Peruvian Journal of Agronomy
http://revistas.lamolina.edu.pe/index.php/jpagronomy/index

Received for publication: 27 September 2022
Accepted for publication: 30 November 2022
Published: 31December 2022
ISSN: 2616-4477

RESEARCH ARTICLE
https://doi.org/10.21704/pja.v6i3.1974

Eclidean distance can recognize the best *Stevia* genotype and environment to produce rebaudioside and stevioside under controlled conditions

La distancia euclidiana puede reconocer el mejor genotipo y entorno de Stevia para producir rebaudiósido y esteviósido en condiciones controladas

María de Lourdes Tapia y Figueroa[1]*;Luz R. Gómez Pando[1]

*Corresponding author:ltapia@lamolina.edu.pe
*https://orcid.org/0000-0002-5000-3504

Abstract

Stevia rebaudiana is considered an important medicinal plant possessing low-calorie glucoside sweeteners. The present work describes the comparison of three stevia genotypes (IBT 1, IBT 2 and IBT 3) in two contrasting environments simulated under controlled conditions: Sullana in Peru; and Misiones in Paraguay (regarded as the center of origin of Stevia). In the study, we explored the Euclidean distance as an *integrating indicator* for simultaneous selection of several stevia traits. Plant scientists often record multiple morphological, physiological and biochemical indicators in their experiments. Common statistical data evaluations involve univariate analyses such as t-test, Mann-Whitney and Analysis of Variance followed by Tukey HSD. However, these analyses do not evaluate integrally the effects of the experimental treatments because each indicator is analyzed independently. Euclidean distance from each treatment combination to the ideal phenotype of the stevia plantlets was calculated. IBT 2 grown in Sullana environmental conditions showed the best integral results, while IBT 1 displayed the worst results. esponse parameters to different contrasting environments. The analysis shown here indicates that the use of the Euclidean distance could contribute to establishing a more integrated evaluation of the contrasting Stevia genotypes. On the other hand, the Euclidean distance, as a non-dimensional indicator, can help to compare different phenotype traits.

Keywords: Biostatistics, research methods, glucoside, phenotypic traits, sweet grass.

Resumen

Stevia rebaudiana se considera una importante planta medicinal que posee edulcorantes glucósidos bajos en calorías. El presente trabajo describe la comparación de tres genotipos de stevia (IBT 1, IBT 2 e IBT 3) en dos ambientes contrastantes simulados bajo condiciones controladas: Sullana en Perú; y Misiones en Paraguay (considerado como el centro de origen de la Stevia). Exploramos la distancia euclidiana como un indicador integrador para la selección simultánea de varios rasgos de stevia. Los científicos de plantas suelen registrar múltiples indicadores morfológicos, fisiológicos y bioquímicos en sus experimentos. Las

How to cite this article:
Tapia, M., & Gómez, L. (2022). Euclidean distance can recognize the best *Stevia* genotype and environment to produce rebaudioside and stevioside under controlled conditions. *Peruvian Journal of Agronomy, 6*(3), 222–228. https://doi.org/10.21704/pja.v6i3.1974

[1] Universidad Nacional Agraria La Molina, Av. La Molina s/n, Lima, Perú

evaluaciones de datos estadísticos comunes implican análisis univariados como la prueba t, Mann-Whitney y el análisis de varianza seguido de Tukey HSD. Sin embargo, dichos análisis no evalúan integralmente los efectos de los tratamientos experimentales porque cada indicador se analiza de forma independiente. Se calculó la distancia euclidiana de cada combinación de tratamientos al fenotipo ideal de las plántulas de stevia. IBT 2 cultivado en condiciones ambientales de Sullana presentó los mejores resultados integrales, mientras que IBT 1 mostró los peores resultados. El análisis indica que el uso de la distancia euclidiana podría contribuir a establecer una evaluación más integrada de los genotipos contrastantes de Stevia. Así también, puede ayudar a comparar diferentes rasgos fenotípicos.

Palabras clave: bioestadística, Métodos de búsqueda, glucósidos, rasgos fenotípicos, hierba dulce.

Introduction

Stevia rebaudiana (Asteraceae; Sweet Grass) is considered an important medicinal plant as it possesses many beneficial effects on type II diabetes, possessing low-calorie glucoside sweeteners (Brahmachari et al., 2011). Also, the plant contains important natural antioxidants such as flavonoids, phenols, tannin and essential oils (Christaki et al., 2013).

It is a native species of the tropical region of South America; and is still found in the wild in Paraguay (Giuffre et al., 2013; Gusmaini et al., 2022). The bushes of this species are perennial and reach 0.9 m height. Its leaves, lanceolate or elliptic and toothed, are alternate, simple, of a shiny dark green color and a rough surface, sometimes somewhat hairy, up to 5 cm long by 2 cm wide. The stems, pubescent and straight, only branch after the first vegetative cycle, with a tendency to lean. The roots are mostly superficial, although a thickened section sinks deeper; fibrous, threadlike and perennial, they are the only part of the plant in which steviosides it is not founded (Ijaz et al., 2015; Hossain et al., 2017).

As for their flowers, they are dioecious plants that at the beginning of spring present small, tubular and white flowers, without perceptible fragrance, in corymboid panicles formed by small axillary chapters; they take more than a month to produce all the flowers (Büyük et al., 2022; Gusmaini et al., 2022). In nature they are pollinated by bees. The fruits are achenes endowed with a hairy pappus that facilitates their transport by the wind. The yield difference in steviosides and rebaudiosides is very pronounced between the different genotypes. Currently, the best quality and most profitable is the "Paraguayan stevia", with up to 4 to 5 annual harvests (Giuffre et al., 2013; Hossain et al., 2017). The present work describes the comparison of three new *Stevia* genotypes (IBT 1, IBT 2 and IBT 3) in two contrasting environments simulated under controlled conditions: Sullana in Peru; and Misiones in Paraguay (considered the center of origin of *Stevia*). In this study we explore Euclidean distance as an *integrating* indicator for simultaneous selection of several *Stevia* traits.

Plant scientists usually record multiple morphological, physiological and biochemical indicators in their experiments. The common statistical data evaluations involve univariate analyses such as t-test, Mann-Whitney and Analysis of Variance (ANOVA) followed by Tukey HSD (Lorenzo et al., 2015). However, such analyses do not evaluate integrally the effects of the experimental treatments because each indicator has been analyzed independently. For this reason, in this study we explore Euclidean distance combined with the *Stevia* phenotypic traits as an *integrating* indicator. Euclidean distance has been widely used in many scientific fields *e.g.* in context of pattern recognition (Ichino, 1988), bioinformatics (Tavazoie et al., 1999), intelligent control systems (Jafar & Zilouchian, 2001; Fliege et al., 2019), spectral identification (Granahan & Sweet 2001), information retrieval (Kogan, 2007), simultaneous selection of several agricultural traits (Gomez-Pando et al., 2009; Haque & Haque, 2018; Angassa & Mohammed, 2022), and plant *in vitro* culture experiments (Lorenzo et al. 2013; Gómez et al., 2018; Villalobos-Olivera et al., 2019). However, as far as we know, the statistical management reported here has not been frequently utilized in studies of interactions of *Stevia* with different environments.

Materials and methods

Three *Stevia* genotypes (IBT1, IBT2 and IBT3) were grown under controlled conditions, simulating two contrasting environments:

Sullana in Peru; and Misiones in Paraguay that is regarded as the center of origin of *Stevia* (Table 1). IBT1 is the Morita II variety, developed in Japan by Toyoshigue Morita and introduced to Peru. IBT2 is the Miskibamba variety derived from the Morita II variety; and IBT3 is a mutant line derived from the Miskibamba variety using gamma radiation at a dose of 20 Gy in the laboratory of the Institute of Biotechnology (IBT) of Universidad Nacional Agraria La Molina. Thirty-day plantlets used in the study were obtained under *in vitro* conditions in MS (Murashige – Skoog) culture medium following established protocols for stevia. Acclimatization was carried out for one month in a greenhouse using the sterilized premix 3 substrate (500 g sterile-PREMIX soil per plantlet).

All data of this study were statistically evaluated using SPSS (Version 8.0 for Windows, SPSS Inc., New York, NY) to perform Two-Way analysis of variance (ANOVA) and Student – Newman - Keuls tests (p=0.05). The overall coefficients of variation (OCV) were used to assess sensitivity and they were calculated as follows:

(Standard desviation/Average)*100 (1)

In this formula, we considered the average values of the six combinations compared (2 environments and 3 *Stevia* genotypes) to calculate the standard deviation and average. Therefore, the higher the difference between the six treatments compared, the higher is the OCV (Lorenzo et al., 2015). The OCVs were categorized as follows: *Low* from 12.37 % to 26.99 %, *Medium* from 26.99 % to 41.61 % and *High* from 41.61 % to 56.2 3%. Ranges for OCV classification as *High*, *Medium* and *Low* depend on OCV values recorded in each experiment (Lorenzo et al., 2015).

In addition, to identify the most significant integral effect of the six combinations studied on *Stevia* phenotype, the euclidean distances to the expert criteria were calculated. As the ideal expert criteria, the maximum values of plant height, numbers of leaves and shoots per plant, root length, plant fresh weight, and rebaudioside and stevioside content were considered. An excel sheet (Microsoft Office) was used. Original data were standardized with 0 to 1 by the min-max normalization (Kantardzic, 2003). Min and max values of each variable were identified and then the following formula was used to standardize data:

(Value to be normalized-Min.value observed in the experiment)/(Max.value observed in the experiment – Min.value observed in the experiment) (2)

After standardization of all treatment data, the Euclidian distance to the best expert criteria was calculated according to equation:

$$\text{Euclidean distance to the expert criteria} = \sqrt{\sum_{i=1}^{n}(qi - pi)^2} \quad (3)$$

n = Number of dependent variables evaluated

q = Indicators evaluated

p= Indicators described as expert criteria: maximum values of the seven dependent variables described

i = Indicators evaluated

Table 1. Environments compared under control conditions

Time (h)	Temperature (°C)	Relative humidity (%)	Radiation	Time (h)	Temperature (°C)	Relative humidity (%)	Radiation (W/m²)
0-0	20.0	80.0	0.0	0-0	21.0	76.0	0.0
0-6	19.7	81.6	0.0	0-6	18.8	83.3	0.0
6-11	20.6	77.7	100.0	6-11	24.4	58.0	200.0
12-5	28.1	52.7	300.0	11-5	35.0	22.5	300.0
6-6	28.7	50.9	100.0	5-8	28.7	34.3	200.0
7-12	23.8	66.6	0.0	8-12	26.3	36.0	0.0

Table 2. Phenotypes of *Stevia* materials in two contrasting environmental conditions simulated under controlled conditions (averages ± SE).

Experimental location simulated	Stevia genotype	Plant height (cm)	Number of leaves per plant	Number of shoots per plant	Root length (cm)	Plant fresh weight (g)	Rebaudioside content (g/100 g DW)	Stevioside content (g/100g DW)
Sullana (Peru)	IBT 1	9.33 ± 0.81 b	34.01 ± 2.94 c	3.66 ± 0.26 c	11.25 ± 1.23 cd	2.91 ± 0.15 c	10.10 ± 1.03 f	3.44 ± 0.26 f
	IBT 2	13.37 ± 1.25 a	57.66 ± 3.87 a	7.00 ± 0.54 a	13.66 ± 1.56 b	7.04 ± 0.64 a	16.05 ± 1.68 e	4.59 ± 0.36 a
	IBT 3	13.37 ± 0.92 a	45.00 ± 3.56 b	5.00 ± 0.48 b	12.33 ± 1.13 bc	3.94 ± 0.26 b	18.48 ± 1.54 c	4.58 ± 0.34 b
Misiones (Paraguay, center of origin of *Stevia*)	IBT 1	12.66 ± 1.14 a	20.31 ± 1.96 d	2.00 ± 0.19 d	10.00 ± 0.98 d	2.48 ± 0.23 cd	16.46 ± 1.13 d	3.61 ± 0.28 e
	IBT 2	12.75 ± 1.11 a	12.50 ± 1.12 d	5.00 ± 0.33 b	26.33 ± 1.35 a	1.73 ± 0.14 d	19.66 ± 1.59 b	4.53 ± 0.34 c
	IBT 3	11.50 ± 1.15 a	37.00 ± 2.98 c	4.50 ± 0.34 bc	14.00 ± 1.38 b	2.39 ± 0.12 cd	20.21 ± 1.96 a	4.27 ± 0.31 d
OCV (%)*		12.72	47.54	36.57	40.69	56.23	21.95	12.37
Classification of OCV**		*Low*	*High*	*Medium*	*Medium*	*High*	*Low*	*Low*

Note. Results with the same *letter* are not statistically different (Two-Way ANOVA, Student-Newman-Keuls, p>0.05). For statistical analysis only, the numbers of leaves and shoots were transformed according to $y'=y^{0.5}$.

* Overall coefficient of variation = (Standard deviation/Average)*100. To calculate this coefficient, average values were considered. The higher the difference among the six treatments compared, the higher the overall coefficient of variation. OCVs were calculated for those indicators with statistically significant difference according to Two-Way ANOVA and Student-Newman-Keuls tests.

** *Low* from 12.37 % to 26.99 %, *Medium* from 26.99 % to 41.61 % and *High* from 41.61 % to 56.23 %.

Results and discussion

Several statistically significant differences among the six combinations studied were recorded (Table 2). However, *High* OCVs were only noted in the number of leaves per plant, and in the plant fresh weight. The highest values were recorded in IBT 2 grown in the chamber that simulates the environment of Sullana (Peru). On the other hand, *Medium* OCVs were observed in the number of shoots per plant and in the root length (Table 2). The IBT 2 grown in Sullana showed the highest values of shoots per plant. This genotype, grown in Misiones showed the longest roots.

The calculation of the Euclidean distance from treatment combination to the ideal phenotype of the *Stevia* plantlets is shown Figure 1. The IBT 2 grown in Sullana showed the best integral results, while IBT 1 displayed the worst results.

A comprehensive picture of the effects of *Stevia* genotypes and environments on the plant phenotypes is shown in Table 2. Each dependent variable was analyzed separately, giving a broad insight into the physiological changes induced by the treatment. However, experimental noise and sensitivity of individual measurements make many of these parameters insufficient to accurately rank the treatments in terms of severity. Therefore, we combined all measurements by calculating Euclidean distances to integrate results.

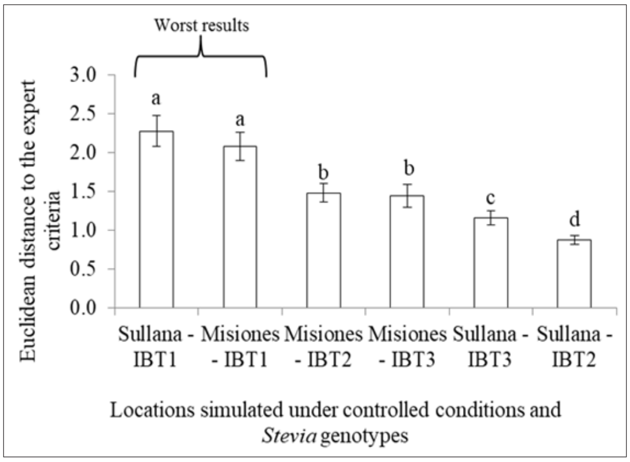

Figure 1. Euclidean distance from each environment simulated and *Stevia* genotype to the expert criteria. Results with the same letter are not statistically different (Two-Way ANOVA, Student-Newman-Keuls, p>0.05), vertical bars represent SE.

Before calculating the Euclidean distances, standardization of variables is important to prevent certain features from dominating distance determinations, merely because they have large numerical values (Duda et al., 2001; Kantardzic, 2003). The following information was considered as the ideal phenotype of *Stevia* to calculate the Euclidean distances: maximum values of plant height; number of leaves per plant; number of shoots per plant; root length; plant fresh weight; rebaudioside content; and stevioside content.

The equation (3) means that the Euclidian distance tends to increase with increase in the number of parameters. The number of recorded variables should be the same for all treatments; and traits must be quantitative. Theoretically, the number of variables evaluated is limitless. We would like to mention other methods available, such as the biplot metrics or NMDS (non-dimensional scaling), to visualize the distances between the treatments. A biplot is a graphical representation of multivariate data, where the elements of a data matrix are represented according to dots and vectors associated with the rows and columns of the matrix. On the other hand, the goal of NMDS is to represent the original position of data in multidimensional space as accurately as possible using a reduced number of dimensions that can be easily plotted and visualized. NMDS relies on rank orders (distances) for ordination (*i.e.* non-metric). The use of distances omits some of the issues associated with using predictor variables alone (*e.g.* sensitivity to transformation). The NMDS allows for much more flexible technique that accepts a variety of data types (Chapman et al., 2001; Gabriel, 2002; Krzanowski, 2004; Faria & Demetrio, 2008; Blasius et al., 2009).

Conclusions

As discussed earlier, the Euclidean distance has been widely used in other scientific fields, but to our knowledge it has not been used to integrate *Stevia* plantlet response parameters to different contrasting environments. The analysis shown here indicates that the use of the Euclidean distance could contribute to establishing a more *integrated* evaluation of the contrasting *Stevia* genotypes. On the other hand, the Euclidean distance, as a non-dimensional indicator, can help to compare different phenotype traits.

Acknowledgements

This research was supported by the Universidad Nacional Agraria La Molina, Lima, Peru.

ORCID and e-mail

Tapia, L. ltapia@lamolina.edu.pe
https://orcid.org/0000-0002-5000-3504
Gómez luzgomez@lamolina.edu.pe
https://orcid.org/0000-0001-5536-5179

References

Angassa, D., & Mohammed, J. (2022). Agro-morphological Variability Study of Ethiopian Barley (*Hordeum vulgare L.*) Accessions for Their Important Agronomical Traits at Hadiya Zone, Southern Ethiopia. *J Plant Sci*., *10*,19–25

Blasius, J., Eilers, P., & Gower, J. C. (2009). Better Biplots. *Computational Statistics and Data Analysis, 53*(8), 3145–3158

Brahmachari, G., Mandal, L. C., Roy, R., Mondal. S., & Brahmachari, A. K. (2011). Stevioside and related compunds - molecules of pharmaceutical promise: a critical overview. *Arch Pharm, 344*(1), 5–19

Büyük, G., Bayram, C. A., İnan, M., & Kırpık, M. (2022). The effect of organic and inorganic fertilizers on plant nutrient content and agronomic performance of stevia. *Journal of Plant Nutrition, 45*(15), 2303–2314. https://doi.org/10.1080/01904167.2022.2063739

Chapman, S., Shenk, P., Kazan, K., & Manners, J. (2001). Using Biplots to interpret gene expression pattern in plants. *Bioinf Applic Note, 18*(1), 202–204

Christaki, E., Bonos. E., Giannenas. I., Karatzia, M. A., & Florou-Paneri, P. (2013). *Stevia rebaudiana* as a novel source of food additives. *Journal of Food & Nutrition Research, 52*(4), 167–178.

Duda, R., Hart, P., & Stork, D. (2001). *Pattern Classification.* John Wiley & Sons Inc, New York.

Faria, J. C., Demétrio, C. G. B., & Allaman, I. B. (2008). *BPCA: Biplot of multivariate data based on principal components analysis* (R package; version 1.02) UESC and ESALQ, Ilheus, Bahia, Brasil and Piracicaba, Sao Paulo, Brasil. http://CRAN.R-project.org/package=bpca

Fliege, J., Qi, H-D., & Xiu, N. (2019). Euclidean distance matrix optimization for sensor network localization. In: C. Gao, G. Zhao & H. Fourati (eds.), *Cooperative Localization and Navigation: Theory, Research and Practice.* CRC Press, Boca Raton, FL

Gabriel, K. R. (2002). Goodness of fit of biplots and correspondence analysis. *Biometrika, 89*(2), 423–436

Giuffre, L., Romaniuk, R., & Ciarlo, E. (2013). Stevia, ka'a he'e, wild sweet herb from South America-An overview. *Emirates Journal of Food and Agriculture, 25*(10), 746–50, https://doi.org/10.9755/ejfa.v25i10.16405

Gomez-Pando, L., Jimenez-Davalos, J., Eguiluz-de la Barra, A., Aguilar-Castellanos, E., Falconí-Palomino, J., Ibañez-Tremolada, M., Varela, M., & Lorenzo, J.C. (2009). Field performance of new in vitro androgenesis-derived double haploids of barley. *Euphytica, 166,* 269–276. https://doi.org/10.1007/s10681-008-9840-0

Gómez, D., Hernández, L., Yabor, L., Beemster, G. T. S., Tebbe, C. C., Papenbrock, J., & Lorenzo, J. C. (2018). Euclidean distance can identify the mannitol level that produces the most remarkable integral effect on sugarcane micropropagation in temporary immersion bioreactors. *Journal of Plant Research, 131*, 719–724. https://doi.org/10.1007/s10265-018-1028-7

Granahan, J., & Sweet, J. (2001). An evaluation of atmospheric correction techniques using the spectral similarity scale. *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium, 5,* 2022–2024. https://doi.org/10.1109/IGARSS.2001.977890

Gusmaini, Kartikawati, A., Nurhayati, H., Permadi, R., & Syakir, M. (2022). The utilization of plant growth-promoting bacteria to enhance stevia (*Stevia rebaudiana*) herb yield at low land. *IOP Conference Series: Earth and Environmental Science, 974*, 012028. https://doi.org/10.1088/1755-1315/974/1/012028

Haque, F., & Haque, S. (2018). Plant recognition system using leaf shape features and minimum Euclidean distance. *Ictact Journal On Image And Video Processing, (9)*2, 1919–1925

Hossain, M., Islam, M., Islam, M., & Akhtar, S. (2017). Cultivation and uses of stevia (*Stevia rebaudiana* Bertoni): A review. *African J. Food Agric. Nut. Dev. 17*(4), 12745–12757. https://doi.org/10.18697/ajfand.80.16595

Ichino, M. (1988). General metrics for mixed features-the cartesian space theory for pattern recognition. *Proceedings of the 1988 IEEE International Conference on Systems, Man, and Cybernetics*, 494–497. https://doi.org/10.1109/ICSMC.1988.754346

Ijaz, M., Pirzada, A. M., Saqib, M., & Latif, M. (2015). Stevia rebaudiana: An alternative sugar crop in Pakistan–a review. *Zeitschrift für Arznei- & Gewürzpflanzen, 20*(2), 88–96

Jafar, M., & Zilouchian, A. (2001). Application of soft computing for desalination technology. In: A. Zilouchian & M. Jamshidi (eds.), *Intelligent Control Systems Using Soft Computing Methodologies*. (pp. 315–353). CRC Press, Boca Raton.

Kantardzic, M. (2003). *Data mining: concepts, models, methods and algorithms.* Wiley-IEEE Press.

Kogan, J. (2007). *Introduction to Clustering Large and High Dimensional Data*. Cambridge University Press, New York,

Krzanowski, W. J. (2004). Biplots for multifactorial analysis of distance. *Biometric,* 60(2), 517–524

Lorenzo, J. C., Varela, M., Hernández, M., Gutiérrez, A., Pérez, A., Loyola, O. (2013). Integrated criteria to identify the best treatment in plant biotechnology experiments. *Acta Physiol Plant, 35*, 3261–

3264. https://doi.org/10.1007/s11738-013-1352-4

Lorenzo, J. C., Yabor, L., Medina, N., Quintana, N. & Wells V. (2015). Coefficient of variation can identify the most important effects of experimental treatments. *Not Bot Horti Agrobo Cluj-Nap,43*(1), 287–291. https://doi.org/10.15835/nbha4319881

Tavazoie, S., Hughes, J. Campbell, M., Cho, R., & Church, G. (1999). Systematic determination of genetic network architecture. *Nat Genet, 22*, 281–285. https://doi.org/10.1038/10343

Villalobos-Olivera, A., Hernández, L., Martínez, J., Quintana, N., Zevallos, B. E., Yabor, L., Martínez-Montero, M. E., González-Olmedo, J., Sershen, J. C. L. (2019) Euclidean distance can recognize the Biojas® concentration that produces the ideal physiological status of pineapple *in vitro*-plantlets. *In Vitro Cell.Dev.Biol.-Plant, 56,* 259–263 (2020). https://doi.org/10.1007/s11627-019-10023-5