

RESEARCH

Open Access



Self-pollinated cannabis seeds lead to less variation in shape: a technological approach of potential commercial interest

Francisco Fernandez Torne^{1,2}, Yanina L. Idaszkin^{2,3}, Gregorio Bigatti^{1,2,4}, Pablo Navarro⁵, Rolando González-José⁵ and Federico Márquez^{1,2*} 

Abstract

Background The breeding process enables plants to inherit desirable traits, such as yield, flowering time, pest resistance, and cannabinoid and/or terpene content. As a result of these intensive genetic improvement practices, where genetically similar individuals or those from the same lineage are crossed, the expression of unfavorable recessive alleles may occur due to homozygosity. This can lead to less productive plants, increased susceptibility to diseases, and reduced quality. Despite the potential negative effects associated with inbreeding, self-pollination (a form of inbreeding) is a necessary cultivation technique used to obtain seeds that produce phenotypically female plants (feminized seeds) for commercialization and/or to fix desirable traits, albeit at the cost of reduced genetic variability. The *Cannabis sativa* L. seed market has grown significantly in recent decades, driven by the legalization and regulation of medicinal and recreational use. Self-pollinated feminized seeds are popular among growers and commercial seed banks because, in most cases, they guarantee that inflorescences will express the cannabinoid and terpene profile of the single parent plant.

The objective of this work is to compare the morphological variation of seeds obtained from the reversal of female clones followed by self-pollination, and seeds obtained from crossing genetically distinct parental. To study seed shape and size, we employed 2D geometric morphometrics (GM) based on landmarks and semilandmarks, coupled with a supervised machine learning approach and multivariate statistical approach for analysis.

Results No direct relationship was observed between size and seed type, although significant differences between varieties were detected. The shape of seeds from crosses between different parents (male and female) showed lower classification accuracy compared to feminized seeds. These results support the hypothesis that inbreeding reduces the variability, as feminized seeds from self-pollination were correctly identified at a high rate using a discriminant function.

Conclusions Our research demonstrates that 2D geometric morphometrics can effectively distinguish and trace feminized and self-pollinated cannabis seeds. These seeds exhibit the least morphological variation, enabling accurate

*Correspondence:
Federico Márquez
fedem@cenpat-conicet.gob.ar

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

identification and providing a reliable foundation for practical applications. The Random Forest classifier's high performance confirms the effectiveness of using morphological traits for seed discrimination. These results open the door for advanced machine-learning techniques aimed to improve scalability and automation.

Keywords Inbreeding, Shape variability, Technological application, Seed traceability, Self-pollination

Introduction

Research on selection and crossbreeding across generations has been essential for understanding the genetics and diversity of *Cannabis sativa* L. This species has been subjected to diverse genetic improvement programs and cultivation, a common practice among growers and seed banks, resulting in a wide diversity of varieties [1]. Controlled crosses have enabled the development of new varieties with specific traits, such as higher yield, disease resistance, and/or targeted cannabinoid content [2, 3]. In some cases, intensive selection and crossbreeding for genetic improvement lead to problems associated with inbreeding due to reduced genetic variability. This often manifests as diminished adaptability and loss of hybrid vigor [4]. Furthermore, inbreeding effects can unmask deleterious recessive alleles, resulting in plants with reduced fitness and undesirable phenotypic traits [5]. Previous studies have linked inbreeding to changes in plant morphology [6]; such changes could have direct implications for variety identification and differentiation, as well as for inflorescence quality and yield [7, 8].

In this context, cannabis seeds' (achenes) size and shape emerge as a suite of phenotypes of interest for studying the effects of artificial selection. Furthermore, geometric morphometrics (GM) has arisen as a multi-disciplinary tool used in the study of cellular structures, ecosystem characterization, identification of evolutionary patterns, and functional adaptations. It addresses research questions across biological disciplines, from biological anthropology to taxonomy. GM is employed to quantify morphological variation in plants [9, 10], enabling detailed analysis of the size and shape of biological structures while providing insights into intra and interspecific variation [11]. In botanical studies, it is commonly used to characterize leaf shape [12] and floral architecture [13], whereas seed morphology has proven useful for species identification and/or characterization [14], assessment of genetic diversity [15], and studies of plant evolution [16].

Recently, automated systems combining GM and machine learning have been developed to classify different grain types [17, 18]. These techniques have also made it possible to distinguish between seeds of different oak species [19] and to discriminate between tomato [20] and rice [21] varieties. Some of the examples mentioned above support the feasibility of using GM in grain and seed identification, which has potential implications

for various industries and scientific disciplines [9, 17]. GM applied to cannabis seeds may be particularly relevant for identifying and characterising varieties [22–26], considering that cannabis plants can exhibit a wide range of shapes and sizes in vegetative and reproductive organs [27]. Notably, certain seed shape characteristics correlate with chemotypes: seeds from THC-dominant plants (Chemotype I) display rounded shapes. In contrast, CBD-dominant plants (Chemotype III) tend toward slender shapes [22, 23]. Regardless of whether this trend is maintained across varieties, the results expose certain potential in the use of 2D geometric morphometry in the study of the shape and size of cannabis seeds.

Therefore, this study tests the central hypothesis that the breeding method, specifically, self-pollination in feminized lines using a standard commercial technique, leaves a quantifiable imprint on seed morphology, resulting in lower shape variation compared to seeds from crosses.

Materials and methods

Material

Regular and feminized seeds were studied. Regular seeds produce either male or female plants, while feminized seeds are designed to produce almost exclusively female plants. A total of 713 seeds from seven varieties were utilized. These comprised seeds from the following lines (abbreviation and sample sizes among parentheses): Malvina (Ma, 129), Pachamama (P, 91), Fruti (F, 97), H (H, 97), K (K, 99), Lemon (L, 96), and Medicinal (Me, 104). Seeds of P and Ma, derived from commercial varieties selected for their specific chemotypes (cannabinoid profiles) and registered at the National Seed Institute (INASE) by CONICET, were feminized by self-pollination. These registered varieties have been maintained through clonal propagation since 2018. The remaining seeds were obtained from crossing varieties donated by recognized cannabis growers. F and L seeds were feminized and originated from female “regular” plants; each female was crossed with the same reversed female parent pollen of the registered variety “CONICET” (INASE). In contrast, H, K, and Me seeds were regular (non-feminized); H and K seeds were derived from different THC varieties and crosses with pollen of the regular “Jaguar Negro”, also a THC variety. Me seeds were obtained by crosses with pollen of the same variety (F2 progenie). Feminized seeds were produced through

sexual reversion via exogenous growth regulators [28]. The seeds analyzed in this study represent a realistic sample of the commercial and non-commercial varieties used by growers. Seeds were derived from crosses involving F1 parents with uncertain or undocumented pedigrees and are therefore not considered genetically stabilized lines. Since harvest, all seeds were stored under controlled conditions (darkness, low humidity, and room temperature) prior to imaging to minimize environmental effects on morphology.

Seed morphometric analysis

Seeds were photographed under a Carl Zeiss stereomicroscope equipped with AxioVision Rel.4.5 software (©Carl Zeiss Imaging Solutions), oriented with the abscission zone upward and the radicle wall suture positioned leftward (Fig. 1). To capture seed shape, a configuration of 3 Type I (anatomical) landmarks [29] and 10 Type III semilandmarks digitized at equidistant positions along the contour (Fig. 1) was applied, following Márquez et al. (2022). Landmarks represent fixed anatomical points defining discrete biological structures, while semilandmarks quantify contour variation between landmarks [30]. Cartesian coordinates were digitized using the TPS software suite: TpsUtil [31]: Converted .jpg files to .tps format. TpsDig2 [32]: Scaled and digitized landmark/semilandmark configurations. TpsRelw [33]: Semilandmarks were slid using an iterative algorithm minimizing thin-plate spline (TPS) bending energy to establish homology. A Generalized Procrustes Analysis (GPA) was performed to remove rotational, translational, and scaling effects, preserving pure shape information [30]. Centroid size (CS), calculated as the square root of the sum of squared distances from all landmarks to their centroid [30], was used as a proxy for seed size. CS retains complete size information and, in the absence of allometry (shape-size covariation) represents the sole size estimator uncorrelated with shape.

Statistical analyses of size and shape

To evaluate differences in the Centroid size (CS) among varieties, a non-parametric Kruskal-Wallis test was performed because the data did not satisfy the normality and homogeneity of variance assumptions of parametric analysis. For significant results ($P < 0.05$), post hoc pairwise comparisons were performed using Dunn's test with Holm adjustment for multiple testing [34]. All analyses were conducted in R (version 4.5.1) using the packages FSA, ggplot2, dplyr, and multcompView.

Procrustes-aligned coordinates and CS values from TpsRelw were exported to MorphoJ [35] for multivariate analyses. Allometry was evaluated via multivariate regression of shape (Procrustes coordinates, dependent variable) against size (CS, independent variable). Shape variation (morphospace) was explored through principal component analysis (PCA) of the Procrustes distance variance-covariance matrix. Descriptive statistics of the first two PCs were calculated to quantify the magnitude of variation within each analyzed variety. These indices included the standard deviation and variance derived from the PC1 and PC2 scores. Subsequently, a hierarchical clustering analysis was conducted to categorize the varieties based on their variability profiles. Clustering of the seed shape patterns were examined using average linkage (UPGMA) hierarchical clustering based on Mahalanobis distances, implemented in InfoStat. This inferential method determines optimal cluster numbers in hierarchical analyses [36]. Finally, discriminant analysis and Hotelling's T^2 tests identified shape components maximizing differences between feminized and regular seeds and tested mean shape equality.

Machine learning

To complement the statistical analysis and evaluate the predictive power of seed morphology, a supervised machine learning approach was implemented. The full dataset, utilizing the Procrustes-aligned coordinates as

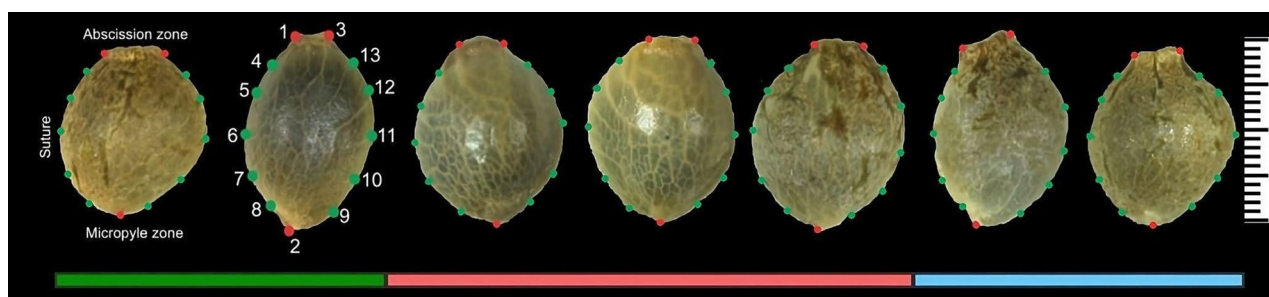


Fig. 1 External view (pericarp) of the longitudinal plane in seeds of Malvina, Pachamama, H, K, Medicinal, Lemon, and Fruti varieties of *C. sativa* L. (ordered left to right). Horizontal green bar: feminized by self-pollination; Red: regular seeds; Blue: feminized by crossing different female parents. All seeds display the landmark (red points) and semilandmark (green points) configuration used to capture contour shape. Landmarks: (1) Inflection point between the abscission zone and the fruit wall suture; (2) Micropyle region; (3) Inflection point between the abscission zone and the contour opposite the fruit wall suture. Semilandmarks: (4–8) equidistantly placed between landmarks 1 and 2 and (9–13) positioned along the cotyledon's external contour between landmarks 2 and 3. Vector scale = 5 mm

features and the seed variety as the target variable, was partitioned into a training subset (70% of the samples) and a testing subset (30% of the samples) to evaluate model generalization.

Four distinct classification algorithms were initially trained and evaluated for their performance based on test accuracy: Random Forest, Quadratic Discriminant Analysis (QDA), Gaussian Naive Bayes, and Decision Tree.

Following this initial comparison, the top-performing algorithm (Random Forest) was selected for comprehensive hyperparameter optimization. This optimization was conducted using a grid search methodology integrated with 5-fold cross-validation (GridSearchCV) on the training data to find the best combination of parameters. The parameter grid explored included: the number of estimators (`n_estimators`: [50, 100, 150]), maximum tree depth (`max_depth`: [5, 10, 15, None]), minimum samples per leaf (`min_samples_leaf`: [1, 2, 4]), minimum samples to split a node (`min_samples_split`: [2, 5, 10]), and the split quality criterion (`criterion`: ['gini', 'entropy']). All machine learning models were implemented using the Scikit-learn library in Python [37]. The code developed for this analysis is available in a GitHub repository

(<https://github.com/Francisco-ft/Self-pollinated-cannabis-seeds-lead-to-less-variation-in-shape>).

Results and discussion

Seed size

Although the multivariate regression of shape on size (allometry) was statistically significant ($P < 0.001$), the association percentage among varieties was negligible (% predicted = 4.64). This aligns with expectations, as seeds develop during a specific life-stage cycle with limited post-maturation growth variation, consistent with findings by Márquez et al. (2022) for five other *C. sativa* L. varieties.

Significant size differences occurred among varieties ($H = 584.7$; $P < 0.001$). However, no relationship existed between seed size and type (regular vs. feminized; Fig. 2). Therefore, size could serve as an indicator that provides information in discriminating varieties, but not to distinguish between feminized or regular seeds.

Seed shape

The first four principal components collectively accounted for 90.2% of total shape variation. Shape

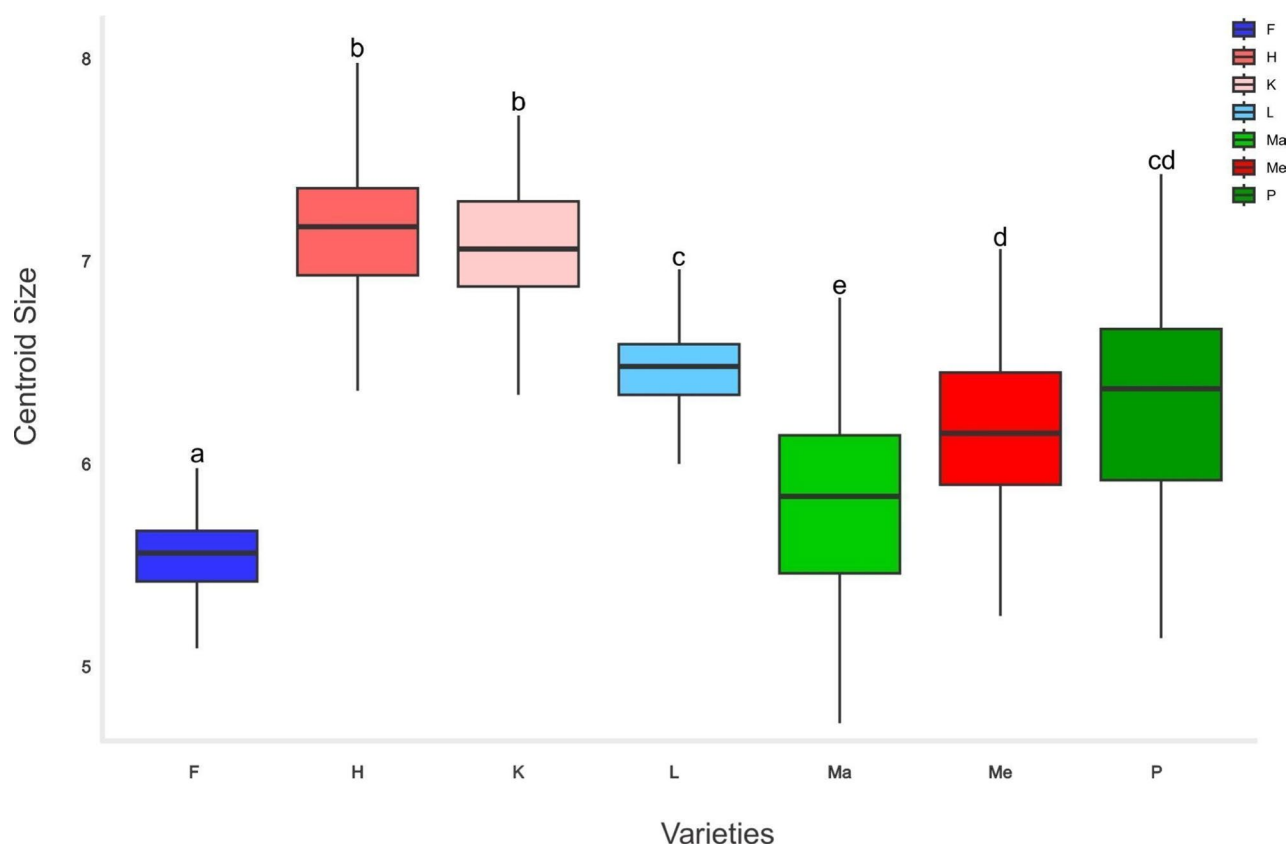


Fig. 2 Size variations among seed types and varieties. The central dot represents the mean; the central line indicates the median; box limits correspond to the first and third quartiles; whiskers show the 95% confidence interval; points beyond whiskers denote outliers. Different lower letters indicate significant differences ($P < 0.05$) in pairwise comparisons. Varieties: Malvina (Ma), Pachamama (P), K, H, Lemon (L), Fruti (F), and Medicinal (Me). Labels with green shades: feminized by self-pollination; Red shades: regular seeds, and Blue shades: feminized by crossing different parents

variation along PC1 (37.96% of total variation) was associated with seed slenderness. Positive extreme values represented elongated forms featuring longitudinal plane expansion, a more projected apex, and abscission zone constriction. Conversely, negative values corresponded to rounded morphologies (Fig. 3). Remaining principal component axes exhibited substantial variety overlap.

Descriptive statistics of the first two PCs revealed unequal dispersion levels among the varieties, with P as

the least variable and L as the most variable (Supplementary Table S1). Notably, P, Me, and Ma formed a distinct cluster characterized by minimal dispersion in the PCA space (Supplementary Figure S1). This group displayed the lowest variability, as indicated by lower variance and standard deviation values compared to the other varieties analyzed.

To quantify the discriminatory power of seed morphology, four distinct supervised machine learning models

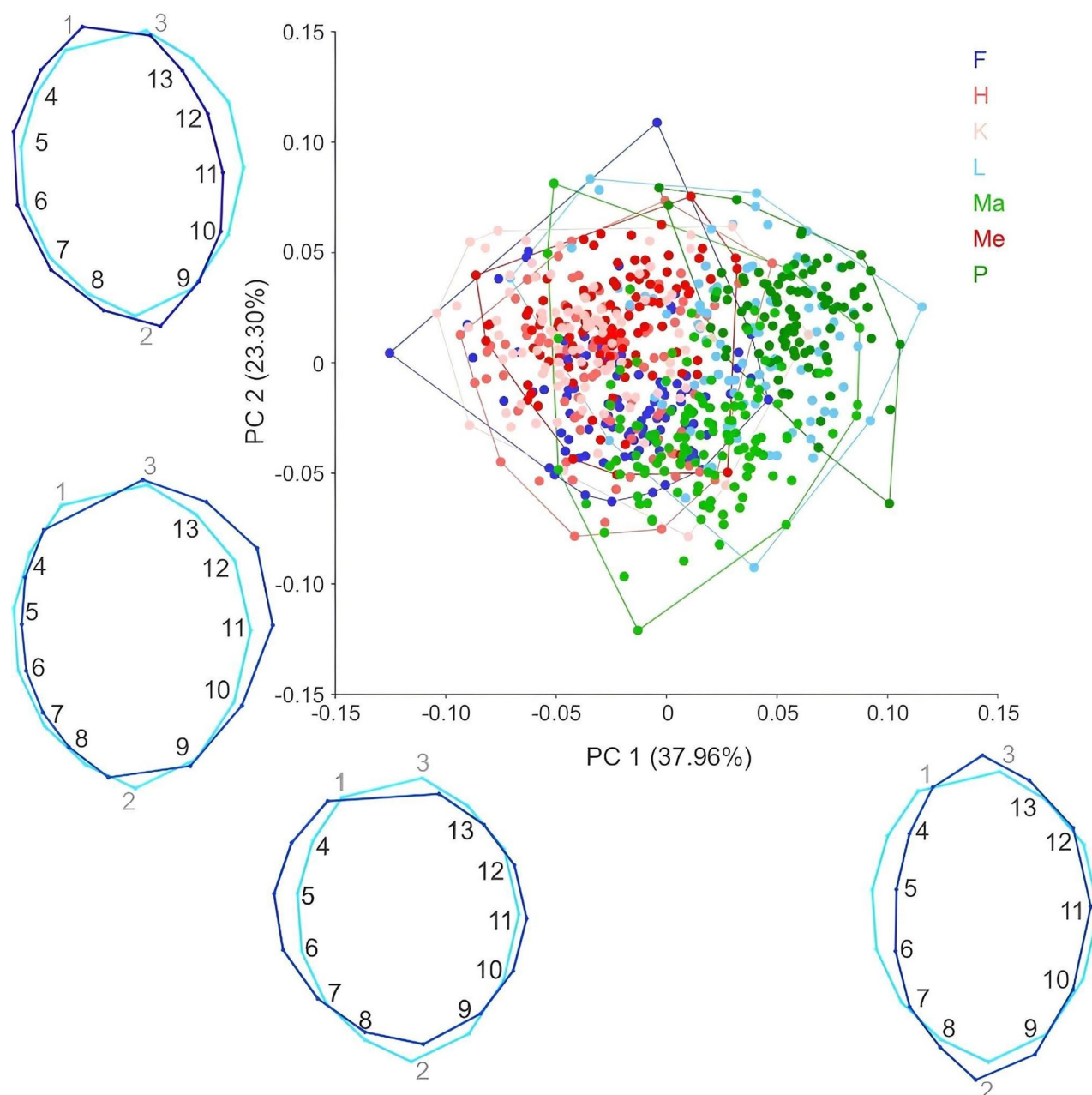


Fig. 3 Principal component analysis of shape variation in seeds from seven *C. sativa* L. varieties. The wireframe of the consensus shape is shown in turquoise, with ± 0.15 deviations depicted in blue. Ellipses represent 95% confidence intervals for observations per variety. Varieties: Malvina (Ma), Pachamama (P), K, H, Lemon (L), Fruti (F), and Medicinal (Me). Labels with green shades show seeds feminized by self-pollination; Red shades, regular seeds, and Blue shades, feminized seeds by crossing different parents

were initially trained and evaluated. When compared to the initially trained test dataset, the models achieved the following accuracies: Random Forest (67.8%), Quadratic Discriminant Analysis (66.4%), Gaussian Naive Bayes (63.6%), and Decision Tree (62.6%). Given that the Random Forest (RF) classifier demonstrated the highest generalization performance, it was selected for comprehensive hyperparameter optimization. The optimal parameters identified via grid search were: criterion='entropy', max_depth=15, min_samples_leaf=1, min_samples_split=5, and n_estimators=150. The cross-validated classification results from this optimized RF model are presented in Table 1. The model achieved an overall correct classification rate of 70%. Notably, feminized P seeds yielded the highest assignment accuracy (96.3%), whereas the regular seeds H (41.4%) and K (50%) showed the lowest classification success. This supports the hypothesis that the lower shape variability found in self-pollinated, inbred varieties (P) improves distinguishability, whereas greater morphological variance in regular seeds from crosses (H and K) reduces classification accuracy (Supplementary Figure S1). Furthermore, re-evaluating the classifier's performance after removing the highly variable H and K varieties increased the overall accuracy to 78%, which further reinforces this hypothesis.

Our results show that feminized seeds obtained through self-pollination (P and Ma) exhibited reduced shape variability. This could be reflecting diminished genetic diversity, as evidenced by inbreeding effects on seed morphological diversity. Conversely, regular seeds showed greater shape variability (Supplementary Figure S1 and Supplementary Table S1). This phenomenon, where inbreeding leads to a reduction in morphological variance of seeds, finds consistency in morphometric studies of other crop species. For instance, in wheat (*Triticum aestivum*), inbred lines exhibit significantly less variation in grain shape compared to genetically diverse landraces or hybrid populations [38]. A similar pattern has been documented in rice (*Oryza sativa*), where the

stabilization of pure lines results in highly uniform grain morphology, a key trait for varietal identification and seed certification [39].

The reduced shape variation in self-pollination feminized seeds improved geometric morphometric (GM) classification accuracy for commercially registered cannabis varieties studied. In particular, H and K seeds from the crossing of different parents expressed a wide variation in seed shape, which made accurate classification difficult. This variation caused feature overlap with other varieties, which reduced the effectiveness of the discriminant function and produced lower correct assignment rates. Hierarchical cluster analysis differentiated two primary groups (Fig. 4): Cluster 1 contained regular seeds (Me, H, and K), while Cluster 2 comprised all feminized seeds (P, Ma, F, and L). Within Cluster 1, H and K formed a distinct subgroup derived from different THC varieties and crosses with pollen of the regular “Jaguar Negro”, also a THC variety. Cluster 2 contained feminized seeds by self-pollination, with F and L forming a primary subgroup originating from female “regular” plants.

These groupings may reflect an association influenced by the genetic variability load and chemotype of each variety [21, 22]. This suggests potential developmental canalization toward specific shapes, intensified by inbreeding effects in self-pollinated feminized seeds, where morphological variation is reduced. Our results reveal that major commercial cannabis seeds worldwide likely share a key characteristic: intra-variatal homogeneity in seed shape, resulting from the loss of variability due to self-pollination. The optimized Random Forest model showed higher classification accuracy (>75% correct assignments; Table 1) for self-pollinated feminized seeds. These findings suggest that feminized seeds could be distinguished from non-stabilized regular seeds through primary screening using a discriminant function trained on feminized versus regular seed morphology (Fig. 5). Furthermore, there was a statistically significant difference in mean seed shape between feminized and regular seeds ($T^2 = 1238.4$; $P < 0.0001$). It is important

Table 1 Cross-validated classification analysis showing seed assignments to each variety using a random forest based on varietal shape morphology. Varieties: Malvina (Ma), Pachamama (P), K, H, lemon (L), fruti (F), and medicinal (Me). The percentage of seed assigned into varieties determined a priori, were highlighted in green

[illegible]

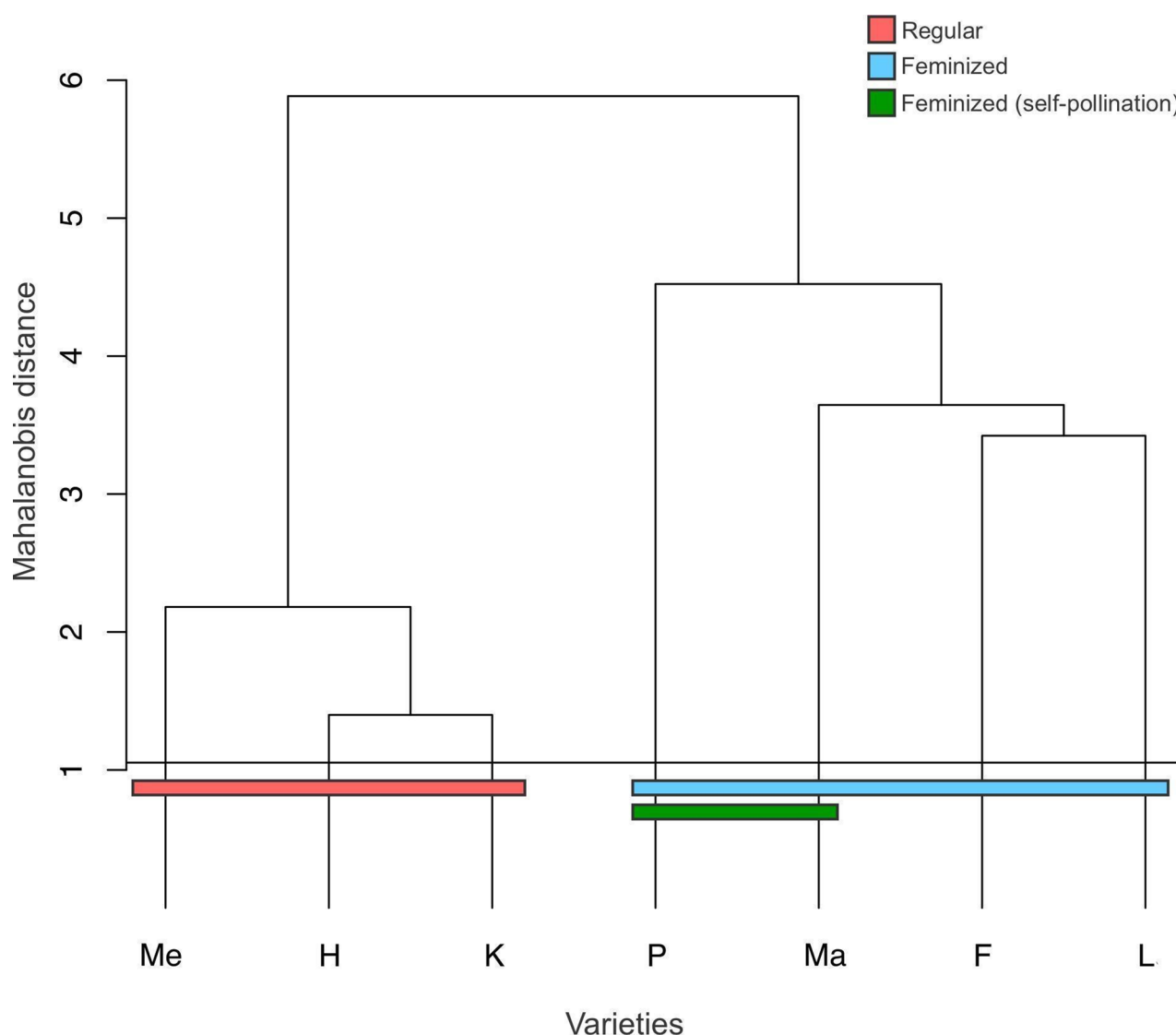


Fig. 4 Hierarchical clustering dendrogram using Ward's minimum variance method. Varieties: Medicinal (Me), H, K, Pachamama (P), Malvina (Ma), Fruti (F) and Lemon (L). The cut-off criterion ($P=0.05$) obtained with the MDGC test is represented by a horizontal line. Horizontal green bar: feminized by self-pollination; Red: regular seeds; Blue: feminized by crossing different female parents

to note that our study directly quantifies morphological uniformity in seed shape, a phenotypic trait. While this reduced variance is consistent with the expected effects of inbreeding, caution is advised against directly equating it with genetic uniformity. Therefore, further research integrating genotyping with morphometrics is needed to establish the relationship between phenotypic and genetic uniformity.

The accurate identification of feminized, self-pollinated seeds suggests that this consistent morphological signature could form the basis for practical applications, where 2D geometric morphometrics is emerging as a potentially cost-effective tool for seed discrimination and traceability.

Conclusions

Our research highlights the potential of 2D geometric morphometrics for the discrimination and traceability of feminized and self-pollinated cannabis seeds. The fact that feminized and self-pollinated cannabis seeds exhibit less morphological variation and that these seeds were very accurately identified makes it possible for this consistent morphological signature to lay the groundwork for future practical applications. The high performance of the Random Forest classifier further supports the reliability of morphological traits for seed discrimination, and suggests that more advanced machine-learning approaches could build upon this foundation to enhance scalability and automation. A logical extension would be

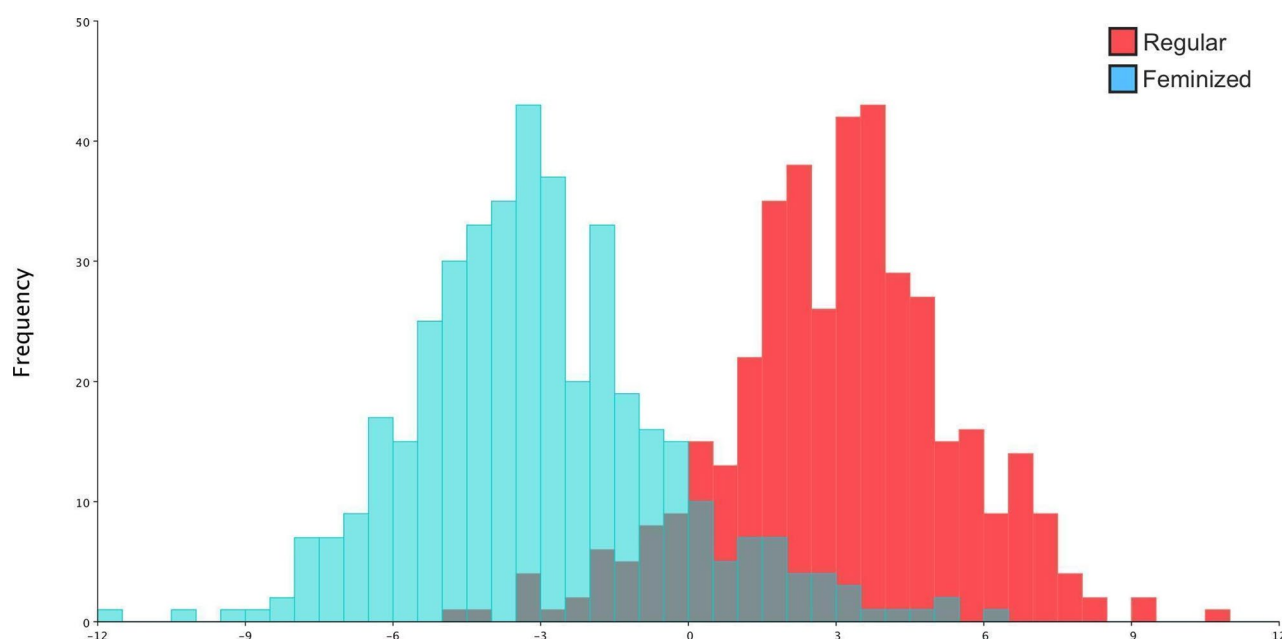


Fig. 5 Discriminant analysis of the seed shape differences between regular and feminized seeds. Frequencies of the discriminant scores predicted by jackknife cross-validation (leave-one-out) of seed shape are shown using histogram bars

to develop deep learning models, such as Convolutional Neural Networks (CNNs), to create a fully automated, end-to-end system capable of classification directly from raw seed images, enhancing the scalability of this method for high-throughput applications. This approach shows significant promise for standardization in seed quality control, although its current limitations, including the number of varieties assessed and the focus on 2D morphology, must be considered for future application and validation.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-025-07989-3>.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.
Supplementary Material 7.

Acknowledgements

We thank CCT CONICET-CENPAT, IBIOMAR, and IPEEC for institutional support, and acknowledge all collaborators of the Interdisciplinary Cannabis Program at CCT CONICET-CENPAT. This research formed part of FFT's doctoral thesis at Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB). FFT is supported by a research grant from CONICET and Secretaría de Ciencia y Tecnología de la Provincia del Chubut.

Authors' contributions

FFT: Conceptualization, data curation, formal analysis, writing original draft, review & editing; YLI: Conceptualization, resources, interpretation of results, review & editing; GB: Sample preparation, resources, writing, review & editing; PN: Formal analysis, writing, review & editing; RG-J: Writing, review & editing; FM: Conceptualization, formal analysis, supervision, resources, writing, review & editing.

Funding

This research was conducted under the Interdisciplinary Cannabis Program (<https://cenpat.conicet.gov.ar/cannabismedicinal/>) with partial support from: Convenio aprobado por Res. 1170/2019. IF-2019-36298756-APN-GVT#CONICET2, Nodos-Chubut denominado "Consolidación de la provincia del Chubut como nodo científico y de transferencia de tecnología asociada al Cannabis: plataforma integral de servicios para el desarrollo de su cadena productiva".

Data availability

To ensure full reproducibility, the Procrustes coordinates and the custom Python code used for the machine learning classification are provided in a public GitHub repository: <https://github.com/Francisco-ft/Self-pollinated-cannabis-seeds-lead-to-less-variation-in-shape>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing financial interests or personal relationships that could influence this work.

Author details

¹Instituto de Biología de Organismos Marinos (IBIOMAR, CONICET), Boulevard Brown 2915, Chubut Puerto Madryn U9120ACF, Argentina

²Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB), Blvd. Brown 3100, Puerto Madryn, Argentina

³Instituto Patagónico para el Estudio de Ecosistemas Continentales (IPEEC, CONICET), Boulevard Brown 2915, Puerto Madryn, Chubut U9120ACF, Argentina

⁴Universidad Espíritu Santo, Samborombón, Ecuador

⁵Instituto Patagónico de Ciencias Sociales y Humanas (IPCSH, CONICET), Boulevard Brown 2915, Puerto Madryn, Chubut U9120ACF, Argentina

Received: 24 September 2025 / Accepted: 16 December 2025

Published online: 20 December 2025

References

- McPartland JM, Guy GW. Models of cannabis taxonomy, cultural bias, and conflicts between scientific and vernacular names. *Bot Rev*. 2017;83:327–81.
- Hillig KW. Genetic evidence for speciation in cannabis (Cannabaceae). *Genet Resour Crop Evol*. 2005;52:161–80.
- Barcaccia G, Palumbo F, Scariolo F, Vannozzi A, Borin M, Bona S. Potentials and challenges of genomics for breeding cannabis cultivars. *Front Plant Sci*. 2020;11:573299.
- Birchler JA, Yao H, Chudalayandi S. Unraveling the genetic basis of hybrid vigor. *PNAS*. 2006;103(35):12957–8.
- Charlesworth D, Willis JH. The genetics of inbreeding depression. *Nat Rev Genet*. 2009;10(11):783–96.
- Keller LF, Waller DM. Inbreeding effects in wild populations. *Trends Ecol Evol*. 2002;17(5):230–41.
- Barrera-Irigoyen CA, Peña-Lomelí A, Magaña-Lira N, Sahagún-Castellanos J, Pérez-Grajales M. Estudio de La Endogamia En tomate de cáscara (*Physalis ixocarpa* Brot. Ex Horn). *Rev Chapingo Ser Hortic*. 2021;27(3):185–98.
- Rauf S, da Silva JT, Khan AA, Naveed A. Consequences of plant breeding on genetic diversity. *Int J Plant Breed*. 2010;4(1):1–21.
- Spani F, Locato V, De Gara L. Unveiling nature's architecture: geometric morphometrics as an analytical tool in plant biology. *Plants*. 2025;14(5):808.
- Zelditch ML, Moscarella RA, Pigliucci M, Preston K. (2004). Form, function, and life history. *Phenotypic Integration: Stud Ecol Evol Complex Phenotypes*, 274–301.
- Klein LL, Svoboda HT. Comprehensive methods for leaf geometric morphometric analyses. *Bio-protocol*. 2017;7(9):e2269–2269.
- Savriama Y. A step-by-step guide for geometric morphometrics of floral symmetry. *Front Plant Sci*. 2018;9:1433.
- Terral JF, Newton C, Ivorra S, Gros-Balthazard M, de Moraes CT, Picq S, Pintaud JC. Insights into the historical biogeography of the date palm (*Phoenix dactylifera* L.) using geometric morphometry of modern and ancient seeds. *J Biogeogr*. 2012;39(5):929–41.
- Hodač L, Karbstein K, Tomasello S, Wäldchen J, Bradican JP, Hörandl E. Geometric morphometric versus genomic patterns in a large polyploid plant species complex. *Biology*. 2023;12(3):418.
- Chitwood DH, Otoni WC. Morphometric analysis of passiflora leaves: the relationship between landmarks of the vasculature and elliptical fourier descriptors of the blade. *GigaScience*. 2017;6(1):giw008.
- Luo T, Zhao J, Gu Y, Zhang S, Qiao X, Tian W, Han Y. Classification of weed seeds based on visual images and deep learning. *IPA*. 2023;10(1):40–51.
- Ghaffari A. Precision seed certification through machine learning. *Technol Agron*. 2024;4:e019.
- Ceyhan M, Kartal Y, Özkan K, Seke E. Classification of wheat varieties with image-based deep learning. *Multim Tools Appl*. 2024;83(4):9597–619.
- Vanhees DJ, Vanderbeken M, Verlinden BE, Nicolai B. (2023). Tomato (*Solanum lycopersicum*) shape classification with deep learning AI-algorithms. In VII International Conference Postharvest Unlimited 1396 (pp. 185–192). <https://doi.org/10.17660/ActaHortic.2024.1396.25>
- Cinar I, Koklu M. (2022). Identification of rice varieties using machine learning algorithms. *J Agric Sci*, 9–9.
- Márquez F, Lozada M, Idaszkin YL, González-José R, Bigatti G. Cannabis varieties can be distinguished by achene shape using geometric morphometrics. *Cannabis Cannabinoid Res*. 2022;7(4):409–14.
- Fernandez Torre F, Idaszkin YL, Bigatti G, Garcés N, Lozada M, Gonzalez-Jose R, Marquez F. La forma de Los aquenios de *Cannabis sativa* L. para La diferenciación de cultivares medicinales comerciales de Argentina. *Revista Cannabis Y Salud*. 2023;2:50–7.
- Hesami M, Pepe M, Jones AMP. Morphological characterization of cannabis sativa L. throughout its complete life cycle. *Plants*. 2023;12(20):3646.
- Babaei M, Nemati H, Arouiee H, Torkamaneh D. Characterization of Indigenus populations of cannabis in iran: a morphological and phenological study. *BMC Plant Biol*. 2024;24(1):151.
- Balant M, Garnatje T, Vitales D, Hidalgo O, Chitwood DH. Intra-leaf modeling of cannabis leaflet shape produces leaf models that predict genetic and developmental identities. *New Phytol*. 2024;243(2):781–96.
- Reed DH, Fox CW, Enders LS, Kristensen TN. Inbreeding–stress interactions: evolutionary and conservation consequences. *Ann NY Acad Sci*. 2012;1256(1):33–48.
- Hall J, Bhattacharai SP, Midmore DJ. Review of flowering control in industrial hemp. *J Nat Fibers*. 2012;9(1):23–36.
- Bookstein FL. Morphometric tools for landmark data: geometry and biology. Cambridge: Cambridge University Press; 1992.
- Zelditch M, Swiderski D, Sheets HD. Geometric morphometrics for biologists: a primer. Academic; 2012.
- Rohlf F. TpsUtility program version 2.30. Department of ecology and evolution. State University of New York, Stony Brook; 2017a.
- Rohlf F. TpsDig2 program version 2.30. Department of ecology and evolution. State University of New York, Stony Brook; 2017b.
- Rohlf F. TpsRw1 program version 1.67. Department of ecology and evolution. State University of New York, Stony Brook; 2017c.
- Herberich E, Sikorski J, Hothorn T. (2010). A robust procedure for comparing multiple means under heteroscedasticity in unbalanced designs. *PLoS ONE*, 5(3), e9788.
- Dinno A. Nonparametric pairwise multiple comparisons in independent groups using dunn's test. *Stata J*. 2015;15(1):292–300.
- Klingenberg CP. MorphoJ: an integrated software package for geometric morphometrics. *Mol Ecol Resour*. 2011;11(2):353–7.
- Valdano SG, Di Rienzo J. Discovering meaningful groups in hierarchical cluster analysis. An extension to the multivariate case of a multiple comparison method based on cluster analysis. *InterStat*. 2007;4:1–28.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Duchesnay É. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825–30.
- Gegas VC, Nazari A, Griffiths S, Simmonds J, Fish L, Orford S, Snape JW. A genetic framework for grain size and shape variation in wheat. *Plant Cell*. 2010;22(4):1046–56.
- Gueye T, Goldberg V. Seed morphology as a tool for quantifying genetic diversity in rice (*Oryza sativa* L.) landraces. *PI Syst Evol*. 2021;307(3):29.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.