



Is there a domestication syndrome in *Vitis* (*Vitaceae*) seed morphology?

Concepción Obón · Diego José Rivera-Obón · Javier Valera · Gonzalo Matilla · Francisco Alcaraz · David Maghradze · Maia Kikvadze · Carlos-Alvar Ocete · Rafael Ocete · Anna Nebish · Javier Abellán · Jose Antonio Palazón · Diego Rivera

Received: 25 April 2024 / Accepted: 10 May 2024 / Published online: 17 June 2024
© The Author(s) 2024

Abstract Characterizing plant genetic resources and understanding the domestication process are key challenges in horticulture. This study focuses on the common grapevine (*Vitis vinifera* L.), employing molecular, ampelographic, and morphometric techniques to explore its domestication syndrome. We propose a combined domestication index for grape seeds, integrating six commonly used indices. Logistic regression and randomForest machine learning

methods were employed to assess the probability of seed domestication. Analysis involved 4029 seeds from 783 samples, spanning live plants (cultivated, wild, feral), archaeological samples, and fossils. Results indicate a distinct domestication syndrome in grapevine seeds, notably characterized by a long stalk (> 1 mm), with combined index values ranging from 0.83 to 1 and domestication probabilities of 0.75 to

C. Obón (✉)
CIAGRO, EPSO, Universidad Miguel Hernández de Elche, 30312 Orihuela (Alicante), Spain
e-mail: cobon@umh.es

D. J. Rivera-Obón
ENS Paris-Saclay, 4 Av. des Sciences, 91190 Gif-sur-Yvette, France
e-mail: dieguitojos@gmail.com

J. Valera · G. Matilla
Departamento de Prehistoria Arqueología e Historia Antigua, Facultad de Letras, Universidad de Murcia, Campus de la Merced, 30001 Murcia, Spain
e-mail: javier.valera.martinez@gmail.com

G. Matilla
e-mail: gmatilla@um.es

J. Valera · F. Alcaraz · D. Rivera
Departamento de Biología Vegetal, Facultad de Biología, Universidad de Murcia, Campus de Espinardo, 30100 Murcia, Spain
e-mail: falcaraz@um.es

D. Rivera
e-mail: drivera@um.es

D. Maghradze
Faculty of Viticulture and Winemaking, Caucasus International University, Chargali Street 73, 0141 Tbilisi, Georgia
e-mail: david.maghradze@ciu.edu.ge; david.maghradze@gmail.com

D. Maghradze · M. Kikvadze
Faculty of Agrarian Sciences and Biosystems Engineering, Georgian Technical University, 17 David Guramishvili Ave, 0192 Tbilisi, Georgia
e-mail: m.kikvadze70@gmail.com

C.-A. Ocete · R. Ocete
La Gualdia 9, 26211 Tirgo, La Rioja, Spain
e-mail: carlos_ocete@hotmail.com

R. Ocete
e-mail: ocete@us.es

A. Nebish
Departamento de Viticultura, Instituto de Ciencias de la Vid y del Vino (ICVV) (CSIC, UR, Gobierno de La Rioja), Finca La Grajera, Ctra. de Burgos Km. 6, 26007 Logroño, La Rioja, Spain
e-mail: annita.nebish@icvv.es; anita.nebish@gmail.com

1. This syndrome is prevalent in modern cultivars but absent in wild grapevines, aiding in the differentiation of archaeological materials. While not exclusive to *Vitis vinifera*, the domestication syndrome is characteristic of it, distinguishing it from wild *Vitis sylvestris* and suggesting isolated domestication events in a few American and Asian *Vitis* species. Fossil and archaeological materials exhibit low domestication values, hinting at the beginnings of domestication. However, dating reliability remains a challenge.

Keywords *Vitis sylvestris* C.C. Gmelin · *Vitis vinifera* L. · Domestication index · Stummer's index · Random forest · Logistic regression

Introduction

Among the challenges of horticultural science is to characterize plant genetic resources for agriculture and food and to reconstruct the phylogeny of domesticated species, which are currently cultivated together with their wild relatives. The common grapevine (*Vitis vinifera* L.) has been studied in this regard for years and by numerous authors using molecular, ampelographic and morphometric techniques.

The name *Vitis vinifera* was lectotypified by Sidiqi (1980) using Herb. Linn. No. 281.1 (LINN 2023a, b a). This name is applied in the strict sense (var. *vinifera*) to cultivated grapes whose berries are provided with (0)1 to 4 seeds. Iamónico et al. (2015) neotypified for Corinth and Sultanina group of ancient seedless cultivars the name *V. vinifera* var. *apyrena* L. using Herb. Linn. No. 281.3 (LINN 2023a, b b). The Eurasian wild grapevine has received very

diverse taxonomic treatments, from the rank of variety to the one of the species. This implies the use of the subsequent alternative names, depending on the accepted level: *V. vinifera* var. *sylvestris* Willd., or *V. vinifera* subsp. *sylvestris* (C.C. Gmel.) Hegi, or, even, *V. sylvestris* C.C.Gmel. (Ferrer-Gallego et al. 2019). However the conservation and typification proposal of Ferrer-Gallego et al. (2019) for *V. sylvestris* was rejected (Applequist 2023) so in the species rank the valid name requires further study, in the meantime we will continue to use *V. sylvestris* C.C.Gmel.

Based on the Bayesian molecular dating, *Vitis* is inferred to have originated in the New World during the late Eocene (32.6–48.6 Ma), then migrated to Eurasia in the late Eocene (30.9–45.1 Ma) (Liu et al. 2016). The North Atlantic land bridges are hypothesized by Liu et al. (2016) to be the most plausible route for the *Vitis* migration from the New World to Eurasia, while intercontinental long-distance dispersal cannot be eliminated as a likely mechanism (Nie et al. 2012).

Seeds of *Vitaceae* are easily identifiable due to their distinctive features, which include a pair of ventral furrows and a dorsal chalaza shield. The shape and positioning of these ventral furrows and chalaza, along with the testa anatomy, are key characteristics that differentiate seed types within the family. Typically, species such as *Leea*, *Cissus*, *Cyphostemma*, *Tetrastigma*, *Rhoicissus*, and *Cayratia* have a thin linear chalaza, while others, including *Vitis*, generally exhibit an oval chalaza (Manchester et al. 2012).

However, some seed forms are shared across multiple genera, and certain genera may possess more than one seed form. This overlapping seed morphology suggests a closer relationship among some genera, while in other cases, it may indicate convergent or parallel evolution (Chen and Manchester 2011).

Vitis seeds present a characteristic differential morphology, however Chen and Manchester (2011) have shown that two of the *Ampelocissus* species (*A. erdvendbergiana* Planch. and *A. robinsonii* Planch.) have smaller, heart-shaped seeds that resemble those of *Vitis* and are inseparable from *Vitis*; these species are, both, endemic to Central America. *Ampelocissus erdvendbergiana* is grouped by Nie et al. (2012) with *Vitis* section *Muscadinia* grapevines, such as *Vitis rotundifolia* and *V. popenoei*. The similarity between

A. Nebish
Department of Genetics and Cytology, Yerevan State
University, A. Manoogian str.1, 0025 Yerevan, Armenia

J. Abellán
Departamento de Física, Facultad de Química, Universidad
de Murcia, Campus de Espinardo, 30100 Murcia, Spain
e-mail: fabellan@um.es

J. A. Palazón
Departamento de Ecología e Hidrología, Facultad de
Biología, Universidad de Murcia, Campus de Espinardo,
30100 Murcia, Spain
e-mail: palazon@um.es

the seeds of certain *Ampelocissus*, *Pterisanthes*, *Nothocissus* and *Vitis* can be explained by the close phylogenetic relationships of these genera (Chen and Manchester 2011).

Vitis seeds usually have short ventral inward folds and an oval chalaza. Their apical and/or basal grooves are sometimes prominent in dorsal view. The seeds are usually smooth; however, there are also species with rough seeds. *V. rotundifolia* has larger seeds with a slightly rough surface; its ventral folds are longer than those of other *Vitis* species (Chen and Manchester 2011).

In early morphological research with *Vitis* seeds, two morphotypes were defined consisting of pear-shaped and oblong to ovoid seeds, which corresponded respectively to the taxonomic groups *Euvinis* and *Muscadinia*. The seeds of cultivated species (*V. vinifera*, *V. labrusca* L., *V. aestivalis* Michx.) are larger and their dimensions are more variable than those of wild species (Martín-Gómez et al. 2020).

Seeds of Eurasian wild (*V. sylvestris*) and domesticated (*V. vinifera*) grapevine display dissimilarities, which allow the discrimination between both taxa: wild grapevine bears small and roundish seeds with short stalks, while pips from cultivars are more elongated, with longer stalks.

These morphological traits have been used to discriminate archaeological pips since Stummer's work of 1911. The co-variation of berry size, number of seeds per berry, pip size and pip shape were explored by Bonhomme et al. (2020) on Euro-Mediterranean traditional cultivars and wild grapevines. They show that for both wild and domesticated grapevine, the longer pip has a "domesticated" shape.

However, the use of morphometric criteria has been criticized because of the overlap between wild and domesticated seed shapes and because of seed deformation due to carbonization. Moreover, morphometric criteria and measurement values defined by Stummer and later by Mangafa and Kotsakis in 1996 are based on a reference model which includes grape cultivars and wild individuals from the Mediterranean and Western and Central Europe, which for Pagnoux et al. (2015) is an excessively restricted area, which cannot be considered as an appropriate representative of grapevine diversity.

Jacquat and Martinoli (1999) made the classification of grapevine seeds, as products of domesticated or wild plants, applying separately Stummer (1911), Facsar (1970) and Facsar and Jerem (1985) in the version by Perret (1997), and Mangafa and Kotsakis (1996) indexes. They carried out different biometric research on Nabataean and Roman seeds found at Petra, Jordan, dated to 150 B.C.—400 A.D. Depending on the identification method selected, the seeds were attributed either to wild grapevines (based on the ratio of breadth over length, and on discriminant analyses of size variables such as pip length, stalk length, and chalaza position), or an archaic variety of vine with seeds morphologically close to those of wild grapevines (ratio of stalk length over total pip length).

Traits associated with domesticated *V. vinifera* are unquestionable, but what is less clear is to what extent the lack of these traits in a seed sample indicates that this comes from wild, uncultivated plants. Whether these seeds, especially in archaeological contexts, correspond to wild or cultivated plants is quite uncertain, since we do not know at what point in the process of grapevine cultivation the seeds acquired traits diagnostic of domestication (Valamoti 2009).

Furthermore, the study of Logothetis (1970, 1974) revealed that after carbonization seeds of both, *V. vinifera* and *V. sylvestris*, became smaller and more rounded, since length decreases more than breadth which in turn decreases more than thickness. These changes were even more profound in wild seeds. However, what is most relevant is that carbonization approached domesticated seed shape to the wild type. Considering that, the B/L indices (Stummer index) increased as a result of carbonization, and for that reason, Mangafa and Kotsakis (1996) suggest that their second and third formulae (see below) are more appropriate for the identification of charred archaeological grape pips, because fewer simple size variables are used with less contribution than allometric variables.

The Eurasian wild grapevine, *V. sylvestris* C.C. Gmel., was seen as the hypothetical dioecious parental of domesticated *V. vinifera* L., which is, usually, hermaphrodite (Rivera and Walker 1989; This et al. 2006; Zohary and Hopf 2000). *V. sylvestris* was found to be sister to all Asian species and to be the oldest living Eurasian species amongst those included in the DNA study by Zecca et al. (2012).

Fossils attributed to *V. sylvestris* appear within sediments dated from the end of the *Pliocene* (Sémah and Renault-Miskovsky 2004). Currently, wild populations of *V. sylvestris* and related taxa are scattered in predominantly riparian natural ecosystems from the Iberian Peninsula through Central Europe, the Balkans and the Caucasus, to the Hindu Kush (Arnold 2002). Some populations of this liana can be also found in the African Maghreb (Ocete et al. 2007). Their main habitats are river-bank forests, river mouths, flood plains, colluvial positions on the slopes of hills and mountains and coasts between the parallels 49° 27' N (Rhine River, Germany) and 31° N (Ourika River, Morocco) (Iriarte et al. 2013). In such places, soils are often renewed by flooding (Arnold 2002; Maghradze et al. 2010).

Our main objective is to analyze and validate the use of domestication indices to distinguish *V. vinifera* and differentiate it from *V. sylvestris* and other *Vitis* wild species. For this purpose, we will study the values of the indices in living populations of European, American and Asian wild grapevines. In parallel we approach the use of logistic regression and random forest tools to estimate the probability of each seed of being domesticated or not and to compare these results with the Domestication Index values.

We also aim to detect the most influential variables and indexes, using machine learning tools.

In addition, we also intend to assess the information that the indices provide on the evolution of *Vitis*, from the first fossil seed remains to the present day, paying particular attention to materials from archaeological sites. Finally, we intend to evaluate the human contribution to the development of the domestication syndrome in *Vitis* seeds and to consider the influence of possible non-anthropogenic factors.

Materials and methods

Seed samples and characters analyzed

The living reference material here used aims to give a comprehensive picture of the diversity within *Vitis* species and cultivars, starting with samples that cover a large number of cultivars, followed of wild and spontaneous vines, in order to compare the morphological differences between them, together with a comprehensive sample collection of fossils *Vitaceae*

and modern samples of related living genera such as *Ampelopsis*, *Nekemias* or *Parthenocissus*. Of the total number of seeds analyzed, 3484 are modern from living species (481 samples), 398 are archaeological (195 samples) where most of the material was conserved carbonized, and 147 are fossilized (107 samples). We have approached the study of seeds of the genus *Vitis* following the most widely accepted concept of *Vitis* and the list of included species adopted in references such as: POWO (2023), GRIN (2023) or GBIF (2023).

As far as it was possible, the most relevant ampelographic data included in the Grapevine Descriptors (IPGRI-UPOV-OIV 1997) have been collected, especially those related to the hairs covering the leaves and the characteristics of the grape berry, either recorded directly in the field, since more than two hundred of these grapevines were and are still grown on farm in Molina de Segura (Spain), or from the databases (FNAGR, VIVC, European *Vitis* Database, Armenian *Vitis* Database) (FNAGR 2023; Maul 2023; European *Vitis* Database 2023; Armenian *Vitis* Database 2007). These ampelographic data have been used to assign the different cultivated grapevines, from which seeds were collected for the study, to the various *proles* following the criteria established by Negru (1946a, b a). Herbarium specimens were prepared and are kept at MUB herbarium of the Universidad de Murcia (Spain).

For each sample, on average 10 seeds were randomly selected and measured to build the data matrix. When the number of seeds per set was less than 10, the analysis was carried out on all available seeds. However, in this article we will deal with single seeds. Each seed was individually described according to 14 characters. Of these, 11 are quantitative direct measurements and three are qualitative.

The quantitative analysis involves several key measurements of the seed morphology as depicted in Fig. 1. This includes assessing the total length, maximum breadth, and thickness of the seed (1, 2, and 3 in Fig. 1, respectively). Additionally, measurements are taken for the breadth of the stalk at its junction and on the seed base (4 and 5 in Fig. 1), as well as the length of the beak in both dorsal and ventral views (6, and 7 in Fig. 1, respectively) and thickness of the beak at the seed base (8 in Fig. 1). Furthermore, the total length and breadth of the chalaza shield are examined

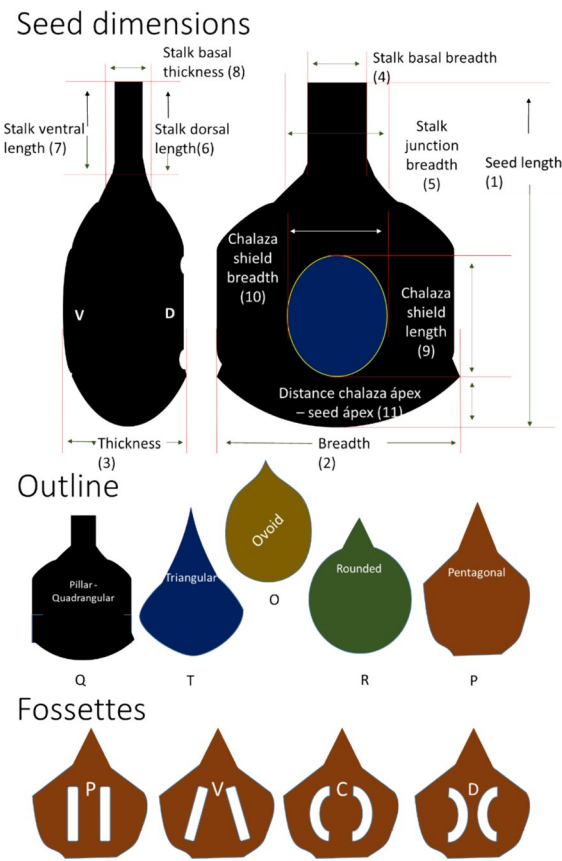


Fig. 1 Seed morphological characters. Note: label color codes: label Blue, used in the Stummer’s index; label Green, used in the indices of Facsar-Perret and Mangafa—Kotsaki’s; label Yellow, used in the Mangafa–Kotsaki’s, indices and label Brown, used in all six indices. Image: D. Rivera

(9 and 10 in Fig. 1). The position of the chalaza denoted as PCH according to Mangafa and Kotsakis, is also considered (9 + 11 in Fig. 1). These comprehensive measurements provide valuable insights into the seed’s structural characteristics (Rivera et al. 2007).

Qualitative analysis involves categorizing seed characteristics into distinct states (Fig. 1). This includes assessing the contour type, which can be ovoid, quadrangular, triangular, rounded, or pentagonal. The arrangement of fossettes is also considered, with options including parallel, furcate, convergent, or divergent. Additionally, the presence or absence of radial furrows is noted as part of the evaluation process.

The primary raw data matrix used for this study consists of 4029 single rows of analyzed seeds, belonging to 783 samples, and 33 columns of variables, among which fourteen are the characters above listed, sixteen are allometric indices, and three are derived.

- Allometric: Stummer’s (Breadth / length), Breadth / depth, Facsar and Perret’s Stalk length / seed length, Seed body length (seed length minus stalk length), Stalk breadth / stalk length, Stalk breadth at junction / stalk depth at junction, Seed depth / stalk depth at junction, Chalaza shield breadth / chalaza shield length, Seed length / chalaza shield length, Seed breadth / chalaza shield breadth, Distance chalaza apex to seed apex / chalaza length, Accumulated length / seed length. And finally, the four by Mangafa and Kotsakis in 1996.
- Derived: Length x breadth x depth, Sphericity and Triangularity.

The quantitative and qualitative characters were measured or determined using digital scaled images. The seeds of each sample were individually placed, on a plasticine support with a built-in scale to be photographed in dorsal, ventral and lateral view and measured using the open-source Fiji software (Schindelin et al. 2012). All photographs were taken under the same zoom conditions. Also, scale images of fossilized and archaeological seeds from specialized literature have been used for measurements. The characters were recorded in an Excel spreadsheet where the allometric relationships were automatically calculated using algorithms.

Data analysis: morphometric indices

Stummer’s index

Stummer (1911) proposed an index based on the seed breadth/length ratio (Fig. 1) adapted to Central European populations (1).

$$STI(X_i) = \frac{100 \times \text{Breadth of the seed}_{X_i}}{\text{Length of the seed}_{X_i}} \quad (1)$$

Table 1 Thresholds of Stummer's index for wild and domesticated grapevine seeds¹

Range of Stummer's index values	Taxa
44–53	<i>V. vinifera</i>
54–75	Intermediate or hybrids
76–83	<i>V. sylvestris</i>

¹Formula: $B/L \times 100$

This index allows a fairly effective differentiation between extreme forms, but intermediate values are found in both wild and cultivated populations (Table 1). Values of the Stummer index between 0.44 and 0.53 would be exclusive to cultivars, whereas between 0.76 and 0.83 would be exclusive to wild populations. Values between 0.54 and 0.75 are intermediate and could point to the presence of “hybrids”, cultivars or wild vines. Levadoux (1956) has suggested that this index would have limited validity in distinguishing wild grapevines from cultivars.

Facsar and Perret's index

Perret (1997), proposed a new index based on the allometric relationship between the length of the stalk or column and the total length of the seed (2) (1, 6 and 7 in Fig. 1).

$$FPI(X_i) = \frac{100 \times (\text{Mean of dorsal and ventral stalk lengths}_{X_i})}{\text{Total length of the seed}_{X_i}} \quad (2)$$

Apparently, this index allows to differentiate quite effectively between wild and cultivated populations, with the threshold situated between 18 and 19 (Table 2). However, the same index was previously proposed by Facsar (1970), Terpó (1976), and Facsar and Jerem (1985). According to the latter the mean index values of the relative stalk length (stalk length/seed length) × 100 of the 47 grapevine cultivars (*V.*

Table 2 Thresholds of Facsar and Perret's index for wild and domesticated grapevine seeds¹

Range of values	taxa
12–18	<i>V. sylvestris</i>
19–30 (35)	<i>V. vinifera</i>

¹Formula: $(LS/L) \times 100$. *LS* stalk length, *L* seed length

vinifera) cultivated in Hungary, range between values 22–35 with a mean of 29. The index of relative stalk length related to the seeds of the native populations of *V. sylvestris* varies between 13 and 23 (mean = 17). Here the overlap between the cultivated and wild grapevine is minimal.

Mangafa and Kotsaki's indices

It is important to remember that these seed-based indices were largely developed to work with archaeological seeds. The formulae proposed by Mangafa and Kotsakis in 1996 were successfully applied to local Greek samples of both modern seeds and archaeological remains. The four formulae (Table 3) are based on the combined use of ratios involving variables such as total seed length (L), seed stalk length (LS) and chalaza position (PCH) (distance from the chalaza base to the base of the seed, involving 1, 9 and 11 in Fig. 1).

Sensitivity and specificity of indices

For biomedical tests, it is common to represent their sensitivity and specificity as parameters that help to assess their usefulness in the diagnosis of diseases. Sensitivity and specificity are measures of the accuracy of a diagnostic test. There is a difference between sensitivity and specificity that depends on the cut-off level chosen for a positive diagnosis (Chu 1999). While a highly sensitive and highly specific test is desirable there is usually a trade-off between sensitivity, and specificity: as one increases, the other decreases (Table 4). In medicine predictive values depend on the prevalence of disease (prior) as well as the sensitivity and specificity of the test within a Bayesian reasoning framework (Chu 1999).

It is precisely these differences in sensitivity and specificity of the different tests (Table 4) that suggest that their combined use may provide information better adapted to the different cases we may encounter in the case of *Vitis*.

It is important to underline that the average of the sensitivity and specificity values varies significantly (Table 4) and provides a tool to distinguish the most informative indices such as the 1st Mangafa and Kotsakis index (=0.87) or the Facsar and Perret index (=0.84).

Table 3 Ranges of Mangafa and Kotsakis's indices for wild and domesticated grapevine seeds

Range of values	Taxonomic information
<i>Range of values (formula 1)¹</i>	
< -0.2	Wild grapes (with probability c. 1)
$-0.2 \leq x < 0.2$	Wild grapes (with probability 0.64)
$0.2 \leq x < 0.8$	Domesticated grapes (with probability 0.76)
> 0.8	Domesticated grapes (with probability c. 1)
<i>Range of values (formula 2)²</i>	
< -0.2	Wild grapes (with probability c. 1)
$-0.2 \leq x < 0.4$	Wild grapes (with probability 0.64)
$0.4 \leq x < 0.9$	Domesticated grapes (with probability 0.76)
> 0.9	Domesticated grapes (with probability c. 1)
<i>Range of values (formula 3)³</i>	
< 0	Wild grapes (with probability c. 1)
$0 \leq x < 0.5$	Wild grapes (with probability 0.9)
$0.5 \leq x < 0.9$	Domesticated grapes (with probability 0.63)
> 0.9	Domesticated grapes (with probability c. 1)
<i>Range of values (formula 4)⁴</i>	
< -0.9	Wild grapes (with probability c. 1)
$-0.9 \leq x < 0.2$	Wild grapes (with probability 0.9)
$0.2 \leq x < 1.4$	Domesticated grapes (with probability 0.63)
> 1.4	Domesticated grapes (with probability c. 1)

¹Formula 1: $-0.3801 + (-30.2 \times LS/L) + 0.4564 \times PCH - 1.386 \times L + 2.88 \times PCH/L + 9.4239 \times LS$

²Formula 2: $0.2951 + (-12.64 \times PCH/L - 1.6416 \times L + 4.5131 \times PCH + 9.63 \times LS/L)$

³Formula 3: $-7.491 + (1.7715 \times PCH + 0.49 \times PCH/L + 9.56 \times LS/L)$

⁴Formula 4: $0.7509 + (-1.5748 \times L + 5.297 \times PCH - 14.47 \times PCH/L)$

LS, stalk length; L, seed length; PCH, chalaza position

Table 4 Sensitivity and specificity of the different morphometric indices used

	D*	Threshold*	W*	Based on the prior** information		Mean
				Sensitivity	Specificity	
Stummer's	\leq	65	$>$	0.52	0.93	0.73
Facsar and Perret's	\geq	19	$<$	0.80	0.88	0.84
Mangafa and Kotsakis' formula 1	\geq	0.2	$<$	0.99	0.75	0.87
Mangafa and Kotsakis' formula 2	\geq	0.4	$<$	0.36	1.00	0.68
Mangafa and Kotsakis' formula 3	\geq	0.5	$<$	1.00	0.19	0.59
Mangafa and Kotsakis' formula 4	\geq	0.2	$<$	0.55	0.98	0.73

*Threshold for ascribing one seed to the domesticated (D) or to the wild (W) type, including intermediate

**Here we refer as prior information the knowledge of the whole set of ampelographic data from the plants from where the example seeds were collected, which allowed us to determine these as wild or cultivars

Domestication index

Since the above indices serve the same purpose, to distinguish wild from domesticated forms, but their results differ from case to case, the combined use of

the six indices may provide a better ability to discriminate seeds from wild or cultivated grapevines.

Here we propose the domestication index (DI) and the wildness index (WI). DI is calculated individually for each seed using the formula (3), where

NIT means the number of indices exceeding, above or below, each threshold value, and NRI is the number of indices considered (in the present case 6). Regardless of the specific values of DI and WI, their sum will always be equal to 1, hence WI is calculated as in formula (4).

$$DI = \frac{\sum_{i=1}^n NIT_i}{\sum_{i=1}^n NRI_i} \quad (3)$$

$$WI = 1 - DI \quad (4)$$

Since it is based on the positive result of each of the six indices, the DI values for a single seed can have the following values: 0 (0/6), 0.17 (1/6), 0.33 (2/6), 0.5 (3/6), 0.67 (4/6), 0.83 (5/6) and 1 (6/6), and likewise for WI. In heterogeneous samples the mean of the DI of the seeds that constitute the sample can present any value between 0 and 1.

Similarly, when calculating mean values for a period or a locality using DI and WI indices, the possible values lie on the continuum between 0 and 1. The wildness index is complementary to the domestication index, so the sum of their values is always equal to one.

We usually work with samples consisting of several seeds, which in the case of living populations, wild or cultivated, usually come from the same bunch, although not always, but then, at least from the same population. There are several relevant parameters when transferring the results from single seeds to the entire sample, however we mainly used those (Valera et al. 2023).

However, as we will see in the next section, other approaches are available to estimate the probability of individual seeds coming from a domesticate or wild grapevine.

Domestication probabilities estimate using logit and random forest

We utilized logistic regression models and the random forest technique to assess, based on morphometric data and comparison collections, the likelihood of individual seeds originating from either domesticated or wild grapevines, or exhibiting intermediate characteristics that hinder clear categorization. Our analyses were conducted using R Studio within the R framework, a

freely available software environment for statistical computing and graphics (R Project 2023). RStudio (2023) provides an integrated development environment (IDE) designed to enhance productivity in R and Python programming languages. These tools are accessible via the Cran R Project (2023).

The logistic regression model links the conditional probability of the binary response variable to the explanatory variables considered (Couronné et al. 2018), in our case, the probability of being ‘domesticated’ based on the six domestication indices. Utilizing the Logit function in R, we estimate logistic regression models using the generalized linear model function. Logit models are specific instances of generalized linear models with a binomial distribution and a logit link function, deriving their name from the logarithm of the likelihood ratios (Paladino 2017). In our study, Logit in R assigns a probability value of ‘domesticated’ to each seed, ranging from 0 to 1, abbreviated as PDILO.

Traditional machine learning algorithms often suffer from low classifier accuracy and are prone to overfitting. In contrast, Random Forest is a composite of machine learning algorithms that control a series of tree classifiers. Each tree contributes a vote for the most popular class, and these results are combined to yield the final classification. Random Forest exhibits high accuracy in classification tasks, robustness to outliers and noise, and mitigates the risk of overfitting. It has emerged as a widely used research method in both data mining and biological fields (Liu et al. 2012).

In our study, we utilized the ‘randomForest’ package version 4.7–1.1 in R (Breiman et al. 2022; randomForest 2023; Rossiter 2023). Random Forest in R generates classifications and regressions based on a forest of trees using random inputs, as described by Breiman (2001). Random forests employ a combination of tree predictors, where each tree depends on values from a randomly sampled vector with the same distribution across all trees. The generalization error of random forests approaches a limit as the number of trees in the forest increases (Breiman 2001). One of its major advantages is the ability to prevent overfitting and handle large numbers of variables, aiding in the identification of important attributes. Notable parameters of the random forest include ‘ntree’, which defaults to 500 trees, and ‘mtry’, which randomly samples a subset of variables for each split to avoid

overfitting (Breiman et al. 2022; R-Bloggers 2023). In our regression-based analysis, weight is set to null and replace is set to true. In our study, ‘randomForest’ in R assigns a probability value of ‘domesticated’ to each seed, ranging from 0 to 1, abbreviated as PDIrF.

We first worked with ‘randomForest’ the six formulas mentioned above, which were used to calculate the domestication index and in a second stage we added all the qualitative and quantitative variables and a series of allometric indices used in the description of the seed (Fig. 2). In both cases, we tried to use the random Forest model to detect the relative influence of the variables in the distinction between wild and domesticated grapevine seeds.

Fig. 2 Random Forest Analysis of Most Influential Variables. Two models are represented: (A) with the six domestication indices and (B) with all descriptive variables and allometric indices. Color codes: Red (indices) and violet (descriptive) most influential variables. Blue, relatively highly influential variables. This figure displays the estimated importance of variables or features obtained through randomForest technique. The bars represent the relative importance of each variable, highlighting those deemed most influential in the analysis. The higher a variable is on the y-axis, the more important it is in predicting the target variable. The order of variables from top to bottom indicates their importance. If a variable is consistently high across a large number of trees, it suggests that it plays a significant role in the predictive power of the random forest. Image: D.J. Rivera-Obón and D. Rivera

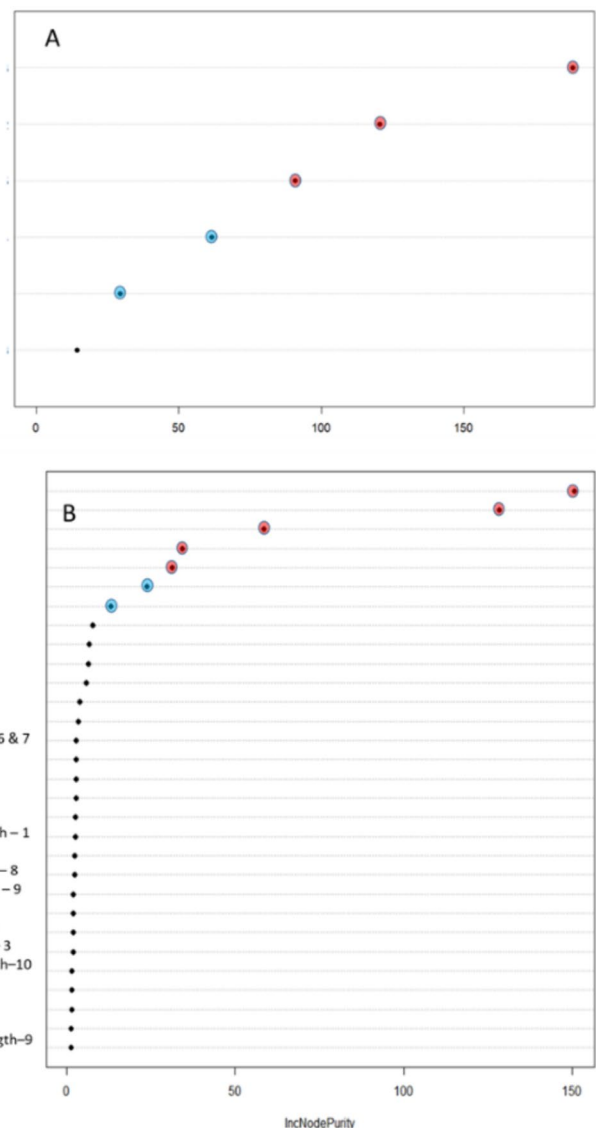
Mangafa & Kotsakis Formula 1
 Facsar & Perret
 Mangafa & Kotsakis Formula 3
 Mangafa & Kotsakis Formula 2
 Stummer
 Mangafa & Kotsakis Formula 4

Mangafa & Kotsakis Formula 1
 Stalk ventral length - 7
 Facsar & Perret
 Mangafa & Kotsakis Formula 3
 Stalk dorsal length - 6
 Mangafa & Kotsakis Formula 2
 Stummer
 Allometric: Dist. - 11 / Chalaza length 9
 Distance apex chalaza to seed apex - 11
 Mangafa & Kotsakis Formula 4
 Stalk breadth at junction - 5
 Seed length - 1
 Seed depth - 3
 Allometric: Stalk breath - 4 / Stalk length - 6 & 7
 Allometric: Prism volume
 Stalk breadth, basal - 4
 Stalk depth basal - 8
 Seed breadth - 2
 Allometric: Accumulated length / Seed length - 1
 Allometric: Seed sphericity
 Allometric: Stalk breadth - 10 / Stalk depth - 8
 Allometric: Seed length - 1 / Chalaza length - 9
 Allometric: Chalaza length - 9
 Allometric: Seed depth - 3 / Stalk depth - 8
 Allometric: Seed breadth - 2 / Seed depth - 3
 Allometric: Seed breadth-2 / Chalaza breadth-10
 Body seed length
 Chalaza breadth - 10
 Allometric: Triangularity
 Allometric: Chalaza breadth-10/Chalaza length-9

Couronné et al. (2018) have shown that the results of logistic and random forest classifications were noticeably dependent on the inclusion criteria used to select the example datasets, thus emphasizing the importance of clear statements regarding this dataset selection process. For this present analysis we classified all 4029 seeds according to the origin of the 783 samples:

Selection criteria for example data sets

- *Domesticated* This category includes the samples attributed on ampelographic grounds to *V. vinifera* Proles and *Subproles* described by Negru (1946a)



(*Occidentalis*, *Pontica*, *Orientalis Caspica* and *Orientalis Antasiatica*).

- **Wild** This includes well-characterized samples from wild Eurasian grapevine (*V. sylvestris*), wild Caucasian grapevine (*V. caucasica*), wild American *Vitis* species, and wild Asian *Vitis* species.

Categories not included in the example datasets are: archaeological samples, feral, fossils and others.

Results

Detection of influential variables and characterization of the domestication syndrome

The ‘randomForest’ methodology is instrumental in identifying the most influential variables (Fig. 2). Notably, the formulas 1 and 3 by Mangafa and Kotsakis, along with the Facsar and Perret index, emerge as prominent indices, whereas the contributions of Stummer and Mangafa and Kotsakis’ formula 2 are comparatively less significant. Conversely, formula 4 by Mangafa and Kotsakis appears to have minimal impact on distinguishing wild from domesticated seeds (Fig. 2A). When incorporating all descriptive variables and calculated allometric indices into the model, it becomes evident that, except for seed stalk (beak) length, the remaining characteristics contribute less than the aforementioned influential indices (Fig. 2B).

Overall, the ‘randomForest’ methodology demonstrates the utility of these indices in discerning signal from noise in the dataset, as noted by Silver (2012). Initially, DI values of 0 and 0.17 are indicative of clearly wild seeds, while values of 0.83 and 1 signify domestic grapevine seeds, leaving 18% of seeds categorized as indeterminate. Regarding the probabilities estimated by ‘randomForest’ (PDIrF), values ranging from 0 to 0.2 denote wild seeds, whereas those from 0.8 to 1 are typical of domestic grapevine seeds, with 12% of seeds falling into an indeterminate or intermediate category.

Beak lengths less than 1 mm are characteristic of wild grapevines, whereas in domesticated grapevines, beak lengths typically range between 1 and 2.5 mm, rarely reaching 3 mm. Only 4.8% of seeds

exhibit intermediate PDIrF values (0.3–0.6), with stalk dimensions around 1 mm in these cases.

Living species, domestication index and domestication probability values

Among the 46 *Vitis* species, hybrids, and subspecific taxa examined (Table 5), 26 exhibit predominantly wild traits, characterized by DI values less than 0.6, as well as PDILO and PDIrF values below 0.23. Conversely, 6 species demonstrate traits typical of domestication, with DI values equal to or greater than 0.6, and PDILO and PDIrF values equal to or greater than 0.8 and 0.75, respectively. Selected species are illustrated in Fig. 3. Additionally, 14 species display intermediate values, possibly resulting from various stages of the evolutionary process, hybridization, or introgression.

In terms of the distribution of domestication traits, it’s noteworthy to highlight the prevalence of such traits in an American species commonly considered wild: *V. aestivalis* var. *linsecornii* (Buckley) L.H.Bailey, exhibiting a DI of 1 and PDILO and PDIrF values of 0.93. These values are comparable only to those of *V. vinifera* L. and the hybrid *V. vinifera* L. × *V. amurensis* Rupr. (Table 5). Additionally, other taxa such as *Vitis peninsularis* M. E. Jones, *V. acerifolia* Raf., *V. nesbittiana* Comeaux, *V. popenoei* J. H. Fennel, *V. shuttleworthii* House, *V. californica* Benth., *V. ×doaniana* Munson ex Viala, and *V. aestivalis* Michx., exhibit intermediate values. It’s noteworthy that the American grapevine species studied generally demonstrate higher domestication index values compared to Asian species (Table 5). However, Asian species like *V. amurensis* Rupr. and notably *V. piasezkii* Maxim. var. *pagnuccii* (Rom. Caill. ex Planch.) Rehder also displays intermediate values close to the “domestication syndrome” (Table 5).

V. rotundifolia, the type species of the Muscadinia section of the *Vitis* genus, shows low domestication index values. Moreover, the domestication index values and probabilities shed light on differences between European and South Caucasian grapevine types and species (Table 5).

In the genus *Ampelopsis*, which is closely related to *Vitis*, seed domestication values are notably low, resembling those observed in typical wild *Vitis* species.

Table 5 Relevant measurements and indices for modern seeds of living *Vitis* species, *proles* and hybrids

S	Taxa	PDILO	PDIrF	Domestication index	L (mm)	B (mm)	Stummer's index	Stalk length (mm)
		Values ($X \pm \sigma_x$)	Values ($X \pm \sigma_x$)	Values ($X \pm \sigma_x$)			Values ($X \pm \sigma_x$)	Values ($X \pm \sigma_x$)
<i>Vitis vinifera</i> hybrids								
I	<i>Vitis caucasica</i> Vavilov \times <i>Vitis vinifera</i> L	0.46 \pm 0.038	0.47 \pm 0.039	0.38 \pm 0.30	5.4	4.1	76.74 \pm 7.22	0.99 \pm 0.28
D	<i>Vitis vinifera</i> L. \times <i>V. amurensis</i> Rupr	0.89 \pm 0.22	0.90 \pm 0.18	0.82 \pm 0.15	6.0	4.2	69.58 \pm 7.19	1.39 \pm 0.26
I	<i>Vitis vinifera</i> L. \times <i>V. sylvestris</i> C.C.Gmelin	0.41 \pm 0.36	0.42 \pm 0.38	0.31 \pm 0.30	5.3	3.8	71.86 \pm 6.86	1.01 \pm 0.24
<i>Domesticated grapevine</i>								
<i>Vitis vinifera</i> L.								
D	Antasiatica	0.99 \pm 0.05	0.99 \pm 0.04	0.96 \pm 0.10	6.5	3.8	59.49 \pm 6.81	1.88 \pm 0.34
D	Caspica	0.98 \pm 0.06	0.98 \pm 0.06	0.94 \pm 0.12	6.7	4.0	60.91 \pm 6.22	1.70 \pm 0.31
D	Pontica	0.93 \pm 0.16	0.93 \pm 0.17	0.82 \pm 0.23	5.9	3.8	65.18 \pm 6.68	1.51 \pm 0.36
D	Occidentalis	0.93 \pm 0.17	0.93 \pm 0.18	0.80 \pm 0.26	5.8	3.7	64.27 \pm 5.37	1.58 \pm 0.41
I	β <i>Vitis vinifera</i> L	0.56 \pm 0.40	0.45 \pm 0.47	0.78 \pm 0.29	5.7	3.6	62.57 \pm 6.78	1.44 \pm 0.41
<i>Wild grapevine</i>								
I	α <i>Vitis caucasica</i> Vavilov	0.50 \pm 0.39	0.47 \pm 0.40	0.45 \pm 0.38	4.9	3.9	80.35 \pm 7.02	1.02 \pm 0.41
W	α <i>Vitis sylvestris</i> C.C.Gmelin	0.06 \pm 0.13	0.05 \pm 0.12	0.05 \pm 0.10	4.8	3.7	76.15 \pm 6.29	0.74 \pm 0.15
W	β <i>Vitis sylvestris</i> C.C.Gmelin	0.02 \pm 0.05	0.01 \pm 0.02	0.15 \pm 0.14	4.2	3.3	76.09 \pm 11.01	1.07 \pm 0.38
<i>American Vitis</i> species used as rootstocks								
W	α <i>Vitis riparia</i> Michx	0.06 \pm 0.14	0.04 \pm 0.09	0.16 \pm 0.23	4.9	3.6	75.57 \pm 10.82	0.58 \pm 0.17
W	α <i>Vitis rupestris</i> Scheele	0.01 \pm 0.03	0.01 \pm 0.02	0.07 \pm 0.09	4.7	3.8	79.92 \pm 5.30	0.47 \pm 0.12
<i>American Vitis</i> species								
D	α <i>Vitis aestivalis</i> var. <i>linsecornii</i> (Buckley) L. H. Bailey	0.93 \pm 0.10	0.93 \pm 0.09	0.90 \pm 0.12	6.4	4.1	64.89 \pm 2.79	1.50 \pm 0.18
D	α <i>Vitis peninsularis</i> M. E. Jones	0.58 \pm 0.33	0.74 \pm 0.30	0.62 \pm 0.26	5.9	4.7	80.36 \pm 9.62	1.10 \pm 0.15
I	α <i>Vitis nesbittiana</i> Comeaux	0.48 \pm 0.41	0.53 \pm 0.44	0.35 \pm 0.37	5.2	3.8	74.01 \pm 7.27	1.01 \pm 0.29
I	α <i>Vitis acerifolia</i> Raf	0.25 \pm 0.30	0.26 \pm 0.31	0.42 \pm 0.18	6.0	4.3	72.89 \pm 5.99	0.87 \pm 0.19
I	α <i>Vitis popenoei</i> J. H. Fennel	0.26 \pm 0.20	0.24 \pm 0.22	0.53 \pm 0.07	6.8	4.6	68.54 \pm 6.80	0.91 \pm 0.10
I	α <i>Vitis shuttleworthii</i> House	0.21 \pm 0.26	0.28 \pm 0.28	0.32 \pm 0.23	5.7	4.4	77.48 \pm 9.59	0.96 \pm 0.11

Table 5 (continued)

S	Taxa	PDILO	PDIrF	Domestication index	L (mm)	B (mm)	Stummer's index	Stalk length (mm)
		Values ($X \pm \sigma_x$)	Values ($X \pm \sigma_x$)	Values ($X \pm \sigma_x$)			Values ($X \pm \sigma_x$)	Values ($X \pm \sigma_x$)
I	<i>αVitis californica</i> Benth	0.20 ± 0.11	0.38 ± 0.13	0.38 ± 0.16	5.7	4.8	84.03 ± 7.88	1.06 ± 0.09
I	<i>αVitis cinerea</i> (Engelm.) Millardet	0.33 ± 0.29	0.40 ± 0.35	0.23 ± 0.20	5.2	4.0	76.16 ± 4.17	1.01 ± 0.14
I	<i>αVitis aestivalis</i> var. <i>aestivalis</i>	0.19 ± 0.27	0.19 ± 0.30	0.37 ± 0.17	5.7	4.0	70.75 ± 3.54	0.79 ± 0.15
I	<i>αVitis × doaniana</i> Munson ex Viala	0.19 ± 0.21	0.26 ± 0.33	0.47 ± 0.07	6.4	4.8	74.68 ± 4.36	0.88 ± 0.16
I	<i>αVitis aestivalis</i> Michx	0.15 ± 0.25	0.17 ± 0.28	0.28 ± 0.23	5.5	3.9	71.19 ± 5.20	0.79 ± 0.19
W	<i>αVitis arizonica</i> Engelm	0.02 ± 0.06	0.03 ± 0.05	0.18 ± 0.15	5.4	4.0	74.61 ± 4.06	0.60 ± 0.09
W	<i>αVitis mustangensis</i> Buckley	0.04 ± 0.03	0.04 ± 0.03	0.45 ± 0.16	6.2	4.7	75.55 ± 6.03	0.75 ± 0.08
W	<i>αVitis labrusca</i> L	0.06 ± 0.12	0.08 ± 0.18	0.08 ± 0.14	4.9	3.6	74.66 ± 3.35	0.71 ± 0.19
W	<i>αVitis monticola</i> Buckley	0.13 ± 0.25	0.15 ± 0.24	0.35 ± 0.21	6.2	5.0	80.69 ± 5.27	0.89 ± 0.18
W	<i>αVitis rotundifolia</i> Michx	0.01 ± 0.01	0.01 ± 0.03	0.24 ± 0.28	5.7	3.6	63.98 ± 7.80	0.52 ± 0.10
W	<i>αVitis palmata</i> Vahl	0.00 ± 0.00	0.01 ± 0.02	0.23 ± 0.12	5.6	4.9	87.94 ± 2.64	0.55 ± 0.11
W	<i>αVitis biformis</i> Rose	0.21 ± 0.25	0.08 ± 0.14	0.08 ± 0.09	4.1	3.3	81.36 ± 7.83	0.79 ± 0.14
W	<i>αVitis blancoi</i> Munson	0.07 ± 0.09	0.06 ± 0.12	0.08 ± 0.14	4.4	3.7	83.48 ± 8.22	0.66 ± 0.15
W	<i>αVitis bloodworthiana</i> Comeaux	0.02 ± 0.03	0.00 ± 0.01	0.00 ± 0.00	4.2	3.7	89.20 ± 5.29	0.66 ± 0.11
W	<i>αVitis cinerea</i> var. <i>helleri</i> (L. H. Bailey) M. O. Moore	0.03 ± 0.08	0.00 ± 0.01	0.07 ± 0.14	5.0	3.8	77.21 ± 8.02	0.69 ± 0.18
W	<i>αVitis girdiana</i> Munson	0.09 ± 0.15	0.11 ± 0.19	0.10 ± 0.14	5.0	3.7	75.13 ± 7.99	0.81 ± 0.14
W	<i>αVitis vulpina</i> L	0.08 ± 0.12	0.09 ± 0.13	0.18 ± 0.20	5.3	4.0	76.39 ± 6.11	0.73 ± 0.26
	<i>Asian vitis species</i>							
D	<i>αVitis amurensis</i> Rupr	0.46 ± 0.34	0.46 ± 0.37	0.48 ± 0.30	5.8	3.8	66.45 ± 8.15	1.00 ± 0.16
D	<i>αVitis piasezkii</i> Maxim. var. <i>pagnuccii</i> (Rom. Caill. ex Planch.) Rehder	0.60 ± 0.26	0.63 ± 0.31	0.53 ± 0.32	5.6	4.0	72.30 ± 5.76	1.10 ± 0.16

Table 5 (continued)

S	Taxa	PDILO	PDIrF	Domestication index	L (mm)	B (mm)	Stummer's index	Stalk length (mm)
		Values ($X \pm \sigma_x$)	Values ($X \pm \sigma_x$)	Values ($X \pm \sigma_x$)			Values ($X \pm \sigma_x$)	Values ($X \pm \sigma_x$)
I	<i>αVitis romanetii</i> Rom. Caill	0.20 ± 0.26	0.22 ± 0.31	0.22 ± 0.27	5.7	4.0	70.47 ± 6.39	0.92 ± 0.19
W	<i>αVitis heyneana</i> Schult	0.21 ± 0.20	0.13 ± 0.20	0.07 ± 0.12	4.6	3.3	72.53 ± 9.55	0.76 ± 0.14
W	<i>αVitis bryonii-folia</i> Bunge	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	4.2	4.3	102.98 ± 2.95	0.43 ± 0.08
W	<i>αVitis ficifolia</i> Bunge	0.02 ± 0.03	0.00 ± 0.01	0.02 ± 0.06	3.5	3.0	86.49 ± 5.64	0.54 ± 0.21
W	<i>αVitis flexuosa</i> Thunb	0.01 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	4.1	3.5	84.77 ± 11.87	0.60 ± 0.09
W	<i>αVitis tsoii</i> Merr	0.01 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	4.3	3.4	78.37 ± 0.00	0.55 ± 0.00
W	<i>αVitis wilsonae</i> hort. Veitch ex R. H. Pearson	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	4.5	3.6	78.83 ± 0.70	0.63 ± 0.71

In addition to the modern materials, and preceded by β , the data from the archaeological samples identified as *Vitis vinifera* in the broad sense and those of *Vitis sylvestris* are summarized to differentiate them from the present-day ones. Abbreviations and Codes: S, Syndromes; D, Domesticated; I, Intermediate; W, Wild. α desiccated modern seeds from wild living taxa, β seeds from living taxa recovered from archaeological levels since the Paleolithic to recent times and identified as wild, β seeds recovered from archaeological levels since the Paleolithic to recent times and identified as domesticated. PDILO, mean domestication probability according to a logistic model. PDIrF, mean domestication probability according to a randomForest model. L, mean length. B, mean breadth

For *Ampelopsis aconitifolia* Bunge, the Domestication Index (DI) stands at 0.02 ± 0.05 , with PDILO and PDIrF both registering at 0.0 ± 0.0 . *Ampelopsis orientalis* (Lam.) Planch. exhibits a DI of 0.0 ± 0.0 , with PDILO at 0.0 ± 0.01 and PDIrF at 0.00 ± 0.0 . *Ampelopsis vitifolia* Planch. and *Ampelopsis glandulosa* var. *brevipedunculata* (Maxim.) Momiy. both display DI, PDILO, and PDIrF values of 0.0 ± 0.0 .

Fossil taxa, domestication index and domestication probability values

Seemingly the oldest *Vitis* seeds were found in the Black Buttes, Wyoming (USA), in connection with or rather in the same clayey sandbank, where *Vitis tricuspudata* leaves are frequent, attributed to an age of 75 Ma and a Maastrichtian chronology, received the name of *Vitis sparsa*, and according to the analyzed image, which is a drawing not a photograph and may not be completely faithful, presents the “domestication syndrome” (Table 6; Fig. 4). However, the attribution to *Vitis* of this seed is clearly in conflict with the Bayesian molecular dating for *Vitis* ancestries

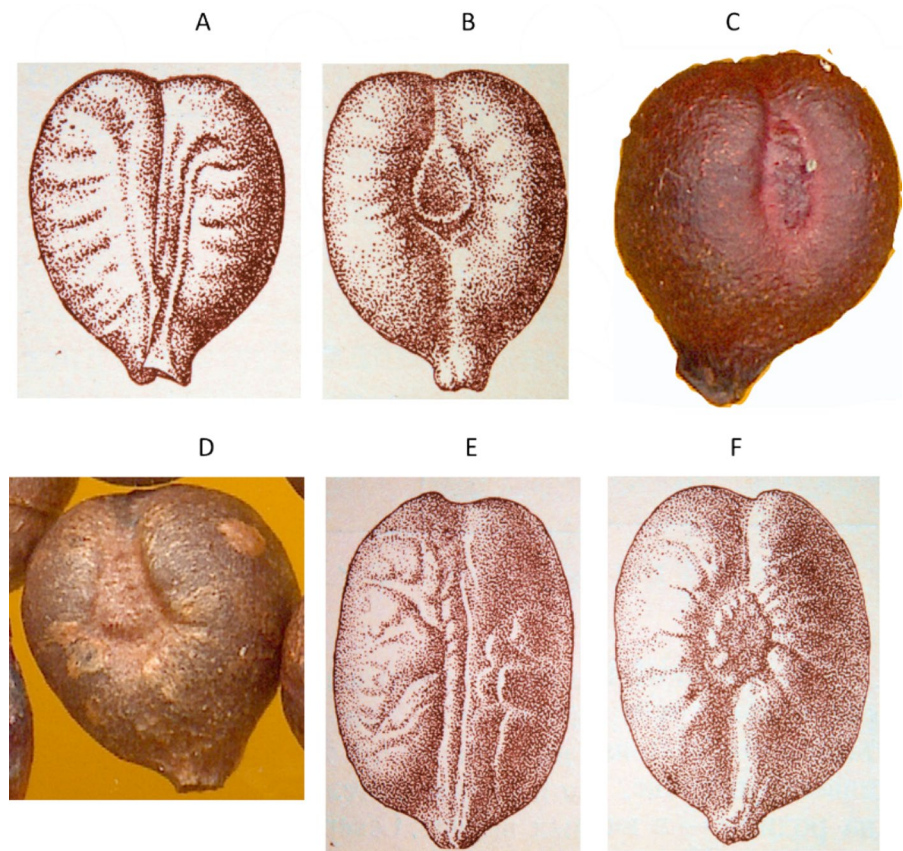
(Liu et al. 2016) where the inferred origin is estimated about thirty Ma later, during the late Eocene (32.6–48.6 Ma).

A more recent discovery comes from the Deccan cherts (sedimentary microcrystalline quartz, SiO₂) of India, dating back to the Late Cretaceous period (70 Ma), where seeds of *Indovitis chittaleya* Manchester and Kapgate which have been found to exhibit distinct wild characteristics (Table 6). However, it's important to note that these seeds do not belong to the *Vitis* genus. Considering the hypothesis of an American origin of the *Vitis* genus followed by its spread to Eurasia during the late Eocene period (30.9–45.1 Ma) (Liu et al. 2016), *Indovitis* would not be an ancestor of the Asian *Vitis* species either.

It's crucial to highlight that only a single seed of *V. lusatica*, discovered in Miocene layers of the salt mine at Wieliczka (Southern Poland) and showing a high probability of being ‘domesticated’, has been studied thus far (Fig. 4; Table 6). Therefore, the interpretation of this finding remains highly uncertain.

It should be noted that most of the fossil seeds show clearly wild characteristics with short beaks

Fig. 3 Modern *Vitis* seeds that present the “Wildness syndrome” opposite to the “Domestication syndrome”. (A, B) *Vitis labrusca* (USA). (C) *V. amurensis* (China). (D), *V. riparia* (U.S.A.). (E, F) *V. rotundifolia* (USA). Images: A, B, E, F, (Negrul 1946b). C, D, Diego Rivera



ranging from 0.1 to 0.8 mm, low combined domestication index values ranging from 0 to 0.5 and very low PDIrF values ranging from 0 to 0.07 (Table 6). Some examples of this type of seeds are shown in Fig. 5.

Discussion

Domestication syndrome and *Vitis vinifera* groups

The Domestication probability distributions based on PDIrF values are highly informative in this respect (Fig. 6) being very useful to characterize the ensembles of Asian, American and European wild species, and to differentiate them from the different geographical groups of cultivated grapevines.

The case of *Vitis* cf. *labrusca* described by Pierce and Tiffney (1986) is very remarkable since the sub-fossil material, c. 700 BC presents domestication characteristics (Table 6, Fig. 4) while the modern

seeds of living *V. labrusca* L. are clearly of the wild type (Table 5). This could suggest the existence of domestication processes that affected some American populations of *Vitis*, and in which Native American populations were involved, prior to the arrival of Europeans, and that possibly became later extinct or reverted to wild forms. An example of these links would be the use by the Cherokee of the fruits of *V. labrusca* as food and its juice as a beverage, as well as its leaves and the bark of its stems as medicine (Moerman 2023). However, we only consistently detected the domestication syndrome in the different geographic groups that include all cultivars of *Vitis vinifera*: *proles Occidentalis*, *Pontica* and *Orientalis*.

Moreover, we detect an east—west gradient in domestication values so that the grapevines of the *proles Orientalis*, that comprises the *Antasiatica* and *Caspica sub-proles*, present slightly higher DI, PDILO and PDIrF values than those of the *proles Pontica* and *Occidentalis* (Table 5, Fig. 6). This may be due to recent differential selective factors that

Table 6 Fossil *Vitis* seeds (selected examples) and their Stummer's and domestication indices, and *PDiLO* and *PDiRF* values

Taxa	Dimensions	Stalk length	Stummer's index	Domestication index	<i>PDiLO</i>	<i>PDiRF</i>	Period	Area
Parameters	L x B x T (mm)	Values (X ± σ _x)	Values (X ± σ _x)	Values (X ± σ _x)	Values (X ± σ _x)	Values (X ± σ _x)		
<i>Wild type</i>								
† <i>Vitis messelensis</i> M.E. Collinson, Manchester and V. Wilde ¹	4.9 × 3.8 × 1.8	0.6 ± 0.1	78.8 ± 9.18	0.03 ± 0.07	0.00 ± 0.00	0.00 ± 0.00	15.97–11.61 Ma	Germany
† <i>Indovitis chittaleya</i> Manchester and Kapgate ¹⁴	5.7 × 3.8 × 3	0.6	66.67	0.17	0.00	0.07	70 Ma	India
† <i>Vitis grayensis</i> Gong, F., Karsai, I., and Liu, Y. S. C. ^{2, 11}	4.1 × 2.7 × 2.1	0.5 ± 0.2	65.66 ± 3.95	0.04 ± 0.08	0.01 ± 0.01	0.00 ± 0.01	7–4.5 Ma	USA
† <i>Vitis lanatoides</i> Gong, F., Karsai, I., and Liu, Y. S. C. ^{2, 11}	4.3 × 3.4 × 2.6	0.6 ± 0.01	79.31 ± 3.1	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	7–4.5 Ma	USA
† <i>Vitis sphaerocarpa</i> Kinkelin in Engelh. ³	5.6 × 4.6 × 3	0.9	82.14	0.17	0.02	0.00	5.3–2.58 Ma	Germany
† <i>Vitis teutonica</i> A. Braun ^{5, 15}	4.1 × 2.7 × 2.1	0.6 ± 0.2	67.37 ± 7.27	0.10 ± 0.12	0.27 ± 0.35	0.05 ± 0.09	30–24 Ma	Germany
† <i>Vitis subglobosa</i> Reid and Chandler ⁶	5 × 5 × 4.4	0.1 ± 0.1	100 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	57.8–36.6 Ma	England
† <i>Vitis semenlabruscoidea</i> Reid and Chandler ⁶	4 × 3.5 × 2.5	0.3 ± 0.1	87.5 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	57.8–36.6 Ma	England
† <i>Vitis minuta</i> Reid and Chandler ⁶	3.5 × 2.8 × 2.1	0.2	80	0.00 ± 0.00	0.00	0.00	57.8–36.6 Ma	England
† <i>Vitis bognorensis</i> Reid and Chandler ⁶	4 × 3 × 2.3	0.2	75	0.00 ± 0.00	0.00	0.00	57.8–36.6 Ma	England
† <i>Vitis brachypoda</i> Miki ⁷	6.5 × 5 × 3.7	0.5 ± 0.0	77.38 ± 8.42	0.08 ± 0.12	0.00 ± 0.00	0.00 ± 0.01	5.3–2.58 Ma	Japan
† <i>Vitis labruscoidea</i> Miki ⁷	6.8 × 5.3 × 4.1	0.8 ± 0.1	78.17 ± 3.45	0.25 ± 0.17	0.03 ± 0.02	0.05 ± 0.06	5.3–2.58 Ma	Japan
† <i>Vitis rotundata</i> Miki ⁷	3.1 × 2.5 × 1.9	0.3 ± 0.1	80.86 ± 7	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.01	5.3–2.58 Ma	Japan
† <i>Vitis tomskiana</i> P.I. Dorof ^{8, 16}	3.8 × 2.6 × 2	0.3 ± 0.1	68.42 ± 12.4	0.07 ± 0.09	0.00 ± 0.00	0.00 ± 0.00	30–24 Ma	Russia
† <i>Vitis pseudorotundifolia</i> Berry ^{9, 10}	5.8 × 3.2 × 2.4	0.5 ± 0.1	55.8 ± 3.18	0.33 ± 0.17	0.01 ± 0.01	0.02 ± 0.03	30–24 Ma	USA
<i>Intermediate type</i>								
† <i>Vitis latisulcata</i> Gong, F., Karsai, I., and Liu, Y. S. C. ^{2, 11}	5.1 × 3.4 × 2.6	0.7 ± 0.3	67.34 ± 6.27	0.13 ± 0.08	0.17 ± 0.31	0.21 ± 0.38	7–4.5 Ma	USA
† <i>Vitis eolabrusca</i> Tiffney and Barghoorn ⁹	5.5 × 3.6 × 2.7	0.9	64.73	0.5	0.38	0.51	30–24 Ma	USA
† <i>Vitis plioaenica</i> Kinkelin in Engelh. ³	6.5 × 4.4 × 2.5	1.1	67.69	0.5	0.48	0.28	5.3–2.58 Ma	Germany
<i>"Domesticated" type</i>								
<i>Cf. Vitis labrusca</i> L. ⁴	7.4 × 4.7 × 3.9	1.5	63.34	0.83	0.93	0.80	800–600 BC	USA
† <i>Vitis sparsa</i> Lesq. ¹²	7 × 5 × 5	1.6	71.43	0.83	0.98	0.95	75 Ma	USA
† <i>Vitis lusatica</i> Lesq. ¹³	7.3 × 4.2 × 3.3	1.1	58.34	0.67	0.74	0.86	15 Ma	Poland

Abbreviations and Codes: *L* length, *B* breadth, *T* thickness. *X* Mean. σ_x standard deviation. (†) extinct taxon. *PDiLO* mean domestication probability according to a logistic model. *PDiRF* mean domestication probability according to a random forest model. Data: ¹Collinson et al. 2012. ²Gong 2009. ³Engelhardt and Kinkelin 1908. ⁴Pierce and Tiffney 1986. ⁵Kirchheimer 1938. ⁶Reid and Chandler 1933. ⁷Miki 1956. ⁸Popova and Yakovleva 2009. ⁹Tiffney and Barghoorn 1976. ¹⁰Berry 1910. ¹¹Gong et al. 2010. ¹²Lesquereux 1878. ¹³Zastawniak, 1996. ¹⁴Manchester et al. 2013. ¹⁵Braun 1854. ¹⁶Dorofeev 1962

Fig. 4 Fossil seeds that present relatively high and intermediate “Domestication syndrome” levels. (A) *Vitis sparsa*, Upper Cretaceous, Wyoming (USA). (B) *V. pliocaenica*, Pliocene, Frankfurt (Germany). (C) *V. lusatica*, Miocene layers of the salt mine at Wieliczka (S. Poland). (D) *V. vinifera* type (sub. *V. sylvestris*), Atlantic postglacial, Stuttgart (Germany); (E) cf. *V. labrusca*, Holocene, Connecticut (USA). Images: A, (Lesquereux 1878). B, (Engelhardt and Kinkelin 1908). C, (Zastawniak 1996). D, (Kirchheimer 1938). E, (Tiffney and Barghoorn 1976)

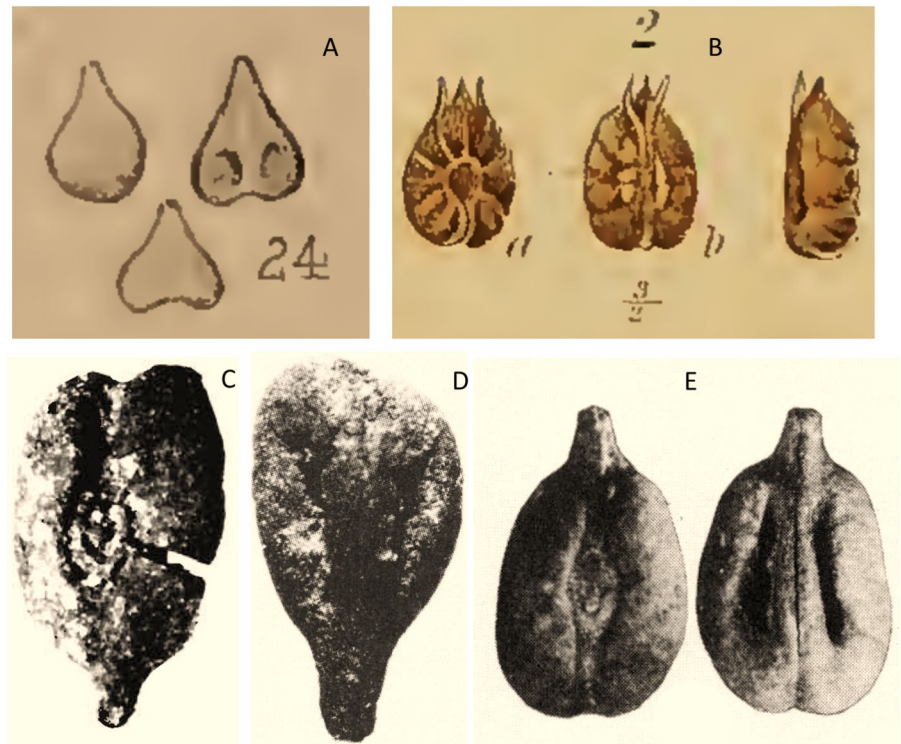
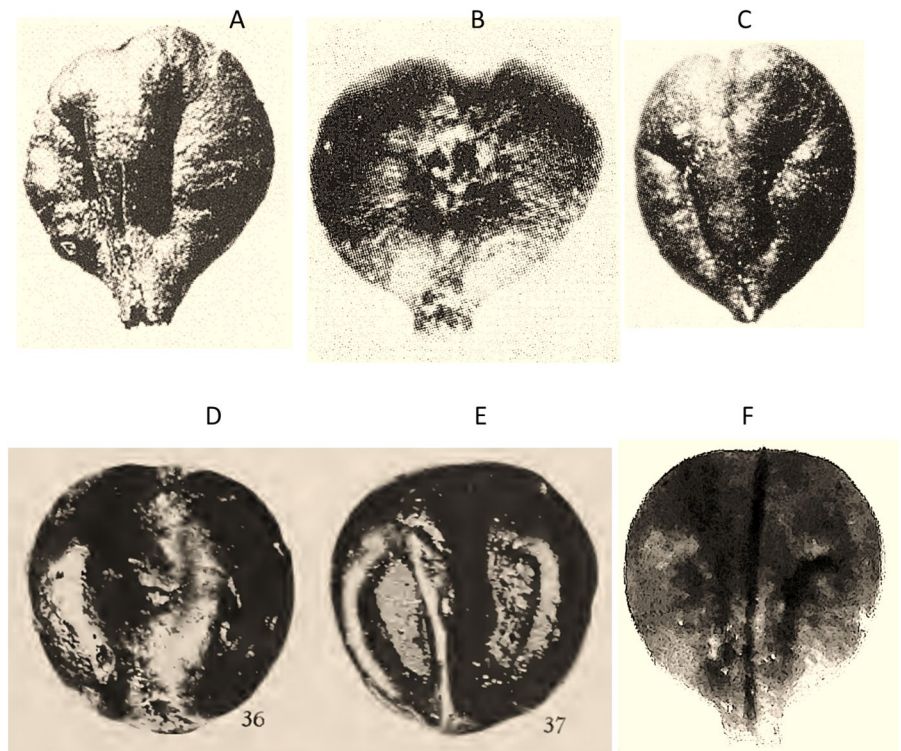


Fig. 5 Fossil seeds that present the “Wildness syndrome”. (A) *Vitis sylvestris*, Upper Pliocene, Limburg (Netherlands). (B) *V. sylvestris*, Upper Pliocene, Kroscienko (Poland). (C) *V. sylvestris*, Upper Pliocene, Limburg (Netherlands). (D, E), *V. subglobosa* (p.p.), Ypresian, Eocene, London Clay (England). (F) *V. labruscoidea*, Pliocene, *Pinus trifolia* bed of central Hondo (Japan). Images: A, B, C, (Kirchheimer 1938). D, E, (Reid and Chandler 1933). F, (Miki 1956)



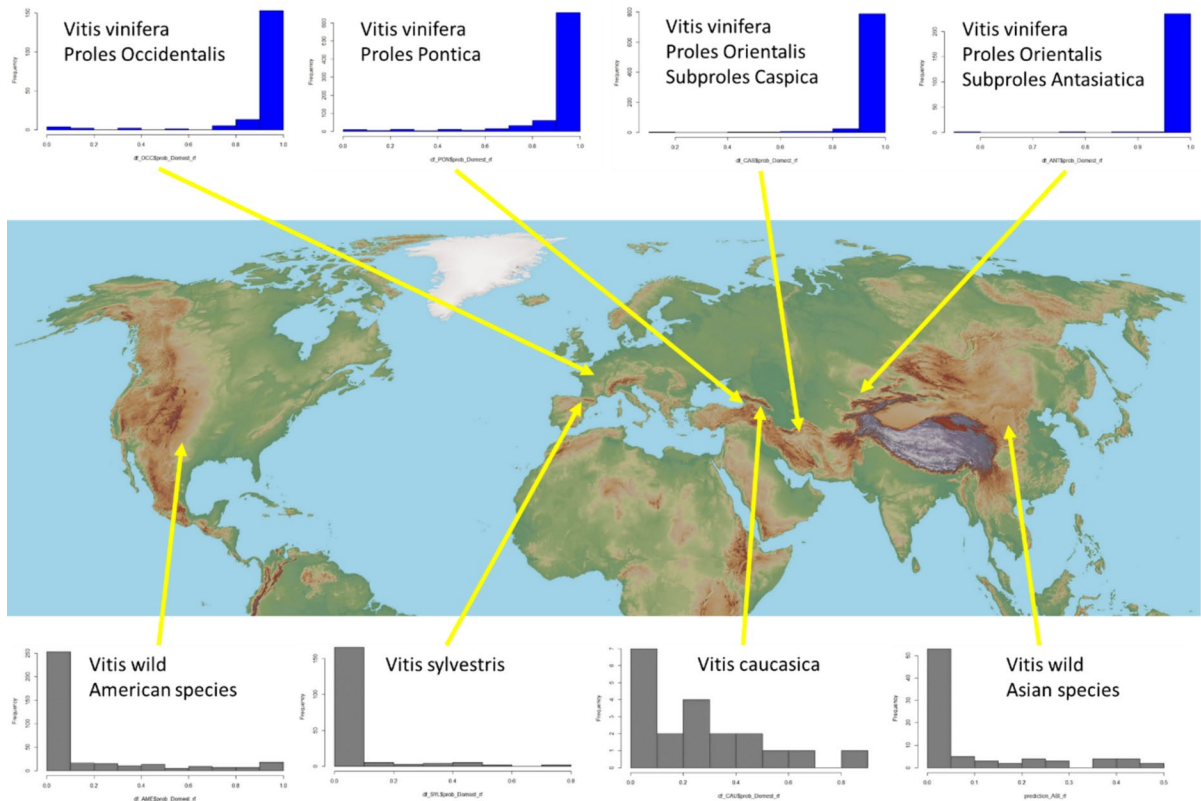


Fig. 6 Distribution of the domestication probability, estimated with randomForest method, within the main geographical groups of domesticated and wild grapevines. Notice: Please note that the units on the x-axis represent probabilities ranging from 0 to 1 overall, but in each individual graph, the scale changes based on the minimum and maximum values of probability. Similarly, the y-axis frequency is represented by the

number of seeds analyzed, so the scale of the y-axis varies in each graph depending on the quantity of seeds analyzed. Total modern seeds analyzed per category: *Proles Orientalis Subproles Antasiatica* 239, *Proles Orientalis Subproles Caspica* 831, *Proles Occidentalis* 180, *Proles Pontica* 808, American wild species 356, Asian wild species 80, *V. sylvestris* 627, *V. caucasica* 20

favor large grape berries with elongated seeds typical of table grapes and raisins, which are predominant in the *Antasiatica sub-proles*. But, more interestingly, it may be related to the chronology gradient of domestication within *V. vinifera*, which would be the oldest in Central Asia and gradually more modern towards the West. The study of Myles et al. (2011) of genome-wide association and genomic selection reflects a clear west–east gradient in *Vitis sylvestris* that is recapitulated in the *V. vinifera* cultivars and that all analyzed *V. vinifera* populations are genetically closer to eastern *V. sylvestris* than to western *V. sylvestris*. However (Arroyo-García et al. 2006) report that over 70% of the Iberian Peninsula cultivars display chlorotypes that are only compatible with their having

derived from western *V. sylvestris* populations, but it could be a more recent domestication event (Dong et al. 2023).

Considering the case detected in the wild grapevines of the South Caucasus, with higher DI and PDIrF values (Kikvadze et al. 2024), it could not also be ruled out that the ancestral grapevines of Central Asia had “more domesticated” starting traits. However, the seeds of six of the nine Asian species analyzed showed wild traits (PDIrF 0–0.13), excepted *Vitis amurensis* Rupr., and *V. piasezkii* Maxim. var. *pagnuccii* (Rom. Caill. ex Planch.) Rehder, with clear traces of the “domestication syndrome” (PDIrF 0.5–0.7), and *V. romanetii* Rom. Caill., which is somewhat intermediate (PDIrF 0.22 ± 0.31) (Table 5).

The seeds of the interspecific hybrids (*V. vinifera* x *V. sylvestris*) and the samples from feral individuals present intermediate values (Table 5).

Did domestication traits precede domestication in *Vitis* species?

In the case of fossils, we dealt with extinct *Vitis* species. Most are based on leaf imprints recovered from different geological levels, but others are based on fossil seeds. The boundaries between *Vitis* and other *Vitaceae* fossil genera are difficult to delimit and the same difficulty is met when deciding whether living species are present in the fossil register or not. Using a method based on conditional probability, Parins-Fukuchi (2018) assigned a series of *Vitaceae* fossils based on morphometric evidence within the cladistics of the *Vitaceae* with satisfactory results, but they show that the assignment made by palaeontologists of fossil remains, in particular seeds, to a specific genus is uncertain, especially when considering *Ampelocissus*, *Cissus*, or *Vitis* boundaries. So, we have followed the consensus of nomenclature and the inclusion criteria of IFPNI (2023) database.

Wan et al. (2013) based on the study of nuclear DNA and Liu et al. (2016), on plastidial and nuclear DNA, deduced that *Vitis* originated in the New World. But their dates disagree, for Liu et al. (2016) it occurred during the late Eocene (39.4 Ma), and then *Vitis* migrated to Eurasia in the late Eocene (37.3 Ma), when the segregation of the subgenera *Muscadinia* and *Vitis* occurred. For Wan et al. (2013) these events are 10 Ma more recent and occurred during the Oligocene. The separation between American and Eurasian grapevine species occurred later, so the divergence of Eurasian and American grapevines would occur at the earliest between 30 and 25 Ma in the model of Liu et al. (2016) and in the Miocene (11 Ma) for the model of Wan et al. (2013). What seems evident in both models is that *Vitis vinifera* is closer to Asian than to American species, and in particular it is close to *V. heyneana* (= *V. Jacquemontii*). However, *Vitis vinifera* can hybridize successfully with all species of the subgenus *Vitis*.

The existence of fossil *Vitis* taxa showing the “domestication syndrome” in their seeds (Table 5, Fig. 4) would suggest that the possibility of expression of the syndrome was present in the evolutionary characteristics of the genus and its expression

is modulated by the selective effect of environmental factors (such as rainfall and temperature) and by the preferences of the dispersion vectors (humans but also many other species of fauna) that transport the seeds. This has evolved throughout the different geological periods (Fig. 7), being particularly high in present living population while presenting a very marked minimum during the glaciations.

A decrease in index values is seemingly detected during the Eocene, especially 50–40 Ma, followed by a stabilized increase, throughout the Oligocene (33.7–23.8 Ma), Miocene (23.8–5.3 Ma) and Pliocene (5.3–1.8 Ma) (Fig. 7). However, these apparent oscillations are due to doubtful samples (marked with red stars in Fig. 7), so it is evident that the domestication syndrome is later than the last glaciation (Fig. 7).

Vitis fruits containing seeds are known to have been consumed by mammals, specifically the palaeotheres *Eurohippus* (47.4–37.7 Ma), a sister taxon of the equids (true horses). It is quite possible that these palaeotheres dispersed the seeds, as they are found intact in the gut and not fragmented (Collinson et al. 2010).

The typical seed morphology of *Vitis* appears to be earlier in nearly 30 Ma than the putative date of origin of the genus. *Vitis* is inferred by Liu et al. (2016) to have originated in the New World during the late Eocene (32.6–48.6 Ma), then migrated to Eurasia in the late Eocene (30.9–45.1 Ma) which contrasts with the earlier age attributed to *V. sparsa* (Table 6).

The least presence of “domesticated” traits (in terms of maximum and mean values) in *Vitis* seeds is reached at the beginning of the last Quaternary glaciation, in the Pleistocene, 100 kyr, but the decrease in values begins two million years earlier, associated with the previous glacial, period, 2.4–2.1 Ma, at the beginning of the Quaternary. It is very important to underline that seeds with fully domesticated traits (Value 1 of the domestication index), with the exception of *Vitis sparsa* in the Late Cretaceous, do not appear until the beginning of the Holocene, after the last glaciation, 10 kyr (Fig. 4).

In short, *Vitis* fossil seeds present a clearly wild syndrome, characterized by rounded, relatively wide seeds and with a very short beak or stalk: mean length 0.5 mm, standard deviation 0.27 mm (Table 6, Fig. 4).

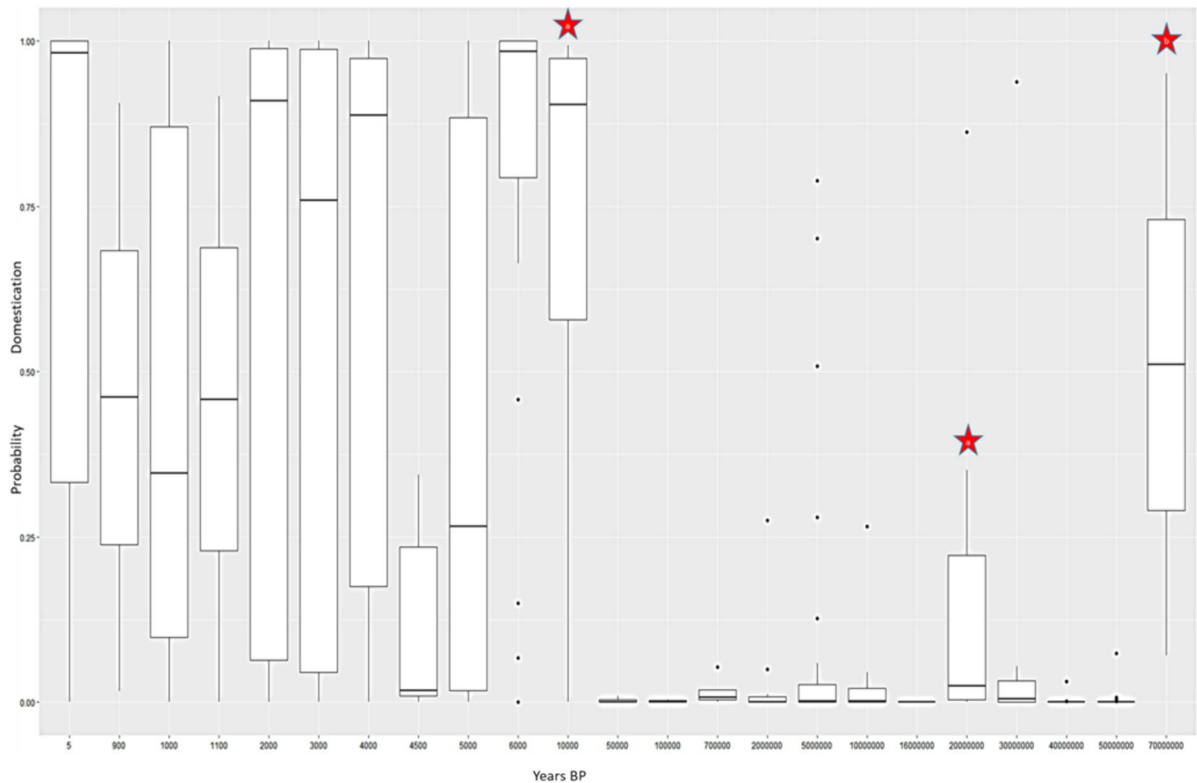


Fig. 7 Chronology of the “Domestication syndrome” in terms of PDIrF. Note: Abbreviations. Years BP, years before present. Box diagram: Whiskers join the boxes with the upper and lower ends, the boxes indicate the upper quartile, the median

The oldest *Vitaceae* seeds linked to human activity were found in the Near East as part of the Acheulian diet at Gesher Benot Ya’aqov, Israel, 700 kyr (Melamed et al. 2016). The seed whose image was analyzed is small (3.4×2.6 mm, stalk length 0.6 mm) and the Stummer’s index value, 74.5, is intermediate between wild and domesticated (Stummer 1911; Valera et al. 2023), the PDILO and PDIrF values are extremely low, 0.07 and 0.00 respectively, clearly pointing to a wild taxon in coincidence with the $DI=0.00$ value. Furthermore, Valera et al. (2023) suggested the close relationship of these seeds with those of endemic *Ampelopsis* species of West Asia.

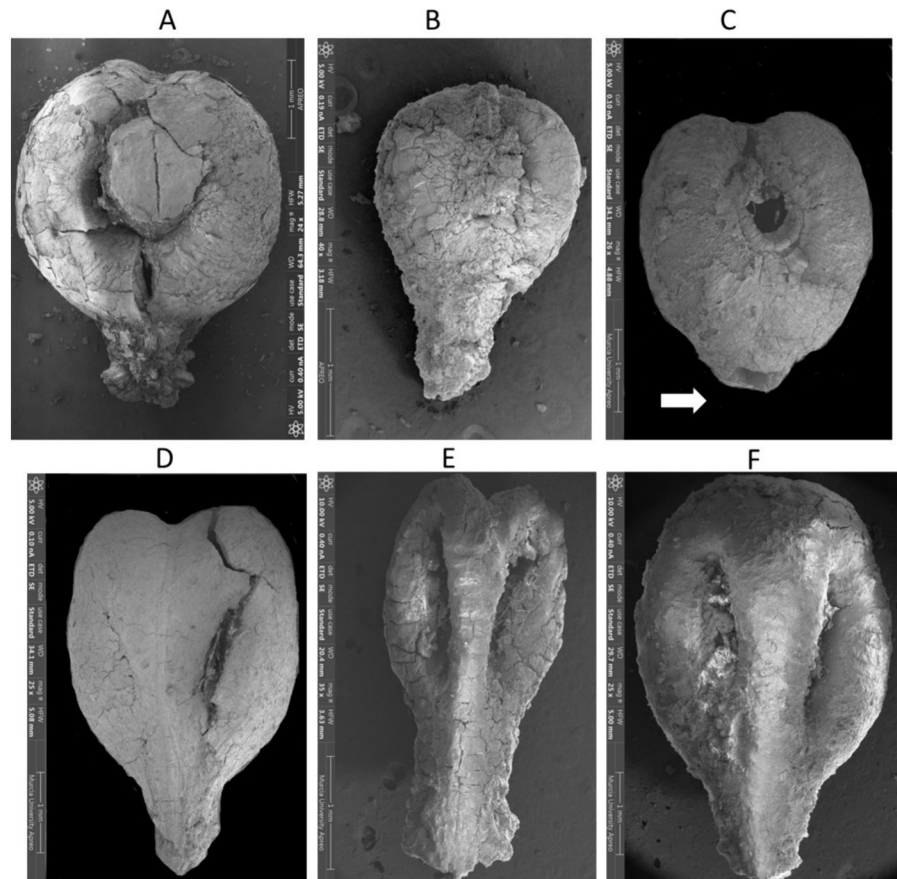
The mean values of the recorded domestication syndrome follow a geographical rather than a chronological pattern (Fig. 6). The highest values occur in Italy and the Caucasus, while the lowest values were recorded in India, Denmark and Central Asia, despite the more recent age of the deposits. It should be

and the lower quartile. The isolated points correspond to Outlier unique data points. Red stars mark doubtful data due to the morphological evidence available or datation. This is the case of several seeds studied from the South Caucasus

noted here that Class 1 and Class 2 seeds from Late Bronze Sapalli levels (Chen et al. 2022) show clearly domesticated traits, although we have not been able to include them in the present study. Other seeds show clearly wild traits and it is even possible that they are, in Class 4, endemic Asiatic species of *Vitis*.

Zhou et al. (2017) based on genomic data from wild and cultivated grapevine samples concluded that domestication can occur rapidly, instead of over millennia. They estimated that *V. sylvestris* and *V. vinifera* diverged c. 22,000 years ago and that the *V. vinifera* lineage experienced a steady decline in population size thereafter. The long decline may reflect low-intensity management by humans before domestication. According to Coito et al. (2019), *V. vinifera* and *V. sylvestris* diverged from a hermaphrodite and hexaploid common ancestor with basic chromosomal number $x=7$, around 22,000 to 30,000 bp, long before the domestication process started. This could

Fig. 8 Archaeological seeds with different degrees the “Domestication syndrome”. (A) *V. sylvestris*, Middle Bronze Age: 185AR Qara Quzaq 1 137–92 (Syria). (B) *V. sylvestris*, Medieval, Alcazaba de Badajoz 2022/4 (Spain). (C), *V. vinifera*, Assyrian 262 Tell Khâmis as 85 5 (Syria) (the beak was broken during the handling for SEM). (D), *V. vinifera*, Assyrian: 266 Tell Khâmis as 108 1 (Syria). (E). *V. vinifera*, Medieval 576 La Graja 2021/2_5 (Spain); (F), *V. vinifera*, Medieval, 646 Fortaleza Isso 2021/7a (Spain). SEM Images: J. Valera and D. Rivera



explain the setting of “domestication syndrome” in terms of high DI, PDILO and PDIrF values, associated to *V. vinifera*, long before cultivation started (Fig. 7). However, if we consider the exclusion of doubtful samples, it seems that the emergence of the domestication syndrome is associated with the development of the exploitation of the species by humans (Fig. 7).

Later on, *V. vinifera* “cultivated” grapevines became increasingly present in anthropogenic contexts (Fig. 4).

Concerning the identification of genes that contribute to domesticated phenotypes. In cultivated grapes, Zhou et al. (2017) identified candidate genes that function in sugar metabolism, flower development, and stress responses. In contrast, candidate genes in the wild sample were limited to abiotic and biotic stress responses. The cost of domestication is reflected by cultivated *V. vinifera* grape accessions containing 5.2% more deleterious variants than wild individuals, and these were more often in

a heterozygous state. However, it seems that genes determining the “domestication syndrome” in seeds remained overlooked (Fig. 8).

In general, archaeological *V. vinifera* seeds with the complete domestication syndrome (DI, PDILO and PDIrF values = 1) have been found in every investigated region (Fig. 9), together with seeds, not necessarily in the same site and period, clearly wild (values below 0.4 for DI, PDILO and PDIrF), and even completely wild, with DI, PDILO and PDIrF = 0 (Fig. 9). It should be noted that in the Caucasus all the archaeological seeds analyzed showed traits of domestication, as did the archaeological Italian seeds. Despite the existence of numerous wild populations of grapevine in the South Caucasus nowadays (Ocete et al. 2018; Maghradze et al. 2020) most of these, recently analyzed from Armenia and Georgia, using the seed morphometry show a high degree of domestication, likely involving feral and hybrid individuals (Kikvadze et al. 2024). This coincides with the

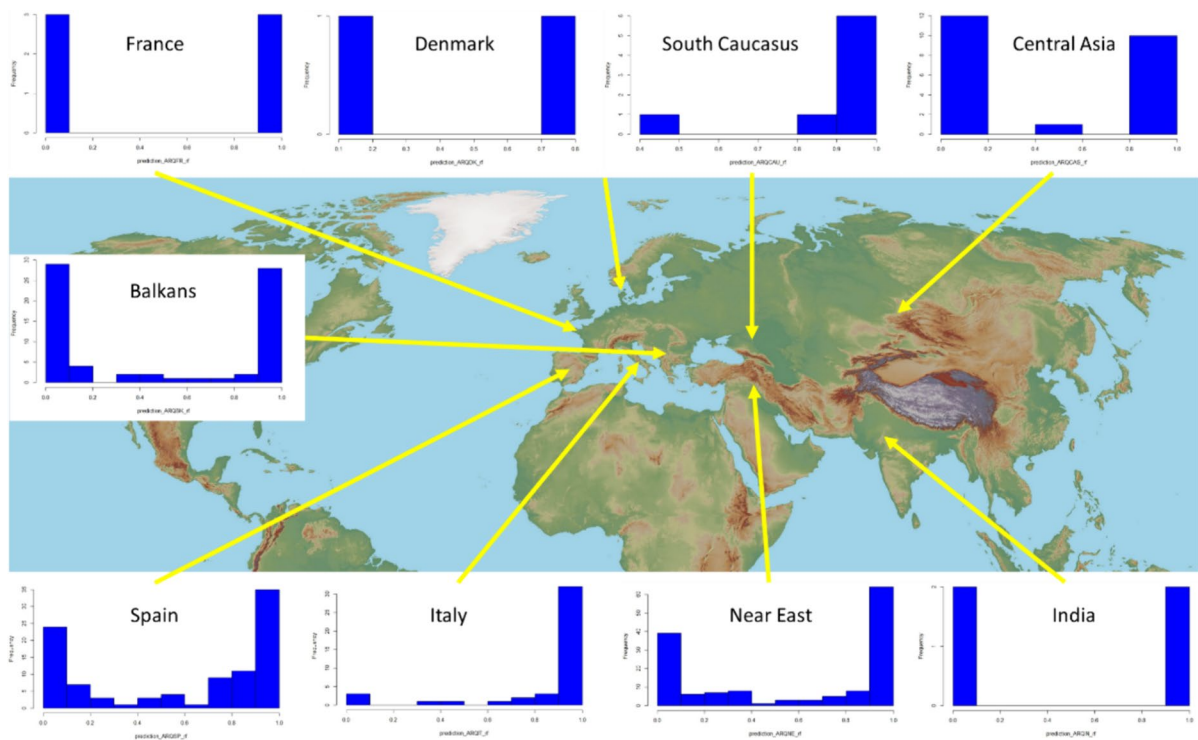


Fig. 9 Archaeology of the “Domestication syndrome”, distribution of probability estimate using random forest (PDIrF). Notice: Please note that the units on the x-axis represent probabilities ranging from 0 to 1 overall, but in each individual graph, the scale changes based on the minimum and maximum values of probability. Similarly, the y-axis frequency is rep-

resented by the number of seeds analyzed, so the scale of the y-axis varies in each graph depending on the quantity of seeds analyzed. Total archaeological seeds analyzed per category: Balkans 70, Central Asia 23, Caucasus 8; Denmark 2, France 6, India 4, Italy 43, Near East 144 and Spain 98

very scarce presence of grapevines with typically wild traits in the archaeological record of the area (Fig. 9).

In addition, populations or sparse individuals of wild grapevine still living in the Georgian natural environments’ present smaller genetic distances with local cultivars than in other European regions. Principal component analysis has also identified special overlapping of the wild set with some cultivars (Imazio et al. 2013).

Conclusion

The domestication syndrome exists in grapevine seeds and is notably characterized by the length of the stalk.

Published domestication indices, based on allometric relationships involving seed length and width, chalaza shield position and stalk length, effectively summarize the information provided by seeds, but unevenly. The combination of the results provided by the six indices into a single index, with values between 0 and 1, represents an advance in the tools available to represent the degree of domestication in numerical terms. The combined domestication index makes it possible to clearly differentiate the wild vines of the *V. sylvestris* – *V. caucasica* complex from the cultivars of *V. vinifera*, regardless of the geographical group of cultivars to which they belong. Its effectiveness is much less significant when it comes to differentiating *V. sylvestris* from American or Asian grapevines.

There are several approaches in terms of conditional probability, based on the availability of a set of

well-characterized reference samples of seeds from clearly domesticated and wild grapevines, which allow estimating the probability that a particular seed comes from a domesticated grapevine. Logistic regression and the randomForest machine learning method provide very similar results that place the threshold for the domestication syndrome at 0.75. At the same time, they show a similar pattern to the combined index.

The randomForest method has made it possible to assess the relevance of the different indices and characters measured, showing the importance of stalk length, the first formula of Mangafa and Kotsakis and the Facsar-Perret index, well above the classic Stummer index.

The domestication syndrome in grapevine seeds, characterized by stalk longer than 1 mm, an elongated shape, and relatively large, is not exclusive to domesticated *Vitis vinifera*, as it is also found in a few American and Asian species of *Vitis*, otherwise considered wild. There are even several fossil species of *Vitis*, now extinct, that show the domestication syndrome in their seeds. In general, however, the domestication syndrome was only clearly established from 2000 BC onward. The age of earlier seeds with domesticated characteristics should be confirmed by specific radiocarbon dating.

Acknowledgements We thank Erika Maul (The Institute for Grapevine Breeding Geilweilerhof, Germany) for sending samples of grapes of contemporary Near Eastern cultivars. We acknowledge the U.S. National Plant Germplasm System (United States Department of Agriculture, Agricultural Research Service), Rancho Santa Ana, Istituto ed Orto Botanico di Palermo, Botanisches Garten Johannes Gutenberg—Universität Mainz, Smithsonian Institution, and Giardino Botanico di Padova for providing us with seeds of American and Asian *Vitis* species and related genera. Also, to the curators for collections of Rioja grapevines (Spain) in Mendavia, that of La Casa de las Vides nurseries in Agullent (Valencia, Spain), that of Rojas Clemente grapevines of the Royal Botanical Garden of Madrid (Spain) and that of the Istituto Agrario di San Michele all'Adige (Trentino, Italy) for allowing us to sample leaves, shoots and grapes. Wild populations were sampled by the authors and Encarna Carreño.

Author contributions Conceptualization, D.R. and C.O.; methodology, D.R.; software, D.R., D.J.R.O, J.A., J.A.P.; validation, F.A., D.M., M.K., C.A.O., R.O., and A.N.; formal analysis, D.R., D.J.R.O, J.A., J.A.P; investigation, J.V., G.M., D.R.; field sampling, C.A.O, R.O., A.N., M.K., D.M.; data curation, D.R., and C.O.; writing—original draft preparation, D.R.; writing—review and editing, C.O. and D.R.;

visualization, D.R.; supervision, C.O.. All authors have read and approved the final version of the manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. AN was funded by a MSCA individual fellowship (IF-EF-ST/0685-896290, GRAPEINNOVATION).

Declarations

Conflict of interest The authors declare no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Applequist WL (2023) Report of the nomenclature committee for vascular plants: 73. *Taxon* 72(1):179–204
- Armenian Vitis Database (2007). Armenian Vitis Database <http://www.vitis.am/eng>. Accessed 12 December 2023
- Arnold C (2002) Ecologie de la vigne sauvage, *Vitis vinifera* L. ssp. *silvestris* (Gmelin) Hegi, dans les forêts alluviales et colluviales d'Europe. *Geobot Helv* 6:1–256
- Arroyo-García R, Ruiz-García L, Bolling L, Ocete R, López MA, Arnold C, Ergul A, Söylemezo L, Uzun HI, Cabello F, Ibáñez J (2006) Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. *Mol Ecol* 15(12):3707–3714
- Berry EW (1910) Additions to the pleistocene flora of New Jersey. *Torreyia* 10(12):261–267
- Bonhomme V, Picq S, Ivorra S, Evin A, Pastor T, Bacilieri R, Lacombe T, Figueiral I, Terral JF, Bouby L (2020) Eco-evo implications and archaeobiological perspectives of trait covariance in fruits of wild and domesticated grapevines. *PLoS ONE* 15(11):e0239863. <https://doi.org/10.1371/journal.pone.0239863>
- Braun A (1854) Einige Beiträge zur flora der tertiär-zeit. *Neues Jahrbuch Für Mineralogie, Geognosie, Geologie* 22(1824):138–147
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Breiman L, Cutler A, Liaw A, Wiener M (2022) Package 'randomForest' <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>

- Chen I, Manchester SR (2011) Seed morphology of *Vitaceae*. *Int J Plant Sci* 172(1):1–35
- Chen G, Zhou X, Khasanov M, Spengler RN, Ma J, Annaev T, Kamarov N, Maksudov F, Wang J, Askarov A, Li X (2022) Morphotype broadening of the grapevine (*Vitis vinifera* L.) from oxus civilization 4000 BP. *Central Asia. Sci Rep* 12(1):16331. <https://doi.org/10.1038/s41598-022-19644-0>
- Chu K (1999) An introduction to sensitivity, specificity, predictive values and likelihood ratios. *Emerg Med* 11(3):175–181. <https://doi.org/10.1046/j.1442-2026.1999.00041.x>
- Coito JL, Silva HG, Ramos MJ, Cunha J, Eiras-Dias J, Amâncio S, Costa MM, Rocheta M (2019) *Vitis* flower types: from the wild to crop plants. *PeerJ* 7:e7879. <https://doi.org/10.7717/peerj.7879>
- Collinson M, Manchester S, Wilde V, Hayes P (2010) Fruit and seed floras from exceptionally preserved biotas in the European Paleogene. *Bull Geosci* 85(1):155–162
- Collinson ME, Manchester S, Wilde V (2012) Fossil fruits and seeds of the Middle Eocene Messel biota, Germany. In: *Abhandlungen der Senckenberg Gesellschaft für Naturforschung Band 570*, Schweizerbart science publishers, Stuttgart
- Couronné R, Probst P, Boulesteix AL (2018) Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics* 19:1–14
- Cran R Project (2023) The comprehensive R archive network. <https://cran.r-project.org/>. Accessed 18 December 2023
- Dong Y, Duan S, Xia Q, Liang Z, Dong X, Margaryan K, Musayev M, Goryslavets S, Zdunić G, Bert PF, Lacombe T (2023) Dual domestications and origin of traits in grapevine evolution. *Science* 379(6635):892–901
- Dorofeev PI (1962) Megaspory, semena i plody iz tretichnykh otlozheny [Megaspores, seeds and fruits from the tertiary sediments. Biostratigraphy of mesozoic and tertiary deposits of Western Siberia]. *Trudy Sibirsk. Nauchno-Issled. Inst. Geol. Geofiz. Miner. Syr.* 22(1):369–415 ((In Russian))
- Engelhardt H, Kinkel E (1908) I. Oberpliocäne flora und fauna des untermaintales: insbesondere des frankfurter klärbeckens. II. Unterdiluviale flora von hainstadt a. M. Im selbstverlage der senckenbergischen. Naturforschenden Gesellschaft 29(3):151–296
- European *Vitis* Database (2023) <http://www.eu-vitis.de>. Accessed 10 December 2023
- Facsar G (1970) Összehasonlító morfológiai vizsgálatok kerti szőlőfajták magjain i (vergleichende morphologische untersuchungen der samen von gartenrebsorten. I.). *Botanikai Közlemenyek* 57:221–231
- Facsar G, Jerem E (1985) Zum urgeschichtlichen weinbau in mitteleuropa. rebkernfunde von *vitis vinifera* l. aus der urnen-felder-, hallstatt-und latenezeitlichen siedlung sopron-krautacker. *Wissenschaftliche Arbeiten Aus Dem Burgenland* 71:121–144
- Ferrer-Gallego PP, Ferrer-Gallego R, Laguna E, Pipia I (2019) (2682) Proposal to conserve the name *Vitis sylvestris* CC Gmel. (*Vitaceae*) against *V. sylvestris* W. Bartram. *Taxon* 68(2):409–410. <https://doi.org/10.1002/tax.12043>
- FNGR (2023) French network of grapevine repositories, version 4 © 2017–2019 INRA - IFV - montpellier supagro, database of the collections https://bioweb.supagro.inra.fr/collections_vigne/SearchS.php?l=EN. Accessed 18 December 2023
- GBIF (2023) GBIF | Global biodiversity information facility, free and open access to biodiversity data <https://www.gbif.org/>. Accessed 18 December 2023
- Gong F, Karsai I, Liu YSC (2010) *Vitis* seeds (*Vitaceae*) from the late neogene gray fossil site, Northeastern Tennessee, USA. *Rev Palaeobot Palynology* 162(1):71–83
- Gong F (2009) *Vitis* Seeds (*Vitaceae*) from the late neogene gray fossil site, Northeastern Tennessee, USA. Department of biological sciences, East Tennessee State University. Electronic theses and dissertations. Paper 1819. <https://dc.etsu.edu/etd/1819> (last accessed 18/12/2023)
- GRIN (2023) Search taxonomy data in GRIN-Global. <https://npgsweb.ars-grin.gov/gringlobal/taxon/taxonomysearch> Accessed 18 December 2023
- Iamónico D, Galasso G, Banfi E, Ardenghi NM (2015) Typification of Linnaean names in the genus *Vitis* (*Vitaceae*). *Taxon* 64(5):1048–1050
- IFPNI (2023) The International fossil plant names index (IFPNI). <http://ifpni.org/index.htm> Accessed 14 January 2023
- Imazio S, Maghradze D, De Lorenzis G, Bacilieri R, Laucou V, This P, Scienza A, Failla O (2013) From the cradle of grapevine domestication: molecular overview and description of Georgian grapevine (*Vitis vinifera* L.) germplasm. *Tree Genet Genomes* 9:641–658
- IPGRI-UPOV-OIV (1997) Descriptors for grapevine (*Vitis* spp.). International Union for the protection of new varieties of plants, Geneva, Switzerland/Office International de la Vigne et du Vin, Paris, France/International Plant Genetic Resources Institute, Rome, Italy. https://www.ecpgr.cgiar.org/fileadmin/bioiversity/More_pubs/393_Descriptors_for_grapevine__Vitis_spp_.pdf Accessed 15 October 2023
- Iriarte MJ, Salinas JA, Pérez MA, Ocete E, Ocete R (2013) La vid silvestre, un taxón amenazado en diversos ecosistemas ibéricos. *Cuadern Soc Esp Cienc for* 38:143–148
- Jacquat C, Martinoli D (1999) *Vitis vinifera* L.: wild or cultivated? Study of the grape pips found at Petra, Jordan 150 BC–AD 40. *Veg Hist Archaeobot* 8(1):25–30
- Kikvadze M, Valera J, Obón C, Ocete R, Ocete CA, Kikilashvili S, Maghradze D, Rivera-Obón DJ, Rivera D (2024) Wild in press. Wild grapevines of Georgia and their relationship with local cultivars: a Bayesian analysis of seed morphology. *Vitis* XX: XXX.
- Kirchheimer F (1938) Beiträge zur näheren kenntnis von vitaceen-samenformen tertiären alters. *Planta* 28(4):582–598
- Lesquereux L (1878) Contributions to the fossil flora of the Western territories: the tertiary flora, vol 2. US Government Printing Office, Washington
- Levadoux L (1956) Les populations sauvages et cultivées de *Vitis vinifera* L. *Ann L'amelioration Des Plantes* 1:59–118
- LINN (2023a) The Linnean collections, LINN 281.1 *Vitis vinifera* (Herb Linn). Available on: <https://linnean-online.org/2850/> Accessed 18 December 2023
- LINN (2023b) The Linnean collections, LINN 281.3 *Vitis corinthiaca* (Herb Linn). Available on: <https://linnean-online.org/2852/> Accessed 18 December 2023

- Liu Y, Wang Y, Zhang J (2012) New machine learning algorithm: random forest. In: Liu B, Ma M, Chang J (eds) ICICA 2012, LNCS 7473. Springer-Verlag, Berlin and Heidelberg, pp 246–252
- Liu XQ, Ickert-Bond SM, Nie ZL, Zhou Z, Chen LQ, Wen J (2016) Phylogeny of the *Ampelocissus-Vitis* clade in *Vitaceae* supports the new World origin of the grape genus. *Mol Phylogenet Evol* 95:217–228
- Logothetis B (1970) Η ανάπτυξη της αμπέλου και της αμπελοουργίας στην Ελλάδα με βάση τα αρχαιολογικά ευρήματα στην περιοχή (the development of the vine and of viticulture in Greece based on archaeological findings in the area, in Greek with English summary). *Epistimoniki Epetiris Tis Georonikis Kai Dasologikis Sholis, University of Thessaloniki* 13:167–249
- Logothetis B (1974) Η συμβολή του αμπελιού και του κρασιού στον πολιτισμό της Ελλάδας και της Ανατολικής Μεσογείου (The contribution of the vine and the wine to the civilization of Greece and Eastern Mediterranean, in Greek with French summary). *Epistimoniki Epetiris Tis Georonikis Kai Dasologikis Sholis, University of Thessaloniki* 17:5–286
- Maghradze D, Ocete ME, Pérez-Izquierdo MA, Chipashvili R, Ocete CA, Ocete R (2010) El patrimonio vitícola de Georgia el estado sanitario de sus poblaciones, silvestres. XXXI Jornadas de Viticultura y Enología de Tierra de Barros. Centro Universitario de Santa Ana, Almedralejo. Spain, pp 113–126
- Maghradze D, Melyan G, Salimov V, Chipashvili R, Íñiguez M, Puras P, Melendez E, Vaca R, Rivera D, Obón C, Valle Melón JM (2020) Wild grapevine (*Vitis sylvestris* CCGmel) wines from the Southern Caucasus region. *OENO One* 54(4):849–862
- Manchester SR, Chen I, Lott TA (2012) Seeds of *Ampelocissus*, *Cissus*, and *Leea* (Vitales) from the paleogene of western Peru and their biogeographic significance. *Int J Plant Sci* 173(8):933–943
- Manchester SR, Kapgate DK, Wen J (2013) Oldest fruits of the grape family (*Vitaceae*) from the late cretaceous deccan cherts of India. *Am J Bot* 100(9):1849–1859
- Mangafa M, Kotsakis K (1996) A new method for the identification of wild and cultivated charred grape seeds. *J Archaeol Sci* 23(3):409–418
- Martín-Gómez JJ, Gutiérrez del Pozo D, Uchescu M, Bacchetta G, Sáenz C, de Santamaría F, Tocino Á, Cervantes E (2020) Seed morphology in the *Vitaceae* based on geometric models. *Agronomy* 10(5):739. <https://doi.org/10.3390/agronomy10050739>
- Maul E (2023) *Vitis* International Variety Catalogue VIVC www.vivc.de Accessed December 2023
- Melamed Y, Kislev ME, Geffen E, Lev-Yadun S, Goren-Inbar N (2016) The plant component of an acheulian diet at gesher benot ya' aqov. *Israel Proc Natl Acad Sci u s a* 113(51):14674–14679. <https://doi.org/10.1073/pnas.1607872113>
- Miki S (1956) Seed Remains of *Vitaceae* in Japan. *J Inst Polytech Osaka City Univ, Ser d, Biol* 7:247–271
- Moerman D (2023) *Vitis labrusca* in native American ethnobotany. A database of foods, drugs, dyes and fibers of native American Peoples, Derived from plants. <http://naeb.brit.org/uses/search/?string=Vitis+labrusca> Accessed 18 December 2023
- Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia JM, Ware D, Bustamante CD (2011) Genetic structure and domestication history of the grape. *Proc Natl Acad Sci U S A* 108(9):3530–3535. <https://doi.org/10.1073/pnas.1009363108>
- Negrul AM (1946a) Proiskhozhdeniye kulturnogo vinograda i yego klassifikatsiya [Origin of the cultivated grapevine and its classification]. In: Varanov A (ed) *Ampelografia SSSR [Ampelography of the USSR]*, vol 1. Pischepromizdat, Moscow (in Russian), pp 159–216
- Negrul AM (1946b) Semeitsvo vitaceae lindley (ampelideae kunth) [familia vitaceae lindley (ampelideae kunth)]. In: Varanov A (ed) *Ampelografia SSSR [Ampelography of the USSR]*, vol 1. Pischepromizdat, Moscow (in Russian), pp 45–132
- Nie ZL, Sun H, Manchester SR, Meng Y, Luke Q, Wen J (2012) Evolution of the intercontinental disjunctions in six continents in the *Ampelopsis* clade of the grape family (*Vitaceae*). *BMC Evol Biol* 12(1):1–13
- Ocete R, Cantos M, López MA, Gallardo A, Pérez MA, Troncoso A, Lara M, Failla O, Ferragut FJ, Liñán J (2007) Caracterización y conservación del recurso fitogenético vid silvestre en Andalucía. *Consejería de medio ambiente. Junta de Andalucía Falcor, Jerez de la Frontera*
- Ocete R, Rivera D, Maghradze D, Salimov V, Melyan G, Musayev M, Ocete CA, Chipashvili R, Failla O, Obón C (2018) Support trees and shrubs for the Eurasian wild grapevine in Southern Caucasus. *Ann Agric Sci* 16(4):427–431. <https://doi.org/10.1016/j.aasci.2018.06.005>
- Pagnoux C, Bouby L, Ivorra S, Petit C, Valamoti S-M, Pastor T et al (2015) Inferring the agrobiodiversity of *Vitis vinifera* L. (grapevine) in ancient Greece by comparative shape analysis of archaeological and modern seeds. *Veg Hist Archaeobot* 24:1–10
- Paladino M (2017) Modelos logit con R https://www.institutomora.edu.mx/testU/SitePages/martinpaladino/modelos_logit_con_R.html Accessed 18 December 2023
- Parins-Fukuchi C (2018) Bayesian placement of fossils on phylogenies using quantitative morphometric data. *Evolution* 72(9):1801–1814. <https://doi.org/10.1111/evo.13516>
- Perret M (1997) Caractérisation et évaluation du polymorphisme des génotypes sauvages et cultivés de *Vitis vinifera* L à l'aide de marqueurs RAPD et de certains traits morphologiques. *Travail de diplôme Université de Neuchâtel, Neuchâtel, Switzerland*
- Pierce LS, Tiffney BH (1986) Holocene fruit, seed and leaf flora from riverine sediments near New Haven. *Connecticut Rhodora* 88(854):229–252
- Popova SS, Yakovleva OV (2009) Seed morphology and anatomy of extant and fossil species of *Vitis* and *Ampelopsis* (*Vitaceae*). *Botanicheskiĭ Zhurnal* 94(11):1647–1657
- POWO (2023) Plants of the world online <https://powo.science.kew.org/> Accessed 18 December 2023
- R Project (2023) The R project for statistical computing <https://www.r-project.org/> Accessed 18 December 2023
- Random Forest (2023). Random forest: breiman and cutler's random forests for classification and regression. <https://>

- cran.r-project.org/web/packages/randomForest/index.html
Accessed 18 December 2023
- R-Bloggers (2023) Random forest in R <https://www.r-bloggers.com/2021/04/random-forest-in-r/> Accessed 18 December 2023
- Reid E, Chandler M (1933) The London clay flora. British Museum (Natural History), London
- Rivera D, Walker MJ (1989) A review of paleobotanical findings of early *Vitis* in the Mediterranean and on the origin of cultivated grapevines, with special reference to new pointers to prehistoric exploitation in the Western Mediterranean. *Rev Paleobot Palyno* 6:205–237
- Rivera D, Miralles B, Obón C, Carreño E, Palazón JA (2007) Multivariate analysis of *Vitis* subgenus *Vitis* seed morphology. *Vitis* 46(4):158–167. <https://doi.org/10.5073/vitis.2007.46.158-167>
- Rossiter DG (2023) Random forests two ways https://www.css.cornell.edu/faculty/dgr2/_static/files/R_html/CompareRandomForestPackages.html Accessed 18 December 2023
- RStudio (2023) RStudio desktop <https://posit.co/download/rstudio-desktop/> Accessed 18 December 2023
- Schindelin J, Arganda-Carreras I, Frise E et al (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9:676–682. <https://doi.org/10.1038/nmeth.2019>
- Sémah AM, Renault-Miskovsky J (2004) L'évolution de la végétation depuis deux millions d'années. Artcom, Paris
- Siddiqi MA (1980) Vitaceae. In: Jafri SMH, El-Gadi A (eds) *Flora of Libya*, vol 83. Al-Faateh University, Tripoli, pp 1–4
- Silver N (2012) *The signal and the noise: the art and science of prediction*. Penguin UK, London
- Stummer A (1911) Zur urgeschichte der rebe und des weinbaues. *Mitteilungen Der Anthropologischen Gesellschaft in Wien* 61:283–296
- Terpó A (1976) The carpological examination of wild-growing vine species of Hungary. *Acta Bot Acad Sci Hung* 22:209–247
- This P, Lacombe T, Thomas MR (2006) Historical origins and genetic diversity of wine grapes. *Trends Genet* 22:511–519. <https://doi.org/10.1016/j.tig.2006.07.008>
- Tiffney BH, Barghoorn ES (1976) Fruits and seeds of the brandon lignite. I Vitaceae *Rev Palaeobot Palynology* 22(3):169–191
- Valamoti SM (2009) Plant food ingredients and 'recipes' from prehistoric Greece the: archaeobotanical evidence. In: Morel JP, Mercuri AM (eds) *Plants and culture: seeds of the cultural heritage of Europe*. Centro Europeo per I Beni Culturali Ravello Edipuglia, Ravello, pp 25–38
- Valera J, Matilla-Seiquer G, Obón C, Alcaraz F, Rivera D (2023) Grapevine in the ancient upper euphrates: horticultural implications of a bayesian morphometric study of archaeological seeds. *Horticulturae* 9(7):803. <https://doi.org/10.3390/horticulturae9070803>
- Wan Y, Schwaninger HR, Baldo AM, Labaet JA, Zhong GY, Simon CJ (2013) A phylogenetic analysis of the grape genus (*Vitis* L.) reveals broad reticulation and concurrent diversification during neogene and quaternary climate change. *BMC Evol Biol* 13(1):1–20
- Zastawniak E (1996) East-Asiatic element in the fossil flora of the salt mine at Wieliczka (S. Poland). *Journal of Palaeosciences* 45:407–415
- Zecca G, Abbot JR, Sun WB, Spada A, Sala F, Grassi F (2012) The timing and the mode of evolution of wild grapes (*Vitis*). *Mol Phylogenet Evol* 62(2):736–747
- Zhou Y, Massonnet M, Sanjak JS, Cantu D, Gaut BS (2017) Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proc Natl Acad Sci U S A* 114(44):11715–11720
- Zohary D, Hopf M (2000) *Domestication of plants in the old world. The origin and spread in West Asia, Europe and the Nile Valley*. Oxford University Press, Oxford

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