Intra-leaf modeling of *Cannabis* leaflet shape produces synthetic leaves that predict genetic and developmental identities

Manica Balant [1,2,3], Teresa Garnatje [1], Daniel Vitales [1,2], Oriane Hidalgo [1,4], Daniel H. Chitwood [3,5]

[1] Institut Botànic de Barcelona, IBB (CSIC-Ajuntament de Barcelona), Passeig del Migdia s.n., 08038 Barcelona, Spain

[2] Laboratori de Botànica (UB), Unitat Associada al CSIC, Facultat de Farmàcia i Ciències de l'Alimentació–Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Joan XXIII 27–31, 08028 Barcelona, Spain

[3] Department of Horticulture, Michigan State University, East Lansing, MI, USA

[4] Royal Botanic Gardens, Kew, Kew Green, Richmond TW9 3AE, UK

[5] Department of Computational Mathematics, Science & Engineering, Michigan State University, East Lansing, MI, USA

Authors for correspondence:

Daniel H. Chitwood

Email: dhchitwood@gmail.com

Manica Balant

Email: manica.balant@ibb.csic.es

Summary

- The iconic, palmately compound leaves of *Cannabis* have attracted significant attention in the past. However, investigations into the genetic basis of leaf shape or its connections to phytochemical composition have yielded inconclusive results. This is partly due to prominent changes in leaflet number between leaves of a single plant during development, which has until this point prevented the proper use of common morphometric techniques.
- Here we present a new method that overcomes the challenge of nonhomologous landmarks in *Cannabis*. By modeling leaflet shape as a function of angle within a leaf,

we create synthetic leaves with comparable numbers of leaflets to which morphometric analysis is applied.

- We analyze 339 leaves from 24 individuals from nine *Cannabis* accessions, including both wild/feral accessions and cultivated varieties. Using 1,995 pseudo-landmarks in synthetic leaves, we visualize a morphospace and accurately predict accession identity, leaflet number, and relative node number.
- Intra-leaf modeling offers a rapid, cost-effective means of identifying *Cannabis* accessions, making it a valuable tool for future taxonomic studies, cultivar recognition, and possibly chemical content analysis and sex identification, in addition to permitting the morphometric analysis of leaves in any species with variable numbers of leaflets or lobes.

Key words: *Cannabis*, development, geometric morphometrics, intra-leaf modeling, leaf morphology, leaf shape

INTRODUCTION

Cannabis sativa L. (hereafter referred to as *Cannabis*) is a versatile crop plant used by humans for a variety of purposes throughout history. Although today it is commonly associated with its psychoactive properties, traditional medicine has relied heavily on *Cannabis,* and it is also a valuable source of food and fibers (Clarke & Merlin, 2013). Genetic and archaeological evidence suggests that *Cannabis* was domesticated around 12,000 years ago in East Asia, initially serving as a multipurpose crop before separate selections for fiber and drug production emerged around 4,000 years ago (Ren *et al.*, 2021). Since then, widespread cultivation has facilitated its global distribution. Throughout the 20th century, *Cannabis* use was largely abandoned due to its illegal status in many parts of the world. However, recent legalization for recreational and/or medicinal purposes in many countries worldwide has led to a surge in the cannabis industry (*The Global Cannabis Report, 3rd Edition*, 2022).

Extensive *Cannabis* use has resulted in the development of numerous cultivars and strains that are well-suited to diverse uses and climates (Small, 2015). This significant

morphological and phytochemical diversity within the *Cannabis* genus poses challenges for taxonomic classification. Over the past two centuries, various taxonomic approaches based on genetics, morphology, and phytochemistry have been proposed (McPartland & Small, 2020). Some scientists advocated for a polytypic classification, recognizing the presence of two (Lamarck & Poiret, 1783; Zhukovskii, 1971; Hillig, 2005a) or three (Emboden, 1974; Schultes *et al.*, 1974; Hillig, 2005b; Clarke & Merlin, 2013) species with multiple subspecies, while others argued for a monotypic genus, considering only a single species, *Cannabis sativa* (Small & Cronquist, 1976; Sawler *et al.*, 2015; Small, 2015; McPartland, 2018; McPartland & Small, 2020; Ren *et al.*, 2021). Hilling introduced a classification system based on biotypes, considering molecular, morphological, and phytochemical data. He proposed dividing *Cannabis* into two species, *C. sativa* and *C. indica* Lam., and six biotypes: *C. indica* as narrow-leaflet drug (NLD), wide-leaflet drug (WLD), hemp and feral biotype, and *C. sativa* as hemp and feral biotype (Hillig, 2005a). Recently, Lapierre *et al.* (2023) conducted a comprehensive taxonomic review of the *Cannabis* genus and based on available genetic data, strongly supported the theory that *Cannabis* is a highly diverse monotypic species.

Apart from taxonomic classification, *Cannabis* is often categorized based on its cultivation purpose, morphology, and chemical composition. Fiber-type plants, commonly known as hemp, are primarily grown for fiber and seed production. These plants contain less than 0.3% of the psychoactive compound THC (Δ9-tetrahydrocannabinol), while drug-type plants, often referred to as marijuana and medicinal cannabis, can contain higher levels of THC (Hurgobin *et al.*, 2021). *Cannabis* plants can also be separated based on the ratio of two major cannabinoids THC and CBD (cannabidiol) into Type I (THC dominant), Type II (balanced CBD/THC ratio), and Type III plants (CBD dominant) (Small & Beckstead, 1973). In the medicinal and recreational cannabis industries plants are normally categorized as '*sativa*', '*indica*', or '*hybrid*'. Taller plants with narrow leaflets and high THC percentage are called '*sativa*', while shorter and bushier plants with wider leaflets and high percentages of both CBD and THC are called '*indica*'. Plants with intermediate characters are called '*hybrids*' (McPartland & Guy, 2017). While the classification of *Cannabis* into *'indica'* and *'sativa'* is not supported by genetic data, the visible differences in leaflet width have long been a significant characteristic used to visually discriminate different types of *Cannabis*.

96 *Cannabis* arguably possesses one of the most iconic leaves among all plants. Its palmately
97 compound leaves with a varying number of leaflets are a popular culture symbol. *Cannabis*
98 exhibits a remarkable degree of phenotypic plasticity, further accentuated by selection
99 pressure during the domestication process (Small, 2015). Extensive variability in leaf
100 morphology has already been described by Quimby *et al.* (1973) and later Anderson (1980),
101 who was the first to quantify the width, length, and ratio of the central leaflet. This or
102 similar methods were then commonly used in studies investigating the morphological
103 characteristics of *Cannabis* species, subspecies, cultivars, biotypes and chemotypes (Small *et*
104 *al.*, 1976; de Meijer *et al.*, 1992; de Meijer & Keizer, 1996; Hillig, 2005a; Clarke & Merlin,
105 2013; Lynch *et al.*, 2016; Karlov *et al.*, 2017; Parsons *et al.*, 2019; McPartland & Small, 2020;
106 Carlson *et al.*, 2021; Islam *et al.*, 2021; Jin *et al.*, 2021a; Vergara *et al.*, 2021; Buzna & Sala,
107 2022; Chen *et al.*, 2022; Murovec *et al.*, 2022), often with contradictory results. Leaf shape
108 has therefore played an important and sometimes controversial role in *Cannabis* taxonomy.
109 While researchers in previous *Cannabis* studies were aware of enormous plasticity and the
110 effect the environment has on leaf shape (Vergara *et al.*, 2021; Murovec *et al.*, 2022), they
111 very rarely paid attention to the effects of developmental processes, even though
112 heteroblastic changes (differences in leaf shape arising from juvenile-to-adult phase
113 transitions in the meristem) profoundly affect the arrangement and shape of *Cannabis*
114 leaves along the shoot. While some studies briefly mention the developmental changes of
115 leaves (Hillig, 2005a; Carlson *et al.*, 2021; Jin *et al.*, 2021a; Spitzer-Rimon *et al.*, 2022), the
116 only study focused on heteroblastic phase changes in leaves along the plant axis was done
117 by Heslop-Harrison and Heslop-Harrison (1958). In the lower part of the shoot *Cannabis*
118 leaves exhibit opposite phyllotaxy and one to three leaflets, transitioning to alternate
119 phyllotaxy and leaves with up to 11 to 13 leaflets in the upper section (Hillig, 2005a; Clarke
120 & Merlin, 2013; Small, 2015). Additionally, the changes in leaflet number are not uniform
121 between different *Cannabis* accessions (Hillig, 2005a). These changes during development
122 not only complicate categorization of plant accessions based on leaf shape, but also prevent
123 the use of morphometric techniques.
124
125 Morphometrics is the quantitative analysis of shape. It includes a wide range of methods,
126 from measuring allometric differences in dimensions like lengths, widths, and angles in
127 relation to size (Niklas, 1994), to geometric techniques that measure shape

comprehensively, like elliptical Fourier (EFDs; Kuhl & Giardina, 1982) and landmark-based analyses (Bookstein, 1997). It can be used to classify species and to separate effects on shape arising from genetic, developmental, and environmental mechanisms (Chitwood & Sinha, 2016). Historically the field of ampelography (ἄμπελος, 'vine' + γράφος, 'writing'; Ravaz, 1902; Galet, 1952; Galet & trans. Morton, 1979) relied heavily on leaf shape to distinguish grapevine varieties, which like *Cannabis*, is a culturally popular plant used for psychoactive, recreational purposes with a charismatic leaf shape. Unlike *Cannabis,* grapevine leaves have a consistent number of lobes, sinuses, and other associated homologous points that can be used for both landmark-based and EFD morphometric analysis (Chitwood *et al.*, 2014; Chitwood, 2021) to disentangle genetic (Demmings *et al.*, 2019), developmental (Chitwood *et al.*, 2016a; Bryson *et al.*, 2020; Migicovsky *et al.*, 2022), and environmental effects (Chitwood *et al.*, 2016b, 2021) embedded in leaf shapes. The variable number of leaflets in *Cannabis* leaves precludes analysis methods that rely on homologous, comparable points to measure shape comprehensively. Vergara *et al.* (2021) used a landmark-based approach but were limited to analyzing the central and two most distal leaflets on each side, features that all *Cannabis* leaves except single-leaflet leaves possess, but which excludes most of the shape variation within a leaf.

Here, we model leaflet shape as a function of angle within *Cannabis* leaves. The resulting continuous models can be used to create synthetic leaves with any desired number of leaflets. Analyzing 339 leaves from across the nodes of 24 individuals representing nine accessions, we create synthetic leaves with five leaflets and 1,995 comparable pseudo-landmarks. Dimension reduction separates leaf types along orthogonal axes and visualizes an underlying morphospace. Linear discriminant analysis (LDA) predicts accession, leaflet number, and relative node number with high accuracy. Intra-leaf modeling allows the application of morphometric techniques to comprehensively measure leaf shape in *Cannabis*, enabling future taxonomic and developmental studies, cultivar recognition, and possibly chemical content analysis and sex identification, in addition to permitting the morphometric analysis of leaves in any species with variable numbers of leaflets or lobes.

158 **MATERIAL AND METHODS**

159

160 **Plant material and growing conditions**

161 This study includes 24 individuals from nine accessions of *Cannabis sativa* (Table **1**; Fig. **1**),
162 encompassing both wild/feral accessions and cultivated varieties with a wide distribution
163 area. The plants were grown from seed in a growth chamber (Fitoclima D1200PLL, Aralab,
164 Portugal) to minimize the influence of the environment. Before sowing, the seeds were
165 sterilized overnight in a 5% H_2O_2 solution with the addition of Inex-A solution (Cosmocel,
166 Spain) at room temperature. Sterilized seeds were then transferred to Petri dishes and
167 placed in the growth chamber for germination. Once the first leaves emerged, the seedlings
168 were transferred to small peat pots with a pre-fertilized soil substrate (Kilomix Atami,
169 Spain). During this phase, the environmental conditions were set to 25°C, with an 18-hour
170 day and 6-hour night photoperiod, and a light intensity of 50 µmol $m^{-2}s^{-1}$ (Philips Master PL-
171 L 55W, Spain). After two weeks the surviving plants were transplanted to 3.5 l pots with the
172 same soil substrate. The light intensity was gradually increased to 300 µmol $m^{-2}s^{-1}$ over the
173 following week, without changing the photoperiod and temperature. After four weeks, the
174 photoperiod was changed to 12 hours of daylight and 12 hours of darkness, and the light
175 intensity was gradually increased to 700 µmol $m^{-2}s^{-1}$ over the following week, while keeping
176 the temperature at 25°C. The plants remained in these environmental conditions until the
177 flowering stage. Plants received daily irrigation with tap water, without any application of
178 nutrient or phytosanitary control.

179

180 **Table 1** Accession details and number of leaves collected and analyzed in the study.

Accession ID	Accession type	Location/Cultivar name	Number of individuals	Number of leaves collected	Number of leaves analyzed
AM15	Wild/feral	Armenia, Sjunik marz, Goris town	5	90	72
BNG	Wild/feral	Bangladesh, Rangpur, Carmichael College Campus	1	14	9
FUT75	Cultivar	Futura 75	2	44	30
HU1	Wild/feral	Hungary, Nyírvasvári	4	83	65
IK	Landrace	India, Kerala	4	99	53
IKL	Landrace	India, Kullu	4	69	59
MAR	Landrace	Morocco, North Morocco	1	18	11
MN9	Wild/feral	Mongolia, Selenge aimag, Baruunburen sum	1	14	12
RO1	Wild/feral	Romania, Mangalija	2	36	28

181

182 **Leaf sampling and imaging**

183 A total of 467 leaves were sampled during the flowering stage, with the exception of
184 individuals from the accession IK, which did not begin to flower during the two-month
185 cultivation period. Leaves along the main axis of the plants were collected and immediately
186 scanned using a flatbed photograph scanner (Epson Perfection V370, Tokyo, Japan) at 1200
187 dpi resolution. A piece of velvet fabric was placed between the leaf and the scanner cover to
188 avoid any shadows. No adjustments to the angle of individual leaflets were made before
189 scanning. Each leaf was scanned with a scale and a label indicating the node it originated
190 from, followed by a sequential lowercase letter, since typically two leaves are present per
191 node. Starting at the base of the plant, the first two leaves were labeled as leaves "a" and
192 "b" from node number 1, and so on, until the shoot apex.

193 *Cannabis* leaves display a marked heteroblastic, or juvenile-to-adult, leaf shape progression.
194 Mature, juvenile leaves located on the first node at the base of the plant usually have a
195 simple, serrated leaf. As node number increases so does the leaflet number, reaching a
196 maximum of 9 to 13 leaflets in young, adult leaves at the growing tip. Eventually leaves
197 transition into an inflorescence type. During this transition, the number of leaflets per leaf
198 starts to decrease again until the top of the inflorescence. Leaves at the shoot base have
199 opposite phyllotaxy and transition to alternate phyllotaxy in the upper section on the stem
200 and inflorescence (Heslop-Harrison & Heslop-Harrison, 1958; Hillig, 2004; Potter, 2009;

201 Spitzer-Rimon *et al.*, 2022). To ensure that only stem leaves were included in our analysis,
202 we separated the two types (i.e., stem and inflorescence leaves) based on the point where
203 the decrease in the number of leaflets appeared. This point determined the "total node
204 number", the number of nodes per plant used for further analysis. Total node number
205 varied among individuals. To compare node positions, a relative node number was
206 calculated, which was defined by the node position divided by the total node number for
207 the individual plant, where zero is at the plant base and one at the last node included in the
208 analysis (Fig. **1**). Because of the nature of plant growth, the leaves at the base of the plant
209 were frequently too senesced to be incorporated in the analysis or were entirely lost.
210 Nevertheless, the nodes could still be identified, which allowed them to be taken into
211 account in the calculation of relative node number.

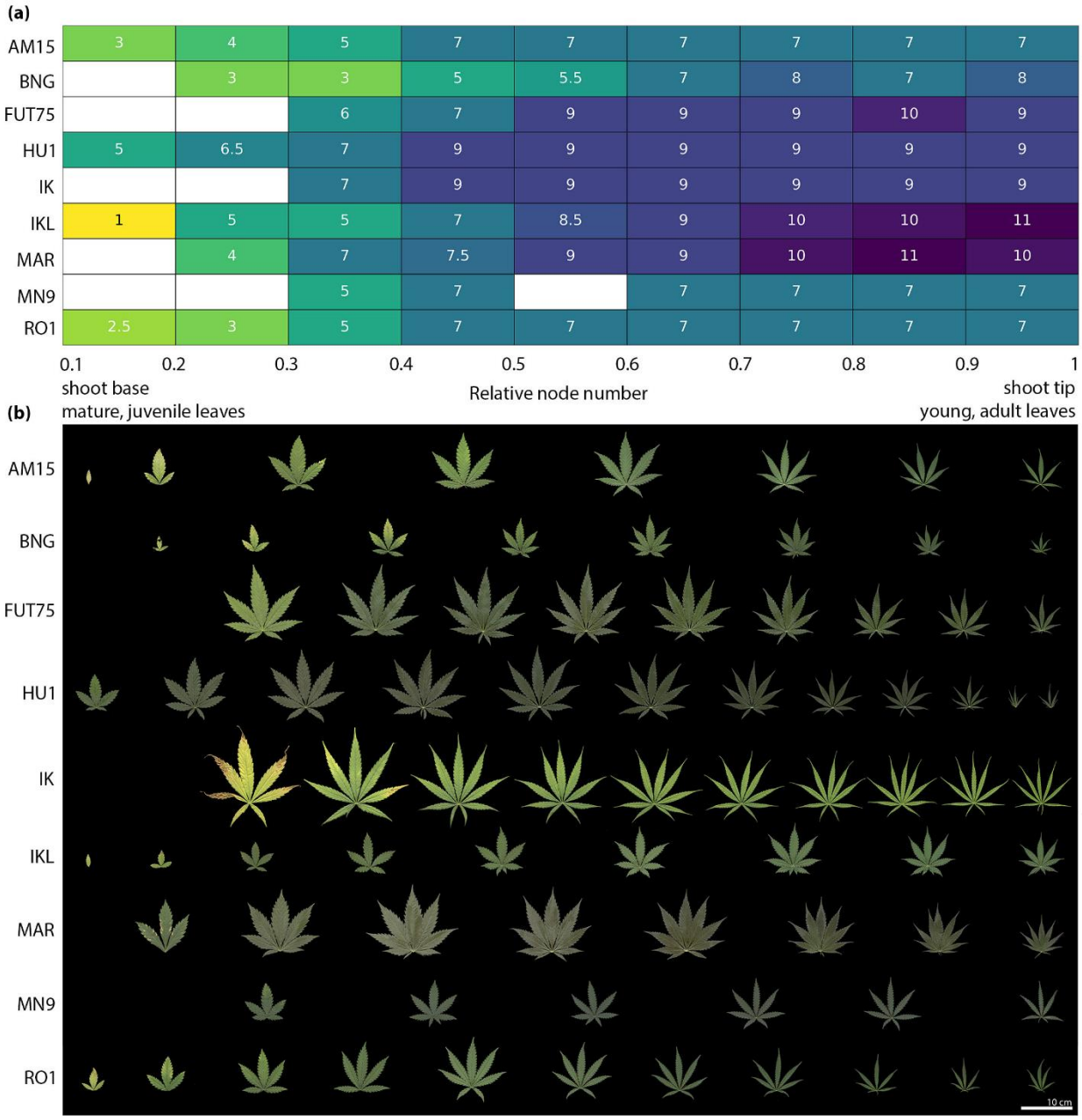
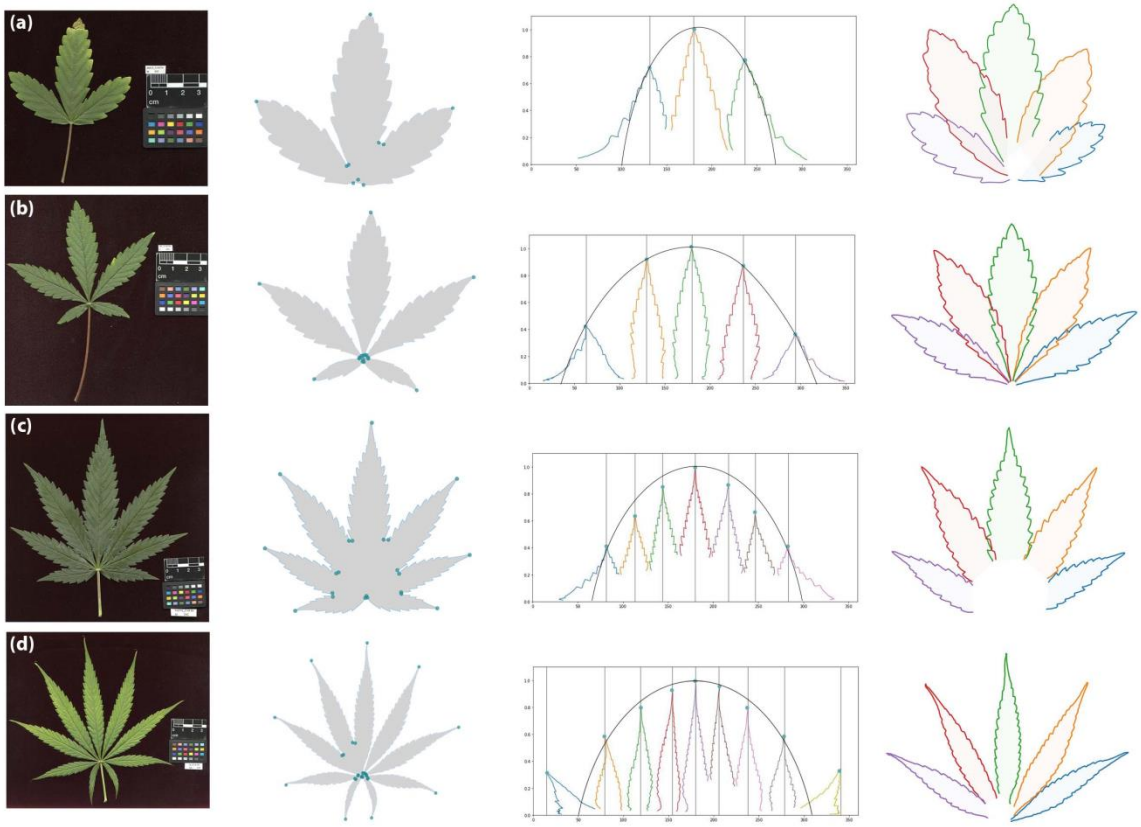


Fig. 1 Changes in the leaf shape and leaflet number during the development in nine *Cannabis* accessions. (a) Median values for all available leaflet number for each relative node number for the nine *Cannabis* accessions. (b) Changes in leaf shape between different developmental stages in different *Cannabis* accessions.

Image analysis and landmarking

After eliminating deformed leaves, simple leaves, leaves with even leaflet numbers, leaves with relative node values above one, inaccurately modeled leaves and outliers detected during dimension reduction, a total of 339 *Cannabis* leaves were analyzed. Photoshop was

used to separate petioles and leaflets smaller than 1 cm from the rest of the leaf. The leaf outlines were then extracted and saved using Python modules NumPy (Harris *et al.*, 2020), Matplotlib (Hunter, 2007) and OpenCV (Bradski, 2000). The code for extracting and plotting the leaf outlines can be found on GitHub (https://github.com/BalantM/Cannabis_leaf_morpho). The *x* and *y* coordinates of blade outlines and landmarks were extracted using ImageJ (Abràmoff *et al.*, 2004). The outline was extracted using the *wand* tool (setting tolerance to 20 and including "smooth if thresholded" option) and the landmarks were placed using the *multi-point* tool. Initially, landmarks were placed at the beginning and end of each leaflet, starting from the lower left side, and continuing to the lower right side of the leaf outline. Subsequently, landmarks were placed in the same order on the tips of the leaflets. The final landmark was positioned at the center of the petiolar junction (Fig. **2**). These landmarks delimit the boundaries of the leaflets so that equidistant pseudo-landmarks can later be placed along the contour. The number of landmarks per leaf ranged from 10 to 28, depending on the leaflet number. The raw data containing the coordinates for leaf outlines and landmarks can be accessed on GitHub (https://github.com/BalantM/Cannabis_leaf_morpho).



238 **Fig. 2** The process of modeling new synthetic leaves for a leaf with (a) three leaflets from
239 accession AM15, (b) five leaflets from accession IKL, (c) seven leaflets from accession FUT75,
240 and (d) nine leaflets from accession IK. The first column shows the scans of the leaves, which
241 we use to extract the outline and place the landmarks on the tip, start, and end of each
242 leaflet and on the petiolar junction (second column). These coordinates are used to
243 generate 200 pseudo-landmarks on each side of each leaflet, sharing the landmark on the
244 tip of the leaflet and are converted into polar coordinates. Second degree polynomials are
245 then fitted through these pseudo-landmarks (third column). A synthetic leaf with five
246 leaflets defined by 1,995 pseudo-landmarks can then be modeled using the collection of
247 polynomial models for each leaf (fourth column).
248
249 **Data analysis**
250 To analyze leaves with different numbers of leaflets, pseudo-landmarks of each leaflet were
251 modeled as polynomial functions of angles within a leaf, in order to use the models to
252 construct a synthetic leaf with a desired number of leaflets. The Python code, presented as a
253 Jupyter notebook with detailed description, is available on GitHub
254 (https://github.com/BalantM/Cannabis_leaf_morpho). In brief, *x* and *y* coordinates of the
255 leaf outline were first interpolated to create an arbitrarily high number of coordinates to
256 increase resolution. The coordinates of manually selected landmarks were then compared
257 against the high-resolution coordinates of the leaf outline and the nearest neighboring point
258 of the high-resolution coordinates to each original landmark was identified and specified as
259 the new landmark point. Next, the outline and new landmark coordinates were rotated,
260 translated, and scaled so that the central leaflet had a length of one and pointed in the
261 same direction. The transformed points were then interpolated to generate 200 pseudo-
262 landmarks on each side of each leaflet, sharing the landmark on the tip of the leaflet (i.e., a
263 total of 399 pseudo-landmarks per leaflet). These pseudo-landmarks were then converted
264 to polar coordinates, where each point was defined by a radius and angle relative to the
265 petiolar junction and tip of the central leaflet.
266
267 Using the polar coordinates of each leaflet, a second-degree polynomial function using the
268 NumPy *polyfit* function was fitted through each of the corresponding pseudo-landmarks for
269 each leaflet, modeling radius as a function of angle. A synthetic leaf with five leaflets was

270 then modeled using the collection of 399 polynomial models for each leaf. The coordinates
271 defining the 1,995 pseudo-landmarks of the synthetic five-leaflet leaves were plotted and
272 visually inspected. Any inaccurately modeled leaves or plots containing noticeable errors or
273 discrepancies were eliminated. A principal component analysis (PCA) was performed using
274 scikit-learn (Pedregosa *et al.*, 2011) module in Python and proportions of explained variance
275 for each principal component and the cumulative variance was calculated. The
276 morphospace was calculated using the inverse transform of PC1 and PC2 scores and
277 visualized using seaborn (Waskom, 2021) and Matplotlib modules in Python. Theoretical
278 eigenleaves were plotted across equal intervals of the PCA morphospace. Points
279 representing the leaves were then plotted over and colored by the accession identity, leaflet
280 number, or relative node number (Fig. **3, 4**). To see which of the first two PCs explains most
281 of the leaf shape variation for accessions, leaflet number and relative node number Kruskal-
282 Wallis test were calculated using *stats.kruskal* function from the scipy.stats module
283 (Virtanen *et al.*, 2020). To visualize an average leaf for each accession, leaflet number, and
284 relative node number, the average coordinate values of synthetic leaves were calculated for
285 each of the categories and plotted using the Matplotlib module in Python.
286
287 Linear discriminant analysis (LDA) was applied to model accession, leaflet number, and
288 relative node number as the function of synthetic leaf coordinate values (Fig. **5**) using the
289 *LinearDiscriminantAnalysis* function from the scikit-learn module in Python. To test the
290 performance of the LDA model, the dataset was divided into two parts. Since most of the
291 analyzed leaves exhibit opposite phyllotaxy, wherein the nodes were represented by two
292 leaves (a and b) in the same developmental phase with the same number of leaflets, the
293 dataset was split into a training dataset (leaf a) comprising 177 leaves and a test dataset
294 (leaf b) containing 162 leaves. The *predict* function from *LinearDiscriminantAnalysis* in the
295 scikit-learn module was used to predict the accession identity, leaflet number, and relative
296 node number. The accuracy of the LDA model was calculated and visualized using the
297 function *confusion_matrix* from scikit-learn. Spearman Rank Correlation was calculated for
298 true and predicted results for relative node number with *spearmanr* function from the
299 scipy.stats module.
300

301 **RESULTS**

302

303 **Heteroblastic changes in leaflet number along the main axis**

304 Over 460 *Cannabis sativa* leaves were collected, scanned, and their leaflet number
305 recorded. The leaves exhibited a profound heteroblastic juvenile-to-adult progression along
306 the axis, but the changes were not uniform between the accessions (Fig. **1**). In the few rare
307 cases where the leaves in the lower nodes were present, the first nodes always started with
308 a simple serrated leaf. The second leaf usually had three leaflets and the most frequent
309 leaflet number in the third node was five. However, the leaflet number in the nodes above
310 varied dramatically between accessions. The number of nodes before the transition into the
311 inflorescence in each of the plants also varied. We therefore calculated relative node
312 number, a fractional number between 0 at the shoot base to 1 at the inflorescence
313 transition, to compare the node leaves between plants.

314

315 **Principal component analysis (PCA) and morphospace**

316 Using the outline and landmark coordinates of 339 leaves, we modeled new synthetic
317 leaves, all with five leaflets (Fig. **2**). Each synthetic leaf is defined by 1,995 pseudo-
318 landmarks, which overcomes the problems associated with variable leaflet numbers and
319 permits dimension reduction using PCA and the visualization of a *Cannabis* leaf
320 morphospace. The first and second PCs account for 69.46% and 10.06% of the shape
321 variation, respectively (Fig. **3, 4**). The eigenleaves plotted in the morphospace show that the
322 first PC mostly accounts for the variability in the width of the leaflets; with increasing values
323 along PC1, the leaflets become wider. The increasing values along PC2 result in eigenleaves
324 with wider and less acuminate leaflets. Wider, less acuminate leaflets represent high PC2
325 value eigenleaves that become narrower and more acuminate as leaflet and relative node
326 numbers increase and PC2 values decrease. A large portion of variance associated with PC1
327 corresponds to narrow vs. wide leaflets that are stereotypical descriptions of *sativa* vs.
328 *indica* or *wide-* vs. *narrow- leaflet* drug varieties.

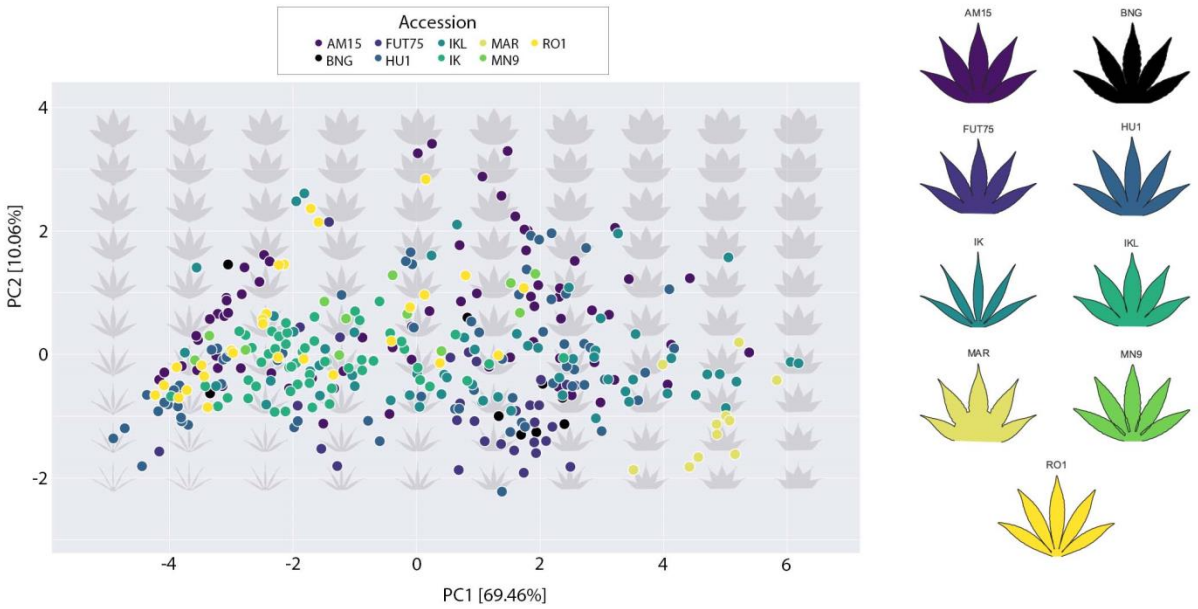


Fig. 3 Principal component analysis (PCA) of the accessions performed on the synthetic leaves using the 1,995 pseudo-landmarks (left) and average synthetic leaf shapes for each of the nine analyzed accessions (right). The leaf shapes in the background of the PCA plot represent the eigenleaves of the morphospace.

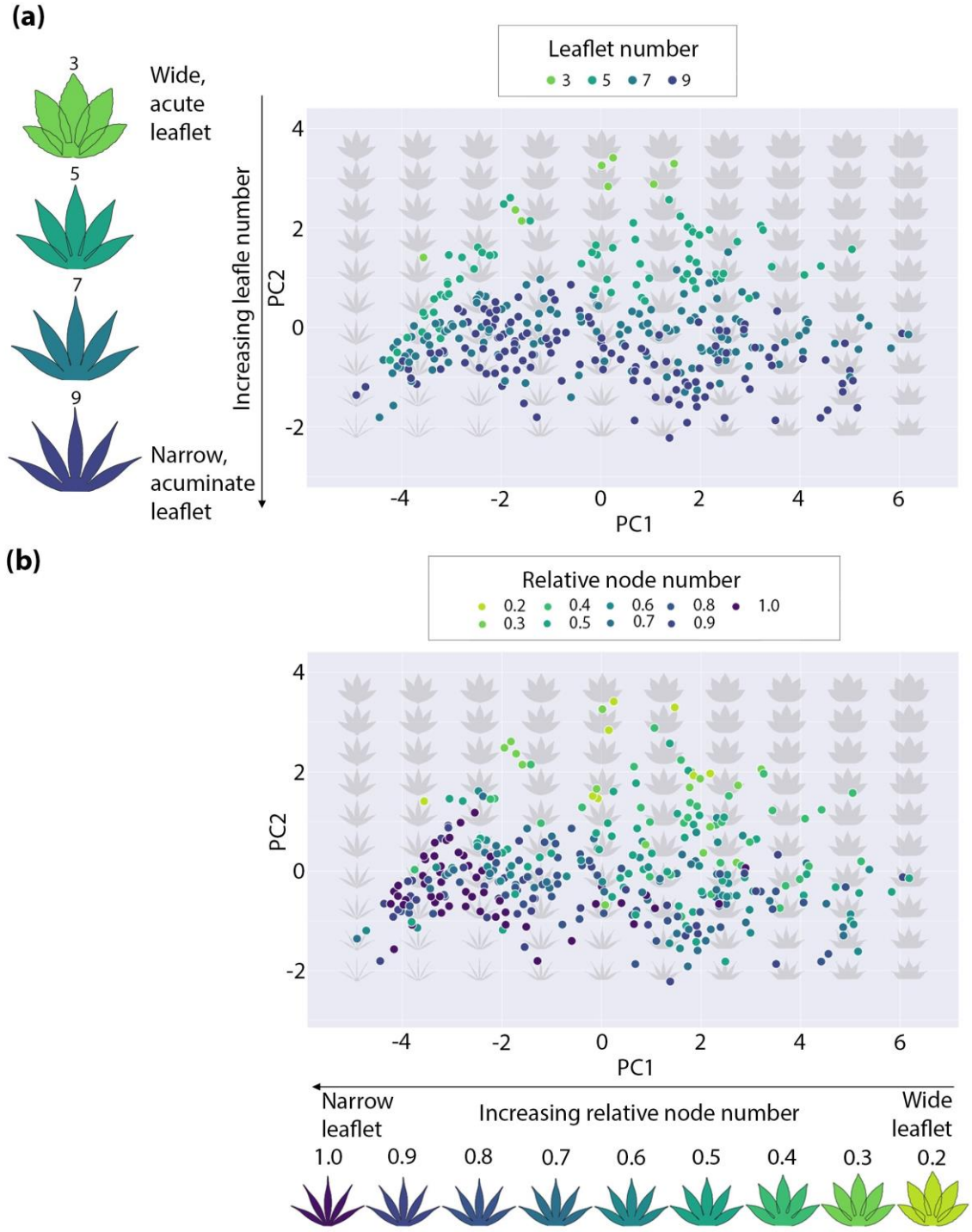


Fig. 4 Principal component analysis (PCA) and average leaf shapes showing the developmental changes caused by (a) leaflet number and (b) relative node number. The first PC explains 69.46% and the second 10.06% of variation.

Examining the distribution of factors within the morphospace using a Kruskal-Wallis test reveals that accession (Table **2**, Fig. **3**) and relative node number (Table **2**, Fig. **4b**)

342 significantly vary by both PC1 and PC2, while leaflet number significantly varies only by PC2
343 (Table **2**, Fig. **4a**). This indicates that the changes in leaf shape between accessions are not
344 independent from developmental variation. That a facet of variation in accession leaf shape
345 covaries with developmental variation across the shoot in leaflet and relative node number
346 suggests a heterochronic mechanism by which accession differences in leaf shape arise from
347 changes in developmental timing.
348
349 **Table 2** Kruskal-Wallis test was used to test the leaf shape variation along PC1 and PC2 for
350 accessions, leaflet number and relative node number.

	PC1		**PC2**	
	H	**p value**	**H**	**p value**
Accession	78.60	**p < 0.0001**	84.43	**p < 0.0001**
Leaflet number	1.27	p = 0.7360	156.16	**p < 0.0001**
Relative node number	91.21	**p < 0.0001**	101.35	**p < 0.0001**

351
352 **Linear discriminant analysis (LDA) and prediction of genetic and developmental identities**
353 We used LDA to model each factor as a function of 1,995 pseudo-landmark points (Fig. **5a-**
354 **c**). Model accuracy was calculated by splitting leaves arising from the same nodes into
355 training (177 leaves) and testing (162 leaves) datasets. To evaluate model accuracy,
356 accession was treated as a categorical variable, as was leaflet number, as it not only has a
357 small number of levels (3, 5, 7, and 9 leaflets), but each level is well separated from the
358 others. To evaluate the accuracy of relative node number, we treated it as a continuous
359 variable, due to a high number of levels (9) that continuously overlap with each other (Table
360 **3**). Models for both accession and leaflet number are highly accurate (74.07% and 92.59%,
361 respectively). The model for relative node number is highly accurate as well, as inferred by a
362 highly significant Spearman's rank correlation coefficient value between actual and
363 predicted values (rho = 0.729, p < 0.0001).
364

365 **Table 3** Predictive power of genetic and developmental identities using the LDA model.

	Correct prediction [n]	False prediction [n]	Prediction accuracy [%]	Correlation coefficient [rho]	p value
Accession	120	42	74.07	NA	NA
Leaflet number	150	12	92.59	NA	NA
Relative node number	NA	NA	NA	0.729	< 0.0001

366

367 A confusion matrix reveals that the LDA model in most cases had a high accuracy for
368 predicting accession identity (Fig. **5d**). Accessions IK, RO1, and MN9 show practically no
369 overlap in LDA space, while AM15, IKL, FUT75, BNG, HU1, and MAR show more overlap (Fig.
370 **5a**). The model showed an almost 93% success in determining leaflet number. Leaves with
371 only 3 leaflets are markedly different from the rest, and the prediction model consistently
372 classified them correctly (Fig. **5b**). Leaves with 5 to 9 leaflets showed less pronounced
373 differences in shape, resulting in a slightly lower accuracy of the prediction model for these
374 cases. However, an examination of the confusion matrix revealed that misclassifications
375 mainly occurred between leaves with neighboring leaflet numbers. Specifically, the model
376 tended to confuse leaves with 5 leaflets for those with 7 leaflets, but not for leaves with 9
377 leaflets (Fig. **5e**). The exceptional ability to predict leaves with 3 leaflets from the rest may
378 suggest that this developmental mechanism is biased towards variation at the base of the
379 shoot. Similar to leaflet number, the confusion matrix for the relative node model reveals
380 high rates of misclassification between the neighboring relative node numbers, as is
381 expected, and leaves from lower nodes were very rarely classified as those from higher
382 nodes (Fig. **5f**). A pronounced change in leaf shape occurs between the relative nodes 0.3
383 and 0.4, while the shape changes in later relative nodes are more gradual (Fig. **5c**).
384

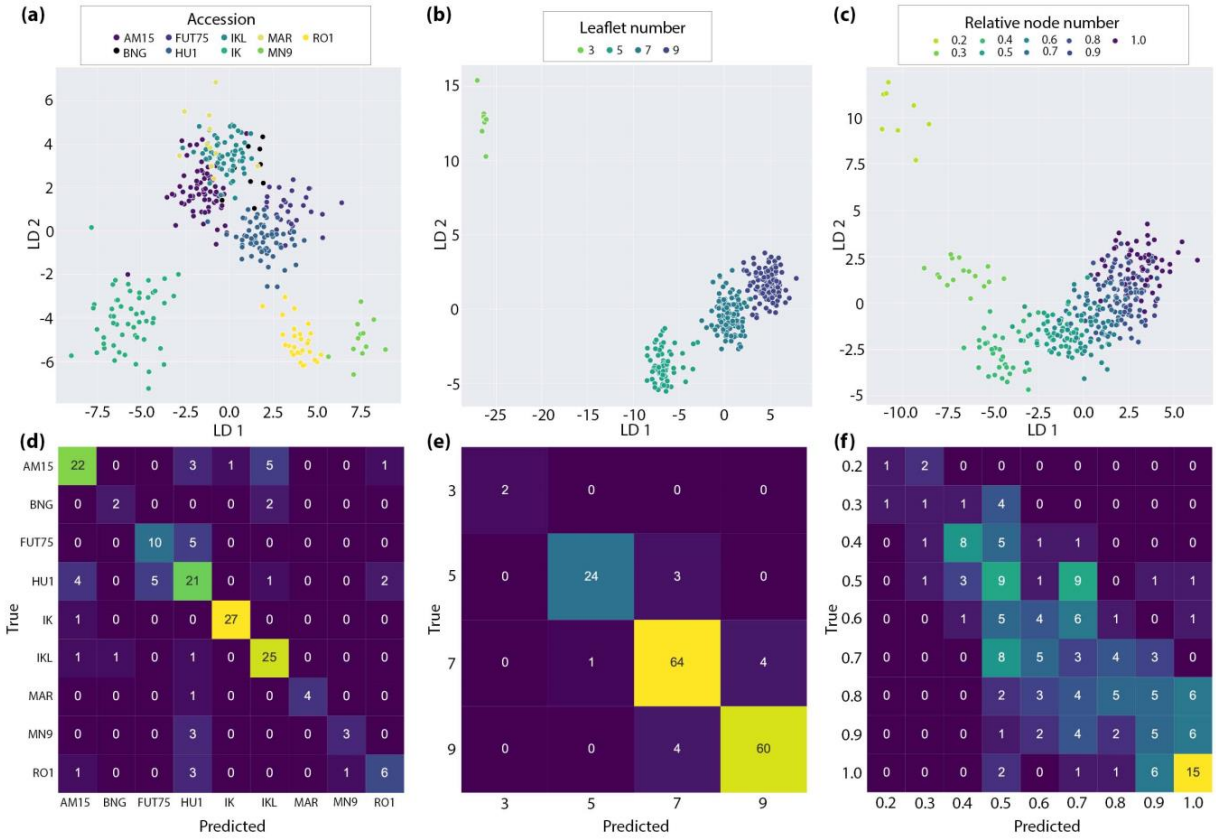


Fig. 5 Accession, leaflet number and relative node numbers of *Cannabis* leaves can be predicted independently of each other. Linear discriminant analysis (LDA) plots for (a) accession, (b) leaflet number and (c) relative node number. In the lower row, the confusion matrices show the true and predicted identities for (d) accessions, (e) leaflet number, and (f) relative node number using the LDA model on the split test and train dataset.

DISCUSSION

To understand how leaf shape can both scientifically and practically aid our classification and biological understanding of *Cannabis*, we first turn to grapevines. Like *Cannabis*, both grapevine rootstocks and wine and table grape scions hold tremendous cultural and economic importance (Ravaz, 1902; Galet, 1952; Galet & trans. Morton, 1979). From enforcing appellation to identifying *Phylloxera*-resistant rootstock varieties, the berries and their associated sensory properties held little taxonomic value compared to leaf shape. Like other features of the shoot, leaf shape can easily be discerned at a glance, and even compared to genotype, it can possess higher power to distinguish varieties due to complicated pedigrees, high heterozygosity, and subtle clonal differences (Myles *et al.*,

403 2011; Zhou *et al.*, 2017). To deal with variation, early ampelographers focused on geometric
404 features that could be quantified, like the degree of lobing, length-to-width ratio, and the
405 angle of the petiolar sinus (Ravaz, 1902; Galet, 1952; Galet & trans. Morton, 1979). That
406 nearly every grapevine leaf has five lobes allowed the application of landmark-based
407 approaches (Chitwood *et al.*, 2014; Chitwood, 2021). With landmarks, leaf shape could be
408 measured comprehensively and statistically. One of the most fundamental properties of
409 grapevine leaf shape is that genetic and node-to-node variation are orthogonal to each
410 other and vary independently (Chitwood *et al.*, 2016a,b). Not only is this convenient
411 because it allows genetic and developmental identities to be independently predicted from
412 each other, but insightful, because it indicates that latent effects of distinct biological
413 mechanisms are embedded within each leaf and can be separated from each other.

414 Like grapevines, striking variation in leaf shape (Fig. **1**) has historically played a significant
415 role in taxonomic classification of *Cannabis*. Leaf shape and differences in phyllotaxy were
416 among the characters Lamarck used to describe a new *Cannabis* species (Lamarck & Poiret,
417 1783). Anderson (1980) introduced a quantitative approach by quantifying the length-to-
418 width ratio of the central leaflet. Further studies using different characters—including plant
419 height, stem diameter, achene shape, and phytochemical profiles—to characterize
420 accessions have only confirmed the importance of leaf characteristics (Small *et al.*, 1976;
421 Hillig, 2005a). The method measuring central leaflet width-to-length ratio has been adopted
422 by researchers as one of the main characters for determining species, subspecies, biotypes
423 and chemotypes of *Cannabis* (Hillig, 2005a; Clarke & Merlin, 2013; McPartland & Small,
424 2020). However, this method is only able to capture a limited aspect of leaf shape variation,
425 neglecting other important characteristics that we measure in this study, such as leaflet
426 outlines, serrations, angles, and relative changes in leaflet shape across the leaf. By
427 modeling leaflet shape as a function of angle, we synthetically create leaves with the same
428 number of leaflets for which high densities of corresponding pseudo-landmarks capture high
429 resolution shape features (Fig. **2**).
430
431 The method presented in this study can accurately determine accession based on leaf
432 shape, regardless of its developmental stage (Fig. **5a, d**). The method not only works
433 effectively on stabilized or cloned cultivar accessions but also on wild or feral accessions

434 cultivated from seed that can exhibit distinct plant phenotypes (Table **1**), indicating its
435 robustness and potential value in future germplasm classification. The variation in leaf
436 shape associated with PC1 (representing 69.64% of variation), mostly accounts for the
437 variability in the width of the leaflets (Fig. **3**). The eigenleaves representing PC1 vary by wide
438 and narrow leaflets, similar to stereotypical classifications of *sativa* and *indica* plants or
439 *wide*- vs. *narrow- leaflet* drug varieties. The reliance on these non-quantitative leaf shape
440 descriptors has led to numerous cultivars with unreliable names, inconsistent genetic
441 origins, and phytochemical profiles (Sawler *et al.*, 2015; Schwabe & McGlaughlin, 2019;
442 Watts *et al.*, 2021; Jin *et al.*, 2021b). Using this new method, any leaf can be projected onto
443 the morphospace, and the resulting PC values used to provide a quantitative measure of the
444 leaf shape would be useful for studies investigating correlation between shape and chemical
445 profiles or sex of the plant. For example, Jin *et al.* (2021a) conducted a study on clones of 21
446 cultivars and found a strong negative correlation between the width and length ratios of
447 central leaflets and CBD, and a positive correlation with THC; however, Vergara *et al.* (2021)
448 and Murovec *et al.* (2022) were unable to confirm these findings. All three studies used low-
449 resolution morphometric approaches. Sex of the plants also plays a crucial role in the
450 cannabis industry, where the presence of male plants and inevitable pollination leads to
451 decreases in cannabinoid production as plants shift the use of energy into seed
452 development. Several methods have been employed to differentiate between male and
453 female plants at early stages, but only genetic methods were successful so far (Toth *et al.*,
454 2020; Prentout *et al.*, 2020; Campbell *et al.*, 2021; Balant *et al.*, 2022; Torres *et al.*, 2022).
455 Our results quantify the variation in leaf shape between accessions that can potentially be
456 used to classify accessions and predict chemical profiles and plant sex faster and more
457 accurately.
458
459 Unlike grapevine, where developmental variance is orthogonal and separate from genetic
460 variance, in *Cannabis* these two factors are correlated. That the developmental source of
461 variation is colinear with accession identity suggests that part of the differences between
462 accession leaf shape is explained by shifts in developmental timing, or heterochrony.
463 Differences, especially in flowering time and growth rates between cultivars have been
464 observed before (de Meijer & Keizer, 1996; Hillig, 2005a; Spitzer-Rimon *et al.*, 2019; Carlson
465 *et al.*, 2021; Naim-Feil *et al.*, 2021; Stack *et al.*, 2021; Chen *et al.*, 2022) and differences in

466 cannabinoid profiles, leaflet index and phenological development were proposed as
467 characteristics to discriminate between them (de Meijer & Keizer, 1996). Heterochronic
468 shifts are apparent in the differential rates that accessions increase leaflet number across
469 nodes, as well as maximum and average leaflet counts across accessions (Fig. **1**).
470 Remarkably, stages in developmental timing are conserved despite being shifted. For
471 example, a significant shape change exhibited between the leaves with 3 and leaves with 5
472 leaflets, with leaflets becoming more acuminate and narrower. In contrast, changes in shape
473 between leaves with a higher number of leaflets were more gradual. Additionally, we
474 observed a similar shift in leaf shape between the nodes 0.3 and 0.4, potentially indicating a
475 transition between the juvenile and adult phases of leaf development. Similar results were
476 obtained in previous research. Spitzer-Rimon *et al.* (2022) demonstrated that flowering buds
477 were initiated at node 7, while Moliterni *et al.* (2004) analyzing a different cultivar, found
478 developing flower buds in the 4th node, suggesting that transitions in growth phases are
479 conserved but not synchronized across cultivars.
480

481 **Conclusions**

482 In grapevine, leaf shape has long been utilized for variety identification. However, in the
483 case of *Cannabis*, previous attempts were hindered by the variability in leaflet numbers. In
484 this study, we present a pioneering method that successfully addresses this issue. By
485 generating synthetic leaves with customizable leaflet counts, we can now employ high-
486 resolution morphometric techniques to accurately classify different wild/feral and cultivated
487 *Cannabis* accessions. Through the use of 1,995 densely placed pseudo-landmarks, we were
488 able to predict the accession identity with over 74% accuracy. The method works well not
489 only on stabilized cultivars, but also on phenotypically more variable wild/feral accessions
490 grown from seed. Unifying the number of leaflets allowed us, for the first time, to make
491 comparisons among several leaves along the main axis, enabling us to investigate
492 developmental changes in leaf shape and detect heterochronic mechanisms influencing the
493 leaf shape in *Cannabis*. The implications of this new high-resolution method in both the
494 cannabis industry and research extend beyond its role in determining *Cannabis* accessions.
495 It also offers a promising tool for developmental studies, and for studying the correlation
496 between leaf shape and phytochemical profiles and the sex of the plants, where lower-
497 resolution methods provided inconclusive results so far. The method presented here offers

a fast, effective, robust and low-cost tool that can aid the future classification of *Cannabis* germplasm. Furthermore, the use of this methodology extends beyond *Cannabis*, and can be applied to numerous other plant species with varying numbers of lobes and leaflets where the use of geometric morphometrics methods was not previously possible.

ACKNOWLEDGEMENTS

This research was supported by projects WECANN (CGL2017-84297-R, Ministerio de Ciencia, Innovación y Universidades), and M. Balant FPI predoctoral contract of the Ministerio de Ciencia, Innovación y Universidades (PRE2018- 083226). This work is also supported by NSF Plant Genome Research Program awards IOS-2310355, IOS-2310356, and IOS-2310357. This project was supported by the USDA National Institute of Food and Agriculture, and by Michigan State University AgBioResearch. We would like to thank Joan Uriach Marsal from Uriach Laboratories for additional financial support. We would also like to acknowledge Cannaflos—Gesellschaft für medizinisches Cannabis mbH for providing seed of accessions IK, IKL and MAR analyzed in the study, Carlos Sáez, Carlos Ribelles, Airy Gras, Joan Vallès, Magsar Urgamal, Shagdar Tsooj, Marine Oganesian and Nina Stepanyan-Gandilyan for help with the sample collection and the cultivation process, and Paula Bruna for helpful advice on improving the quality of scanned images.

COMPETING INTERESTS

None declared.

AUTHOR CONTRIBUTIONS

MB and DHC conceived the study, with OH, TG and DV inputs in a preliminary design phase of the project. MB, TG and DV cultivated the plants. MB and OH selected the method for imaging *Cannabis* leaves. MB and TG scanned the leaves used in the study. MB and DHC developed the morphometric method and MB analyzed the data. MB and DHC wrote the first draft of the paper, that all authors read, commented, and edited.

DATA AVAILABILITY

The datasets and code for the method developed here are freely available on GitHub (https://github.com/BalantM/Cannabis_leaf_morpho).

ORCID

Manica Balant https://orcid.org/0000-0002-4316-8511

Teresa Garnatje https://orcid.org/0000-0001-6295-6217

Daniel Vitales https://orcid.org/0000-0002-3109-5028

Oriane Hidalgo https://orcid.org/0000-0002-1547-8627

Daniel H. Chitwood https://orcid.org/0000-0003-4875-1447

SUPPORTING INFORMATION

The datasets and code generated and analyzed in this study are available on GitHub (https://github.com/BalantM/Cannabis_leaf_morpho).

REFERENCES

Abràmoff MD, Magalhães PJ, Ram SJ. **2004**. Image processing with ImageJ Part II. *Biophotonics International* **11**: 36–43.

Anderson LC. **1980**. Leaf Variation among Cannabis Species from a Controlled Garden. *Botanical Museum leaflets, Harvard University* **28**: 61–69.

Balant M, González Rodríguez R, Garcia S, Garnatje T, Pellicer J, Vallès J, Vitales D, Hidalgo O. **2022**. Novel Insights into the Nature of Intraspecific Genome Size Diversity in *Cannabis sativa* L . *Plants* **11**: 2736.

Bookstein FL. **1997**. *Morphometric Tools for Landmark Data: Geometry and Biology*. ambrige: Cambrige University.

Bradski G. **2000**. The OpenCV Library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer* **25**: 120–123.

Bryson AE, Wilson Brown M, Mullins J, Dong W, Bahmani K, Bornowski N, Chiu C, Engelgau P, Gettings B, Gomezcano F, *et al.* 2020. Composite modeling of leaf shape along shoots discriminates *Vitis* species better than individual leaves. *Applications in Plant Sciences* **8**:

561 e11404.
562 **Campbell LG, Peach K, Wizenberg SB**. **2021**. Dioecious hemp (*Cannabis sativa* L.) plants do
563 not express significant sexually dimorphic morphology in the seedling stage. *Scientific*
564 *Reports* **11**: 1–8.
565 **Carlson CH, Stack GM, Jiang Y, Taşklran B, Cala AR, Toth JA, Philippe G, Rose JKC, Smart**
566 **CD, Smart LB**. **2021**. Morphometric relationships and their contribution to biomass and
567 cannabinoid yield in hybrids of hemp (*Cannabis sativa*). *Journal of Experimental Botany* **72**:
568 7694–7709.
569 **Chen X, Guo H, Zhang Q, Wang L, Guo R, Zhan Y, Lv P, Xu Y, Guo M, Zhang Y, *et al.* 2022**.
570 Whole-genome resequencing of wild and cultivated cannabis reveals the genetic structure
571 and adaptive selection of important traits. *BMC Plant Biology* **22**: 371.
572 **Chitwood DH**. **2021**. The shapes of wine and table grape leaves: An ampelometric study
573 inspired by the methods of Pierre Galet. *Plants People Planet* **3**: 155–170.
574 **Chitwood DH, Klein LL, O'Hanlon R, Chacko S, Greg M, Kitchen C, Miller AJ, Londo JP**.
575 **2016a**. Latent developmental and evolutionary shapes embedded within the grapevine leaf.
576 *New Phytologist* **210**: 343–355.
577 **Chitwood DH, Mullins J, Migicovsky Z, Frank M, VanBuren R, Londo JP**. **2021**. Vein-to-blade
578 ratio is an allometric indicator of leaf size and plasticity. *American Journal of Botany* **108**:
579 571–579.
580 **Chitwood DH, Ranjan A, Martinez CC, Headland LR, Thiem T, Kumar R, Covington MF,**
581 **Hatcher T, Naylor DT, Zimmerman S, *et al.* 2014**. A modern ampelography: A genetic basis
582 for leaf shape and venation patterning in grape. *Plant Physiology* **164**: 259–272.
583 **Chitwood DH, Rundell SM, Li DY, Woodford QL, Yu TT, Lopez JR, Greenblatt D, Kang J,**
584 **Londo JP**. **2016b**. Climate and developmental plasticity: Interannual plasticity in grapevine
585 leaf morphology. *Plant Physiology* **170**: 1480–1491.
586 **Chitwood DH, Sinha NR**. **2016**. Evolutionary and Environmental Forces Sculpting Leaf
587 Development. *Current Biology* **26**: R297–R306.
588 **Clarke R, Merlin M**. **2013**. Cannabis*: Evolution and ethnobotany*. Berkley, Los Angeles,
589 London: University of California Press.
590 **Demmings EM, Williams BR, Lee C-R, Barba P, Yang S, Hwang C-F, Reisch BI, Chitwood DH,**
591 **Londo JP**. **2019**. Quantitative Trait Locus Analysis of Leaf Morphology Indicates Conserved
592 Shape Loci in Grapevine. *Frontiers in Plant Science* **10**: 1373.

Emboden WA. **1974**. *Cannabis* - a polytypic genus. *Economic Botany* **28**: 304–310.

Galet P. **1952**. *Précis d'Ampélographie Pratique*. Digitized by Google Books from Cornell University.

Galet P, trans. Morton L. **1979**. *A Practical Ampelography: Grapevine Identification*. Cornell University Press (English translation).

Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, *et al.* 2020. Array programming with NumPy. *Nature* **585**: 357–362.

Heslop-Harrison J, Heslop-Harrison Y. **1958**. Studies on flowering-plant growth and organogenesis-III. Leaf shape changes associated with flowering and sex differentiation in *Cannabis sativa*. *Proceedings of the Royal Irish Academy. Section B: Biological, Geological, and Chemical Science* **59**: 257–283.

Hillig KW. **2004**. A multivariate analysis of allozyme variation in 93 *Cannabis* accessions from the VIR germplasm collection. *Journal of Industrial Hemp* **9**: 5–22.

Hillig KW. **2005a**. *A Systematic Investigation of Cannabis [PhD Thesis]*. Indiana University.

Hillig KW. **2005b**. Genetic evidence for speciation in *Cannabis* (Cannabaceae). *Genetic Resources and Crop Evolution* **52**: 161–180.

Hunter JD. **2007**. Matplotlib: a 2D graphics environment. *Computing in Science and Engineering* **9**: 90–95.

Hurgobin B, Tamiru-Oli M, Welling MT, Doblin MS, Bacic A, Whelan J, Lewsey MG. **2021**. Recent advances in *Cannabis sativa* genomics research. *New Phytologist* **230**: 73–89.

Jin D, Henry P, Shan J, Chen J. **2021a**. Identification of phenotypic characteristics in three chemotype categories in the genus *Cannabis*. *HortScience* **56**: 481–490.

Jin D, Henry P, Shan J, Chen J. **2021b**. Classification of cannabis strains in the Canadian market with discriminant analysis of principal components using genome-wide single nucleotide polymorphisms (T-Y Chiang, Ed.). *PLOS ONE* **16**: e0253387.

Kuhl FP, Giardina CR. **1982**. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing* **18**: 236–258.

Lamarck J-BPA de M de, Poiret J-L-M. **1783**. *Encyclopédie méthodique. Botanique*. Paris,Liège, Panckoucke;Plomteux.

Lapierre É, Monthony AS, Torkamaneh D. **2023**. Genomics-based taxonomy to clarify cannabis classification. *Genome*.

McPartland JM. **2018**. *Cannabis* Systematics at the Levels of Family, Genus, and Species. *Cannabis and Cannabinoid Research* **3**: 203–212.

McPartland JM, Guy GW. **2017**. Models of Cannabis Taxonomy, Cultural Bias, and Conflicts between Scientific and Vernacular Names. *Botanical Review* **83**: 327–381.

McPartland JM, Small E. **2020**. A classification of endangered high-THC cannabis (Cannabis sativa subsp. indica) domesticates and their wild relatives. *PhytoKeyes* **144**: 81–112.

de Meijer EPM, Keizer LCP. **1996**. Patterns of diversity in Cannabis. *Genetic Resources and Crop Evolution* **43**: 41–52.

Migicovsky Z, Swift JF, Helget Z, Klein LL, Ly A, Maimaitiyiming M, Woodhouse K, Fennell A, Kwasniewski M, Miller AJ, *et al.* 2022. Increases in vein length compensate for leaf area lost to lobing in grapevine. *American Journal of Botany* **109**: 1063–1073.

Moliterni VMC, Cattivelli L, Ranalli P, Mandolino G. **2004**. The sexual differentiation of *Cannabis sativa* L.: A morphological and molecular study. *Euphytica* **140**: 95–106.

Murovec J, Eržen JJ, Flajšman M, Vodnik D. **2022**. Analysis of Morphological Traits, Cannabinoid Profiles, THCAS Gene Sequences, and Photosynthesis in Wide and Narrow Leaflet High-Cannabidiol Breeding Populations of Medical Cannabis. *Frontiers in Plant Science* **13**: 786161.

Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia JM, Ware D, *et al.* 2011. Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 3530–3535.

Naim-Feil E, Pembleton LW, Spooner LE, Malthouse AL, Miner A, Quinn M, Polotnianka RM, Baillie RC, Spangenberg GC, Cogan NOI. **2021**. The characterization of key physiological traits of medicinal cannabis (*Cannabis sativa* L.) as a tool for precision breeding. *BMC Plant Biology* **21**: 1–15.

Niklas KJ. **1994**. *Plant allometry: the scaling of form and process*. University of Chicago Press.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, *et al.* 2011. Generating the blood exposome database using a comprehensive text mining and database fusion approach. *Journal ofMachine Learning Research* **12**: 2825–2830.

Potter DJ. **2009**. *Optimisation Of* Cannabis sativa *L. As A Phytopharmaceutical [PhD Thesis]*.

657 London: King's College London, Department of Pharmaceutical Science Research.
658 **Prentout D, Razumova O, Rhoné B, Badouin H, Henri H, Feng C, Käfer J, Karlov G, Marais**
659 **GAB**. **2020**. An efficient RNA-seq-based segregation analysis identifies the sex chromosomes
660 of *Cannabis sativa*. *Genome Research* **30**: 164–172.
661 **Quimby MW, Doorenbos NJ, Turner CE, Masoud A**. **1973**. Mississippi-Grown Marihuana -
662 *Cannabis sativa* Cultivation and Observed Morphological Variations. *Economic Botany* **27**:
663 117–127.
664 **Ravaz L**. **1902**. *Les vignes américaines: Porte-greffes et producteurs directs*. Goulet,
665 Montpellier and Paris: Digitized by Google Books from Cornell University.
666 **Ren G, Zhang X, Li Y, Ridout K, Serrano-Serrano ML, Yang Y, Liu A, Ravikanth G, Nawaz MA,**
667 **Mumtaz AS, *et al.* 2021**. Large-scale whole-genome sequencing unravels the domestication
668 history of *Cannabis sativa*. *Science Advances* **7**: eabg2286.
669 **Sawler J, Stout JM, Gardner KM, Hudson D, Vidmar J, Butler L, Page JE, Myles S**. **2015**. The
670 Genetic Structure of Marijuana and Hemp. *PLOS ONE* **10**: e0133292.
671 **Schultes RE, Klein WM, Plowman T, Lockwood TE**. **1974**. Cannabis: An example of
672 taxonomic neglect. *Botanical Museum Leaflets, Harvard University* **23**: 337–367.
673 **Schwabe AL, McGlaughlin ME**. **2019**. Genetic tools weed out misconceptions of strain
674 reliability in *Cannabis sativa*: implications for a budding industry. *Journal of Cannabis*
675 *Research* **1**: 3.
676 **Small E**. **2015**. Evolution and Classification of *Cannabis sativa* (Marijuana, Hemp) in Relation
677 to Human Utilization. *Botanical Review* **81**: 189–294.
678 **Small E, Beckstead HD**. **1973**. Cannabinoid Phenotypes in *Cannabis sativa*. *Nature* **245**: 147–
679 148.
680 **Small E, Cronquist A**. **1976**. A Practical and Natural Taxonomy for Cannabis. *Taxon* **25**: 405–
681 435.
682 **Small E, Jui PY, Lefkovitch LP**. **1976**. A Numerical Taxonomic Analysis of Cannabis with
683 Special Reference to Species Delimitation. *Systematic Botany* **1**: 67–84.
684 **Spitzer-Rimon B, Duchin S, Bernstein N, Kamenetsky R**. **2019**. Architecture and Florogenesis
685 in Female *Cannabis sativa* Plants. *Frontiers in Plant Science* **10**: 350.
686 **Spitzer-Rimon B, Shafran-Tomer H, Gottlieb GH, Doron-Faigenboim A, Zemach H,**
687 **Kamenetsky-Goldstein R, Flaishman M**. **2022**. Non-photoperiodic transition of female
688 cannabis seedlings from juvenile to adult reproductive stage. *Plant Reproduction* **35**: 265–

277.

Stack GM, Toth JA, Carlson CH, Cala AR, Marrero-González MI, Wilk RL, Gentner DR, Crawford JL, Philippe G, Rose JKC, *et al.* 2021. Season-long characterization of high-cannabinoid hemp (*Cannabis sativa* L.) reveals variation in cannabinoid accumulation, flowering time, and disease resistance. *GCB Bioenergy* **13**: 546–561.

The Global Cannabis Report, 3rd Edition. **2022**. (https://prohibitionpartners.com/reports/the-global-cannabis-report-edition-3/) [accessed 13 August 2023]

Torres A, Pauli C, Givens R, Argyris J, Allen K, Monfort A, Gaudino RJ. **2022**. High-throughput methods to identify male *Cannabis sativa* using various genotyping methods. *Journal of Cannabis Research* **4**: 57.

Toth JA, Stack GM, Cala AR, Carlson CH, Wilk RL, Crawford JL, Viands DR, Philippe G, Smart CD, Rose JKC, *et al.* 2020. Development and validation of genetic markers for sex and cannabinoid chemotype in *Cannabis sativa* L. *GCB Bioenergy* **12**: 213–222.

Vergara D, Feathers C, Huscher EL, Holmes B, Haas JA, Kane NC. **2021**. Widely assumed phenotypic associations in *Cannabis sativa* lack a shared genetic basis. *PeerJ* **9**: e10672.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, *et al.* 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**: 261–272.

Waskom M. **2021**. seaborn: statistical data visualization. *Journal of Open Source Software* **6**: 3021.

Watts S, McElroy M, Migicovsky Z, Maassen H, van Velzen R, Myles S. **2021**. *Cannabis* labelling is associated with genetic variation in terpene synthase genes. *Nature Plants* **7**: 1330–1334.

Zhou Y, Massonnet M, Sanjak JS, Cantu D, Gaut BS. **2017**. Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proceedings of the National Academy of Sciences of the United States of America* **114**: 11715–11720.

Zhukovskii PM. **1971**. *Cultivated plants and their wild relatives*. Leningrad, USSR, Kolos.