



OPEN

Multivariate analysis methods improve the selection of strawberry genotypes with low cold requirement

Eneide Barth¹, Juliano Tadeu Vilela de Resende², Keny Henrique Mariguela³, Marcos Deon Vilela de Resende⁴, André Luiz Biscaia Ribeiro da Silva⁵✉ & Sushan Ru⁵

Methods of multivariate analysis is a powerful approach to assist the initial stages of crops genetic improvement, particularly, because it allows many traits to be evaluated simultaneously. In this study, heat-tolerant genotypes have been selected by analyzing phenotypic diversity, direct and indirect relationships among traits were identified, and four selection indices compared. Diversity was estimated using K-means clustering with the number of clusters determined by the Elbow method, and the relationship among traits was quantified by path analysis. Parametric and non-parametric indices were applied to selected genotypes using the magnitude of genotypic variance, heritability, genotypic coefficient of variance, and assigned economic weight as selection criteria. The variability among materials led to the formation of two non-overlapping clusters containing 40 and 154 genotypes. Strong to moderate correlations were found between traits with direct effect of the number of commercial fruit on the mass of commercial fruit. The Smith and Hazel index showed the greatest total gains for all criteria; however, concerning the biochemical traits, the Mulamba and Mock index showed the highest magnitudes of predicted gains. Overall, the K-means clustering, correlation analysis, and path analysis complement the use of selection indices, allowing for selection of genotypes with better balance among the assessed traits.

World widely grown, strawberries (*Fragaria* × *ananassa* Duchesne) are popular fruits used for fresh or processing markets^{1–3}. In 2020, strawberry production was 9 million tons, with China and United States contributing to more than 50% of this world production⁴. The most common strawberry cultivars in the world are often poorly adapted to tropical conditions as they are mainly developed by breeding programs in temperate regions, such as the United States, Spain, and Italy⁵. In South America, most strawberry seedlings are grown in nurseries of Argentina and Chile, significantly raising the cost of production in countries like Brazil, where strawberries are grown in 4,300 hectares and more than 75% of seedlings are imported from these nurseries⁴.

Strawberry cultivars are usually selected based on yield, resistance to pests and disease, adaptability to semi-hydroponic systems, and fruit quality such as firmness, sweetness, acidity, and aroma^{6–8}. Assembling all the favorable characteristics into a single cultivar is complex, especially because strawberries have an octaploid genome highly interactive with the environment⁹. After making biparental crosses, strawberry cultivar development takes several rounds of selection to identify the best offspring as potential cultivars. The process of selecting among hundreds or thousands of seedlings is arduous, time-consuming, and costly. Traditionally, the selection is conducted for a single trait either through direct selection, which is directly based on the trait of interest, or indirect selection, which uses a correlated characteristic to select the trait of interest⁸. However, single trait selection may result in cultivars that do not adequately meet the demands of producers and consumers¹⁰. Combining direct and indirect effects can better evince the importance of traits on an independent variable, such as production of commercial fruits^{11,12}. Using the path analysis to seek a better understanding of cause and effect, along with multivariate techniques, such as selection indices, to simultaneously evaluate multiple traits may be

¹Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina (Epagri), Rua XV de Novembro, 525, Pomerode, SC 89107-000, Brazil. ²Departamento de Agronomia, Universidade Estadual de Londrina/UUEL, Rodovia Celso Garcia, Km 380, Londrina, PR 86051-900, Brazil. ³Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina (Epagri), Rodovia Antônio Heil, 6800, Itajai, SC 88318-112, Brazil. ⁴Departamento de Estatística, Embrapa Café/Universidade Federal de Viçosa, Campus Universitário, Viçosa, MG 36570-900, Brazil. ⁵Department of Horticulture, Auburn University, 101 Funchess Hall, Auburn, AL 36849, USA. ✉email: azb0207@auburn.edu

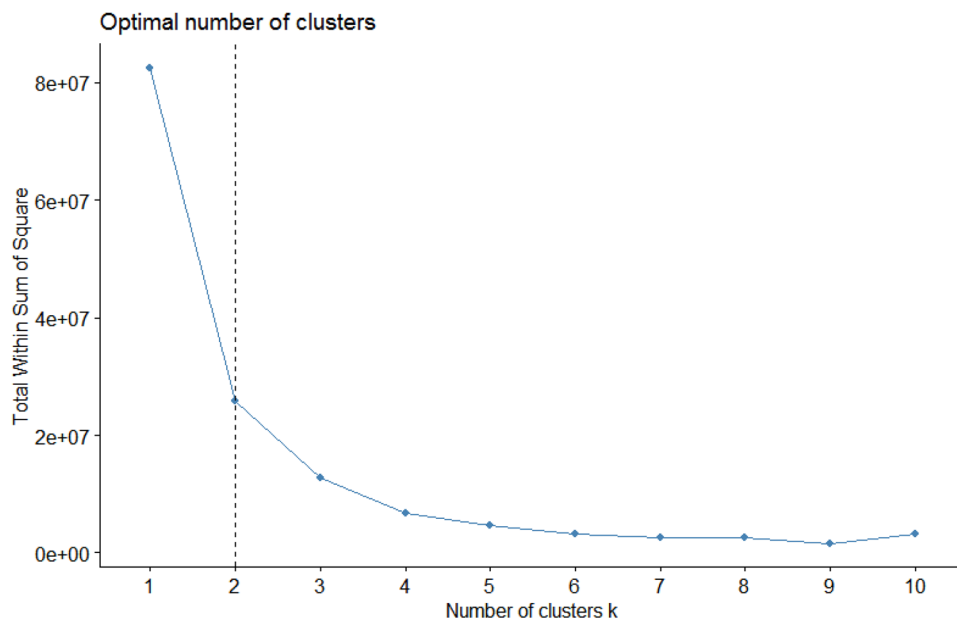


Figure 1. Identification of the elbow point for the evaluated dataset.

the best strategy for obtaining genotypes that balance agronomic, biometric, and biochemical characteristics, especially in the early stages of a breeding program^{10,13,14}.

Thus, the variability of a population can be explored through cluster analysis, among other statistical tools. The K-means algorithm is a non-hierarchical data exploration method that maximizes the variation component between the formed K-groups while minimizing the variation within each group¹⁵. Still, determining the number of groups, which is defined a priori, is complex and can generate imprecise analyses. To minimize this possibility, the Elbow Method determines the ideal number of clusters¹⁶.

This study focused on the classic parametric index presented by Smith¹⁷ and Hazel¹⁸ and the base index established by Williams¹⁹. The classic Smith-Hazel index employs matrices of genetic and phenotypic variances and covariances estimated in the analysis of variance. This index consists of obtaining the maximum correlation between the genotypic aggregate (H) and the index (I). The H is a linear combination of the analyzed traits, pondered by a coefficient established by the economic weights previously assigned to each trait. The I consist of a linear combination of the “x” values of each trait, pondered by a coefficient to be estimated. The base index¹⁹ is established by the linear combination of the average phenotypic values of traits pondered directly by their respective economic weights. The index is estimated by: $I = a^1y^1 + a^2y^2 + \dots + a^ny^n = y'a$, where y_j is the mean of the jth trait and p_j the economic weight. As for the non-parametric indices, the sum of ranks from Mulamba and Mock²⁰ and the genotype-ideotype²¹ indices were used. The index of Mulamba and Mock ranks genotypes in relation to each trait individually, assigning absolute values, according to the classification direction determined by the breeder, from highest to lowest, or vice versa, depending on how this direction favors the genetic improvement. The values assigned in the classification of each trait are added, resulting in the selection index. The genotype-ideotype index²¹ estimates the distance of the evaluated genotypes in relation to an ideotype previously defined by the breeder. The first step is to identify the favorable value for improvement based on the average, maximum, and minimum values informed by the statistical computer program. This favorable value, called optimal value (OV_j), must be within an upper (UL) and lower (LL) limit for each trait ($LL_j \leq X_{ij} \leq UL_j$). The OV_j is corrected by a constant concerning the depreciation of the genotype average, C_j ($C_j = UL_j - L_j$), resulting in the Y_{ij} value. This process guarantees that any value of X_{ij} that is outside the optimal range is not selected. Subsequently, the Y_{ij} values obtained with the transformation are standardized and pondered by weights previously assigned by the breeder for each trait. The OV_j for each trait is standardized and pondered.

Overall, studies have shown that applying the aforementioned selection indices, which associate information from various characters, increases the success rate in crop improvement programs, including alfalfa²², passion fruit²³, soybean^{24,25}, sweet potato²⁶, among others. Nevertheless, the use of indices in the strawberry genetic improvement process has been just recently shown in the literature¹³ and still needs to be better understood. The objective of this study was to evaluate and select intraspecific strawberry genotypes, to assess their phenotypic diversity, to compare different selection indices, and determine the direct and indirect relationship among yield and biochemical traits, using multivariate analysis methods.

Results

The optimal K value for the population was determined to be 2 according to the Elbow Method (Fig. 1). Two clusters were generated without overlapping in the K-means clustering (Fig. 2). The control ‘Camino Real’ and 40 seedling genotypes were in group 1. The control ‘Camarosa’ and 154 seedling genotypes were in group 2. Group 1 was better than group 2, according to the 5% confidence interval, for yield-related traits mass of commercial

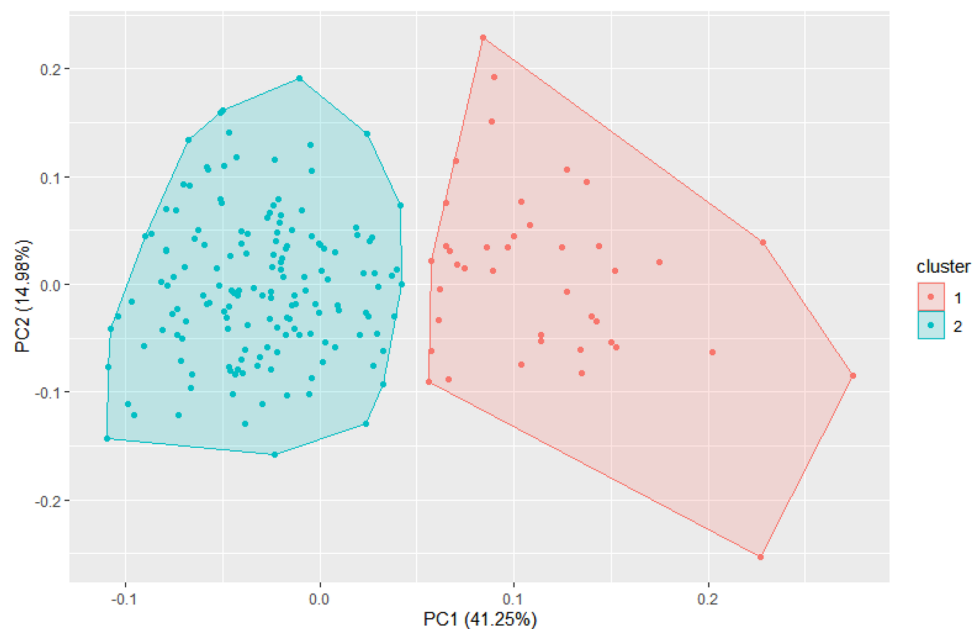


Figure 2. K-means cluster analysis for yield and biochemical traits evaluated in 10 populations of *Fragaria × ananassa* Dusch.

K-means clusters	1	2
Number of genotypes	41	155
MCF	[1474.90; 1831.82]	[324.81; 409.81]
NCF	[95.67; 117.69]	[24.55; 30.79]
AMCF	[14.24; 17.64]	[12.09; 12.95]
Ratio	[11.94; 14.00]	[9.09; 9.79]
TP	[1.73; 2.45]	[1.91; 2.19]
AA	[75.30; 81.86]	[67.80; 71.06]
ANT	[43.44; 50.20]	[35.71; 39.12]

Table 1. Confidence interval of the mean values of the variables evaluated in the two K-means clusters with the respective number of genotypes ($p < 0.05$). MCF mass of commercial fruits (g plant^{-1}), AMCF average mass of commercial fruits (g fruit^{-1}), NCF number of commercial fruits (fruits plant^{-1}), Ratio, soluble solids (Brix°)/titratable acidity ($\text{g citric acid } 100 \text{ g}^{-1} \text{ pulp}$); TP total pectin ($\text{g total pectin } 100 \text{ g}^{-1} \text{ pulp}$), AA ascorbic acid ($\text{mg ascorbic acid } 100 \text{ g}^{-1} \text{ pulp}$), and ANT anthocyanins ($\text{mg cyanidin-3-glucoside } 100 \text{ g}^{-1} \text{ pulp}$).

fruits (MCF), number of commercial fruits (NCF), and average mass commercial fruits (AMCF), as well as for the biochemical characteristics Ratio, ascorbic acid (AA), and anthocyanins (ANT). Contrarily, no significant difference in total pectin (TP) was measured between groups 1 and 2 (Table 1).

Twelve significant and positive correlations were found by the t-test ($p < 0.05$) among the 21 pairs of traits evaluated (Fig. 3). The most robust correlations were obtained for yield-related characteristics. High phenotypic correlations ($r > 0.66$) were measured between MCF and NCF ($r = 0.96$). Medium correlations ($0.33 < r < 0.67$) were measured between MCF and AMCF; MCF and Ratio; and NCF and Ratio, presenting r values of 0.55, 0.53, and 0.53, respectively. Low correlations ($r < 0.34$) were measured between NCF and AMCF ($r = 0.34$), between yield- and biochemical- related traits (MCF and AA ($r = 0.32$); MCF and ANT ($r = 0.29$); NCF and AA ($r = 0.34$); NCF and ANT ($r = 0.32$); and AMCF and Ratio ($r = 0.27$)), and between the biochemical traits Ratio and AA ($r = 0.18$) and Ratio and ANT ($r = 0.32$).

By unfolding the correlations through path analysis for a single causal diagram, direct and indirect effects for the independent trait MCF and the other characteristics were identified (Table 2). Biochemical-related traits had no direct or indirect effect on MCF. The NCF had a direct effect (0.88) on MCF and an indirect effect on AMCF (0.086). Contrarily, AMCF had an indirect effect on NCF (0.30) greater than the direct effect on itself (0.25). Ratio had a small direct effect on MCF, while presented a greater indirect effect on NCF (0.46) and AMCF (0.068).

The total percentage gains obtained by the four indices under the four criteria (GV, h^2 , GCV, and EW) ranged from 366 to 386% (Table 3) for simultaneous selection of yield and biochemical traits. The Smith-Hazel index showed the highest total gains with 386%, for all criteria, followed by the Genotype-Ideotype distance index with 384% for GCV, the Mulamba and Mock index with 384% and 383% for EW and GCV, respectively, and by the

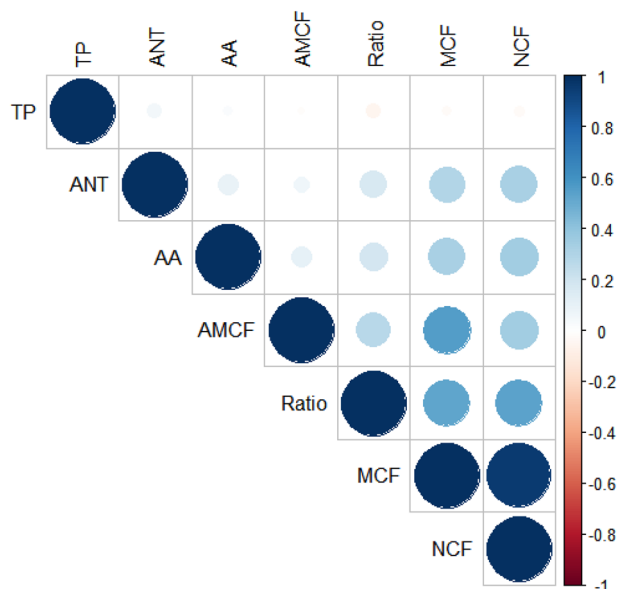


Figure 3. Pearson's phenotypic correlations among yield and biochemical traits evaluated in 10 populations of *Fragaria* × *ananassa* Dusch. *NCF* number of commercial fruits (fruits plant⁻¹), *AMCF* average mass of commercial fruits (g fruit⁻¹), *Ratio*—soluble solids (Brix°)/titratable acidity (g citric acid 100 g⁻¹ pulp), *TP* total pectin (g total pectin 100 g⁻¹ pulp), *AA* ascorbic acid (mg ascorbic acid 100 g⁻¹ pulp), and *ANT* anthocyanins (mg cyanidin-3-glucoside 100 g⁻¹ pulp).

	NCF	AMCF	Ratio	TP	AA	ANT
NCF	0.87785	0.08590	-0.00086	-0.00027	-0.00011	-0.00156
AMCF	0.29847	0.25266	-0.00044	-0.00009	-0.00031	-0.00028
Ratio	0.46526	0.06822	-0.00016	-0.00045	-0.00059	-0.00080
TP	-0.02633	-0.00253	0.00008	-0.00091	-0.00006	-0.00023
AA	0.29847	0.02526	-0.00031	0.00018	-0.00314	-0.00047
ANT	0.28969	0.01516	-0.0028	0.00045	-0.00031	-0.00471
R ² = 0.98						
Residual effect = 0,14,752						

Table 2. Direct (on the main diagonal) and indirect (on the upper and lower diagonals) effects of the independent variables on the mass of commercial fruits in 10 populations of *Fragaria* × *ananassa* Dusch. *NCF* number of commercial fruits (fruits plant⁻¹), *AMCF* average mass of commercial fruits (g fruit⁻¹), *Ratio*, soluble solids (Brix°)/titratable acidity (g citric acid 100 g⁻¹ pulp); *TP* total pectin (g total pectin 100 g⁻¹ pulp), *AA* ascorbic acid (mg ascorbic acid 100 g⁻¹ pulp), and *ANT* anthocyanins (mg cyanidin-3-glucoside 100 g⁻¹ pulp).

Williams index with 380% for h² (Table 3). Regarding yield and biochemical traits, the Mulamba and Mock index provided greater gains for the biochemical-related traits, in relation to h² (86%) and GCV (81%), followed by the Genotype-Ideotype index, under the criteria h² (79%) and GCV (75%). The Smith and Hazel index, despite having shown the greatest gains for yield traits, showed the lowest gains of biochemical traits, in which the four indices under the four criteria selected 53 genotypes (Table 4). From this total, 38 are located in group 1 and 15 in group 2, according to the K-means clustering. A total of 28 genotypes were selected by all indices for all criteria, in which only one belonged to group 1 of the K-means cluster analysis. Eleven genotypes were selected in some indices for all criteria and only in some criteria in other indices. Another eleven genotypes were selected in some indices for some criteria, and three genotypes were selected by some indices for all criteria.

The crosses 'Camarosa' × 'Aromas' and 'Camarosa' × 'Sweet Charlie' stood out, with 14 and nine selected hybrids, respectively. The other crosses showed the following number of selected hybrids: 'Dover' × 'Aromas'—6, 'Osó Grande' × 'Tudla'—5, 'Festival' × 'Aromas'—5, 'Aromas' × 'Sweet Charlie'—3, 'Tudla' × 'Aromas'—3, 'Dover' × 'Sweet Charlie'—3, 'Tudla' × 'Sweet Charlie'—3, and 'Festival' × 'Sweet Charlie'—2.

According to Dindex (Fig. 4), genotypes can be grouped into four groups by a significant knee in the plot of index values against the number of clusters. The circular hierarchical dendrogram (Fig. 5) obtained from the

Traits	Smith and Hazel				Mulamba and Mock			
	GV %	h ² %	GCV %	EW %	GV %	h ² %	GCV %	EW %
MCF	162.5	162.5	162.5	162.5	163.5	142.5	151.9	162.7
AMCF	13.4	13.4	13.4	13.4	14.5	11.7	10.9	14.6
NCF	154.3	154.25	154.3	154.3	142.9	128.0	139.5	141.7
Total for yield traits	330.1	330.1	330.1	330.1	320.9	282.2	302.3	319.1
Ratio	26.9	26.9	26.9	26.9	25.9	28.6	28.8	27.8
TP	1.1	1.1	1.1	1.1	4.8	19.5	18.7	5.3
AA	11.5	11.5	11.5	11.5	10.9	12.0	10.2	11.5
ANT	16.8	16.8	16.8	16.8	17.2	26.3	23.4	20.0
Total for biochemical traits	56.4	56.4	56.4	56.4	58.9	86.3	81.1	64.5
Total	386.5	386.5	386.5	386.5	379.8	368.6	383.4	383.6

Traits	Williams				Genotype-ideotype			
	GV	h ²	GCV	EW	GV	h ²	GCV	EW
MCF	163.5	163.4	163.5	163.5	155.4	150.1	154.4	154.6
AMCF	14.5	13.8	14.3	14.3	13.9	11.2	9.7	14.0
NCF	142.9	144.9	143.8	143.8	137.4	137.1	145.0	136.2
Total of yield traits	320.9	322.0	321.5	321.5	306.6	298.4	309.0	304.8
Ratio	25.9	25.6	26.7	26.7	27.4	28.5	28.3	29.3
TP	4.8	2.9	3.0	3.0	5.6	13.7	11.6	6.0
AA	10.9	12.1	12.0	12.0	10.4	12.0	12.4	11.1
ANT	17.2	17.5	16.1	16.1	16.7	25.2	22.6	19.4
Total of biochemical traits	58.9	58.1	57.7	57.7	60.1	79.4	74.9	65.8
Total	379.8	380.1	379.3	379.3	366.7	377.8	383.9	370.5

Table 3. Estimates of percentage gains obtained by simultaneous selection with application of four indices based on four criteria of economic weights for seven traits evaluated in 10 populations of *Fragaria × ananassa* Dusch. *MCF* mass of commercial fruits (g plant⁻¹), *AMCF* average mass of commercial fruits (g fruit⁻¹), *NCF* number of commercial fruits (fruits plant⁻¹), *Ratio*, soluble solids (Brix°)/titratable acidity (g citric acid 100 g⁻¹ pulp); *TP* total pectin (g total pectin 100 g⁻¹ pulp), *AA* ascorbic acid (mg ascorbic acid 100 g⁻¹ pulp), and *ANT* anthocyanins (mg cyanidin-3-glucoside 100 g⁻¹ pulp). *GV* genotypic variance, *h²* heritability, *GCV* genetic coefficient variation and *EW* economic weight assigned by the breeder.

analysis of the 53 genotypes selected by the indices and the controls ('Camarosa' and 'Camino Real'), generated groups with 32, 20, 2 and 1 genotypes, whose cophenetic correlation value was 0.827 ($p < 0.05$).

Discussion

Brazilian strawberry production depends almost entirely on cultivars developed in foreign breeding programs that, due to aspects related to genotype × environment interactions, may present lower yield, lower biochemical quality, greater susceptibility to pests and diseases, increasing production costs⁵. Nonetheless, these imported cultivars have the potential to be explored in intraspecific crosses aiming to express the existing variability in the species^{13,27}.

Strawberry is an octoploid species that has gone through various levels of ploidization throughout evolutionary history²⁸. Strawberry also harbors millions of DNA variants of the subgenomes of the species that gave rise to actual strawberry fruit²⁹. In general, strawberry presents great variability in hybrids obtained from crosses, which favors the selection of new cultivars³⁰. Significant variability with identification of superior hybrids has been found in phenotypic analyses for yield and physicochemical traits in populations obtained from crosses between commercial strawberry cultivars in Brazil^{9,12,13,31–33}. In addition, genetic studies with hybrids and commercial cultivars based on molecular markers have shown that the germplasm of the Brazilian strawberry breeding program has genetic variability and divergence; therefore, it has a high potential for launching new cultivars^{9,34}.

For the population analyzed in this study, the Elbow method established two clusters, which presented no overlap in the K-means clustering, showing variability in the analyzed population and complete dissimilarity between the two groups formed. The highest phenotypic correlation for the independent variable mass of commercial fruits (MCF) was obtained for the number of commercial fruits (NCF) (0.96), which had a high direct effect (0.88) in the path analysis. The average mass of commercial fruits (AMCF), which also had a medium and positive correlation (0.55) with the MCF, demonstrated in the path analysis that its indirect effect (0.29) on NCF is superior to the direct effect (0.25). Diel et al.³⁵ found a direct effect of the total number of fruits (0.81), and an indirect effect of the mass of commercial fruits, via the total number of fruits (0.71), while the average fruit mass showed a direct relationship of 0.22. Authors results corroborate with our study and these positive findings suggest that direct selection via number of commercial fruits has a greater effect on yield and indirectly benefits the average mass of commercial fruits.

Selection groups	Hybrids	K- means	MCF	NCF	AMCF	Ratio	TP	AA	ANT	
Hybrids selected by all indices for all criteria	RVOT11	1	1097.80	80.0	13.7	12.8	4.8	89.9	58.1	
	RVOT12	1	1188.0	80.0	14.9	13.5	4.7	92.1	48.5	
	RVOT21	1	1857.1	120.0	15.5	14.4	0.9	82.1	56.1	
	RVOT22	1	2524.9	200.0	12.6	9.5	1.4	88.4	52.2	
	RVFA02	1	1968.3	120.0	16.4	12.1	1.5	71.7	58.1	
	RVFA04	1	1984.5	110.0	18.0	10.4	3.6	85.1	31.8	
	RVFA14	1	1670.1	120.0	13.9	16.5	0.4	80.0	49.9	
	RVFA16	1	1518.5	90.0	16.9	13.3	0.9	88.9	50.0	
	RVSA14	1	1503.0	110.0	13.7	11.2	1.8	78.7	49.1	
	RVDA01	1	1732.4	100.0	17.3	12.6	3.5	94.9	48.1	
	RVDA12	1	1489.0	90.0	16.5	18.4	3.1	81.0	41.9	
	RVCS01	1	1643.3	120.0	13.6	12.9	2.5	67.8	51.3	
	RVCS04	1	2503.4	170.0	14.7	18.6	1.0	74.0	51.3	
	RVCS07	1	2005.8	120.0	16.7	11.7	1.6	73.9	39.1	
	RVCS09	1	1500.5	130.0	11.5	18.6	1.1	84.2	48.5	
	RVCS10	1	2722.9	180.0	15.1	24.8	0.4	92.1	54.7	
	RVCS11	1	1486.0	110.0	13.5	17.2	0.5	88.5	50.4	
	RVTA12	1	2046.1	110.0	18.6	9.4	1.5	82.1	30.0	
	RVTA16	1	2393.6	160.0	15.0	8.8	2.9	67.2	59.7	
	RVTA20	1	1336.1	100.0	13.4	15.6	3.1	85.7	44.3	
	RVCA02	1	1042.1	70.0	14.9	15.1	1.7	65.7	61.8	
	RVCA05	2	1162.9	60.0	19.4	7.8	1.3	61.3	47.9	
	RVCA06	1	2247.4	150.0	16.3	12.5	1.9	56.0	63.0	
	RVCA08	1	1133.8	80.0	14.2	15.3	1.5	68.9	27.4	
	RVCA10	1	1042.9	70.0	14.9	13.2	2.7	80.0	37.5	
	RVCA11	1	924.4	70.0	13.2	17.4	2.9	87.1	29.9	
	RVCA14	1	1037.9	60.0	17.3	11.8	3.2	78.4	61.1	
	RVCA16	1	2109.9	120.0	17.6	13.1	2.6	80.2	49.9	
	Hybrids selected by some indices for all criteria and by other indices for only a couple of criteria	RVFS01	1	2060.9	130.0	15.9	15.1	1.6	92.5	32.5
		RVFS24	1	1302.5	100.0	13.0	14.6	3.0	89.8	28.4
RVDS24		1	1350.0	110.0	12.3	7.9	1.6	80.0	54.2	
RVSA08		1	1388.0	110.0	12.6	14.3	3.5	68.4	25.5	
RVDA03		1	1532.6	100.0	15.3	11.8	2.6	91.3	38.5	
RVDA04		1	3370.7	200.0	16.9	11.1	2.6	96.6	41.3	
RVDA11		1	2158.7	150.0	14.4	14.6	0.4	59.9	65.1	
RVDA18		1	2987.3	60.0	49.8	13.2	1.9	61.2	31.1	
RVCS03		2	753.7	60.0	12.6	12.5	1.7	75.0	44.6	
RVCS12		2	723.5	50.0	14.5	15.7	1.1	65.5	25.5	
Hybrids selected by some indices for all criteria	RVTS08	1	1094.6	80.0	13.7	12.6	3.0	82.4	45.2	
	RVDS14	1	1372.4	90.0	15.2	11.1	1.5	62.5	47.6	
	RVCS14	2	668.1	60.0	11.1	14.8	1.7	70.0	38.6	
Hybrids selected by some indices for some parameters	RVCA01	2	900.2	60.0	15.0	10.7	1.8	74.1	49.2	
	RVOT28	2	894.6	70.0	12.8	10.5	4.3	73.7	41.3	
	RVDS26	1	1126.1	80.0	14.1	13.8	3.6	82.5	61.3	
	RVFA01	2	741.1	40.0	18.5	6.4	2.9	75.3	39.5	
	RVSA12	2	929.1	70.0	13.3	10.9	2.9	55.9	48.1	
	RVTS01	2	1112.3	80.0	13.9	8.5	3.0	75.6	32.0	
	RVTS15	2	1070.9	90.0	11.9	7.8	0.9	77.9	56.3	
	RVCA03	2	744.5	60.0	12.4	12.1	1.7	71.7	50.4	
	RVCA07	2	451.5	40.0	11.3	10.3	2.9	58.2	51.4	
	RVCA12	2	720.3	50.0	14.4	14.8	2.1	65.8	51.8	
Controls	RVCA13	2	170.9	10.0	17.1	11.3	3.1	70.6	50.0	
	RVCA15	2	597.6	30.0	19.9	12.5	3.5	55.9	31.6	
	Camarosa	2	776.2	45.0	16.9	10.8	1.9	74.5	39.8	
	Camino Real	1	1173.0	64.0	18.3	11.4	2.2	77.4	42.3	

Table 4. Hybrids selected by Smith-Hazel, Mulamba and Mock, Williams, and Genotype-Ideotype indices and K-means clustering for yield and biochemical traits in 10 populations of *Fragaria × ananassa* Dusch. *MCF* mass of commercial fruits (g plant⁻¹), *AMCF* average mass of commercial fruits (g fruit⁻¹), *NCF* number of commercial fruits (fruits plant⁻¹), *Ratio*, soluble solids (Brix°)/titratable acidity (g citric acid 100 g⁻¹ pulp); *TP* total pectin (g total pectin 100 g⁻¹ pulp), *AA* ascorbic acid (mg ascorbic acid 100 g⁻¹ pulp), and *ANT* anthocyanins (mg cyanidin-3-glucoside 100 g⁻¹ pulp).

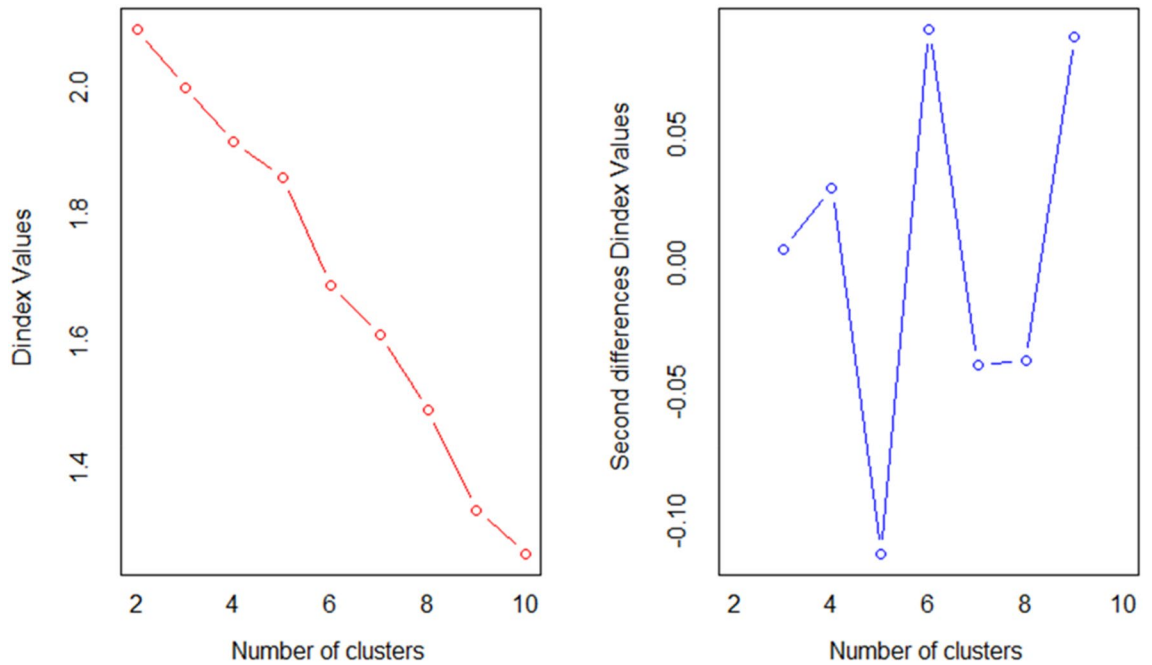


Figure 4. Dindex graphic for determining the best number of clusters in 53 selected genotypes and two controls of *Fragaria × ananassa* Dusch.

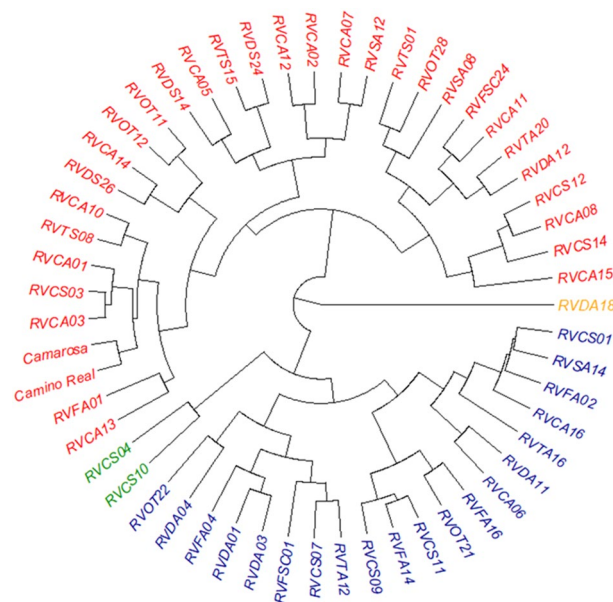


Figure 5. Circular dendrogram based on yield and biochemical traits of the 53 selected genotypes and two controls of *Fragaria × ananassa* Dusch.

The balance between soluble solids and titratable acidity (Ratio) represents the equilibrium between sweetness and acidity. This balance combined with aroma and other biochemical traits makes up flavor, which has great importance in sensory perception and consumer preference^{5,6}. In the present study, Ratio showed a moderate and positive phenotypic correlation with the mass of commercial fruits (0.52) and number of commercial fruits (0.53); however, when unfolding this correlation, a negative direct effect was observed, while the indirect effect was positive via NCF. In agreement with the present study, Diel³⁵ found a negative direct effect (−0.10) and a positive indirect effect (0.15) of Ratio via the total number of fruits on the total fruit mass. Direct effects of the number of strawberry fruit on production per plant were also reported by Ara et al.³⁶ and Garg¹¹, while Sighn et al.³⁷ stated that the greatest direct positive effects came from flower number and fruit length. These results evince that the selection of strawberry genotypes for mass of commercial fruits can be directly performed via the number of commercial fruits and that genotypes with numerous fruits, but of medium size, tend to have a better Ratio than genotypes with large fruits.

Selecting genotypes that balance yield and biochemical traits simultaneously is a complex task¹⁰. The use of selection indices, both parametric and non-parametric, has been useful to identify more balanced hybrids of diverse crops, such as sweet potato²⁶, alfalfa³⁸, soybean^{25,39,40}, potato⁴¹, maize⁴², acai⁴³, passion fruit^{23,44}, and, more recently, strawberry^{13,27}.

In the present study, the Mulamba and Mock and Genotype-Ideotype indices were more sensitive to the use of different criteria, showing greater differences between gains. Cruz et al.⁴⁵ recommend the use of statistics obtained from the analysis of experimental data as economic weights (EW) since it relates to the genotypic variance, they are dimensionless and maintain a certain proportionality among the evaluated traits. In the present study, the greatest gains for yield traits were obtained by the Smith and Hazel index (330.14%); however, it showed no difference between the statistical criteria or assigned weights. Contrarily, the greatest gains for the biochemical-related traits were obtained by the Mulamba and Mock and Genotype-Ideotype indices, under the criteria of h^2 and GCV with 86.34% and 81.06%; 79.41% and 74.87%, respectively. Vieira et al.²⁷, evaluating strawberry genotypes, also reported the greatest increments for yield traits with the Smith and Hazel index and for biochemical characteristics applying the Mulamba and Mock index. It occurs because parametric tests use the distribution parameters to calculate the statistics, while non-parametric tests use ranks assigned to ordered data and are uninfluenced by the probability distribution of the data evaluated⁴⁶. Thus, the non-parametric Mulamba and Mock index is less sensitive, mathematically, to traits that present wide variance, such as number of fruits.

From the 194 genotypes analyzed, 28 were selected for all indices, under all criteria, in which 27 belong to group 1 of the K-means clustering. The use of different indexes and criteria tend to present very similar results for the initial positions of the selected genotypes. Bernardo et al.⁴⁷ analyzed studies in several agronomic crops and concluded that, if the population is large enough, any selection index applied judiciously is useful for the simultaneous improvement of multiple traits, regardless of the method used. Nevertheless, the indices start to select different hybrids for the different criteria with the progress of positions.

The crosses with the highest number of selected hybrids were ‘Camarosa’ × ‘Aromas’ and ‘Camarosa’ × ‘Sweet Charlie’. Similarly, Galvão et al.²⁸ identified the best hybrids for yield traits in the cross between ‘Camarosa’ × ‘Aromas’. Camarosa has been reported as a highly productive cultivar, with large, firm, and tasty fruits⁴⁸, being one of the most planted short-day cultivars in the world⁴⁹. The presence of large number of favorable alleles in ‘Camarosa’ and ‘Aromas’³³ and their high productive potential^{50,51} make them promising parents for strawberry breeding programs⁵. Camargo et al.³² also found and selected the best hybrids coming from the crosses ‘Camarosa’ × ‘Aromas’ and ‘Camarosa’ × ‘Sweet Charlie’, concerning biochemical traits.

The dendrogram generated from the 53 selected genotypes led to the formation of five groups, demonstrating that this population still has variability that can be further investigated.

Conclusion

K-means clustering, correlation analysis, and path analysis complement the use of selection indices, leading to the selection of hybrids with better balance between yield- and biochemical-related traits in strawberry. This combined approach is more promising than the direct selection based on only one or a few traits. Furthermore, the multivariate analysis methods were efficient in selecting strawberry genotypes for multi-characters.

The number of commercial fruits was more relevant to the mass of commercial fruits than the average mass of commercial fruits. Therefore, NCF is a trait of greater importance for the selection of strawberry genotypes aiming at yield. The Smith and Hazel index showed the greatest gain for yield traits. Possibly because it is mathematically more influenced by characteristics with greater variability such as yield. The Mulamba and Mock and Genotype-Ideotype indices, both non-parametric, showed the highest estimated gains for biochemical traits under the criteria of h^2 and GCV. The crosses with the highest number of selected hybrids were ‘Camarosa’ × ‘Aromas’ and ‘Camarosa’ × ‘Sweet Charlie’. The selected population of 53 hybrids still has variability with potential to be exploited.

Material and methods

The material and methods of our study was performed in accordance with the relevant guidelines and regulations. Plant material and replications followed the regulations of the Ministry of Agriculture, Cattle and Supplying of Brazil.

Plant material. Ten populations were obtained from biparental crosses among strawberry cultivars traditionally grown in South America (Table 5). All parents are public commercial cultivars available at the Brazilian Agricultural Research Corporation (EMBRAPA) from the Ministry of Agriculture, Cattle and Supplying of Brazil, and they were grown with Multiplanta Tecnologia Vegetal (Andradas, MG, Brazil). Parents are short-day

Population	Parents	
	♀	♂
1—DA	Dover	Aromas
2—CA	Camarosa	Aromas
3—DS	Dover	Sweet Charlie
4—OT	Oso Grande	Tudla
5—FS	Florida Festival	Sweet Charlie
6—AS	Sweet Charlie	Aromas
7—TA	Tudla	Aromas
8—TS	Tudla	Sweet Charlie
9—CS	Camarosa	Sweet Charlie
10—FA	Florida Festival	Aromas

Table 5. Intraspecific crosses used to obtain 10 segregating strawberry (*Fragaria × ananassa* Dusch.) populations.

cultivars based on photoperiod responses, except by Aromas which is a day-neutral cultivar⁵². Hybridization was performed following Chandler et al.⁵³. The choice between cultivars to carry out the crosses to obtain segregating populations was based on the genetic dissimilarity study carried out by Morales et al.³⁴. After crossing, achenes present in the fruits were removed and germinated in vitro, as described by Galvão et al.³¹. At 60 days after germination, the seedlings were transplanted to 72-cell polypropylene trays containing biostabilized substrate. A total of 2000 plants (about 200 seedlings per population) were transplanted to low-tunnel-covered beds in an augmented block design. Seedlings were transplanted in April 2015 and the genotypes evaluated until November of the same year. Based on agronomic (total and commercial fruit production, average fruit mass), phytosanitary [no symptoms of anthracnose (*Colletotrichum acutatum* and *C. fragariae*), *Botrytis cinerea* and *Mycosphaerella fragariae*], SS (Content of soluble solids above 8° Brix)], firmness traits and distribution of production on cycle, 194 genotypes were selected, grown in the greenhouse, cloned, and transplanted to the experimental field. Strawberry runners were transplanted in trays with substrates to obtain seedlings in sufficient numbers to be used as replications.

Experimental area. The experimental area is located in the city of Guarapuava, Paraná, Brazil (25° 23' 36" S and 51° 27' 19" W). The area has a humid mesothermal subtropical climate, type Cfb, with moderate winter and summer with average temperatures around 22 °C according to the Köppen's classification⁵⁴. The soil is classified as a typical dystroferic Bruno Latosol⁵⁵.

Seedlings were obtained from the stolons emitted by the parent plants, kept in a greenhouse. Rooting took place in 46-cell polypropylene trays filled with commercial substrate. At 50 days after planting, seedlings were transplanted in the experimental area, with evaluations occurring between May and November 2016.

Strawberry transplanting was performed in a low-tunnel system 0.8 m high with beds 1 m wide × and 0.25 m high surface covered with a black polyethylene film 30- μ m thick. To cover the tunnels, 120- μ m thick transparent polyethylene film was used. The plant spacing was 0.30 × 0.30 m between plants and 0.40 m between rows.

Beds were fertilized with 1,650 kg ha⁻¹ of simple superphosphate, 250 kg ha⁻¹ of potassium chloride, and 295 kg ha⁻¹ of urea, based on the soil chemical analysis in accordance with the recommendations for the strawberry crop⁵². Nutritional replacement was performed via fertigation twice a week. Irrigation water was provided using a micro-drip system and followed the crop water demand. Additionally, for phytosanitary preventive control, applications of Thiamethoxam and Azoxystrobin + Difenoconazole were carried out. Strawberry fruits were harvested at maturity stage when 75% of fruit were red.

The experiment was conducted using the randomized block design with three replications and ten plants per plot. There was a total of 194 F₁ experimental hybrids and two commercial controls ('Camarosa' and 'Camino Real').

Yield and biochemical traits evaluated. Traits that showed significant differences in the analysis of variance were used in the further analyses, namely: mass of commercial fruits (MCF) (g plant⁻¹), number of commercial fruits (NCF) (fruit plant⁻¹), average mass of commercial fruits (AMCF) (g fruit⁻¹), ratio between soluble solids (SS) (Brix°) and titratable acidity (TA) (g citric acid 100 g⁻¹ pulp (Ratio)), total pectin (g total pectin 100 g⁻¹ pulp), ascorbic acid content (AA) (mg ascorbic acid 100 g⁻¹ pulp), and anthocyanin content (ANT) (mg cyanidin-3-glucoside 100 g⁻¹ pulp). The biochemical traits were assessed in samples of commercial ripe strawberries (above 10 g), stored at -2 °C right after harvest. Strawberries were thawed, crushed, and homogenized. Using the homogenized pulp, soluble solids content was measured with an Optech bench refractometer. Titratable acidity was determined by the titration method, with aliquots of 10 g of strawberry pulp plus 100 mL of distilled water 0.1 mol L⁻¹ NaOH standard solution up to pH 8.2, which corresponds to the turning point of phenolphthalein⁵⁶. The total pectin was determined by the method described by McCready and McComb⁵⁷, and calorimetrically determined while using the carbazole reaction, according to the methodology that was described by Bitter and Muir⁵⁸. Ascorbic acid was obtained by the standard titration method of the Association of Official Analytical Chemists (AOAC), modified by Benassi and Antunes⁵⁹. Whereas anthocyanin was deter-

Variables	Parameters			
	Genotypic variance	Herdability	GCV	Assigned weight
MCF	462,808.65	87.59	104.89	500
AMCF	10.19	53.84	24.24	100
NCF	1799.28	93.46	93.61	100
Ratio	9.72	82.91	30.49	100
TP	1.53	87.87	59.97	50
AA	128.37	82.92	15.86	25
ANT	191.17	93.73	34.99	25

Table 6. Economic weights criteria used in the application of selection indices for trait analysis in 10 populations of *Fragaria* × *ananassa* Dutch. MCF mass of commercial fruits (g plant⁻¹), AMCF average mass of commercial fruits (g fruit⁻¹), NCF number of commercial fruits (fruits plant⁻¹), Ratio, soluble solids (Brix°)/titratable acidity (g citric acid 100 g⁻¹ pulp); TP total pectin (g total pectin 100 g⁻¹ pulp), AA ascorbic acid (mg ascorbic acid 100 g⁻¹ pulp), and ANT anthocyanins (mg cyanidin-3-glucoside 100 g⁻¹ pulp).

mined by the differential pH method described by Giusti and Wroslad⁶⁰, with adaptations for strawberry. All biochemical analyses were performed in triplicates.

Statistical analyses. The variability of the 194 genotypes/hybrids and the two controls was analyzed using the R software (<http://cran-rc3sl.ufpr.br>). First, the number of clusters was determined by the Elbow Method using the factextra v.1.0.7 package⁶¹. A graph was used to indicate the ideal cluster number to represent a data set, where the value of “K” to be used is the point of the curve that looks like an elbow (inflection). Subsequently, the K-means non-hierarchical cluster analysis was performed based on the Euclidean distance, with the stats R Core Team⁶², dplyr v.0.8.5⁶³, ggplot2⁶⁴, and ggfortify⁶⁵ packages. The relationship between traits was performed using a Pearson correlation map thoughtout the corrplot v.0.84 package⁶⁶, while a path analysis was performed with agricolae v.1.3-2⁶⁷. The genotypic correlations were classified as high ($r > 0.66$), medium ($0.33 < r < 0.67$) and low ($r < 0.33$)⁶⁸.

Variance component analysis was performed with the Genes software^{69,70} to estimate genotypic variance (GV), heritability (h^2), and genotypic coefficient variation (GCV). Economic weights (EW) were assigned (Table 6). Subsequently, two parametric indices, the classic index from Smith¹⁷ and Hazel¹⁸ and the base index¹⁹, and two non-parametric indices, the rank-sum-based index²⁰, and the genotype-ideotype distance index²¹ were used to select.

The genotypic aggregate (H) in the classic Smith-Haze index it is obtained by the expression $H = a_1g_1 + a_2g_2 + \dots + a_n g_n$, where “a” is the $n \times 1$ dimension vector of the economic weights and “g” is the $p \times n$ dimension matrix of unknown genetic values of the “n” traits for the “p” families or progenies evaluated. The index (I) consists of a linear combination of the “x” values measured of each trait, pondered by a coefficient. It is obtained by the expression: $I = b_1x_1 + b_2x_2 + \dots + b_n x_n$, where the coefficient “b” is an $(n \times 1)$ vector estimated from the expression $b = P^{-1} Ga$, where “P⁻¹” is the inverse of the phenotypic covariance matrix; “G” is the genetic covariance matrix and “a” is the $(n \times 1)$ vector of the economic weights assigned to the traits^{17,18}.

This index of Mulamba and Mock is obtained by the expression: $I = r_1 + r_2 + \dots + r_n$, where “I” is the index value for a given individual, r_j is the rank of an individual in relation to the j-th variable, and “n” is the number of traits considered in the index. This procedure allows the ranking order of traits to have different weights, as specified by the breeder. Thus, we have that $I = p_1r_1 + p_2r_2 + \dots + p_n r_n$, with p_j being the economic weight attributed by the breeder to the j-th trait²⁰.

To obtain the genotype-ideotype index, the values that will express the distance between genotypes and the ideotype are calculated by the expression: $I_{DGI} = \sqrt{1/n \sum (y_{ij} - v_o)^2}$. The best genotypes were identified, and selection gains were estimated based on I_{DGI} . Based on the values of the ideotype (Y_{ij}), the principal components analysis was performed to obtain the eigenvalues and eigenvectors associated with the correlation matrix between the analyzed variables. The distances of the genotypes in relation to the ideotype were estimated. This process allows the selection of genotypes closer to the optimal pattern defined by the breeder (ideotype)²¹.

Selection gains [SG (%)] in the base index¹⁹ were estimated with the following expression: $SG (\%) = 100 h^2 (X_s - X_o)/X_o$, where X_s is the average genotypic value of selected hybrids, X_o is the average genotypic value of all hybrids, and h^2 is the heritability of the trait of interest. Heritability was obtained by the ratio between genotypic and phenotypic variance, as $h^2 = \hat{\sigma}_g^2 / \hat{\sigma}_p^2$, where $\hat{\sigma}_g^2$ is the genotypic variance and $\hat{\sigma}_p^2$ is the phenotypic variance¹⁹.

Lastly, the optimal number of clusters was identified by Dindex index with R package NbClust⁷¹ to generate a circular hierarchical dendrogram created with all selected hybrids and controls, in all parameters and indices using the R packages vegan v.2.5-6⁷², for the standardization of data, ape v.5.0⁷³, and cluster v.2.1.0⁷⁴.

Data availability

The datasets used and/or analyzed during the current study is available from the corresponding author on reasonable request.

Received: 15 December 2021; Accepted: 28 June 2022

Published online: 06 July 2022

References

- Giampieri, F. *et al.* Strawberry consumption improves aging-associated impairments, mitochondrial biogenesis and functionality through the AMP-activated protein kinase signaling cascade. *Food Chem.* **34**(1), 464–471. <https://doi.org/10.1016/j.foodchem.2017.05.017> (2017).
- Juric, S. *et al.* Stimulation of plant secondary metabolites synthesis in soilless cultivated strawberries (*Fragaria × ananassa* Duchesne) using zinc-alginate microparticles. *Turk. J. Agric. For.* **45**, 324–334. <https://doi.org/10.3906/tar-2011-68> (2021).
- Urün, I. *et al.* Comparison of polyphenol, sugar, organic acid, volatile compounds, and antioxidant capacity of commercially grown strawberry cultivars in Turkey. *Plants* **10**(8), 1654. <https://doi.org/10.3390/plants10081654> (2021).
- FAOSTAT—Food and Agriculture Organization Corporate Statistical Database. FAO Online Database. Retrieved 21 Mar 2021 from <http://www.fao.org/faostat/es/#data/SC>
- Zeist, A. R. & Resende, J. T. V. Strawberry breeding in Brazil: Current momentum and perspectives. *Hort. Bras* **37**, 7–16 (2019).
- Resende, J. T. V., Camargo, L. K. P., Argandona, E. J. S., Marchese, A. & Camargo, C. K. Sensory analysis and chemical characterization of strawberry fruits. *Hort. Bras.* **26**, 371–374 (2008).
- Shaw, D. V. & Larson, K. D. Performance of early-generation and modern strawberry cultivars from the University of California breeding programme in growing systems simulating traditional and modern horticulture. *J. Hortic. Sci. Biotechnol.* **83**(5), 648–652 (2008).
- Whitaker, V. M., Hasing, T., Chandler, C. K., Plotto, A. & Baldwin, E. Historical trends in strawberry fruit quality revealed by a trial of university of Florida cultivars and advanced selections. *HortScience* **46**(4), 553–557 (2011).
- Corrêa, J. V. W., Weber, G. G., Zeist, A. R., Resende, J. T. V. & Silva, P. R. ISSR analysis reveals high genetic variation in strawberry three-way hybrids developed for tropical regions. *Plant Mol. Biol. Rep.* **39**(3), 566–576 (2021).
- Cruz, C. D., Carneiro, P. C. S. & Regazzi, A. J. Modelos biométricos aplicados ao melhoramento genético. (ed. Cruz, C. D.) 668p. (UFV: Viçosa, 2014)
- Garg, S., Sharma, G., Lata, S. & Yadav, A. Correlation and path analysis among different vegetative, floral and fruit characters in strawberry (*Fragaria × ananassa* Duch.). *Ecoscan* **6**, 379–384 (2014).
- Barth, E. *et al.* Yield and quality of strawberry hybrids under subtropical conditions. *Genet. Mol. Res.* **18**, 01–10 (2019).
- Barth, E. *et al.* Selection of experimental hybrids of strawberry using multivariate analysis. *Agronomy* **10**(4), 598 (2020).
- Kang, M. S. Efficient SAS programs for computing path coefficients and index weights for selection indices. *J. Crop. Improv.* **29**(1), 6–22 (2015).
- Turchetto-Zolet, A. C., Turchetto, C., Zanella, C. M. & Passaia, G. Marcadores moleculares na era genômica: metodologias e aplicações 181p, <https://www.lume.ufrgs.br/bitstream/handle/10183/206114/001056131.pdf?sequence=1> (2017)
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S. & Satoto, B. D. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conf. Ser. Mater. Sci. Eng.* **336**(1), 012017 (2018).
- Smith, H. F. A discriminant function for plant selection. *Ann Eugen* **7**, 240–250 (1936).
- Hazel, L. N. The genetic basis for constructing selection indexes. *Genetics* **28**, 476–490 (1943).
- Williams, J. S. The evaluation of a selection index. *Biometrics* **18**, 375–393 (1962).
- Mulamba, N. N. & Mock, J. J. Improvement of yield potential of the Eto Blanco maize (*Zea mays* L.) population by breeding for plant traits. *Egypt. J. Genet. Cytol.* **7**, 40–51 (1987).
- Cruz, C. D. *Programa GENES: Aplicativo Computacional em Genética e Estatística Versão Windows* 382 (UFV, 2006).
- Vasconcelos, E. S. D. *et al.* Estimativas de ganho genético por diferentes critérios de seleção em genótipos de alfafa. *Rev. Ceres* **57**(2), 205–210 (2010).
- Rosado, L. D. S., Santos, C. E. M. D., Bruckner, C. H., Nunes, E. S. & Cruz, C. D. Simultaneous selection in progenies of yellow passion fruit using selection indices. *Rev. Ceres* **59**(1), 95–101 (2012).
- Vianna, V. F. *et al.* The multivariate approach and influence of characters in selecting superior soybean genotypes. *Afr. J. Agric. Res.* **8**(30), 4162–4169 (2013).
- Leite, W. S. *et al.* Estimativas de parâmetros genéticos, correlações e índices de seleção para seis caracteres agrônômicos em linhagens F8 de soja. *Com. Sci.* **7**(3), 302–310 (2016).
- Camargo, L. K. P., Resende, J. T. V., Mógior, A. F., Camargo, C. K. & Kurchaidt, S. M. Uso de índice de seleção na identificação de genótipos de batata doce com diferentes aptidões. *Hortic. Bras.* **34**, 514–519 (2016).
- Vieira, S. D. *et al.* Selection of experimental strawberry (*Fragaria × ananassa*) hybrids based on selection indices. *Genet. Mol. Res.* **16**, 1–11 (2017).
- Edger, P. P. *et al.* Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* **51**(3), 541–547 (2019).
- Hardigan, M. A. *et al.* Genome synteny has been conserved among the octoploid progenitors of cultivated strawberry over millions of years of evolution. *Front. Plant Sci.* **10**, 1789 (2020).
- Nunes, C. F. *et al.* The genetic diversity of strawberry (*Fragaria ananassa* Duch.) hybrids based on ISSR markers. *Acta Sci. Agron.* **35**(4), 443–452 (2013).
- Galvão, A. G. *et al.* Breeding new improved clones for strawberry production in Brazil. *Acta Sci. Agron.* **39**(2), 149–155 (2017).
- Camargo, L. K. P. *et al.* Post-harvest characterization of strawberry hybrids obtained from the crossing between commercial cultivars. *Rev. Bras. Frut* **40**, 1–4 (2018).
- Vieira, S. D. *et al.* Heritability and combining ability studies in strawberry population. *J. Agric. Sci.* **11**, 457–469 (2019).
- Morales, R. G. F. *et al.* Genetic similarity among strawberry cultivars assessed by RAPD and ISSR markers. *Sci. Agric.* **68**(6), 665–670 (2011).
- Diel, M. I. *et al.* Linear relationships between yield, quality and phenological traits of strawberry cultivars. *J. Agric. Stud.* **8**(3), 737–755 (2020).
- Ara, T., Haydar, A., Hayatmahamud, K. K. & Hossain, M. M. Analysis of the different parameters for fruit yield and yield contributing characters in strawberry. *Int. J. Sustain. Crop Prod.* **4**(5), 15–18 (2009).
- Singh, G., Kachwaya, D. S., Kumar, R., Vikas, G. & Singh, L. Genetic variability and association analysis in strawberry (*Fragaria × ananassa* Duch). *Electron. J. Plant Breed* **9**(1), 169–182 (2018).
- Santos, I. G., Cruz, C. D., Nascimento, M., Rosado, R. D. S. & Ferreira, R. P. Direct, indirect, and simultaneous selection as strategies for alfalfa breeding on forage yield and nutritive value. *Pesq. Agropec. Trop.* **48**, 178–189 (2018).
- Bizari, E. H., Val, B. H. P., Pereira, E. M., Mauro, A. O. D. & Unêda-Trivisoli, S. H. Selection indices for agronomic traits in segregating populations of soybean. *Ciênc. Agron.* **48**, 110–117 (2017).
- Teixeira, F. G. *et al.* Genetic parameters and selection of soybean lines based on selection indexes. *Genet. Molec. Res.* **16**(3) (2017).
- Terres, L. R., Lenz, E., Castro, C. M. & Pereira, A. S. Estimativas de ganhos genéticos por diferentes índices de seleção em três populações híbridas de batata. *Hort. Bras.* **33**(3), 305–310 (2015).
- Rangel, R. M., Amaral Júnior, A. T. D., Gonçalves, L. S. A., Freitas Júnior, S. D. P. & Candido, L. S. Análise biométrica de ganhos por seleção em população de milho pipoca de quinto ciclo de seleção recorrente. *Ciênc. Agron.* **42**(2), 473–481 (2011).

43. Teixeira, D. H. L., Oliveira, M. D. S. P. D., Gonçalves, F. M. A. & Nunes, J. A. R. Índices de seleção no aprimoramento simultâneo dos componentes da produção de frutos em açaizeiro. *Pesqui. Agropecu. Bras.* **47**(2), 237–243 (2012).
44. Freitas, J. P. X., Oliveira, E. J., Jesus, O. N., Cruz Neto, A. J. & Santos, L. R. Formação de população base para seleção recorrente em maracujazeiro-amarelo com uso de índices de seleção. *Pesqui. Agropecu. Bras.* **47**, 393–401 (2012).
45. Cruz, C. D., Carneiro, P. C. S. & Regazzi, A. J. *Modelos Biométricos Aplicados ao Melhoramento Genético* 668 (UFV, 2012).
46. Reis, G. M. & Ribeiro Júnior, J. I. Comparação de testes paramétricos e não paramétricos aplicados em delineamentos experimentais. In: *Simpósio Acadêmico de Engenharia de Produção*, 3., Viçosa. Anais. pp. 1–13 (2007).
47. Bernardo, R. *Breeding for Quantitative Traits in Plants* (Stemma Press, 2002).
48. Voth, V., Shaw, D. V. & Bringham, R. S. *Strawberry Plant Called 'Camarosa'*. U.S. Patent 8708. U.S. Patent and Trademark Office (1994).
49. Samtani, J. B. *et al.* The status and future of the strawberry industry in the United States. *HortTechnology* **29**(1), 11–24 (2019).
50. Resende, J. T. V. *et al.* Produtividade e teor de sólidos solúveis de frutos de cultivares de morangueiro em ambiente protegido. *Hort. Bras.* **28**(2), 185–189 (2010).
51. Munaretto, L. M., Botelho, R. V., Resende, J. T. V., Schwarz, K. & Sato, A. J. Productivity and quality of organic strawberries pre-harvest treated with silicon. *Hort. Bras.* **36**(1), 40–46 (2018).
52. Antunes, L. E. C., Reisser Junior, C. & Schwengber, J. E. Morangueiro. Pelotas, RS. (Embrapa Clima Temperado, 2016). <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1092843/morangueiro>
53. Chandler, C. K., Folta, K., Dale, A., Whitaker, V. M. & Herrington, M. Strawberry. In *Fruit Breeding Handbook of Plant Breeding* Vol. 8 (eds Badenes, M. & Byrne, D.) (Springer, 2012).
54. Wrege, M. S., Steinmetz, S., Reisser Junior, C. & Almeida, I. R. *Atlas climático da Região Sul do Brasil: Estados do Paraná, Santa Catarina e Rio Grande do Sul*. 333 p. (Pelotas: Embrapa Clima Temperado, Colombo: Embrapa Florestas 2012)
55. Santos, H. G. *et al.* *Sistema Brasileiro de Classificação de Solos* (Embrapa, 2018).
56. IAL. INSTITUTO ADOLFO LUTZ. Ministério da Saúde. Agência Nacional de Vigilância Sanitária. *Métodos físico-químicos para análise de alimentos*; Ministério da Saúde (2005).
57. McCready, R. M. & McComb, E. A. Extraction and determination of total pectin materials in fruits. *Anal. Chem.* **24**, 1986–1988 (1952).
58. Bitter, T. & Muir, H. M. A modified uronic acid carbazole reaction. *Anl. Biochem.* **4**, 330–334 (1962).
59. Benassi, M. T. & Antunes, A. J. A comparison of methaphosphoric and oxalic acids as extractant solutions for the determination of vitamin C in selected vegetables. *Braz. Arch. Biol. Technol.* **31**, 507–513 (1988).
60. Giusti, M. M. & Wrolstad, R. E. Characterization and measurement of anthocyanins by UV-Visible spectroscopy. *Curr. Protoc. Food Analyt.* **1**, 1–2 (2001).
61. Kassambara, A. & Mundt, F. Factoextra: Extract and visualize the results of multivariate analyses. R package version 1.0.7. (2020)
62. R Core Team. *R: A Language and Environment for Statistical Computing* (2019)
63. Wickham, H., François, R., Henry, L. & Müller, K. dplyr: A grammar of data manipulations. R package version 0.8.5. (2020)
64. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
65. Tang, Y., Horikoshi, M. & Li, W. ggfortify: Unified interface to visualize statistical result of popular R packages. *R J.* **8**(2), 478–489 (2016).
66. Wei, T. & Simko, V. R package “corrplot”: Visualization of a correlation matrix. version 0.84. (2017).
67. Mendiburu, F. Agricolae: Statistical procedures for agriculture research. R package version 1.3-2. (2020).
68. Resende, M. D. V. & Alves, R. S. Linear, generalized, hierarchical, bayesian and random regression mixed models in genetics/genomics in plant breeding. *Funct. Plant Breed. J.* **2**(2) (2020).
69. Cruz, C. D. Genes: A software package for analysis in experimental statistics and quantitative genetics. *Acta Sci. Agron.* **35**, 271–276 (2013).
70. Cruz, C. D. Genes software: Extended and integrated with the R Matlab and Selegen. *Acta Sci.* **38**(4), 547–552 (2016).
71. Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A. NbClust: An R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* **61**(6), 1–36 (2014).
72. Oksanen, J., *et al.* Vegan: Community ecology package. R package version 2.5-6, (2019).
73. Paradis, E. & Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2018).
74. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. Cluster: Cluster analysis basics and extensions. R package version 2.1.0. (2019).

Author contributions

R., R., and da Silva experimental design and methodology; B. and de R. supervise field experiments and data collection. B., da Silva, Ru, and M. conducted the statistical analysis and prepared figures and tables. B., R., and da Silva. wrote the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.L.B.R.S.

Reprints and permissions information is available at 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022