



HAL
open science

Exploring the carbonization effect on the interspecific identification of cotton (*Gossypium* spp.) seeds using classical and 2D geometric morphometrics

Juliette Milon, Charlène Bouchaud, Christopher Viot, Michel Lemoine,
Thomas Cucchi

► To cite this version:

Juliette Milon, Charlène Bouchaud, Christopher Viot, Michel Lemoine, Thomas Cucchi. Exploring the carbonization effect on the interspecific identification of cotton (*Gossypium* spp.) seeds using classical and 2D geometric morphometrics. *Journal of Archaeological Science: Reports*, 2023, 49, pp.104007. 10.1016/j.jasrep.2023.104007. hal-04102316

HAL Id: hal-04102316

<https://hal.science/hal-04102316v1>

Submitted on 22 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PRE-PRINT

Exploring the carbonization effect on the interspecific identification of cotton (*Gossypium* spp.) seeds using classical and 2D geometric morphometrics

Juliette Milon^(a), Charlène Bouchaud^(a), Christopher Viot^(b), Michel Lemoine^(a) & Thomas Cucchi^(a)

^(a) Archaeozoology, Archaeobotany, Societies, Practices, Environments (AASPE-UMR7209), CNRS, National Museum of Natural History (MNHN), Paris, France

^(b) UMR AGAP Institut, CIRAD, Univ Montpellier, INRAE, Institut Agro, F-34398 Montpellier, France

Corresponding author: Juliette Milon, juliette.milon@edu.mnhn.fr / juliette-milon@hotmail.fr

Keywords: cotton seeds, traditional morphometrics, geometric morphometrics, experimental carbonization, interspecific identification

Abstract

The history of cotton (*Gossypium* sp), which has become along past millennia a plant of major economic interest, relies on the dried or charred cotton seeds found in archaeobotanical assemblages. Our capacity to document the dispersal routes of cotton, however, is hampered by the lack of taxonomic resolution at specific levels based on current seed anatomic features. In order to explore the taxonomic signal at the interspecific level of cotton seeds, and the impact of carbonization on this signal, we used 2D geometric morphometrics (GM) comparing modern dried cotton seeds of the four domesticates (*G. arboreum*, *G. herbaceum*, *G. barbadense*, *G. hirsutum*) and experimentally charred seeds. Linear measurements were also considered (length, width and thickness). The shape and form (size + shape) of dried and charred cotton seeds were captured using an outline analysis with a sliding semi-landmarks approach and compared after a Generalized Procrustes Superimposition (GPS). Under the chosen experimental charring protocol, the reaction of cotton seeds to charring was globally characterized by a size reduction but no major shape deformation; however, this study also shows that the outline form of the seeds rather than their shape provides more stringent taxonomic discriminations than linear measurements. This pilot study emphasizes the relevance of the GM approach while underlying the limit of this method for the taxonomical study of archaeological cotton seeds.

1. Introduction

Cotton (*Gossypium* sp.) is the world's leading source of plant textile fiber; however, the history leading to its global economic success remains poorly understood. Cotton belongs to the Malvaceae family and the genus *Gossypium*, which comprises approximately 48 wild species growing at tropical and sub-tropical latitudes and four domesticated species divided into two phylogenetic groups: two diploid species coming from the Old World, *G. herbaceum* L. and *G. arboreum* L., and two tetraploid species coming from the New World *G. hirsutum* L. and *G. barbadense* L. (Kulkarni et al., 2009; Page et al., 2013; Wendel et al., 2009; Wendel and Grover, 2015). Currently, the most widely cultivated cotton is *G. hirsutum*, a highly productive species native to Middle America ou Mesomaerica and potentially domesticated around 5000–4000 years ago (Brubaker et al., 1999; Smith and Stephens, 1971). A second tetraploid American species, *G. barbadense*, already used around 7000 years ago in South America (Peru), is known as "Pima" or "Egyptian" cotton and is characterized by its extra-long fibers (Dillehay et al., 2007; Stephens, 1975; Stephens and Moseley, 1974). Two diploid species were domesticated in the Old World, *G. arboreum* on the Indian subcontinent, hypothetically between 8000 and 6000 BP (Moullhérat et al., 2002), and *G. herbaceum* in Africa, at a time and place still under debate (Bouchaud et al., 2018; Chowdhury and Buth, 1971; Palmer et al., 2012; Viot, 2019).

Cotton seeds found in archaeobotanical assemblages are almost always dried or charred (e.g. Bouchaud et al., 2018; Clapham and Rowley-Conwy, 2007; Fuller, 2008; Stephens and Moseley, 1974), and their identification classically relies on anatomic features (Fuller, 2008). Unfortunately, our capacity to document the dispersal routes of cotton is hampered by the lack of taxonomic resolution at specific levels of the seeds' anatomic features, despite many attempts being made for the distinct assignment of cotton seeds found in archaeological contexts. One study (Chowdhury and Buth, 1971) compared the modern and archaeological measurements of seed coat traits (number of colored cells, height of palisade cells). Despite a low number of measurements, the study concludes that the archaeological seed remains, found in goat pellets at a Nubian site, Afyeh, and dated to 2600–2400 BC (uncalibrated), show the greatest similarity with *G. arboreum* and *G. herbaceum* (Chowdhury and Buth, 1971). In fact, to date, only one archaeogenomic study has clearly assigned the species of the cotton seeds, *G. herbaceum*, to the well-preserved desiccated cotton. (Palmer et al., 2012). Nevertheless, as archaeological cotton seeds are generally preserved by carbonization (Bouchaud et al., 2019), preventing the conservation of long-chain biomolecules used in paleogenetic studies, morphometry is then the commonly used tool for the reconstruction of the history of cotton dispersal and domestication from archaeological seeds.

Charring is the commonest form of plant material preservation on archaeological sites. It involves heating, but not combustion, and occurs as a consequence of warming food, other technical uses of fire or accidents. Charred seeds are less susceptible to degradation and microbial attacks and to rodent and insect gnawing or bird eating, as they have little or no food value (Miller, 1989) At temperatures of around 200 to 300°C, water and volatile organic compounds (syngas) are removed from the matrix. Dramatic morphological alterations are usually associated with the carbonization of archaeobotanical remains (Bonhomme et al., 2017; Bouby, 2010). One published study (Stephens and Moseley, 1974) that includes the measurements of cotton seeds before and after carbonization indicates a 3 to 7% reduction in seed size but no change in shape for one species only (*G. barbadense*), but there are no details of the morphometric data. Therefore, in order to explore the taxonomic signal at the interspecific level of cotton seeds and the impact of carbonization on this signal we used 2D geometric morphometrics (GM) comparing modern fresh cotton seeds and experimentally charred seeds following previous studies (Bonhomme et al., 2017; Bouby et al., 2018, 2013; Jesus et al., 2021; Ros et al., 2014; Tarongi et al., 2021; Terral et al., 2021, 2010). In the present study, we report on the evaluation of GM for the specific assignment of cotton seeds before and after experimental carbonization, on the deformation induced by the carbonization and its impact on the taxonomic signal.

2. Materials and methods

2.1. Cotton accessions and sample sizes

A total of 386 seeds were analyzed, representing cultivars of each of the four domesticated species of cotton *G. hirsutum*, *G. barbadense*, *G. herbaceum* and *G. arboreum* (Fig. 1). Most of the seeds came from irrigated cotton plants grown in greenhouses in 2016 and 2017 at CIRAD (French Agricultural Research Centre for International Development) Montpellier, France; three cultivars by species were studied (Tab. 1). In addition, one cotton sample was obtained in 2017 from the company Algodonera del Sur, that had been grown in an irrigated cotton field at Las Chapatales, near Sevilla, Spain. For each of the 13 samples, 27 to 30 seeds were randomly selected (Tab. 1). All residual fibers (mainly fuzz) around the seeds were removed by hand to ease contouring and measurement. The sample 0014_Hirs was used as a test sample to establish the temperature and duration of the carbonization experiments that gave the most similar results to archaeological charred seeds.

Tab. 1.

Details of the accessions and sample sizes of the study.

Taxa	Sample reference number	Seed source	Sample size	Germplasm origin	Name
<i>Gossypium arboreum</i>	0001_Arb_BG1557	Montpellier, CIRAD	30	India	B 10
	0012_Arb_BG7502	Montpellier, CIRAD	30	Cameroon	SCHW 1289 (plant growing wild)
	0011_Arb_BG7408	Montpellier, CIRAD	30	Cameroon	AR0001 (plant growing wild)
<i>Gossypium herbaceum</i>	0008_Herb_BG7510	Montpellier, CIRAD	27	Cameroon	SS 1314 (plant growing wild)
	0003_Herb_BG1515	Montpellier, CIRAD	29	Iran	Boumi, <i>G. herbaceum</i> race <i>persicum</i>
	0004_Herb_BG1524	Montpellier, CIRAD	30	Vietnam	Phu Yen
<i>Gossypium barbadense</i>	0002_Barb_BG1382	Montpellier, CIRAD	30	Egypt	Giza 12 aka Wafeer
	0007_Barb_BG2011	Montpellier, CIRAD	30	USA	C2 IC PNS 63
	0006_Barb_BG1406	Montpellier, CIRAD	30	USA	Pima S57-4
<i>Gossypium hirsutum</i>	0010_Hirs_Acala	Spain (Las Chapatales)	30	USA	DP Acala 90
	0013_Hirs_BG0023	Montpellier, CIRAD	30	Syria	Aleppo 1
	0005_Hirs_BG2166	Montpellier, CIRAD	30	Paraguay	IAN 425
	0014_Hirs_BG2141 (charring test)	Montpellier, CIRAD	30	Ivory Coast	R405.5




2.2. Experimental carbonizations

Our carbonization experiments intended to create artificially charred seeds with an appearance similar to that of the archaeological specimens, without reaching the threshold of major deformations. The carbonization tests were realized in a muffle furnace oven (Nabertherm GmbH, Archaeobotany Laboratory, MNHN). Tests carried out in an oxidizing atmosphere resulted in deformations or destructions too severe for these heating conditions to be retained. The only known publication on the experimental carbonization of cotton seeds used a temperature of 200°C, but full details on the heat up time were not provided (Stephens and Moseley, 1974). We determined the temperature and time reduction conditions (eliminating the supply of atmospheric oxygen) for this study after several tests using one of our 30 cotton seed samples (see Tab. 2) and a review of literature on the experimental carbonization of seeds (Bonhomme et al., 2017; Bouby, 2010; Bouby et al., 2018; Braadbaart, 2007; Braadbaart et al., 2007; Charles et al., 2015; García-Granero et al., 2016; Ros et al., 2014). Each cotton seed was wrapped individually in a 0.02mm thick aluminium foil and covered with a 3cm layer of sand to recreate a reducing atmosphere. At the end of the process, the seeds were immediately removed from the sand and instantly cooled. At 200°C some

of the seeds were roasted but not carbonized. At 300°C for 2 hours the deformations prevented the recording of measurements. Seeds experimentally charred for 2.5 hours at 250°C produced only light deformations that were considered similar to those of charred archaeological seeds, permitting morphometric studies (Tab. 2). All the 356 remaining seeds were carbonized using the same conditions.

Tab. 2.

Photographs of the experimental charred seeds at different temperatures and times – 3 view of *G. hirsutum* seed.

Carbonization temperatures	200°C	250°C	300°C
Charring time	3h	2h30	2h
			

2.3. Seed anatomy and classical morphometrics methodology

All *Gossypium* species produce fruits called capsules or bolls, which open into between three and five rigid locules at maturity. Each locule contains five to nine seeds bearing fibers on their epidermal surface (Lagière, 1966). Cotton seeds have shapes ranging from ovoid to piriform with a pointy hilum. Their length and mass vary from 3 to 12 mm and from 0.07 to 0.17 g, respectively (Lagière, 1966). A portion of the funicle of cotton seeds is fused with the integument, forming a longitudinal ridge, or raphe, just above the hilum. The raphe connects the hilum to the chalaza (Fig. 1). The almond has an embryo with two cotyledons folded back on themselves. These cotyledons have a high protein and fat content. For instance, the oily content of cotton seed can reach 18 to 25% for *G. hirsutum* (Saxena et al., 2011). Pure cottonseed oil has a smoking point, aka burning point, at around 220–230°C, suggesting that cotton seed fatty acids begin to burn at temperatures above 230°C. In natural drying conditions, the percent humidity of cotton seeds is

less than 5 at 10% relative atmospheric humidity and about 18 at 90% relative humidity (Altschul, 1948), which corresponds to prevailing dry and humid tropical environments, respectively. Charring evaporates this humidity, reducing the size of the internal tissues and the seed globally.

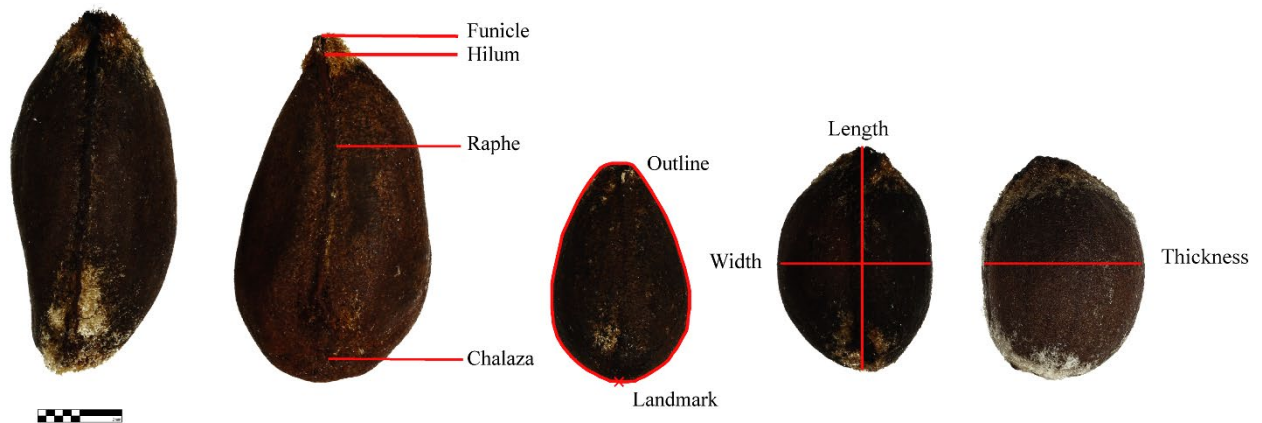


Fig. 1 Photographs of the four domesticated species of cotton (from left to right: *G. hirsutum* ventral view, *G. barbadense* ventral view, *G. arboreum* ventral view, *G. herbaceum* ventral and lateral view). The different anatomical characteristics are indicated on a *G. barbadense* seed. The outline is measured with 1 landmark at the intersection of the seed's contour and the end of the raphe and 50 semi-landmarks on the outline of the ventral view. Traditional morphometry measurements (length, width and thickness) are depicted on a *G. herbaceum* seed's ventral and lateral views.

2.4. Geometric morphometrics

2.4.1. Image acquisition

Each cotton seed was given a unique identification code to facilitate one-to-one comparisons of pre- and post-charring morphologies. Each seed was photographed in ventral and lateral views before and after the charring process (Fig. 1) with two photographic systems: a microscope (Leica Z6 APO DMC 4500, software Leica LAS Core_LAS V4.8) and a handheld digital microscope (Dino-Lite RK-10, software Dino-Capture 2.0). In order to ensure that the two acquisition systems did not introduce any bias in the measurements, and therefore in the results, we performed reproducibility tests. Two samples (60 seeds) of two different species (*G. hirsutum* and *G. herbaceum*) were photographed on the two acquisition systems for the three linear measurement and form variables. When compared, the results did not show any significant differences (MANOVA, Pillai= 0.89, F= 10.79, $p < 0.001$), suggesting that the association of two

photographic systems for image acquisition does not significantly impact the variation in measurements.

2.4.2. Acquisition of linear measurements

The traditional morphometry of cotton seeds relies on three linear measurements: length, width and thickness. The length is measured on the ventral view by drawing a line between the chalaza and the base of the funicle (Fig. 1). The seed width is also measured on the ventral view, along an axis perpendicular to the length and on the part of the seed, often central, where it is the widest. The thickness is measured on the lateral view by drawing a line perpendicular to the raphe, the thickest point of the seed (Fig. 1). The width and thickness are therefore measured at the point in the seed where it is most important. These measurements were made using a stereo microscope coupled with the NIS Elements 3.1 Nikon image acquisition software (UMR 7209 Archaeobotany Platform) and the Image J software.

2.4.3. Acquisition of 2D cartesian coordinates for Geometric Morphometrics

The outline of the seed was quantified using 51 cartesian coordinates along the outline of the ventral view with one landmark at the junction point between the raphe and the chalaza, which is easily recognizable and systematically present on well-preserved archaeological specimens, and 50 equidistant semi-landmarks along the outline (Fig. 1). Landmark and semi-landmarks have been sampled from the outline using TPS dig v.2 (Rohlf, 2005). In order to obtain the shape and centroid size variables of each seed's outline, we performed a Generalized Procrustes Superimposition (GPS) of all 51 cartesian coordinates using a Bending Energy Minimization (BEM) to slide the semi-landmarks. The centroid size is the square root of the sum of square distances between each point of the outline and its center of gravity (centroid). Finally, form variables were obtained by the concatenation of the log of the centroid sizes with the shape variable (Procrustes coordinates) following Mitteroecker and Gunz (2009).

2.4.4. Statistical analyses

The difference in centroid size or linear measurement variations between the domestic species of cotton were tested with ANOVAs and visualized on a violin plot, before and after carbonization, with pairwise comparisons using the Bonferroni correction. To test differences in seed shape and form among domestic taxa before and after carbonization, we used a MANOVA (Pillai's test). MANOVAs were performed on PCA scores on shape or form data. We visualized the effect of carbonization over the seed shape globally and for each taxon using the deformation vectors between seed mean shapes before and after carbonization. In order to predict the correct

assignment of cotton taxa before and after carbonization and compare the taxonomic signal between the linear measurements, shape and form variables, we used a Linear Discriminant Model (LDA) with a leave-one-out cross-validation. To reduce the variables dimensionality and avoid spurious results from the LDA (Mitteroecker and Bookstein, 2011) we performed the LDA on the Principal Component Analysis (PCA) scores performed on shape and form datasets representing 95% of variance. Comparisons of taxonomic resolutions were first made between the classifications after cross-validation of the diploid species and tetraploid species independently, and then among global distinctions on the four domesticated taxa. The comparisons of correct classification rates were performed for each data set on the three linear measures, on shape and form (shape and centroid sizes).

GPS and statistical analyses were performed under the R software using Geomorph and MASS libraries (Adams and Ota, 2013). For traditional morphometry analysis, the packages ggplot2, forcats, stats and MASS were used (R Core Team, 2022; Venables and Ripley, 2002; Wickham, 2021, 2016). For geometric morphometry analysis on shape and form, the R packages used were geomorph, ggplot2, rgl, MASS, scales, ggfortify, tidyverse, dplyr, gridExtra, gridGraphics, grid, grepl and ggpubr with the functions arrayspecs/aov/pairwise.t.test/lda/gpagen/two.d.array/prcomp/manova/mshape. (Adams et al., 2022; Adams and Ota, 2013; Auguie, 2017; Baken et al., 2021; Horikoshi and Tang, 2016; Kassambara, n.d.; Murdoch and Adler, 2022; Murrell and Wen, 2020; Venables and Ripley, 2002; Wickham, 2016; Wickham et al., 2022, 2019; Wickham and Seidel, 2022)

3. Results

3.1. Variation in linear measurements and centroid sizes of uncharred and charred cotton seeds

The length of the diploid cotton species (African and Asian origins, *G. herbaceum* and *G. arboreum*) seeds were significantly shorter than the tetraploid cottons (American origin, *G. hirsutum* and *G. barbadense*, Fig. 2a). The length for *G. arboreum* and *G. herbaceum* ranged from 5.19 mm to 7.44 mm with a mean of 6.56 mm, while the length for *G. barbadense* and *G. hirsutum* ranged from 7.59 mm to 10.79 mm with a mean of 9 mm. Sizes overlapped within each of the subgroups (diploid and tetraploid) making species discrimination difficult (see below). For each species, the length decreased after carbonization (Fig. 2a). *G. barbadense* exhibited the smallest size reduction rate (3.66%), while *G. arboreum* underwent the greatest (8.42%). *G. herbaceum* and *G. hirsutum* evidenced similar ratios of length reduction (6.96% and 6.47%, respectively). Width and thickness were also distributed into two groups (diploid and tetraploid)

but their size variations before and after carbonization were not statistically different (Fig. 2b/ Fig. 2c). Centroid sizes followed the same general trend, forming two groups (diploid and tetraploid). The length reduction rates induced by the charring process were all statistically significant and ranged from 2.95 to 5.26% (log centroid size significantly differed between all species before carbonization: $F = 883.5$, $p < 0.0001$, Fig. 2d).

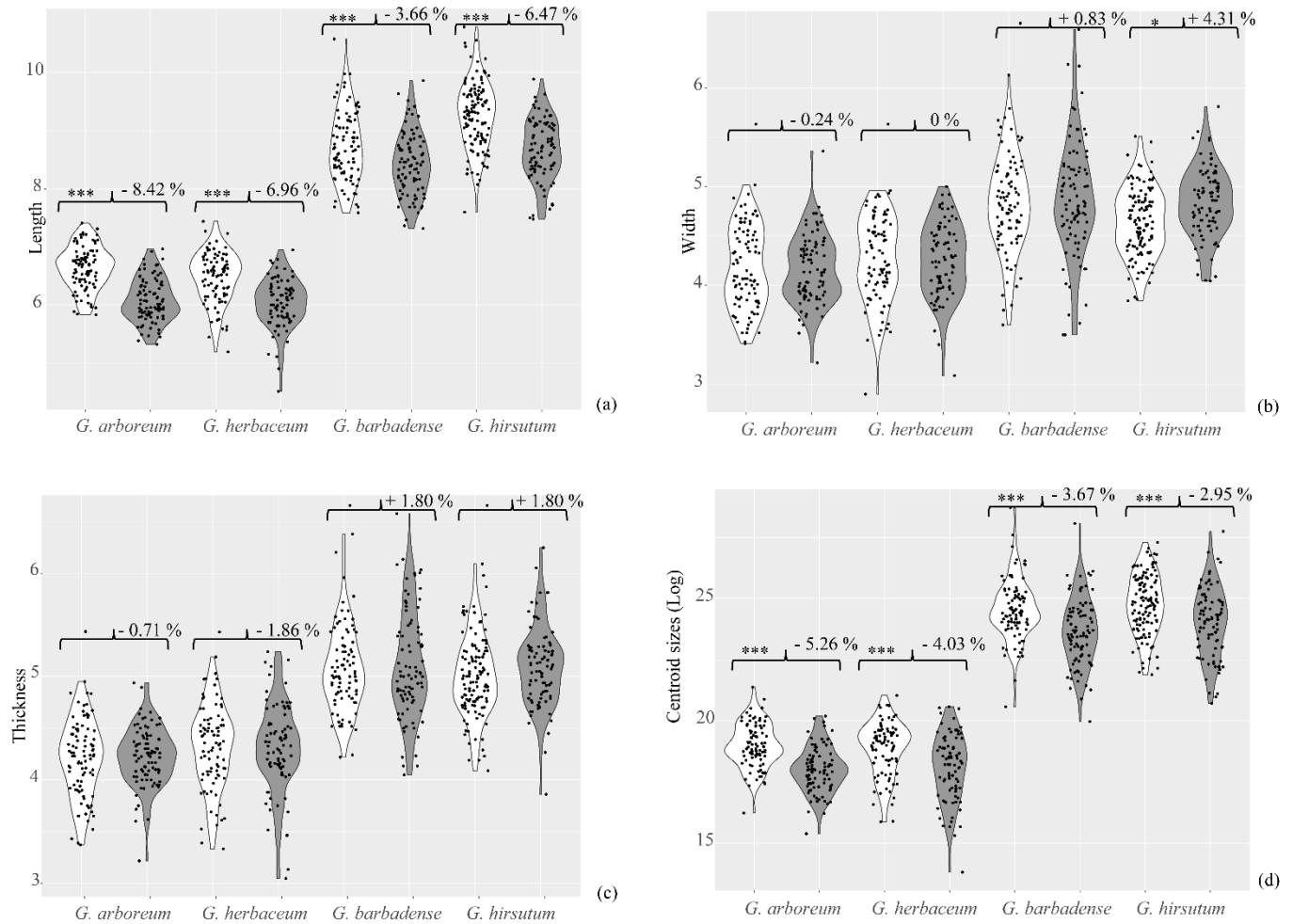


Fig. 2. Violin plots of ventral view length (a), width (b), lateral thickness (c) and log centroid sizes (d) variations for each domesticated cottons species before (white) and after (grey) charring, with the percentage of mean difference (pairwise comparisons were performed using ANOVA) before and after carbonization and the pairwise comparison symbolized by stars (***: p-value < 0.001, **: p-value < 0.01, *: p-value < 0.05, • : p-value < 0.1).

3.2. Shape deformation of the cotton seeds during carbonization

Deformation patterns linked to carbonization displayed varying responses to charring (Fig. 3); however, no significant differences in shape, before or after carbonization, were observed for any species (Pairwise MANOVA test, $p > 0.05$). Among the carbonized seeds, the only significant difference was between *G. arboreum* and *G. barbadense* ($p < 0.05$, $P = 0.016$), while no significant differences appeared in the other pairwise species comparisons. A dilation of the width was observed for *G. barbadense* and *G. herbaceum*, and a decrease in width was noted for *G. arboreum*. *G. hirsutum* displayed almost no shape deformation (Fig. 3).

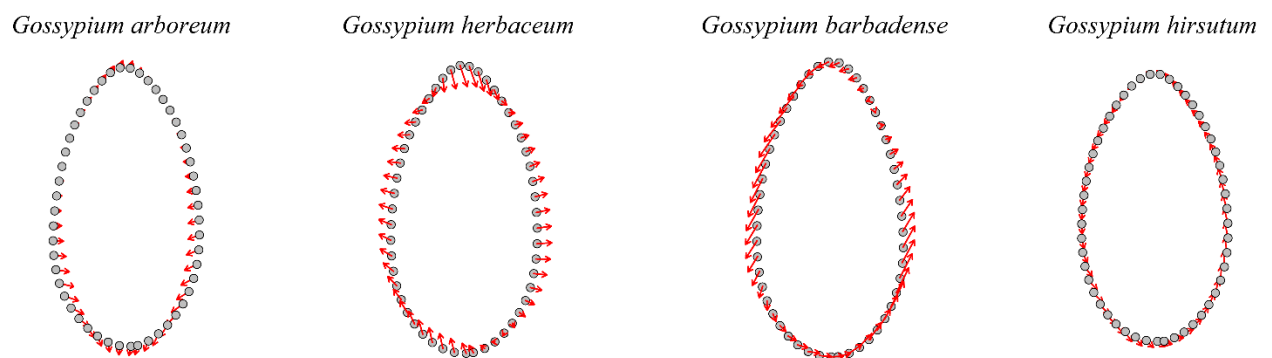


Fig. 3. Shape deformation due to carbonization for each domestic species using vectors to illustrate the direction of deformation.

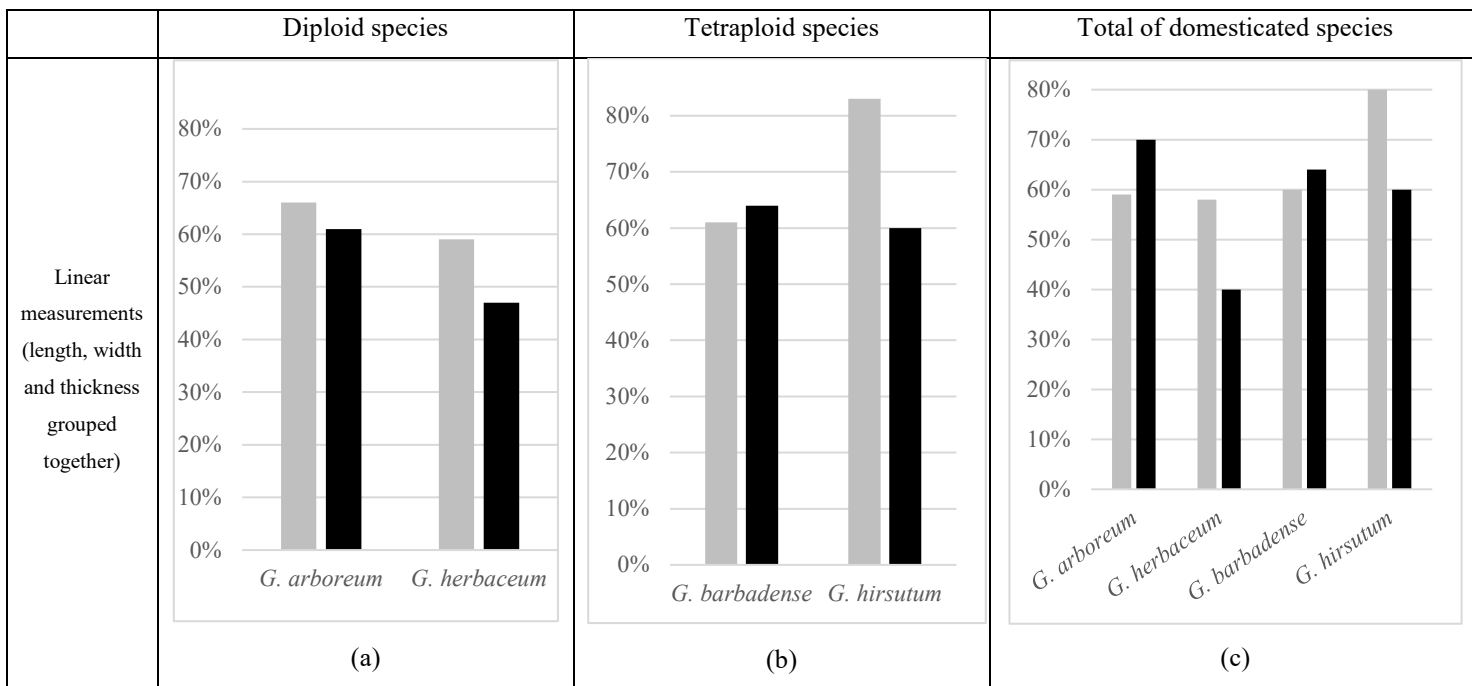
3.3. Comparing taxonomic signals of cotton seeds from linear measurement, shape and form variables

The rates of correct classification after cross-validation on the linear measurements (length, width and thickness) of uncharred seeds range from 58 to 80%, if we consider the four species (Fig. 4g). The rates slightly increase if *G. arboreum* and *G. herbaceum* and *G. barbadense* and *G. hirsutum* are analyzed separately (Fig. 4a, 4d). Carbonization tends to reduce the rate of correct classification with the exception of *G. barbadense* (Fig. 4b) and *G. arboreum* (Fig. 4d), if all species are considered together (Fig. 4g). There are significant differences between the cotton species before and after carbonization in terms of linear measurements.

None of the differences between the species before and after carbonization are significant on shape data. Despite this, the percentages of correct species assignment after cross-validation for the shape data were higher than the rates recorded through linear measurements. For uncharred seeds, the percentages range from 62 to 84% if all species are considered (Fig. 4h). This increases considerably if diploid and tetraploid species are analyzed separately, except for *G. barbadense*

(Fig. 4b, e). Charring did not have much impact on these rates. This result is even more striking when diploid cottons are observed (the rate of correct classification remains the same before and after carbonization for *G. arboreum* (Fig. 4b)).

The rates of correct classification after cross-validation on the form data of uncharred seeds ranges from 43 to 76%, if we consider the four species (Fig. 4i). The rates increase considerably if diploid and tetraploid species are analyzed separately, with a range from 61% (*G. barbadense*) to more than 80% for *G. arboreum* and *G. hirsutum* (Fig. 4c, 4f). Charred seeds had the lowest rates of correct classification. The differences in correct classification are also significant between *G. arboreum* and *G. herbaceum* before carbonization ($p=1.542E-14$) and after carbonization ($p=5.596E-11$) (Fig. 4c). For the diploid cottons, carbonization had an impact on the species assignment after cross-validation. While the differences are significant between *G. hirsutum* and *G. barbadense* before carbonization ($p=2.021E-09$), those differences are not significant after carbonization ($p=0.09988$) (Fig. 4f).



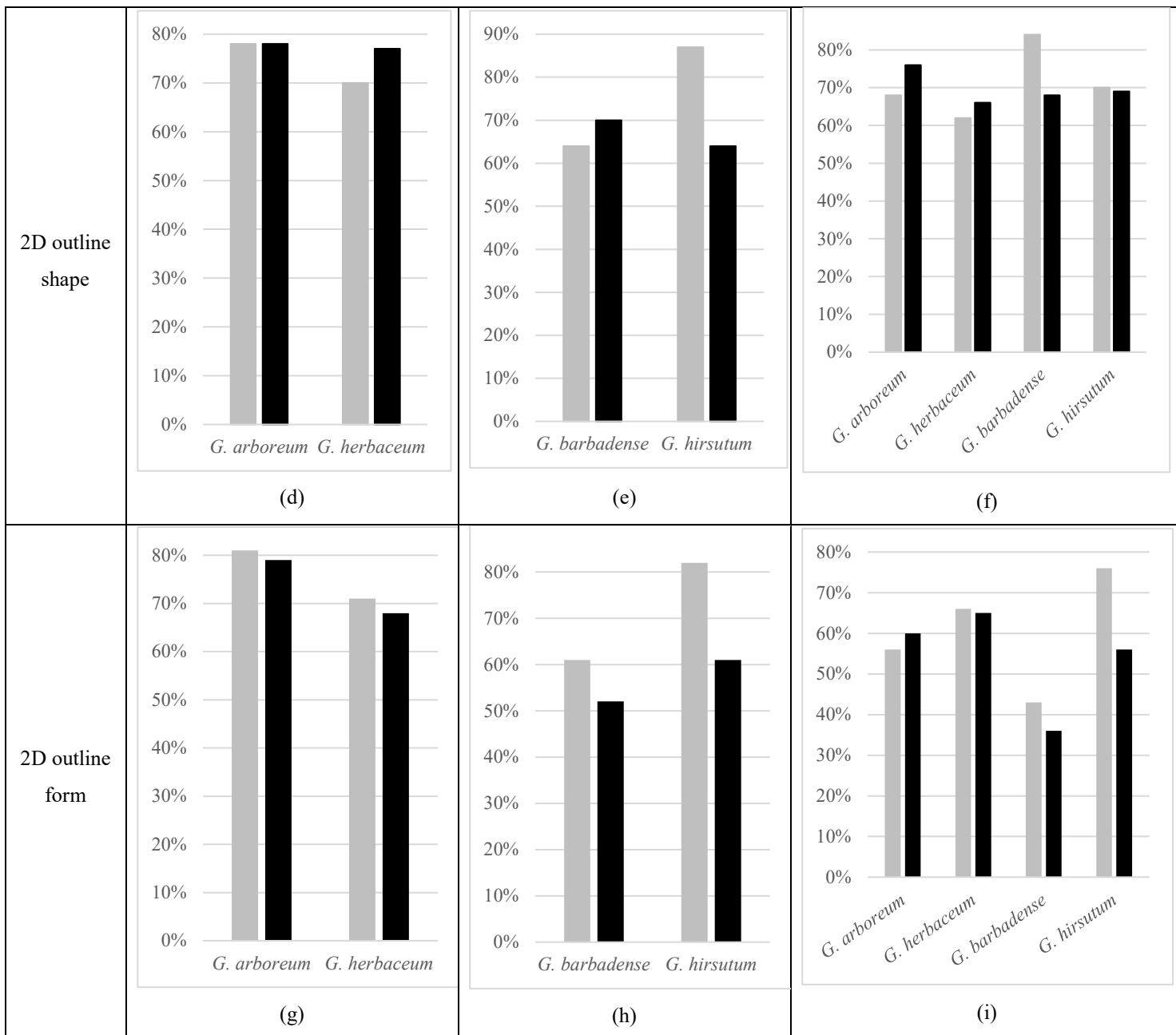


Fig. 4. Graphs illustrating the correct classification rates after cross-validation in traditional morphometry (a, d, g) and geometrical morphometry for shape (b, e, h) and form (centroid sizes and shape) (c, f, i), by diploid species (a, d, g), tetraploid species (b, e, h), and all four domesticated species (c, f, i) before (grey) and after (black) carbonization.

4. Discussion

This pilot morphometric study conducted on fresh and experimentally charred modern seeds offers unprecedented data on carbonization impact and taxonomical discrimination between the four domesticated cotton species. It provides a solid scientific basis to better interpret archaeological and ethnobotanical assemblages.

First, simple measurements of length allowed us to separate the diploid Asian and African cultivated species (*G. herbaceum*, *G. arboreum*,) from the American tetraploid (*G. barbadense*, *G. hirsutum*) cultivated species. Carbonization under our experimental protocol significantly affected the size of the seeds with a mean reduction of 3.66 to 8.42% without preventing the characterization of these two subgroups. Note that our reduction size ratios match with the only other morphometric study conducted on *G. barbadense* charred seeds showing a 3 to 7% reduction in seed size (Stephens and Moseley, 1974). While these data are of little interest for ancient times, when the species of the Old and New World evolved independently, they may be important for distinguishing more recent seeds, after American cottons were introduced into Africa for instance (Seignobos, 2019).

Secondly, the reaction of cotton seeds to charring under our testing protocol is globally characterized by a size reduction but no major shape deformation. However, this reaction is highly variable from one taxon to another, whether we are looking at length reduction, shape or form. Furthermore, the rates of correct classification before and after carbonization follow different trajectories for each species. This heterogeneity of reaction to carbonization has been already reported for other plant models, between species and varieties and different charring conditions (Bouby et al., 2018; Ros et al., 2014). In our case, under homogeneous charring conditions, we can hypothesize, following Ros et al., 2014, that these differentiated responses to the carbonization process are due to grain composition. Indeed, the variation in oil content of the seeds between cotton species (Saxena et al., 2011) may explain the different responses to charring.

Thirdly, the major problem highlighted by this study concerns the discrimination between *G. arboreum* and *G. herbaceum* on the one hand and between *G. hirsutum* and *G. barbadense* on the other. Shape analysis did not allow us to significantly separate the species, while linear measurements and form analysis showed significant differences before and after carbonization. Form analyses gave higher rates of species classification than traditional distance measurements and a better resolution (up to 82% on dried seeds), when diploid and tetraploid groups were considered independently rather than being analyzed together. Geometric morphometry requires a greater investment in time and more complex analysis methods but also provides a better percentage of correct classification on form, especially for diploid species before and after carbonization. Nevertheless, the outline form showed a large overlap between genetically close species, making the correct species assignment of cotton seeds highly complex. These overlaps are possibly related to the numerous hybridizations that have taken place within the genus *Gossypium* in recent decades (Hutchinson et al., 1947; Stephens, 1949), though this should be less of an issue

for ancient periods. There is therefore a bias in the comparative approach due to the interpretation of the data through the projection of modern cultivars.

This result must be strongly nuanced, however, due to the small size of the corpus. Increasing the corpus would maximize the differences between both methods because geometric morphometry can provide more reliable and accurate results. Furthermore, increasing the amount of information, such as the number of seed views – two (ventral and lateral) as in Jesus et al., 2021; Pagnoux et al., 2015; Tarongi et al., 2021; Terral et al., 2021, 2012 or three (ventral, lateral and proximal) as in Bonhomme et al., 2017 – would refine identification and increase species determination. These preliminary results permit to consider an application to a larger corpus. The extension of the corpus should be based on the use of the form on the combination of the ventral and lateral views.

Applying this to both charred and archaeological specimens would make it possible to identify archaeological seeds at species level, enabling a better description of their domestication history and more precise mapping of the cotton trade routes. The results of this study also allow us to consider the application of these techniques on a larger corpus of archaeological seeds, as these are often found in a charred state. Our study therefore appears as the first step of a larger study.

Acknowledgements

We warmly thank Elena Peláez Andérica and the company Semillas Sostenibles Ibérica S.L.U., grupo Algodonera del Sur S.A for providing us with cotton plants, and the Centre de Ressources Biologiques Tropicales (CRB-T) of CIRAD (Montpellier, France) for cotton seeds from its collections. We would also like to thank the morphometric platform of the Museum national d’Histoire Naturelle (Paris, France) and Amandine Blin (Museum national d’Histoire Naturelle - UMS 2700 2AD) for her help in the image acquisition of cotton seeds.

Bibliography

- Adams, D.C., Collyer, M.L., Kaliontzopoulou, A., Baken, E.K., 2022. Geomorph: Software for geometric morphometric analyses.
- Adams, D.C., Ota, E., 2013. Geomorph: an r package for the collection and analysis of geometric morphometric shape data. *Methods Ecol. Evol.* 7.
- Altschul, A.M., 1948. Biological processes of the cottonseed, in: Bailey, A.E. (Ed.), *Cottonseed and Cottonseed Products*. NY, pp. 157–212.
- Auguie, B., 2017. gridExtra: Miscellaneous Functions for “Grid” Graphics.
- Baken, E.K., Collyer, M.L., Kaliontzopoulou, A., Adams, D.C., 2021. Geomorph v4.0 and gmShiny: enhanced analytics and a new graphical interface for a comprehensive morphometric experience. *Methods in Ecology and Evolution*. *Methods in Ecology and Evolution*.

- Bonhomme, V., Forster, E., Wallace, M., Stillman, E., Charles, M., Jones, G., 2017. Identification of inter- and intra-species variation in cereal grains through geometric morphometric analysis, and its resilience under experimental charring. *J. Archaeol. Sci.* 60–67.
- Bouby, L., 2010. Agriculture dans le bassin du Rhône, du Bronze final à l'Antiquité, Agrobiodiversité, économie, cultures. EHESS, Paris.
- Bouby, L., Bonhomme, V., Ivorra, S., Pastor, T., Rovira, N., Tillier, M., Pagnoux, C., Terral, J.F., 2018. Back from burn out: are experimentally charred grapevine pips too distorted to be characterized using morphometrics? *Archaeol. Anthropol. Sci.* 10, 943–954. <https://doi.org/10.1007/s12520-016-0425-x>
- Bouchaud, C., Clapham, A., Newton, C., Tallet, G., Thanheiser, U., 2018. Cottoning on to Cotton (*Gossypium* spp.) in Arabia and Africa During Antiquity, in: Mercuri, A.M., D'Andrea, A.C., Fornaciari, R., Höhn, A. (Eds.), *Plants and People in the African Past*. Springer International Publishing, Cham, pp. 380–426. https://doi.org/10.1007/978-3-319-89839-1_18
- Bouchaud, C., Yvanez, E., Wild, J.P., 2019. Tightening the thread from seed to cloth. New enquiries in the archaeology of Old World cotton: A case for inter-disciplinarity. *Rev. D'ethnoécologie* 33. <https://doi.org/10.4000/ethnoecologie.4501>
- Braadbaart, F., 2007. Carbonization and morphological changes in modern dehusked and husked *Triticum dicoccum* and *Triticum aestivum* grains. *Veg. Hist. Archaeobotany* 17, 155–166. <https://doi.org/10.1007/s00334-007-0134-6>
- Braadbaart, F., Wright, P.J., van der Horst, J., Boon, J.J., 2007. A laboratory simulation of the carbonization of sunflower achenes and seeds. *J. Anal. Appl. Pyrolysis* 78, 316–327. <https://doi.org/10.1016/j.jaap.2006.07.007>
- Brubaker, C.L., Paterson, A.H., Wendel, J.F., 1999. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors 42, 20.
- Charles, M., Forster, E., Wallace, M., Jones, G., 2015. “Nor ever lightning char thy grain”¹: establishing archaeologically relevant charring conditions and their effect on glume wheat grain morphology. *STAR Sci. Technol. Archaeol. Res.* 1, 1–6. <https://doi.org/10.1179/2054892315Y.0000000008>
- Chowdhury, K.A., Buth, G.M., 1971. Cotton seeds from the Neolithic in Egyptian Nubia and the origin of Old World cotton. *Biol. J. Linn. Soc.* 3, 303–312. <https://doi.org/10.1111/j.1095-8312.1971.tb00524.x>
- Clapham, A.J., Rowley-Conwy, P.A., 2007. New discoveries at Qasr Ibrim, Lower Nubia, in: *Fields of change, progress in African archaeobotany*. Groningen, pp. 157–164.
- Dillehay, T.D., Rossen, J., Andres, T.C., Williams, D.E., 2007. Preceramic Adoption of Peanut, Squash, and Cotton in Northern Peru. *Science* 316, 1890–1893. <https://doi.org/10.1126/science.1141395>
- Fuller, D.Q., 2008. The spread of textile production and textile crops in India beyond the Harappan zone: an aspect of the emergence of craft specialization and systematic trade, in: Osada, T., Uesugi, A. (Eds.), *Linguistics, Archaeology and the Human Past*, Indus Project Research Institute for Humanity and Nature. Kyoto, p. 27.
- García-Granero, J.J., Arias-Martorell, J., Madella, M., Lancelotti, C., 2016. Geometric morphometric analysis of *Setaria italica* (L.) P. Beauv. (foxtail millet) and *Brachiaria ramosa* (L.) Stapf. (browntop millet) and its implications for understanding the biogeography of small millets. *Veg. Hist. Archaeobotany* 25, 303–310. <https://doi.org/10.1007/s00334-015-0541-z>
- Horikoshi, M., Tang, Y., 2016. ggfortify: Data Visualization Tools for Statistical Analysis Results.
- Hutchinson, J., Silow, R., Stephens, S., 1947. *The evolution of Gossypium L. and the differentiation of the cultivated cottons*, Empire cotton growing corporation. Oxford university press, Oxford.

- Jesus, A., Bonhomme, V., Evin, A., Ivorra, S., Soteras, R., Salavert, A., Antolín, F., Bouby, L., 2021. A morphometric approach to track opium poppy domestication. *Sci. Rep.* 11, 9778. <https://doi.org/10.1038/s41598-021-88964-4>
- Kassambara, A., n.d. ggpubr: “ggplot2” Based Publication Ready Plots.
- Kulkarni, V.N., Khadi, B.M., Maralappanavar, M.S., Deshapande, L.A., Narayanan, S.S., 2009. The Worldwide Gene Pools of *Gossypium arboreum* L. and *G. herbaceum* L., and Their Improvement, in: *Genetics and Genomics of Cotton, Plant Genetics and Genomics: Crops and Models*. Springer, New York, pp. 69–97.
- Lagière, R., 1966. *Le cotonnier, Techniques agricoles et productions tropicales*. Maisonneuve et Larose, Paris.
- Miller, N.F., 1989. What mean these seeds: A comparative approach to archaeological seed analysis. *Hist Arch* 23, 50–59. <https://doi.org/10.1007/BF03374108>
- Mitteroecker, P., Bookstein, F., 2011. Linear Discrimination, Ordination, and the Visualization of Selection Gradients in Modern Morphometrics. *Evol Biol* 16.
- Mitteroecker, P., Gunz, P., 2009. Advances in geometric morphometrics. *Evol. Biol.* 36, 235–247. <https://doi.org/10.1007/s11692-009-9055-x>
- Moulhérat, C., Tengberg, M., Haquet, J.-F., Mille, B., 2002. First Evidence of Cotton at Neolithic Mehrgarh, Pakistan: Analysis of Mineralized Fibres from a Copper Bead. *J. Archaeol. Sci.* 29, 1393–1401. <https://doi.org/10.1006/jasc.2001.0779>
- Murdoch, D., Adler, D., 2022. rgl: 3D visualization using OpenGL.
- Murrell, P., Wen, Z., 2020. gridGraphics: Redraw Base Graphics Using “grid” Graphics.
- Page, J.T., Huynh, M.D., Liechty, Z.S., Grupp, K., Stelly, D., Hulse, A.M., Ashrafi, H., Van Deynze, A., Wendel, J.F., Udall, J.A., 2013. Insights into the Evolution of Cotton Diploids and Polyploids from Whole-Genome Re-sequencing. *Genes|Genomes|Genetics* 3, 1809–1818. <https://doi.org/10.1534/g3.113.007229>
- Pagnoux, C., Bouby, L., Ivorra, S., Petit, C., Valamoti, S.-M., Pastor, T., Picq, S., Terral, J.-F., 2015. Inferring the agrobiodiversity of *Vitis vinifera* L. (grapevine) in ancient Greece by comparative shape analysis of archaeological and modern seeds. *Veg. Hist. Archaeobotany* 24, 75–84. <https://doi.org/10.1007/s00334-014-0482-y>
- Palmer, S.A., Clapham, A.J., Rose, P., Freitas, F.O., Owen, B.D., Beresford-Jones, D., Moore, J.D., Kitchen, J.L., Allaby, R.G., 2012. Archaeogenomic evidence of punctuated genome evolution in *Gossypium*. *Mol. Biol. Evol.* 29, 2031–2038. <https://doi.org/10.1093/molbev/mss070>
- R Core Team, 2022. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing. ed. R Foundation for Statistical Computing, Vienne.
- Rohlf, F., 2005. TpsRelw 1.41-Thin Plate Spline Relative Warp. *Ecol. Evol.*
- Ros, J., Evin, A., Bouby, L., Ruas, M.-P., 2014. Geometric morphometric analysis of grain shape and the identification of two-rowed barley (*Hordeum vulgare* subsp. *distichum* L.) in southern France. *J. Archaeol. Sci.* 41, 568–575. <https://doi.org/10.1016/j.jas.2013.09.015>
- Saxena, D.K., Sharma, S.K., Sambhi, S.S., 2011. Comparative extraction of cottonseed oil by n-hexane and ethanol. *ARPN J. Eng. Appl. Sci.* 6, 6.
- Seignobos, C., 2019. Du coton traditionnel au coton colonial, le coup de force du progrès (Nord-Cameroun). *Rev. D’ethnoécologie* 26.
- Smith, C.E., Stephens, S.G., 1971. Critical identification of Mexican archaeological cotton remains. *Econ. Bot.* 25, 160–168. <https://doi.org/10.1007/BF02860076>
- Stephens, S.G., 1975. A reexamination of the cotton remains from Huaca Prieta, North Coastal Peru. *Am. Antiq.* 4, 406–419.
- Stephens, S.G., 1949. The cytogenetics of speciation in *Gossypium* L. Selective elimination of the donor parent genotype in interspecific backcrosses. *Genetics* 34, 627.
- Stephens, S.G., Moseley, M.E., 1974. Early Domesticated Cottons from Archaeological Sites in Central Coastal Peru. *Am. Antiq.* 39, 109–122. <https://doi.org/10.2307/279225>

- Tarongi, M., Bonhomme, V., Evin, A., Ivorra, S., López, D., Alonso, N., Bouby, L., 2021. A new way of seeing pulses: preliminary results of geometric morphometric analyses of Iron Age seeds from the site of La Font de la Canya (Barcelona, Spain). *Veg. Hist. Archaeobotany*. <https://doi.org/10.1007/s00334-020-00801-2>
- Terral, J.-F., Bonhomme, V., Pagnoux, C., Ivorra, S., Newton, C., Paradis, L., Ater, M., Kassout, J., Limier, B., Bouby, L., Cornet, F., Barazani, O., Dag, A., Galili, E., 2021. The Shape Diversity of Olive Stones Resulting from Domestication and Diversification Unveils Traits of the Oldest Known 6500-Years-Old Table Olives from Hishuley Carmel Site (Israel). *Agronomy* 11, 2187. <https://doi.org/10.3390/agronomy11112187>
- Terral, J.-F., Newton, C., Ivorra, S., Gros-Balthazard, M., de Morais, C.T., Picq, S., Tengberg, M., Pintaud, J.-C., 2012. Insights into the historical biogeography of the date palm (*Phoenix dactylifera* L.) using geometric morphometry of modern and ancient seeds: Historical biogeography of date palm agrobiodiversity. *J. Biogeogr.* 39, 929–941. <https://doi.org/10.1111/j.1365-2699.2011.02649.x>
- Terral, J.-F., Tabard, E., Bouby, L., Ivorra, S., Pastor, T., Figueiral, I., Picq, S., Chevance, J.-B., Jung, C., Fabre, L., Tardy, C., Compan, M., Bacilieri, R., Lacombe, T., This, P., 2010. Evolution and history of grapevine (*Vitis vinifera*) under domestication: new morphometric perspectives to understand seed domestication syndrome and reveal origins of ancient European cultivars. *Ann. Bot.* 105, 443–455. <https://doi.org/10.1093/aob/mcp298>
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*. Springer, New York.
- Viot, C., 2019. Domestication and varietal diversification of Old World cultivated cottons (*Gossypium* sp.) in the Antiquity. *Rev. D'ethnoécologie*. <https://doi.org/10.4000/ethnoecologie.4404>
- Wendel, J.F., Brubaker, C., Alvarez, I., Cronn, R.C., Stewart, J., 2009. Evolution and Natural History of the Cotton Genus, in: Paterson, A.H. (Ed.), *Genetics and Genomics of Cotton*. Springer, New York, pp. 3–22.
- Wendel, J.F., Grover, C.E., 2015. Taxonomy and Evolution of the Cotton Genus, *Gossypium*, in: Fang, D.D., Percy, R.G. (Eds.), *Agronomy Monographs*. American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc., Madison, WI, USA, pp. 25–44. <https://doi.org/10.2134/agronmonogr57.2013.0020>
- Wickham, H., 2021. *forcats: Tools for Working with Categorical Variables (Factors)*.
- Wickham, H., 2016. *GGplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.
- Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Milton Bache, S., Müller, K., Ooms, J., Robinson, D., Paige Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. *J. Open Source Softw.* Vol. 4 N°43 1686.
- Wickham, H., François, R., Henry, L., Müller, K., 2022. *dplyr: A Grammar of Data Manipulation*.
- Wickham, H., Seidel, D., 2022. *scales: Scale Functions for Visualization*.