.3 $\pm$ 3.4	-0.43 $\pm$ 0.02	4.4 $\pm$ 0.6	4.9 $\pm$ 0.3	3.9 $\pm$ 0.3	0.7 $\pm$ 0.3	50.7 $\pm$ 1.9	
			14.0 $\pm$ 6.8	2.8 $\pm$ 0.5	3.1 $\pm$ 0.5	1.1 $\pm$ 0.1	24.7 $\pm$ 8.9	43.9 $\pm$ 1.0	1.03 $\pm$ 0.01	3.7 $\pm$ 0.8	9.4 $\pm$ 0.2	4.8 $\pm$ 0.1	0.6 $\pm$ 0.2	45.6 $\pm$ 1.3	
	L5	66	7.9 $\pm$ 4.1	2.3 $\pm$ 0.4	2.1 $\pm$ 0.3	0.9 $\pm$ 0.1	24.9 $\pm$ 4.7	51.9 $\pm$ 1.2	1.04 $\pm$ 0.02	3.1 $\pm$ 0.7	10.3 $\pm$ 0.4	4.8 $\pm$ 0.4	0.3 $\pm$ 0.3	43.9 $\pm$ 1.0	
			6.0 $\pm$ 3.0	2.1 $\pm$ 0.3	2.1 $\pm$ 0.5	1.0 $\pm$ 0.9	20.2 $\pm$ 4.7	39.7 $\pm$ 0.5	0.84 $\pm$ 0.01	2.4 $\pm$ 0.5	7.6 $\pm$ 0.2	4.6 $\pm$ 0.2	0.8 $\pm$ 0.3	38.7 $\pm$ 0.7	
	L8	54	2.4 $\pm$ 1.2	1.8 $\pm$ 0.2	1.7 $\pm$ 0.2	0.9 $\pm$ 0.1	27.7 $\pm$ 5.0	45.0 $\pm$ 1.1	0.93 $\pm$ 0.01	2.4 $\pm$ 0.5	7.1 $\pm$ 0.2	4.7 $\pm$ 0.2	0.7 $\pm$ 0.2	37.5 $\pm$ 1.7	
			2.4 $\pm$ 1.2	1.8 $\pm$ 0.2	1.6 $\pm$ 0.2	0.9 $\pm$ 0.0	14.3 $\pm$ 5.4	38.0 $\pm$ 3.4	1.02 $\pm$ 0.02	2.2 $\pm$ 0.4	6.0 $\pm$ 0.4	4.2 $\pm$ 0.2	0.6 $\pm$ 0.2	36.4 $\pm$ 0.9	
	SBG	L1	9	17.1 $\pm$ 1.8	3.1 $\pm$ 0.2	3.0 $\pm$ 0.7	0.9 $\pm$ 0.1	13.8 $\pm$ 0.6	40.6 $\pm$ 0.9	1.06 $\pm$ 0.03	-	7.5 $\pm$ 0.8	4.7 $\pm$ 0.7	0.3 $\pm$ 0.1	54.4 $\pm$ 6.0
				14.4 $\pm$ 0.7	3.2 $\pm$ 0.5	3.7 $\pm$ 0.7	1.2 $\pm$ 0.1	14.1 $\pm$ 0.1	42.3 $\pm$ 0.1	1.05 $\pm$ 0.02	-	6.0 $\pm$ 1.2	4.7 $\pm$ 0.7	0.3 $\pm$ 0.1	56.1 $\pm$ 3.1
F1		10	13.2 $\pm$ 2.6	2.5 $\pm$ 0.3	2.4 $\pm$ 0.2	1.0 $\pm$ 0.1	14.6 $\pm$ 0.9	44.0 $\pm$ 4.3	1.03 $\pm$ 0.05	-	7.5 $\pm$ 0.2	4.1 $\pm$ 0.6	0.3 $\pm$ 0.1	48.9 $\pm$ 2.6	
			10.8 $\pm$ 4.6	2.6 $\pm$ 0.4	2.6 $\pm$ 0.4	1.0 $\pm$ 0.1	18.1 $\pm$ 4.7	39.1 $\pm$ 1.8	1.11 $\pm$ 0.13	-	7.3 $\pm$ 1.4	4.7 $\pm$ 0.5	0.3 $\pm$ 0.1	56.0 $\pm$ 7.0	
R5		44	12.4 $\pm$ 3.1	2.6 $\pm$ 0.3	2.9 $\pm$ 0.4	1.1 $\pm$ 0.1	19.7 $\pm$ 7.5	38.4 $\pm$ 1.1	1.12 $\pm$ 0.09	-	7.8 $\pm$ 0.8	4.7 $\pm$ 0.7	0.3 $\pm$ 0.1	55.9 $\pm$ 5.9	
			10.9 $\pm$ 1.4	2.6 $\pm$ 0.2	2.5 $\pm$ 0.1	1.0 $\pm$ 0.1	18.5 $\pm$ 5.3	39.0 $\pm$ 1.3	1.11 $\pm$ 0.11	-	8.2 $\pm$ 0.8	4.6 $\pm$ 0.7	0.5 $\pm$ 1.3	54.5 $\pm$ 5.1	

TABLE 2/TABLA 2

Pearson's coefficients of correlation (under the principal diagonal) and their p-values (above the principal diagonal) between each pair of the fruit traits weight (W), diameter (D), height (H), shape index (SI, ratio H/D), shelf life (SL), reflectance percentage (L), chroma index (ratio a/b), locule number (LN), soluble solids content (SS), pH, titratable acidity (TA), and firmness (F) in the set of Recombinant Inbred Lines. / *Coeficientes de correlación de Pearson (debajo de la diagonal principal) y sus valores p (arriba de la diagonal principal) entre cada par de caracteres del fruto peso (W), diámetro (D), altura (H), índice de forma (SI, relación H/D), vida útil (SL), porcentaje de reflectancia (L), índice de croma (relación a/b), número de lóculos (LN), contenido de sólidos solubles (SS), pH, acidez titulable (TA) y firmeza (F) en el conjunto de Líneas Endocriadas Recombinantes.*

	W	D	H	SI	SL	L	a/b	LN	SS	pH	TA	F
W	-	<0.0001	<0.0001	n.s.	n.s.	n.s.	n.s.	<0.0001	n.s.	0.0006	n.s.	<0.0001
D	0.97	-	<0.0001	n.s.	n.s.	n.s.	n.s.	<0.0001	n.s.	0.0005	0.0315	<0.0001
H	0.89	0.93	-	0.0238	0.0469	n.s.	n.s.	<0.0001	0.0441	0.0206	0.0076	<0.0001
SI	-	-	0.28	-	n.s.	n.s.	n.s.	n.s.	n.s.	0.0391	n.s.	n.s.
SL	-	-	0.25	-	-	0.0007	<0.0001	n.s.	0.0010	0.0069	<0.0001	0.0079
L	-	-	-	-	-0.41	-	0.0009	n.s.	<0.0001	0.0256	n.s.	n.s.
a/b	-	-	-	-	-0.84	0.40	-	0.0055	0.0003	0.0034	<0.0001	0.0018
LN	0.49	0.56	0.50	-	-	-	-0.34	-	0.0281	n.s.	0.0132	0.0002
SS	-	-	-0.23	-	-0.40	0.70	0.43	-0.27	-	0.0011	<0.0001	n.s.
pH	-0.41	-0.41	-0.28	0.25	-0.33	-0.33	0.36	-	-0.39	-	<0.0001	0.0117
TA	-	0.27	0.33	-	0.56	-0.27	-0.60	0.30	-0.52	-0.31	-	0.0248
F	0.71	0.72	0.71	-	0.32	-	0.38	0.44	-	-0.65	0.28	-

n.s.: non statistically significant, -: coefficient correlation non statistically significant from 1.00 if they are in the principal diagonal of from 0.00 if they are down the principal diagonal.  
*n.s.: estadísticamente no significativo, -: correlación suficiente no estadísticamente significativa desde 1,00 si están en la diagonal principal o desde 0,00 si están hacia abajo de la diagonal principal.*

TABLE 3/TABLA 3

Pearson's coefficients of correlation (above the principal diagonal) and their p-values (under the principal diagonal) between each pair of the fruit traits weight (W), diameter (D), height (H), shape index (SI), shelf life (SL), reflectance percentage (L), chroma index (ratio a/b), soluble solids content (SS), pH, titratable acidity (TA), and firmness (F) in the set of six basic generations (parental RILs 1 and 18, and their F1, F2 and backcross generations). / *Coeficientes de correlación de Pearson (debajo de la diagonal principal) y sus valores p (arriba de la diagonal principal) entre cada par de caracteres del fruto peso (W), diámetro (D), altura (H), índice de forma (SI, relación H/D), vida útil (SL), porcentaje de reflectancia (L), índice de croma (relación a/b), contenido de sólidos solubles (SS), pH, acidez titulable (TA) y firmeza (F) en el conjunto de seis generaciones básicas (L1: RIL 1 parental, L18: RIL 18 parental, F1; híbrido de segundo ciclo L18 x L1, F2: generación F2 autofecundada a partir de F1, R5: retrocruza F1 x L18, R6: retrocruza F1 x L1).*

	W	D	H	SI	SL	L	a/b	SS	pH	TA	F
W	-	<0.0001	<0.0001	n.s.	0.0099	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
D	0.80	-	<0.0001	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.0315	n.s.
H	0.72	0.74	-	<0.0001	0.0086	n.s.	n.s.	0.0181	n.s.	0.0076	n.s.
SI	-	-	0.44	-	n.s.	n.s.	n.s.	0.0459	0.0205	n.s.	n.s.
SL	0.18	-	0.18	-	-	0.0163	0.0055	n.s.	n.s.	<0.0001	n.s.
L	-	-	-	-	-0.16	-	0.0014	0.0499	n.s.	0.0256	n.s.
a/b	-	-	-	-	0.19	-0.22	-	<0.0001	n.s.	<0.0001	n.s.
SS	-	-	-0.16	-0.14	-	-0.13	0.28	-	n.s.	n.s.	n.s.
pH	-	-	-	0.16	-	-	-	-	-	n.s.	n.s.
TA	-	-0.02	-0.08	-	-0.04	-0.02	-0.03	-	-	-	n.s.
F	-	-	-	-	-	-	-	-	-	-	-

n.s.: non statistically significant, -: coefficient correlation non statistically significant from 1.00 if they are in the principal diagonal of from 0.00 if they are down the principal diagonal.  
*n.s.: estadísticamente no significativo, -: correlación suficiente no estadísticamente significativa desde 1,00 si están en la diagonal principal o desde 0,00 si están hacia abajo de la diagonal principal.*

TABLE 4/TABLA 4

Principal Components Analysis in the RILs population: coefficients for composing (CC) the two first eigenvectors (PC1 and PC2) and their respective correlations (C) with the original fruit traits weight (W), diameter (D), height (H), shape index (SI, ratio H/D), shelf life (SL), reflectance percentage (L), chroma index (ratio a/b), locule number (LN), soluble solids content (SS), pH, titratable acidity (TA), and firmness (F). / *Análisis de Componentes Principales en la población de RILs: coeficientes para conformar (CC) los dos primeros vectores propios (PC1 y PC2) y sus respectivas correlaciones (C) con los caracteres originales del fruto peso (W), diámetro (D), altura (H), índice de forma (SI, relación H/D), vida útil (SL), porcentaje de reflectancia (L), índice de croma (relación a/b), número de lóculos (LN), contenido de sólidos solubles (SS), pH, acidez titulable (TA) y firmeza (F).*

		Fruit Traits											Statistical parameters			
		W	D	H	SI	SL	L	a/b	LN	SS	pH	TA	F	$\lambda$	PTEV	APTEV
PC1	CC	0.38	0.40	0.39	0.02	0.27	-0.05	-0.28	0.29	-0.17	-0.25	0.27	0.38	4.72	0.40	0.40
	C	0.82	0.87	0.85	0.05	0.58	-0.11	-0.61	0.63	-0.37	-0.54	0.59	0.82			
PC2	CC	0.23	0.20	0.13	-0.15	-0.34	0.49	0.34	0.07	0.46	-0.24	-0.30	0.19	2.69	0.24	0.64
	C	0.37	0.33	0.22	-0.24	-0.55	0.81	0.55	0.11	0.75	-0.40	-0.49	0.30			

$\lambda$ : Eigenvalue (variance), PTEV: Proportion of Total Explained Variance, APTEV: Accumulated PTEV of each PC.  
 $\lambda$ : Autovalor (variancia), PTEV: Proporción de Variancia Total Explicada, APTEV: PTEV acumulada de cada PC.

TABLE 5/TABLA 5

Principal Components Analysis in a set of tomato breeding generations (parental RILs 1 and 18, and their F1, F2 and backcross generations): coefficients for composing (CC) the two first eigenvectors (PC1 and PC2) and their respective correlations (C) with the original fruit traits weight (W), diameter (D), height (H), shape index (SI, ratio H/D), shelf life (SL), reflectance percentage (L), chroma index (ratio a/b), soluble solids content (SS), pH, titratable acidity (TA), and firmness (F). / *Análisis de Componentes Principales en la población de seis generaciones básicas (L1: RIL 1 parental, L18: RIL 18 parental, F1; híbrido de segundo ciclo L18 x L1, F2: generación F2 autofecundada a partir de F1, R5: retrocruza F1 x L18, R6: retrocruza F1 x L1): coeficientes para conformar (CC) los dos primeros vectores propios (PC1 y PC2) y sus respectivas correlaciones (C) con los caracteres originales del fruto peso (W), diámetro (D), altura (H), índice de forma (SI, relación H/D), vida útil (SL), porcentaje de reflectancia (L), índice de croma (relación a/b), contenido de sólidos solubles (SS), pH, acidez titulable (TA) y firmeza (F).*

		Fruit traits											Statistical parameters		
		W	D	H	SI	SL	L	a/b	SS	pH	TA	F	$\lambda$	PTEV	APTEV
PC1	CC	0.55	0.54	0.56	0.16	0.13	0.10	-0.11	-0.15	-0.01	-0.07	0.03	2.62	0.24	0.24
	C	0.89	0.88	0.92	0.25	0.22	0.15	-0.17	-0.25	-0.02	-0.11	0.05			
PC2	CC	-0.04	-0.04	0.12	0.18	0.46	-0.42	0.56	0.32	0.29	-0.07	0.26	1.62	0.15	0.39
	C	-0.05	-0.05	0.15	0.22	0.58	-0.53	0.71	0.40	0.37	-0.19	0.33			

$\lambda$ : Eigenvalue (variance), PTEV: Proportion of Total Explained Variance, APTEV: Accumulated PTEV of each PC.  
 $\lambda$ : Autovalor (variancia), PTEV: Proporción de Variancia Total Explicada, APTEV: PTEV acumulada de cada PC.

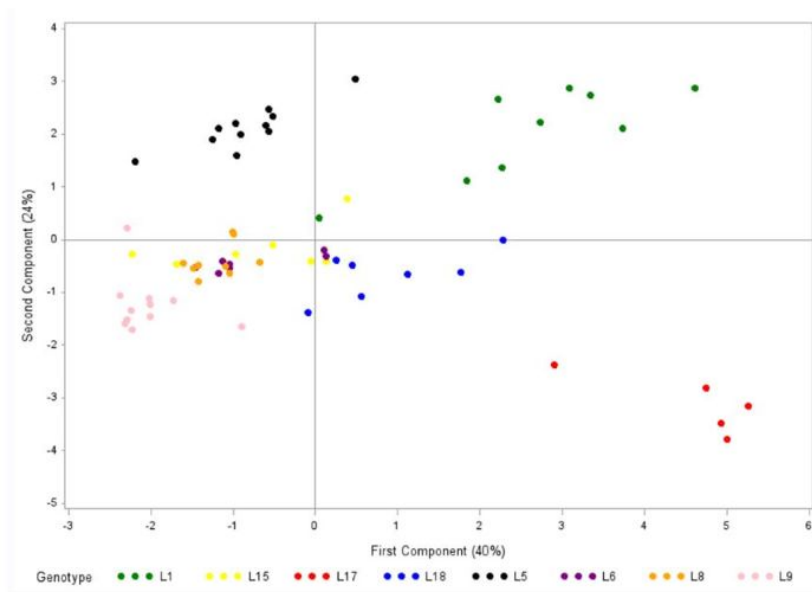


FIGURE 1/FIGURA 1

Tomato plants from a RILs population valorized in the First and the Second Principal Components. Percentage of explained variance by each PC is indicated within parenthesis and different plants belonging to each RIL are indicated by different colors. / *Plantas de tomate de una población de RILs valorizadas en la Primera y Segunda Componentes Principales. El porcentaje de varianza explicada por cada PC se indica entre paréntesis y las diferentes plantas que pertenecen a cada RIL se indican con diferentes colores.*

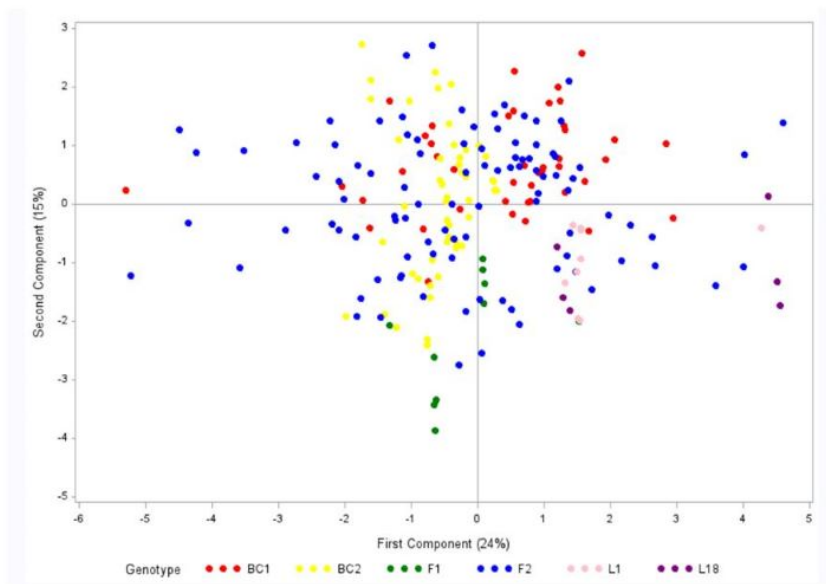


FIGURE 2/FIGURA 2

Tomato plants from a set of tomato breeding generations (L1: parental RIL 1, L18: parental RIL 18, F1; second cycle hybrid L18 x L1, F2(18x1): selfed F2 generation from F1, RC5: backcross F1 x L18, RC6:



backcross F1 x L1, valorized in the First and the Second Principal Components. Percentage of explained variance by each PC is indicated within parenthesis and different plants belonging to each generation are indicated by different colors. / *Plantas de tomate de una población de seis generaciones básicas (L1: RIL 1 parental, L18: RIL 18 parental, F1; híbrido de segundo ciclo L18 x L1, F2: generación F2 autofecundada a partir de F1, R5: retrocruza F1 x L18, R6: retrocruza F1 x L1) valorizadas en la Primera y Segunda Componentes Principales. El porcentaje de varianza explicada por cada PC se indica entre paréntesis y las diferentes plantas que pertenecen a cada generación se indican con diferentes colores.*

## DISCUSSION

A wide range of variation was detected in this research, which indicates a noticeable effect of genetic structure on phenotypic diversity in these populations derived from the same parental lines but being in different breeding stages. Similar results were reported in fennel (*Foeniculum vulgare* Mill.) at the molecular level by Scariolo et al. (2022). Otherwise, though some correlations were conserved among populations (for instance, between weight, diameter and height) others were specific for one populations (chroma index and shelf life in RIL population, as an example). As reported by Iqbal et al. (2022), these differential phenotypic correlations indicate different genetic structure among population. Moreover, not only differences in mean values of traits but also their variance and covariance support divergence in genetic structure of both populations. Genetic components such as heritabilities and genetic correlations, reported in Pratta et al. (2011b) and Cabodevila et al. (2021) agree to this finding. In fact, changes in genetic structure and evaluated by means, heritabilities and trait correlations have been widely reported in plant breeding (Pressoir and Berthaud, 2004; Rodríguez et al., 2006; Thormann et al., 2018). It is interesting to note, however, than within the same gene pool contributed by cv. Caimanta and LA 0722 as original parents of both populations, intermediate Pearson's correlation coefficients were mainly detected in RIL population, i.e. in the final step of a plant breeding program, probably due to strong selection for a desired traits combination (Aditya et al., 2011).

Accordingly to these proposals, the high proportion of total variance explained by PC1 and PC2 in the RILs population suggests a noticeable correlation among traits and agrees to results found by univariate analyses. This covariation in the genetic structure may be due to the selection process followed for obtaining the RILs, which represent a final stage in a plant breeding program. The antagonist-divergent selection reported by Rodríguez et al. (2006) favored the development of phenotypes showing discrepant combinations of traits to generate a set of plant material having high levels of diversity to satisfy the different requirements of consumers. The high contribution of traits to PC1 and PC2 also reflects the effect of selection tending to generate diversity in commercial tomato attributes that are differentially preferred by the fresh market. As examples, some consumers prefer small tomatoes and others prefer big tomatoes. Therefore, traits related to size attribute (W, D and H) were relatively high coefficients and correlations with both PC1 and PC2. However, luminous fruits are more commonly preferred than obscure ones, and the trait related to this attribute (L) had a small coefficient and correlation with PC1. Though they were higher in respect to PC2, this component explained a lesser amount of total variance than PC1. Jointly, this results indicates a reduce diversity for L in the RILs population, which would be due to the great selection pressure to obtain the desired luminous tomatoes. Hence, a strong effect on plant breeding in the genetic structure of the RILs population was evidenced.

On the other hand, in the SBG population none artificial selection process was yet applied, hence its genetic structure is mainly defined by segregation and recombination the F2 and both BC generations of alleles contributed by L18 and L1 at loci that were heterozygous in the F1. Accordingly, traits that were more discrepant among L18 and L1 (W, D and H) had high coefficients and correlations just with PC1, while other traits that were more similar among parents (Sl; L, a/b) were more associated to PC2. In agreement, the correlation among traits had a small contribution to genetic structure of this population, as shown by

the low proportion of total variance explained by PC1 and PC2. All these observations are in concordance with those from the univariate Statistics.

When comparing the PCA results from both populations, it is clear that the genetic structure is very different. It is interesting to note the great difference in the number of PC that are necessary to explain a same percentage of total explained variance in both populations. For instance, to account for approximately 3/4 of total variance, just three PC must be retained in the RILs population, while 6 PC must be retained in the SBG population. Provided that in the RILs population a smaller number of PC explained the same percentage of total variance than in the other, a higher amount of covariance among traits was characterizing its genetic structure. Hence, great differences for covariation in genetic structure between both populations were found. Moreover, considering differences in coefficients and correlations of traits and PC, also differences for mean and variances in genetic structure are verified, as pointed out in the previous paragraphs, when independently analyzing each population.

Finally, valorization of plants in these graphs was distinguished according to their genotype/generation in each set. As homogenization of the distribution of plants in each graph is considered as an assessment of the genetic structure of the corresponding population, the more noticeable grouping of plants from RILs population in relation to plants from the SBG one pointed out a higher differentiation in the corresponding genetic structure of the first population. This observation agrees to the fact already mentioned that RILs were differentially selected in response to human requirements and they are composed by uniformly homozygous genotypes, hence they are well differentiated. Meanwhile, F2 and backcross generations from the other population are segregating and recombining alleles contributed by both parents and are composed by both homozygous and heterozygous genotypes, thus they have a great range of variation. In fact segregation in F2 is greater than in backcrosses, and F1 and parental L18 and L1 are –as the RILs population– uniform genotypes the former heterozygote and the two latter ones homozygote so their phenotypic variation is expected to be reduced. Similar results for an F2 and its parental genotypes' distribution in a graph from PCA were reported by Pratta et al. (2010) in the analysis of protein profile expression in an inter-varietal cross between *S. lycopersicum* var *lycopersicum* and *S. lycopersicum* var. *cerasiforme*. In contrast, Pratta et al. (2000) reported similar position of regenerated plants and the corresponding explant's donors in different tomato species, thus indicating the absence of changes in phenotypic and genotypic population structure and that somaclonal variation had not occurred during *in vitro* culture. Also, Del Medico et al. (2019, 2020) proposed the use of Multiple Factorial Analysis, a 3-ways technique which is a generalization of PCA, for characterizing consecutive segregating generations from a SCH and estimating multivariate heritability for various fruit traits, finding similar genetic structure among F2 and F3 populations for some quantitative traits such as weight, diameter and height but different for others (pH, shelf life, and chroma index, as examples).

According to James et al. (2021), results from PCA indicate a different structure in the amount of total variation and covariation of each population and agree to univariate results. In fact, though both populations are composed by the same gene pool formerly contributed by cv. Caimanta and LA0722, alleles are recombined and rearranged in different ways hence they represent different genetic structures according to genotypic frequencies, linkage disequilibrium and inbreeding coefficients. These discrepant genetic structures are due to the effects of Plant Breeding since each population represents a different step in a breeding program, the final one being the result of artificial selection and the initial one, the result of generating genetic variability by crosses among selected parental lines.

## CONCLUSION

Principal Component Analysis is an adequate statistical technique to evaluate the genetic composition of two different tomato breeding populations derived for the same original gene pool, satisfactorily describing their variance and covariance structure for agronomic important fruit traits.

## FUNDING STATEMENT

This work was financed by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) grant P-UE: 22920160100043CO (IICAR), and by the National University of Rosario, Projects of Technological Linking and Productive Development “Inclusive Linking” 2015. GRPratta is a research career member of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

## ACKNOWLEDGMENTS

Authors are grateful to Drs. Sabina Mahuad and Victoria Cabodevila for generating phenotypic data during their respective Doctoral Thesis, and to Secretaría de Ciencia, Tecnología e Innovación de la Provincia de Santa Fe, Argentina, for financial supporting.

## REFERENCES

- Acquaah, G. (2015). *Conventional Plant Breeding Principles and Techniques*. In: J., Al-Khayri, S. Jain and D. Johnson, (eds.). *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools* (pp 115-158). Springer, Cham. [https://doi.org/10.1007/978-3-319-22521-0\\_5](https://doi.org/10.1007/978-3-319-22521-0_5).
- Aditya, J.P., Bhartiya, P., Bhartiya, A. (2011). Genetic variability, heritability and character association for yield and component characters in soybean (*G. max* (L.) Merrill). *Journal of Central European Agriculture* 12, 27-34. <https://doi.org/10.5513/JCEA01/12.1.877>.
- Cabodevila, V.G., Cambiaso, V., Rodríguez, G.R., Picardi, L.A., Pratta, G.R., Capel, C., Lozano, R., Capel, J. (2021). A segregating F<sub>2</sub> population from a tomato second cycle hybrid allows the identification of novel QTL for fruit quality traits. *Euphytica* 217, 453-461. <https://doi.org/10.1007/s10681-020-02731-6>.
- Del Medico, A.P., Cabodevila, V.G., Vitelleschi, M.S., Pratta, G.R. (2019). Multivariate estimate of heritability for quality traits in tomato by the multiple factor analysis. *Pesquisa Agropecuária Brasileira* 54. <https://doi.org/10.1590/S1678-3921.pab2019.v54.00064>.
- Del Medico, A.P., Cabodevila, V.G., Vitelleschi, M.S., Pratta, G.R. (2020). Characterization of tomato (*Solanum lycopersicum* L.) generations according to three-way data analysis. *Bragantia* 79, 8-18. <https://doi.org/10.1590/1678-4499.20190047>.
- Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., Guarino, L. (2017). Past and future use of wild relatives in Crop Breeding. *Crop Science* 57, 1070-1082. <https://doi.org/10.2135/cropsci2016.10.0885>.
- Food and Agriculture Organization – Statistics [FAOSTAT]. (2017). *Production - Crops - Area harvested / Production quantity - Tomatoes - 2014*. Available at [www.fao.org/faostat/en](http://www.fao.org/faostat/en). [Accessed July 21, 2021].
- Gerszberg, A., Hnatuszko-Konka, K., Kowalczyk, T., Kononowicz, A.K. (2015). Tomato (*Solanum lycopersicum* L.) in the service of biotechnology. *Plant Cell Tissue and Organ Culture* 120, 881–902. <https://doi.org/10.1007/s11240-014-0664-4>.
- Iqbal, Z., Hu, D., Nazeer, W., Ge, H., Nazir, T., Fiaz, S., Gul, A., Shahid Iqbal, M., El-Sabrou, A.M., Maryum, Z., Pan, Z., Du, X. (2022). Phenotypic correlation analysis in F<sub>2</sub> segregating populations of *Gossypium hirsutum* and *Gossypium arboreum* for boll-related traits. *Agronomy* 12, 330. <https://doi.org/10.3390/agronomy12020330>.
- James, G., Witten, D., Hastie, T., Tibshirani, R. (2021). *An Introduction to Statistical Learning With Applications in R*. Springer, New York, NY. USA.
- Kearsey, M.J., Pooni, H.S. (1996). *The Genetical Analysis of Quantitative Traits*. Chapman and Hall, London, UK.
- Lu, D., Xu, S. (2013). Principal component analysis reveals the 1000 Genomes Project does not sufficiently cover the human genetic diversity in Asia. *Frontiers in Genetics, Section Applied Genetic Epidemiology* 4. <https://doi.org/10.3389/fgene.2013.00127>.

- Mahuad, S.L., Pratta, G.R., Rodriguez, G.R., Zorzoli, R. and Picardi, L.A. (2013). Preservation of *Solanum pimpinellifolium* genomic fragments in recombinant genotypes increased tomato fruit quality. *Journal of Genetics* 92, 195-203.
- McVean, G. (2010). A genealogical interpretation of Principal Components Analysis. *PLoS Genetics* 5. e1000686. doi: 10.1371/journal.pgen.1000686.
- Nachimuthu, V.V., Robin, S., Sudhakar, D., Raveendran, M., Rajeswari, S., Manonman, S. (2014). Evaluation of rice genetic diversity and variability in a population panel by Principal Component Analysis. *Indian Journal of Science and Technology* 7, 1555–1562. 10.17485/ijst/2014/v7i10.14.
- Pereira da Costa, J.H., Rodríguez, G.R., Pratta, G.R., Picardi, L.A., Zorzoli, R. (2014). Pericarp polypeptides and SRAP markers associated with fruit quality traits in an interespecific tomato backcross. *Genetics and Molecular Research* 13, 2539-2547. <https://doi.org/10.4238/2014.January.24.10>.
- Pratta, G., Zorzoli, R., Picardi, L.A. (2000). Multivariate analysis as a tool for measuring the stability of morphometric traits in *Lycopersicon* plants from *in vitro*. *Genetics and Molecular Biology* 23, 470-483. <https://doi.org/10.1590/S1415-47572000000200039>.
- Pratta, N.N., Quaglino, M., Rodríguez, G.R., Zorzoli, R., Pratta, G.R. (2010). A multivariate approach to the Proteomics of tomato fruit ripening. *Genes, Genomes and Genomics* 4, 48-51.
- Pratta, G.R., Rodriguez, G.R., Zorzoli, R., Valle, E.M., Picardi, L.A. (2011a). Molecular markers detect stable genomic regions underlying tomato fruit shelf life and weight. *Crop Breeding and Applied Biotechnology* 11, 157-164. <https://doi.org/10.1590/S1984-70332011000200008>.
- Pratta, G.R., Rodriguez, G.R., Zorzoli, R., Valle, E.M., Picardi, L.A. (2011b). Phenotypic and molecular characterization of selected tomato recombinant inbred lines derived from a cross *Solanum lycopersicum* x *S. pimpinellifolium*. *Journal of Genetics* 90, 229-237. <https://doi.org/10.1007/s12041-011-0063-0>.
- Pressoir, G., Berthaud, J. (2004). Population structure and strong divergent selection shape phenotypic diversification in maize landraces. *Heredity* 92, 95–101. <https://doi.org/10.1038/sj.hdy.6800388>.
- Rodriguez, G.R., Pratta, G.R., Zorzoli, R., Picardi, L.A. (2006). Recombinant lines obtained from an interspecific cross among *Lycopersicon* species selected by fruit weight and fruit shelf life. *Journal of the American Society for Horticultural Science* 131, 651-656. <https://doi.org/10.21273/JASHS.131.5>.
- Scariolo, F., Palumbo, F., Barcaccia, G. (2022). Molecular characterization and genetic structure evaluation of breeding populations of fennel (*Foeniculum vulgare* Mill.). *Agronomy* 12, 542, <https://doi.org/10.3390/agronomy12030542>.
- Thormann, I., Reeves, P., Thumm, S., Reilley, A., Engels, J., Biradar, C., Lohwasser, U., Borner, A., Pillen, K., Richards, C.M. (2018). Changes in barley (*Hordeum vulgare* L. subsp. vulgare) genetic diversity and structure in Jordan over a period of 31 years. *Plant Genetic Resources: Characterization and Utilization* 16, 112-126. doi:10.1017/S1479262117000028