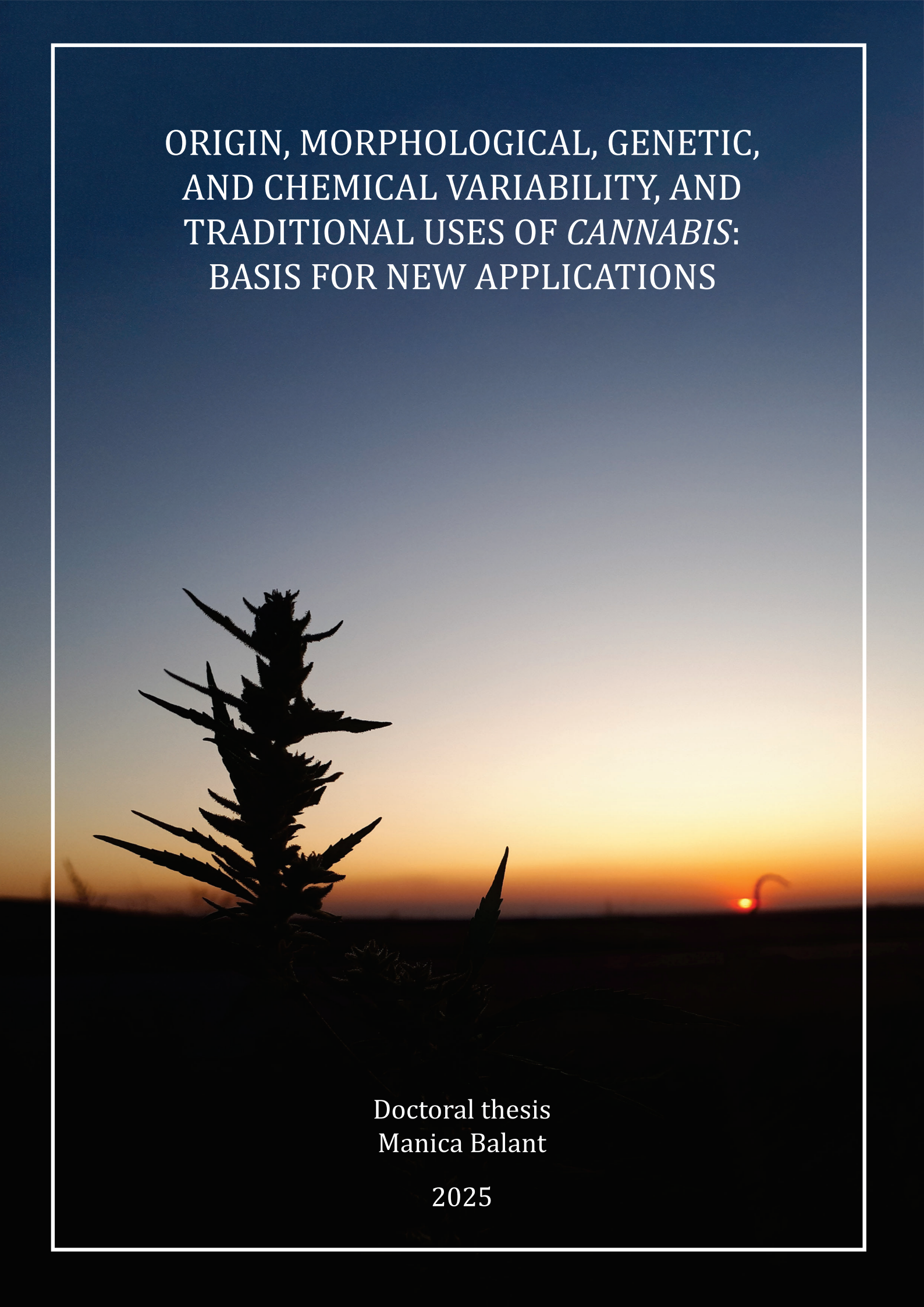
UNIVERSITAT DE BARCELONA

Origin, morphological, genetic, and chemical variability, and traditional uses of *Cannabis*: basis for new applications

Manica Balant

ORIGIN, MORPHOLOGICAL, GENETIC, AND CHEMICAL VARIABILITY, AND TRADITIONAL USES OF *CANNABIS*: BASIS FOR NEW APPLICATIONS



Doctoral thesis

Manica Balant

2025

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

ORIGIN, MORPHOLOGICAL, GENETIC, AND CHEMICAL VARIABILITY, AND TRADITIONAL USES OF *CANNABIS*: BASIS FOR NEW APPLICATIONS

ORIGEN, VARIABILITAT MORFOLÒGICA, GENÈTICA I QUÍMICA, I USOS TRADICIONALS DEL *CANNABIS*: BASE PER A NOVES APLICACIONS

RECERCA, DESENVOLUPAMENT I CONTROL DE MEDICAMENTS

MANICA BALANT

2025

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

RECERCA, DESENVOLUPAMENT I CONTROL DE MEDICAMENTS

Departament de Biologia, Sanitat i Medi Ambient

Secció de Botànica

ORIGIN, MORPHOLOGICAL, GENETIC, AND CHEMICAL VARIABILITY, AND TRADITIONAL USES OF *CANNABIS*: BASIS FOR NEW APPLICATIONS

ORIGEN, VARIABILITAT MORFOLÒGICA, GENÈTICA I QUÍMICA, I USOS TRADICIONALS DEL *CANNABIS*: BASE PER A NOVES APLICACIONS

Memòria presentada per Manica Balant per optar al títol de doctor per la Universitat de Barcelona

Dra. Teresa Garnatje Roca

Directora

Manica Balant

Doctoranda

Dr. Daniel Vitales Serrano

Director

Prof. Joan Vallès Xirau

Tutor

MANICA BALANT

2025

This thesis was funded by the Spanish Ministry of Science, Innovation and Universities under the project WECANN, grant number: CGL2017- 84297-R; and an FPI predoctoral contract associated to the project, grant number: PRE2018-083226.

Dolga, dolga je cesta iz vasi do mesta.
Dolga, dolga je reka iz plenic do človeka.
Polna lukenj in hrepenenja,
tolmunov, želja in ihtenja,
dolga, kot je lahko le noč,
dolga je ta pot od nekje do nekoč.

Vlado Kreslin

ACKNOWLEDGEMENTS

This thesis has proven to be a much bigger challenge than I imagined in my first year (as I suppose is almost always the case with first-year PhD students). Nonetheless, I am extremely grateful for this experience. It gave me an opportunity to study this extraordinary plant (with both its positive and negative aspects), learn so much while doing it, and meet many incredible people along the way. Without your support, this thesis would not be what it is, nor would I have become who I am today.

First and foremost, I would like to thank my mentors, Dr. Teresa Garnatje and Dr. Daniel Vitales, for giving me this opportunity, for your support and trust. You did much more for me than simply being my mentors, for which I will be forever grateful. You helped me with the paperwork from the very beginning and did not give up, when the Spanish bureaucratic wheel decided not to make it easy on us. I am extremely grateful to have you as my mentors and for your guidance during my thesis – you always made sure to point me in the right direction so that learning from my mistakes was possible, but never too painful. Thank you for always being there for me, whenever I needed you, both professionally and personally.

I would also like to thank my tutor at the University of Barcelona Dr. Joan Vallès. I am very grateful for your help during my first months here, trying to make sense of all the ministries and departments – Spanish or Catalan. Without you I would still be standing in some line now, asking for a homologation of my degree. I am very grateful to be able to work with you during the thesis.

I would also like to express my gratitude to the collaborators and co-authors of the publications with whom I had the privilege of working over the past five years. A special thank you goes to all the team members from the WECANN project, especially to: Dr. Teresa Garnatje, Dr. Daniel Vitales, Dr. Joan Vallès, Dr. Oriane Hidalgo, Dr. Airy Gras, Dr. Sònia Garcia, Dr. Jaume Pellicer, Dr. Guillermo Quintas, Carlos Saez and Alexandra Papamichail. Together we successfully resolved all the legal and bureaucratic issues that *Cannabis* brought to the table and managed to carry out so many interesting experiments on this fascinating plant! I learned so much from you all and I feel very lucky to have been part of this team. Without you, this thesis would not be finished.

I would also like to thank to Dr. Sonja Siljak-Yakovlev, Dr. Daniel Chitwood, and Dr. Lisa Pokorny for accepting me and hosting me during my research stays at your institutions. Because of your warm welcome, I really enjoyed my stays with you and learned a lot from each one.

A special thanks also goes to all past and present colleagues from Botanical Institute of Barcelona: Aurora, Beñat, David, Irem, Iván, Joan Pere, Leonardo, Lucía, María Luisa, Pol, Roger, Roi, Teodora, Uxue, and many others. Thank you for your help and support, listening to me complain when things were not great and laughing with me when they were, fighting the cytometer, extracting DNA from samples, helping me solve the errors in CESGA that did not make any sense to me, etc. Sharing lunch breaks and coffees below the *Quercus* sp. will always be one of my best memories of this time.

I would also like to thank my colleagues at University of Primorska in Slovenia, who gave me my first opportunity when I was fresh out of university with no prior research experience. Boštjan, Peter and Živa, with you three, I took my first steps into the research world and learned that all the hard work, frustrations, questions, and uncertainties that come with it are much easier to embrace when you're working with friends.

Thanks also to my parents Mira and Marko and my sister Nataša. I would not be here without your unconditional love, encouragement, and support. You instilled in me the belief that I could become whatever I wanted and encouraged me to pursue studies and work that make me happy. I could always count on you to have my back, which made it so much easier to seize opportunities I might not have been able to otherwise.

And finally, Jan, I am not sure if I have thanked you enough for coming here with me. I hope you know that this thesis would not happen without your support. Having you here with me, made all the difference in the world. I know I promised it's only for three years, and its already the sixth one, but I am hoping you don't mind the extension… I guess it fits with the millions of 'zdej pačas kančam' and 'čak sam še tu nardim' in the past five years, for which I am sorry. I can't wait to see what our next challenge will be, but with you by my side, I know it will be great!

ABSTRACT

Cannabis sativa L. (hereafter referred to as *Cannabis*) is one of the earliest domesticated plants, with human use spanning over millennia for a variety of purposes. Given its widespread use and significance, *Cannabis* has been the focus of research in diverse scientific areas, from medicine, pharmacy, and agronomy to environmental and social sciences. However, most of the research has concentrated on modern hemp cultivars and drug strains, with limited attention given to wild-growing populations or traditional landraces across the plant's natural distribution area, leaving significant gaps in understanding their diversity. Similarly, while numerous publications have documented the traditional uses of *Cannabis*, this body of knowledge remains fragmented across various types of literature. This fragmentation, much like the limited research on wild-growing populations, complicates efforts to conduct comprehensive statistical analyses and gain a holistic understanding of the plant's diversity and cultural significance.

This doctoral thesis aims to fill this gap by exploring the genetic, morphological, and phytochemical diversity of *Cannabis* and its traditional uses worldwide through seven interconnected studies. The research integrated various interdisciplinary methods, such as flow cytometry, Hyb-Seq genomic analysis, high-resolution geometric morphometrics, phytochemical analysis using UPLC-MS, and systematic ethnobotanical data collection, offering a comprehensive view of this versatile plant.

In the first chapter we explored the genetic and cytogenetic diversity of *Cannabis.* Using flow cytometry, we examined genome size diversity across a wide range of wild-growing, landrace, and cultivated *Cannabis* accessions. The analysis revealed minimal variation in ploidy levels, with only one triploid individual identified, while the rest were diploid. A 1.189-fold variation in genome size was observed, but no strong correlation was found between genome size and geographical distribution or between wild-growing and cultivated plants. This suggests that the geographical expansion and domestication of *Cannabis* had a limited effect on its genome size. Notably, significant differences in genome size were observed between male and female plants. In another work of this chapter, we examined the phylogeographic and population genetics structure of wild-growing and landrace *Cannabis* populations, by employing a Hyb-Seq approach, combining shotgun sequencing and target capture with the Angiosperms353 enrichment panel. The results supported the classification of *Cannabis* as a monotypic genus, structured into three primary genetic groups: E Asia, which is sister to both Paleotropis group and Boreal group. Further subdivisions within these groups revealed phylogeographic patterns driven more by geographical distribution than by use-type.

In the second chapter we focused on the morphological diversity of *Cannabis*. Due to significant changes in leaflet number during the plants' development, a novel method was developed that allows, for the first-time, the application of a high-resolution geometric morphometrics approach on the entire leaf. By using polynomial models across more than 3,000 pseudo-landmarks, we modelled theoretical leaves with

comparable leaflet numbers. This innovative approach enables the accurate prediction of genetic and developmental traits, providing new insights into *Cannabis* leaf morphology and uncovering heterochronic mechanisms responsible for changes in leaf shape during the development.

Phytochemical diversity was the focus of the third chapter, which analysed cannabinoid composition in different *Cannabis* tissues, sexes, and across different phylogeographic groups identified in the first chapter. Significant differences in cannabinoid profiles were found between male and female inflorescences, but not between the leaves of either sex. While the cannabinoid profiles did not precisely correlate with phylogeographic groups, there was a clear differentiation between cultivated drug-type landraces and wild-growing populations, indicating the influence of selective breeding during domestication.

The last, forth chapter explored traditional knowledge of *Cannabis* use. The first study in this chapter involved the development of the CANNUSE database, which consolidated 2,330 entries from 649 publications across 41 countries, providing a valuable resource for understanding traditional *Cannabis* uses. The second study analysed the standardized dataset from CANNUSE database and found that medicinal applications dominated, treating over 200 human ailments. Strong associations were observed between specific *Cannabis* plant parts and their corresponding uses and in treating different body systems. The third ethnobotanical study investigated the traditional knowledge of *Cannabis* use in Armenia through ethnobotanical surveys and bibliographic review. The results revealed a significant decline in medicinal and fibre uses but highlighted the continued importance of *Cannabis* seeds in alimentary uses, particularly in symbolic dishes, demonstrating the persistence of *Cannabis* as an important cultural and historical resource.

In conclusion, the results of this thesis highlight the extraordinary genetic, morphological, and phytochemical diversity of *Cannabis,* particularly evident when focusing on wild-growing populations and traditional landraces. The findings support the classification of *Cannabis* as a monotypic genus, with marked intraspecific variability. The newly gathered comprehensive dataset of traditional uses provided novel insights into its rich traditional heritage, and revealed new potential applications rooted in traditional knowledge. These results underscore the importance of further research and the preservation of diversity within wild-growing populations of this versatile species.

RESUM

Cannabis sativa L. (en endavant, *Cannabis)* és una de les primeres plantes domesticades, usades pels humans des de fa mil·lennis amb diverses finalitats. Donat el seu ús estès i la seva rellevància, el *Cannabis* ha estat objecte de recerca en diverses àrees científiques, des de medicina, farmàcia i agronomia fins a ciències ambientals i socials. Tanmateix, la major part de la recerca s'ha centrat en varietats modernes de cànem i estirps psicoactives, deixant més o menys de banda les poblacions silvestres o les races tradicionals en la seva àrea natural de distribució i generant buits significatius en el coneixement de la seva diversitat. De manera similar, malgrat que els usos tradicionals de *Cannabis* han estat recollits en nombroses publicacions, aquest coneixement es troba dispers en diferents formats de publicació. Aquesta fragmentació, juntament amb una investigació limitada sobre poblacions silvestres, dificulta les anàlisis estadístiques i la comprensió holística de la diversitat i la significació cultural de la planta.

Aquesta tesi doctoral pretén omplir aquest buit explorant la diversitat genètica, morfològica i fitoquímica de *Cannabis* i els seus usos tradicionals arreu del món a través de set estudis complementaris i estretament relacionats. Aquesta recerca integra diverses metodologies interdisciplinàries, com la citometria de flux, l'anàlisi genòmica Hyb-Seq, la morfometria geomètrica d'alta resolució, l'anàlisi fitoquímica mitjançant UPLC-MS i la recopilació sistemàtica de dades etnobotàniques, i permet oferir una visió integral d'aquesta planta versàtil.

Al primer capítol es va explorar la diversitat genètica i citogenètica de *Cannabis*. Mitjançant citometria de flux, es va estimar la diversitat de la mida del genoma en una àmplia gamma de poblacions silvestres, races tradicionals i accessions cultivades. L'anàlisi va revelar una variació mínima en els nivells de ploïdia, amb només un individu triploide identificat, essent la resta diploides. Es va observar una variació d'1,189 vegades en la mida del genoma, però no es va trobar una correlació estreta entre la mida del genoma i la distribució geogràfica, ni entre plantes silvestres, ni en cultivades. Això suggereix que l'expansió geogràfica i la domesticació de *Cannabis* van tenir un efecte limitat en la mida del genoma. S'observaren, però, diferències significatives en la mida del genoma entre plantes masculines i femenines. En un altre treball d'aquest capítol, per a examinar l'estructura filogeogràfica i genètica de les poblacions de *Cannabis* silvestres i tradicionals, es va utilitzar l'enfocament Hyb-Seq, que combina la captura dirigida i la seqüenciació massiva amb el *kit Angiosperms353*. Els resultats obtinguts donen suport a la consideració de *Cannabis* com un gènere monotípic, estructurat en tres grups genètics principals: Àsia oriental, que és el grup germà dels grups paleotropical i boreal. Les subdivisions dins d'aquests grups van revelar patrons filogeogràfics condicionats per la distribució geogràfica més que pel tipus d'ús.

El segon capítol es va centrar en la diversitat morfològica de *Cannabis*. A causa de canvis significatius en el nombre de folíols de les fulles que tenen lloc durant el creixement de la planta, es va desenvolupar un mètode nou, que permet, per primera vegada, l'aplicació d'un enfocament de morfometria geomètrica d'alta resolució. Utilitzant models polinòmics en més de 3.000 pseudopunts de referència, es van modelar fulles teòriques amb nombres de folíols comparables. Aquest enfocament innovador permet predir amb

precisió trets genètics i de desenvolupament, de manera que proporciona noves perspectives sobre la morfologia de les fulles i permet descobrir mecanismes heterocrònics responsables dels canvis en la forma de les fulles durant el desenvolupament.

La diversitat fitoquímica va ser l'objecte del tercer capítol, en el qual s'analitzà la composició de cannabinoides en diferents teixits, sexes i grups filogeogràfics identificats en l'estudi anteriorment esmentat de *Cannabis*. Es van trobar diferències significatives en els perfils de cannabinoides entre les inflorescències masculines i femenines, però no entre les fulles dels dos sexes. Tot i que els perfils de cannabinoides no es correlacionava exactament amb els grups filogeogràfics, hi havia una clara diferenciació entre les races tradicionals cultivades per a droga i les poblacions silvestres, la qual cosa indica la influència de la selecció durant la domesticació.

Finalment, el quart capítol va explorar el coneixement tradicional dels usos de *Cannabi*s. El primer estudi d'aquest capítol va significar la creació de la base de dades CANNUSE, que inclogué 2.330 registres corresponents a 649 publicacions de 41 països, la qual cosa proporciona un recurs valuós per a comprendre els usos tradicionals de *Cannabis.* El segon estudi analitzà el conjunt de dades estandarditzades de CANNUSE i posà de manifest que les aplicacions medicinals eren les predominants, i que la planta s'usava per al tractament de més de 200 malalties humanes. S'observaren correspondències entre parts de la planta i els seus usos per a tractar diferents sistemes corporals. El tercer estudi etnobotànic investigà el coneixement tradicional sobre els usos de *Cannabis* a Armènia mitjançant entrevistes etnobotàniques i revisions bibliogràfiques. Els resultats revelaren un descens significatiu al llarg del temps pel que fa als usos medicinals i de fibra, però posaren de manifest la importància actual de les llavors de *Cannabis* en usos alimentaris, especialment en plats simbòlics, demostrant-ne la persistència com un recurs cultural i històric important.

En conclusió, els resultats d'aquesta tesi posen de manifest l'extraordinària diversitat genètica, morfològica i fitoquímica de *Cannabis*, especialment palesa en poblacions silvestres i races tradicionals. Els resultats donen suport a un gènere monotípic, amb una marcada variabilitat intraespecífica. Les dades exhaustives sobre usos tradicionals proporcionen noves perspectives sobre el seu ric patrimoni tradicional i revelen noves aplicacions potencials arrelades en aquest coneixement. Aquests resultats subratllen la importància de la recerca i la preservació de la diversitat en les poblacions silvestres d'aquesta espècie tan versàtil.

TABLE OF CONTENTS

INTRODUCTION

INTRODUCTION

The overview of *Cannabis* biology

Cannabis sativa L. (hereafter referred to as *Cannabis*) is one of the most versatile plants known to humanity, deeply embedded in the cultural traditions of many societies around the world. A remarkable phenotypic plasticity has allowed *Cannabis* to play a vital role in various aspects of human life, ranging from medicinal and ritualistic practices to everyday applications in food and fibre production (Clarke & Merlin, 2013). *Cannabis* belongs to the diverse Cannabaceae family, which comprises 10 genera and over 100 species. Among them, two closely related species are particularly notable for their economic importance: hops (*Humulus lupulus* L.), a key ingredient in the beer industry, and *Cannabis*, widely utilized in both medical and recreational markets (Fu *et al.*, 2023). *Cannabis* is an annual, herbaceous and wind-pollinated plant. Its natural habitats are open steppes, with plenty of sun, nutrient rich and with well-drained soil across Eurasia (Clarke & Merlin, 2013; Small, 2015). However, due to its long history of human use and remarkable adaptability, it is nowadays one of the most widely distributed plants, found in different environments ranging from ruderal habitats, steppes, valleys and riverbanks to fallow fields and even forests (Vavilov, 1992; Small, 2015). Its current global distribution, both as cultivated and naturalized plant, spans across all continents, except Antarctica (Clarke & Merlin, 2013).

Cannabis plants are predominately dioecious, though some monoecious cultivars also exist. The species has a diploid genome with 20 chromosomes—18 autosomes and one pair of sex chromosomes. Female plants have two X chromosomes (XX) whereas male plants have X and Y (XY) chromosomes (Hirata, 1925; Faux *et al.*, 2014). While monoecious plants generally have two X chromosomes, some cases of monoecious plants with X and Y chromosomes have also been reported (Heer *et al.*, 2024).

Male and female *Cannabis* plants cannot be morphologically distinguished until their reproductive organs start to develop. However, once the plants reach the reproductive stage, many morphological and developmental differences between male and female individuals become evident. To ensure successful reproduction in their natural habitats, male plants grow taller and faster, developing loosely branched terminal inflorescences that produce enormous amounts of pollen (Figure 1). Contrary, female plants grow slower and develop compact racemose terminal inflorescences. Each flower contains a single ovule which, upon pollination, develops into an achene or nut, containing a single seed (in both scientific and non-scientific literature, the *Cannabis* achene is commonly referred to as a "seed", a term that will also be used here in continuation). Male plants die shortly after pollination, while seeds continue to mature on the female plants (Small & Cronquist, 1976). The female flower, and later the developing seed are protected with a bract densely covered with various types of glandular trichomes (bulbous, sessile, and stalked), where secondary metabolites, such as cannabinoids and terpenoids are being stored. Glandular

trichomes are also present on other plant parts, including leaves, stems, and male flowers, but are generally less abundant (Livingston *et al.*, 2020). Following fertilization, female *Cannabis* plants shift their energy toward seed development, leading to a decline in the production of secondary compounds like cannabinoids and terpenoids. Since cannabinoid- and terpene-rich female inflorescences are the primary product for medicinal and recreational use, early detection and removal of male *Cannabis* plants from cultivation is crucial to prevent pollination. This ensures maximum production of cannabinoids and terpenoids in the trichomes of female inflorescences (Clarke & Merlin, 2013).

Figure 1. Differences in *Cannabis* inflorescences: fertilised **(A)** and unfertilised **(B)** female terminal inflorescences, male terminal inflorescence **(C)** and male and female flowers of a monoecious *Cannabis* cultivar **(D)**. Photo: Manica Balant.

Sex determination system in *Cannabis* has been a focus of many studies, however, the exact mechanism is still not clear (Kovalchuk *et al.*, 2020). It is believed to be determined by an XY chromosome pair (Hirata, 1925; McPhee, 1926; Sakamoto *et al.*, 1998), or by the X to autosome ratio (Westergaard, 1958; Faux *et al.*, 2014). Recent study found that sex-biased gene expression in *Cannabis* is already established early during vegetative development (Shi *et al.*, 2025). They identified key candidate genes for sex determination that include transcription factors from the REM, bZIP, and MADS families, which drive distinct morphological differences between sexes. Environmental factors or manipulation with hormones and different chemical treatments can also influence sex expression (Atal, 1956; Heslop-Harrison, 1956; Ram & Jaiswal, 1970; Mohan Ram & Sett, 1982; Flajšman *et al.*, 2021). So far, several male-associated DNA markers have been identified for genetic sex determination (Sakamoto *et al.*, 1995; Mandolino *et al.*, 1999; Flachowsky *et al.*, 2001; Törjék *et al.*, 2002; Peil *et al.*, 2003; Divashuk *et al.*, 2014), although accuracy and reproducibility of some of them have been questioned (Mandolino & Carboni, 2004; Toth *et al.*, 2020). An attempt by Heslop-Harrison & Heslop-Harrison (1958) aimed to differentiate male and female *Cannabis* plants before flowering by examining differences in developmental timing, but this approach showed limited success.

Origin of the genus and beginning of domestication

Despite its long history of use, the origin of *Cannabis* genus remains elusive. Various theories suggest that the genus may have originated in C Asia (de Candolle, 1883; Clarke & Merlin, 2013) or S Asia (Zhang *et al.*, 2018). However, current archaeobotanical evidence points to the northeastern Tibetan Plateau as its most likely place of origin (McPartland *et al.*, 2019). From this region, *Cannabis* likely spread westward reaching Europe around 6 million years ago and eastward, reaching E Asia approximately 1.2 million years ago. Clarke and Merlin (2013) proposed that Pleistocene glaciations separated *Cannabis* populations into two glacial refugia: one in the Hengduan Mountains in Asia and the other in the Caucasus in Europe. As the climate warmed, *Cannabis* expanded from these refugia across Asia and Europe. Although *Cannabis* is widely used in India today, it likely arrived on the Indian subcontinent relatively late, by around 32,600 years ago (McPartland *et al.*, 2019; Rull, 2022).

The precise area of *Cannabis* domestication also remains uncertain. The oldest archaeological evidence of human use (i.e., seeds showing domesticated traits) was found in Japan, and dates back about 10,000 years (Kudo *et al.*, 2009). However, the exact centre of domestication is still debated. Some researchers proposed a single domestication event (Schultes *et al.*, 1974; Clarke & Merlin, 2013), while others proposed multiple independent domestication events (Long *et al.*, 2017; McPartland *et al.*, 2018; Zhang *et al.*, 2018). Recent genetic and archaeological evidence suggests that domestication occurred in E Asia around 12,000 years ago, where *Cannabis* was first cultivated as a multipurpose crop (Long *et al.*, 2017; Ren *et al.*, 2021). By around 4,000 years ago, cultivation practices began to diverge, selecting plants for either fibre or drug use (Ren *et al.*, 2021).

After domestication, humans have played a crucial role in the dispersal and evolution of *Cannabis* through selective cultivation for fibre, seeds, or psychoactive properties. The spread of *Cannabis* intensified with the establishment of trade routes, such as the Silk Road, and the expansion of various empires (Warf, 2014). Hindu traders helped spread *Cannabis* to SE Asia and E Africa, while Arab traders likely introduced it across E and N Africa to Morocco. *Cannabis* reached Americas with the European colonization, with settlers primarily bringing European landraces adapted for fibre production. The psychoactive *Cannabis* plants, however, were introduced to the Americas later, likely after 1800, by slaves from W Africa. Following the abolition of slavery in British colonies, new drug-type plants and cultivation techniques were introduced through the West Indies by Indian indentured workers (Clarke & Merlin, 2013). The last significant dispersal period occurred after the Second World War, when new accessions from S Asia reached W Africa, and plants from Afghanistan gained attention from illicit marijuana growers in Europe and America. Today, the dispersal is happening in the opposite direction, with modern fibre and drug cultivars of hybrid origins being reintroduced to regions with traditional landraces, such as Mexico, Morocco, Nepal, Jamaica, Colombia, and Thailand. This has resulted in gene flow back into local landraces

and wild-growing *Cannabis* populations potentially compromising the preservation of their ancestral genetic integrity (Abel, 1980; Clarke, 1998; Clarke & Merlin, 2013; Warf, 2014).

Overview of previous *Cannabis* taxonomical treatments

The long relationship between *Cannabis* and the people who have cultivated it led to the development of a wide array of cultivars, varieties, and strains, suited to different climates and uses (Small, 2015). This extensive cultivation has contributed to the genetic, morphological, and phytochemical diversity, making taxonomic classification within *Cannabis* particularly challenging (Clarke & Merlin, 2013). The classification has been further complicated by different cultural influences and legal considerations, resulting in the inconsistent use of synonyms across various regions (McPartland & Guy, 2017).

One of the earliest written distinctions between European and Asian *Cannabis* was made by Ibn-al-Baitār around 1240 noting the difference between Egyptian and Spanish plants, calling the former Indian hemp (al-qinnab al-hindī; Lozano Cámara, 2017; McPartland & Guy, 2017). Formal scientific descriptions of these species followed in the 18th century by Linnaeus (*C. sativa* L.; 1753) and Lamarck (*C. indica* Lam.; 1783). Since then, various taxonomic approaches using genetic, morphological, and phytochemical data have been proposed (see McPartland & Guy, 2017; Koren *et al.*, 2020; McPartland & Small, 2020; Lapierre *et al.*, 2023; for detailed revision of the topic).

Some scholars considered *Cannabis* as a polytypic genus, generally recognizing either two species (*C. sativa* and *C. indica*; e.g., Lamarck & Poiret, 1783; Hillig, 2005a) or three species (*C. sativa, C. indica, C. ruderalis* Janisch.; e.g., Emboden, 1974; Schultes *et al.*, 1974; Anderson, 1980; Hillig, 2005b; Clarke & Merlin, 2013). Two prominent Russian researchers, Vavilov and Janischevsky, introduced two new taxa: *C. sativa* var. *spontanea* Vav. and *C. ruderalis* Janisch. (sometimes also referred to as *C. sativa* var. *ruderalis* Janisch.), respectively (McPartland & Guy, 2017). Others, however, have supported the idea of a monotypic genus with *Cannabis sativa* as a single species (e.g., Linnaeus, 1753; de Candolle, 1883; Small, 2015; Ren *et al.*, 2021), sometimes further divided into subspecies or varieties such as *C. sativa* subsp. *sativa* and *C. sativa* subsp. *indica* (Lam.) E.Small & Cronquist (Small & Cronquist, 1976; McPartland, 2018; McPartland & Small, 2020) or *C. sativa* subsp. *sativa*, *C. sativa* subsp. *indica* and *C. sativa* subsp. *ruderalis* Janisch. (Zhang *et al.*, 2018). A recent taxonomic review by Lapierre *et al.* (2023), using available genetic data, strongly supported the classification of *Cannabis* as a highly diverse monotypic species.

During the domestication of *Cannabis*, different traits were selected for based on cultivation purpose. Plant height and branching patterns varied depending on whether the plants were cultivated for fibre and seed, or drug production (Romero *et al.*, 2020; Figure 2). Seeds in cultivated plants became larger, lost the perianth and elongated abscission zone, developed thinner walls, and exhibited a more uniform

germination (Small & Cronquist, 1976). Influenced by these morphological differences, some authors divided the plants into various infraspecific taxa based on the plant's cultivation status (Small & Cronquist, 1976; Vavilov, 1992; McPartland & Small, 2020). McPartland and Small (2020) conducted a large-scale review of these morphological traits, building on earlier genetic and phytochemical studies. Following the two-subspecies model by Small and Cronquist (1976), they described four varieties within *C. sativa* subsp. *indica*: two wild varieties (var. *himalayensis* (Cazzuola) McPartl. & E.Small, and var. *asperrima* (Regel) McPartl. & E.Smal) and two cultivated (var. *indica* (Lam.) Persoon and var. *afghanica* (Vav.) McPartl. & E.Small).

Figure 2. Morphological diversity in *Cannabis*. The plants cultivated for fibre are planted in high density, they generally grow over two meters high and develop few branches **(A,** Photo: Manica Balant**)**. Plants cultivated for drug production are planted further apart, develop more branches and many dense inflorescences **(B,** Photo: Manica Balant**)**. Wild-growing plants vary in height and typically develop many branches, with less compact inflorescences **(C**, Photo: Airy Gras**)**.

One of the earliest comprehensive studies of wild and landrace *Cannabis* accessions worldwide with taxonomic focus was conducted by Hillig (2005a). Integrating molecular, morphological, and phytochemical data, Hillig proposed an informal taxonomic treatment that divided *Cannabis* into two species, *C. sativa* and *C. indica*, along with six biotypes, which he later suggested should be considered as varieties: *C. indica* as narrow-leaflet drug (NLD; *C. indica* Lam. var. *indica*), wide-leaflet drug (WLD; *C. indica* Lam. var. *anasha* Hillig), hemp from East Asia (*C. indica* Lam. var. *chinensis* (Delile) Hillig), feral plants from India and Nepal (*C. indica* Lam. var. *kafiristanica* Vav.), and *C. sativa* as hemp (*C. sativa* var. *sativa*) and feral biotypes (*C. sativa* L. var. *spontanea* Vav. (= *C. ruderalis* Janisch.); Hillig, 2005a). He also suggested the possible existence of a third species, *C. ruderalis* (feral plants from C Asia), though the sample size was too small to confirm its status. Based on Hillig's findings (2005a), Clarke and Merlin (2013) adopted a similar classification, recognizing three species and six subspecies. Their classification included *C. ruderalis* as putative ancestor of both *C. sativa* and *C. indica*, with early distribution range in NC Asia. Within *C. sativa* they recognised two subspecies: *C. sativa* subsp. *spontanea* (narrow-leaf hemp ancestor)

and *C. sativa* subsp. *sativa* (narrow-leaf hemp); and within the *C. indica*, they include *C. indica* subsp. *chinensis* (broad-leaf hemp), *C. indica* subsp. *kafiristanica* (narrow-leaf drug ancestor), *C. indica* subsp. *indica* (narrow-leaf drug from S and SE Asia, Middle East) and *C. indica* subsp. *afghanica* (broad-leaf drug from N Afghanistan and Pakistan). They also mentioned the broad-leaf hemp ancestor, however, did not assign it a scientific name.

In addition to scientific and taxonomic classifications, *Cannabis* is often categorized based on its cultivation purpose, morphology, and phytochemical composition. Fiber-type plants, commonly referred to as hemp, are mainly grown for fibre and seed production. They contain less than 0.3% of the psychoactive compound THC (Δ9-tetrahydrocannabinol), whereas drug-type plants, often called marijuana or medicinal cannabis, can contain higher levels of THC (Hurgobin *et al.,* 2021). Further classification of drug-type plants based on the ratio of two major cannabinoids, THC and cannabidiol (CBD) was proposed by Small and Beckstead (1973): type I plants are THC-dominant, type II have a balanced THC/CBD ratio, and type III are CBD-dominant.

Another popular classification, widely used in the recreational and medicinal cannabis industries, categorises *Cannabis* plants into 'sativa,' 'indica,' or 'hybrids.' 'Sativa' refers to taller plants with narrow leaflets and high THC levels, while 'indica' describes shorter, bushier plants with wider leaflets and higher levels of both CBD and THC. Plants that show a mix of these characteristics are classified as 'hybrids' (McPartland & Guy, 2017). However, this classification was criticised by many authors, as it is sometimes wrongly compared with the taxonomical nomenclature of *C. sativa* and *C. indica*. Numerous studies have shown that these popular classifications are not supported by genetic nor phytochemical data and do not reflect true genetic ancestry (McPartland & Small, 2020; Watts *et al.*, 2021). Additionally, inconsistent labelling practices have made strain names unreliable in identifying genetically distinct plants. Research has shown that individual plants with the same strain name were often genetically closer to plants with different names, indicating that strain names and reported ‘sativa’ and ‘indica’ ancestries are poor indicators of genetic identity (Sawler *et al.*, 2015; Schwabe & McGlaughlin, 2019).

Due to the unreliability of strain names and the classification into ‘sativa’, ‘indica’, and ‘hybrid’ categories, many authors started to advocate for labelling *Cannabis* plants based on their phytochemical profiles (i.e., chemovars). In this classification system, monoterpenoids, sesquiterpenoids, and minor cannabinoids are the compounds that are thought to best differentiate between the various chemovars (Hazekamp & Fischedick, 2012; Hazekamp *et al.*, 2016; Birenboim *et al.*, 2022; Herwig *et al.*, 2024).

Cannabis genetic diversity

Genome size and polyploidy in Cannabis

Genome size, also known as the C-value (Swift, 1950), refers to the total amount of DNA in the holoploid genome of an organism (Greilhuber *et al.*, 2005). Within species it is generally considered to be fairly stable, however high-resolution techniques for genome size estimation (e.g., flow cytometry) have provided compelling evidence of intraspecific genome size differences across various taxa (Bennett & Leitch, 2005). These variations are generally linked to factors such as hybridization (Pellicer *et al.*, 2021), polyploidy (Fernández *et al.*, 2022), B-chromosomes (González & Poggio, 2021), changes in repetitive non-coding DNA (Zhang *et al.*, 2020), the presence or absence of specific DNA sequences (Becher *et al.*, 2021), heteromorphic sex chromosomes (Doležel & Göhde, 1995), and illegitimate recombination (Devos *et al.*, 2002). Additionally, intraspecific genome size variation has been associated with other factors like temperature, altitude and latitude, as well as phenological and morphological traits (Walker *et al.*, 2006; Achigan-Dako *et al.*, 2008; Pellicer *et al.*, 2009; Becher *et al.*, 2021).

Polyploidization, both natural and artificial is common in many economically important cultivated plants (e.g., *Triticum* sp. (Peng *et al.*, 2011), *Brassica rapa* (Qi *et al.*, 2021), *Avena sativa* (Peng *et al.*, 2022), and *Ipomoea batatas* (Yang *et al.*, 2017)), because it increases allelic diversity, heterozygosity, and enhances meiotic recombination, leading to greater adaptive plasticity and evolutionary success (Salman-Minkov *et al.*, 2016). However, in *Cannabis*, a widely cultivated plant, polyploidy is relatively uncommon. Small (1972) examined over 200 *Cannabis* accessions from different geographic origins and found that all were diploid (2n=20) individuals. Nevertheless, reports of naturally occurring triploid (Philbrook *et al.*, 2023) and tetraploid (Sharma *et al.*, 2015) *Cannabis* plants exist. Although natural polyploidy appears to be rare in *Cannabis*, artificial methods using chemical treatments have successfully produced triploid, tetraploid, and mixed-ploidy plants in several laboratories (Bagheri & Mansouri, 2015; Mansouri & Bagheri, 2017; Parsons *et al.*, 2019; Galán-Ávila *et al.*, 2020; Kurtz *et al.*, 2020).

Studies analysing the genome size of *Cannabis*, were mostly done on cultivated individuals. For diploids, genome size estimates ranged from 1.42 to 1.95 pg/2C (Sakamoto *et al.*, 1998; Lee *et al.*, 2003; Kubešová *et al.*, 2010; Faux *et al.*, 2014; Parsons *et al.*, 2019). Part of the intraspecific variation in *Cannabis* can be attributed to differences in genome size between male and female plants, as the Y chromosome is approximately 47 Mbp larger than the X chromosome (Sakamoto *et al.*, 1998). Additionally, Lee *et al*. (2003) suggested that part of the variation may also be linked to the different geographic origins of the studied accessions.

Genetic and genomic studies exploring the genetic diversity of Cannabis

In the past years several reference genomes (e.g., van Bakel *et al.*, 2011; Braich *et al.*, 2020; Gao *et al.*, 2020; Grassa *et al.*, 2021; Wei *et al.*, 2024; Ryu *et al.*, 2024) and recently even a draft of pangenome (Lynch

et al., 2024) became available for *Cannabis*. However, these data predominantly originate from modern hemp cultivars or drug strains with unknown geographic origins and limited genetic diversity (with the exception of Gao *et al.*, 2020). Additionally, many whole genome sequences and transcriptomes have been published in recent years (Lynch *et al.*, 2016; McGarvey *et al.*, 2020; Liu *et al.*, 2021; Severson & Adams, 2023). Similarly, a large proportion of them belongs to modern hemp cultivars or drug strains with unknown geographic origins (but see Soorni *et al.*, 2017; Ren *et al.*, 2021; Busta *et al.*, 2022; Chen *et al.*, 2022 as some of the exceptions), while comprehensive studies including wild-growing and landrace individuals across its natural distribution area remain few (Hillig, 2005a; Kovalchuk *et al.*, 2020).

Hillig (2005b) conducted a comprehensive genetic study including wild-growing plants, traditional *Cannabis* landraces and modern cultivars with worldwide distribution. Using allozyme variation at 17 gene loci, he identified two distinct genetic groups within *Cannabis* and classified them as separate species. The first one, named *C. sativa,* contained feral and hemp-type accessions from the Levant, Europe, and N Asia, while the second, named *C. indica,* included both hemp- and drug-type accessions from W and E Asia, and Africa and feral plants from S Asia. Hillig also detected a smaller third group with feral plants from C Asia, but the limited number of individuals prevented its confirmation as a distinct species.

The most recent comprehensive genomic study by Ren *et al.* (2021) analysed primarily hemp cultivars and drug strains, along with some feral individuals from Asia. In contrast to Hillig (2005b), they identified four separate genetic groups interpreted mainly based on use type and domestication status, but did not find sufficient genetic differentiation between them to consider more than one species within *Cannabis*.

Other studies using either whole genome sequencing (WGS), genotyping-by-sequencing (GBS) and microsatellite markers have also found differences between hemp-type and drug-type plants (Sawler *et al.*, 2015; Lynch *et al.*, 2016; Dufresnes *et al.*, 2017; Soorni *et al.*, 2017). Furthermore, within the drug-type accessions, researchers identified two (Sawler *et al.*, 2015; Schwabe & McGlaughlin, 2019) or even three distinct groups (Lynch *et al.*, 2016). Hemp-type cultivars were also differentiated into European and E Asian hemp, with the latter generally being more closely related to drug-type plants (Lynch *et al.*, 2016; Chen *et al.*, 2022).

Cannabis **morphological diversity**

Cannabis displays remarkable phenotypic plasticity, and its overall morphology can vary significantly based on growing conditions (Small & Cronquist, 1976; Islam *et al.*, 2021; Hesami *et al.*, 2023). Throughout its domestication, *Cannabis* underwent changes similar to other cultivated plants, typical of the domestication syndrome. Compared to wild-growing plants, domesticated *Cannabis* plants produce significantly larger seeds, which have a thinner shell. The marbled perianth is absent and the abscission

zone is less pronounced, which enables mature seeds to fall off the plant more easily. Seeds from cultivated plants also exhibit a more uniform germination, a trait uncommon in wild plants. The overall morphology of the plant has also been altered through domestication (Small & Cronquist, 1976). Selective breeding for fibre, seed, or drug production over thousands of years has resulted in a wide range of plant forms, each shaped by its intended use (e.g., fibre cultivars are typically bred to be tall with minimal lateral branching, while drug strains are selected for multiple lateral branches with many dense inflorescences; Clarke & Merlin, 2013).

The cultivated *Cannabis* plants often escape the cultivation and become naturalised in their surrounding environments. Within just 50 generations (i.e., 50 years) these plants can lose many of the traits acquired through domestication and revert to the morphological characters typical for wild plants. This rapid reversion makes it impossible to distinguish between truly wild plants and naturalized (or feral) plants based solely on their morphology in its presumed natural habitats (Small & Cronquist, 1976).

Many studies focused on the agriculturally important morphologic *Cannabis* traits, such as biomass production, hight and internode length, growth rate, branching pattern, bast fibre content, trichome density, inflorescence weight, etc. (Flajšman *et al.*, 2016; Petit *et al.*, 2020; Naim-Feil *et al.*, 2021; Stack *et al.*, 2021, 2023; Jin *et al.*, 2021; Babaei *et al.*, 2024). One part of the *Cannabis* plant that was not directly selected for during domestication, but shows significant variability, is its palmately compound leaf, characterized by a varying number of leaflets. Differences in leaf shape and arrangement were noted and used by some authors in the past to distinguish between different taxa and cultivars (Lamarck & Poiret, 1783; Quimby *et al.*, 1973; Schultes *et al.*, 1974). However, Anderson (1980) was the first to quantify these differences by measuring the width, length, and ratio of the central leaflet. This or similar methods have since been widely used in studies examining the morphological traits of *Cannabis* leaves (Small *et al.*, 1976; de Meijer *et al.*, 1992; de Meijer & Keizer, 1996; Hillig, 2005a; Vergara *et al.*, 2021; Jin *et al.*, 2021; Buzna & Sala, 2022; Murovec *et al.*, 2022; Hesami *et al.*, 2023).

Previous research has highlighted the significant plant plasticity in response to environmental changes (Danziger & Bernstein, 2021a; Islam *et al.*, 2021; Linder *et al.*, 2022; Lyu *et al.*, 2025), but few studies so far have examined the impact of developmental processes. During development, substantial heteroblastic changes—shifts in leaf shape due to the transition from juvenile to adult phases in the meristem—occur along the shoot. In the lower part of the shoot, *Cannabis* leaves exhibit opposite phyllotaxy with one to five leaflets, transitioning to alternate phyllotaxy and leaves with up to 11 or 13 leaflets in the upper regions (Figure 3; Hillig, 2005a; Clarke & Merlin, 2013; Small, 2015). The number of leaflets also varies among different *Cannabis* accessions (Hillig, 2005a). To date, only two studies have specifically addressed heteroblastic changes along the plant axis: Heslop-Harrison & Heslop-Harrison (1958) and Hesami *et al.* (2023), while others have only briefly mentioned it (Hillig, 2005a; Carlson *et al.*, 2021; Jin *et al.*, 2021; Spitzer-Rimon *et al.*, 2022).

Leaf morphology can be studied and quantified using many different methods. From basic quantitative analysis of shape, encompassing a range of techniques from allometric measurements (e.g., lengths, widths, angles) relative to size (Niklas, 1994) to more advanced geometric methods like elliptical Fourier descriptors (EFDs; Kuhl & Giardina, 1982) and landmark-based analyses (Bookstein, 1997). These more advanced methods rely on homologous points to support landmark-based and EFD analyses and are useful for classifying species and distinguishing shape variations resulting from genetic, developmental, and environmental influences (Chitwood *et al.*, 2016, 2021; Chitwood & Sinha, 2016; Demmings *et al.*, 2019; Bryson *et al.*, 2020; Chitwood, 2021; Migicovsky *et al.*, 2022).

Figure 3. Changes in *Cannabis* leaf shape and leaflet numbers along the main stem. Photo: Manica Balant

However, the developmental variability in *Cannabis*, particularly the absence of homologous landmarks due to changing number of leaflets, complicates efforts to classify plant accessions based on leaf shape. Consequently, most studies have relied on basic morphometric techniques using length, width, and their ratios, taking into account only the central leaflet, which is the most consistent and easily identifiable part. The first attempt to apply a landmark-based approach was made by Vergara *et al.* (2021), but they were only able to analyse the central leaflet and the two most distal leaflets on each side—features common to all *Cannabis* leaves except single-leaflet ones—thereby excluding much of the shape variation present in the entire leaf.

Cannabis **phytochemical diversity**

Cannabis is a plant with a diverse array of secondary metabolites. The most characteristic compounds are cannabinoids, a group of non-volatile secondary compounds first identified in and named after the *Cannabis* plant itself. To date, researchers have discovered over 170 different cannabinoids, of which some are an artefact, as they result from natural degradation processes (Hanuš *et al.*, 2016). Beyond cannabinoids, over 120 terpenoids, 20 flavonoids, and other compounds like sterols, vitamins, and fatty acids were identified in *Cannabis* (ElSohly & Slade, 2005; Flores-Sanchez & Verpoorte, 2008; ElSohly *et al.*, 2017; Jin *et al.*, 2020; Liktor-Busa *et al.*, 2021).

The most abundant and well-studied cannabinoids are THC and CBD, though many minor cannabinoids, including cannabigerol (CBG), cannabinol (CBN), cannabichromene (CBC), cannabigerol (CBG), cannabielsoin (CBE), cannabidivarin (CBDV), tetrahydrocannabivarin (THCV), and cannabigerol monomethylether (CBGM), have also been identified (Hillig & Mahlberg, 2004; Hazekamp *et al.*, 2010). Although cannabinoids were initially discovered in *Cannabis*, they have since also been found in other plant genera (e.g., *Helichrysum*, *Amorpha*, and *Glycyrrhiza*), liverworts (e.g., *Radula*), and even some fungi (e.g., *Cylindrocarpon*; Quaghebeur *et al.*, 1994; Hanuš *et al.*, 2016; Andre *et al.*, 2024).

Cannabinoids are synthesized in the glandular trichomes, present on all aerial parts of the plant, but most abundant on the bracts of female flowers (Livingston *et al.*, 2020). The main role of cannabinoids for the plant is still unclear, but it is possible that they protect the plant against UV radiation and/or herbivores, as some cannabinoids have been observed to cause apoptosis (cell death) in various organisms, potentially deterring herbivores (Sirikantaramas *et al.*, 2005; Clarke & Merlin, 2013).

Cannabinoid biosynthesis begins with the precursors olivetolic acid and geranyl pyrophosphate (GPP), which are converted to cannabigerolic acid (CBGA) by the enzyme CBGA synthase (Luo *et al.*, 2019). The CBGA is then secreted into the extracellular storage cavity of the glandular trichomes, where it is further converted in either tetrahydrocannabinolic acid (THCA), cannabidiolic acid (CBDA), and cannabichromenic acid (CBCA) by THCA synthase, CBDA synthase and CBCA synthase, respectively. If exposed to heat, light and atmospheric oxygen and during prolonged storage, the acidic forms of cannabinoids synthesised in the plants undergo non-enzymatic decarboxylation, losing the carboxylic acid (COOH) (Romero *et al.*, 2020; van Velzen & Schranz, 2021). This is a continuous process that starts already in early vegetative plant phase and continues through the plant growth and later on during storage (Kajima & Piraux, 1982). The acidic form of THC, the THCA, is non-psychoactive. To achieve the desired intoxicating effects, THCA is generally heated before or during the consumption (e.g., through smoking, vaping, or baking) to undergo the decarboxylation, which converts it to THC. This compound can further degrade into cannabinol (CBN) during prolonged storage (Romero *et al.*, 2020; Hazekamp *et al.*, 2010).

The psychoactive effects of THC and the broader impact of other cannabinoids in humans and other vertebrates occur through their binding to cannabinoid receptors 1 and 2 (CB1 and CB2) within the endocannabinoid system (ECS). These receptors are found throughout the body, including in the central nervous system, immune system, and digestive system, where they play critical roles in regulating physiological functions such as appetite, inflammation, pain, and mood (Romero *et al.*, 2020; Kovalchuk *et al.*, 2020). The ECS is regulated by endocannabinoids, which are endogenous ligands synthesised in the body in response to neural activity. The two primary endocannabinoids are anandamide or N-arachidonoylethanolamine (AEA) and 2-arachidonoylglycerol (2-AG). Endocannabinoids (cannabinoids synthesised by the human body) and phytocannabinoids (cannabinoids synthesised by plants) have similar structure, which allows them to bind to the same cannabinoid receptors (Hazekamp *et al.*, 2010; Stasiłowicz *et al.*, 2021).

Not only cannabinoids but also other secondary metabolites exert significant effects on the human body. The second most studied group of secondary compounds in *Cannabis* are the terpenoids—a highly volatile group of compounds responsible for the plant's distinct aroma and flavour (Hazekamp *et al.*, 2010). While terpenoids are found across many other plant species, *Cannabis* contains a particularly diverse array, with over 120 identified terpenoids (ElSohly & Slade, 2005), which contributes to its unique phytochemical profile. Like cannabinoids, terpenoids are most concentrated in the glandular trichomes on the female inflorescences. They are produced via distinct biosynthetic pathways, however in both pathways the GPP is one of the precursors (Jin *et al.*, 2020; Kovalchuk *et al.*, 2020). The terpenoids can be further classified in different subclasses based on their molecular structure, among which monoterpenoids and sesquiterpenoids are the most abundant in *Cannabis* (Hazekamp *et al.*, 2010; Liktor-Busa *et al.*, 2021).

Beyond aroma and flavour, terpenoids exhibit various pharmacological properties, including anti-inflammatory, analgesic, and antibacterial effects (Jin *et al.*, 2020; Liktor-Busa *et al.*, 2021). They have been shown to non-selectively bind to different receptors, among them the CB1 and CB2 receptors (Hazekamp *et al.*, 2010; Liktor-Busa *et al.*, 2021), and therefore often interact synergistically with cannabinoids, modulating their effects in what is known as the entourage effect (Russo, 2011). This phenomenon can amplify or temper the physiological impact of cannabinoids on the human body, indicating a complex interplay between these bioactive compounds. Two different entourage effects were described in *Cannabis*: intra-entourage effects, where different cannabinoids or different terpenoids have synergistic effects, and inter-entourage effect, where enhanced biological activity is caused by an interaction between cannabinoids and terpenoids (Koltai & Namdar, 2020).

Terpenoids' volatility makes them easily detectable by humans, contributing to the sensory classification of *Cannabis* strains in the recreational industry. Although hundreds of secondary metabolites have been identified in *Cannabis*, only a specific subset is typically present within an individual plant. Consequently, phytochemical composition has often been utilized in the classification of *Cannabis* varieties, with the ratio of CBD to THC serving as a primary distinguishing characteristic, along with the profiles of minor cannabinoids and terpenoids (Hillig, 2004; Hillig & Mahlberg, 2004; Hazekamp & Fischedick, 2012; Hazekamp *et al.*, 2016; Herwig *et al.*, 2024; see section 'Overview of previous *Cannabis* taxonomical treatments' for further details). While the presence or absence of certain secondary compounds is largely genetically determined, their quantities can vary significantly in response to environmental conditions and other biotic and abiotic factors, making the classification of *Cannabis* plants solely on the phytochemical profile questionable (Booth & Bohlmann, 2019; Stack *et al.*, 2021; van Velzen & Schranz, 2021; Park *et al.*, 2022). Furthermore, different parts of the plant contain varying amounts and, even distinct profiles of secondary compounds, that furthermore change during the development of the plant (Abdollahi *et al.*, 2020; Jin *et al.*, 2020; Noppawan *et al.*, 2022). Differences were also found between male and female plants and between plants of the same sex within a population (Busta *et al.*, 2022).

However, studies examining the phytochemical variability of wild-growing *Cannabis* plants under controlled conditions remain limited. Most recent research has focused either on analysing wild-growing *Cannabis* plants within restricted geographic areas (Busta *et al.*, 2022; Mostafaei Dehnavi *et al.*, 2022; Ghosh *et al.*, 2024) or investigating the phytochemical diversity of cultivated *Cannabis* varieties and strains (Aizpurua-Olaizola *et al.*, 2016; Calvi *et al.*, 2018; Bautista *et al.*, 2021; Eržen *et al.*, 2021; Ahmed *et al.*, 2021; Danziger & Bernstein, 2021b; Birenboim *et al.*, 2022; Ghosh *et al.*, 2024).

Traditional uses of *Cannabis*

Throughout history, plants have been utilized in countless ways by humans. Among these, *Cannabis* undoubtedly stands out as one of the most widely used plants, deeply embedded in the traditions of numerous cultures across the globe (Clarke & Merlin, 2013). The knowledge surrounding traditional uses of plants is the product of generations of experience and innovation. Indigenous and local communities worldwide have transmitted these practices through the ages, tailoring them to their specific cultural and environmental contexts ('Convention on Biological Diversity', 2011). The scientific study of these traditional uses is known as ethnobotany, a term introduced by John W. Harshberger in the 19th century. He defined ethnobotany's primary goals as: i) revealing the cultural significance of plants used by tribes for food, shelter, or clothing; ii) clarifying the historical distribution of plants; iii) tracing ancient trade routes; and iv) suggesting modern applications (Harshberger, 1896).

Cannabis has been recognized not only for its psychoactive and medicinal uses but also for its role in producing fibres for cordage, textiles, and paper. Its seeds have been a significant dietary component, especially for oil production. Additionally, *Cannabis* has historically held an important place in various shamanic and religious practices over the centuries (Abel, 1980; Clarke & Merlin, 2013). However, despite its valuable and widespread use, the early 20th century saw a significant decline in *Cannabis* use, cultivation, and research, as it became classified as an illegal drug, with most information on its use confined to local traditional knowledge (Pisanti & Bifulco, 2019). In recent years, with the relaxation of restrictions in many countries, interest in *Cannabis* use and research has significantly increased. This resurgence has led to the scientific validation and development of several medicinal uses originally discovered through traditional knowledge (Malfait *et al.*, 2000; Mechoulam & Hanuš, 2001; Wright *et al.*, 2005; Kupczyk *et al.*, 2009; Blake *et al.*, 2017; Mondino *et al.*, 2019; Pellesi *et al.*, 2019; Choi *et al.*, 2020; Aviram *et al.*, 2020). These applications have been transformed into effective medicines (e.g., Abuhasira *et al.*, 2018), innovative fibre products (e.g., Vandepitte *et al.*, 2020), and various food products (e.g., Callaway, 2004; Cerino *et al.*, 2021), rapidly propelling *Cannabis* into a billion-dollar industry, with over 200 million users across the world (Kang *et al.*, 2016; UNDOC, 2023).

Psychoactive and ritualistic uses

Today, *Cannabis* is most widely known for its psychoactive recreational use, primarily due to the presence of THC, which is responsible for its mind-altering effects. The origins of the discovery of *Cannabis*'s psychoactive properties remain unclear, but it is believed that the plant was used in various ritualistic and religious contexts since the early Palaeolithic period (Clarke & Merlin, 2013). One of the earliest pieces of evidence of ritualistic *Cannabis* use are charred seeds, wooden braziers, and stones found in the Pamir Mountains, dating back approximately 2,500 years (Ren *et al.*, 2019). Similarly, prehistoric evidence from West China around the same time also suggests ritualistic use of *Cannabis* (Jiang *et al.*, 2016). More recently, ritualistic use of *Cannabis* has also been confirmed at the Judahite Shrine of Arad in Israel, which dates to the 8th century BCE (Arie *et al.*, 2020). Since then, the psychoactive use has spread around the world, with various cultures developing their own methods to harness its effects. In India, for instance, *Cannabis* is used in three primary preparations: 'bhang', 'charas', and 'ganja'. The 'bhang' is mostly prepared from *Cannabis* leaves and flowering shoots. In contrast, for production of 'charas' and 'ganja', primarily plant's female inflorescences. 'Charas' is compacted resin that is hand-rubbed from the fresh plants, while 'ganja' refers to the term for the dried female inflorescences that are usually smoked (Chopra & Chopra, 1957; Clarke & Merlin, 2013). Psychoactive use of *Cannabis* was also common in the Arab countries, where traditionally 'sieved hashish' was produced. 'Sieved hashish' is a resin harvested from dried *Cannabis* plants. It is processed by filtering through multiple sieves to separate the trichomes rich in cannabinoids and terpenoids from the rest of the plant (Clarke, 1998).

Medicinal use

Just as in recreational use, *Cannabis* nowadays plays a significant role in the medicinal and pharmaceutical industry and research. The plant's remarkable ability to produce a wide range of secondary compounds with valuable properties—such as anti-inflammatory, antimicrobial, and neuroprotective activities, among others—has been well-documented (Hanuš *et al.*, 2016; Bonini *et al.*, 2018; Jin *et al.*, 2020). This versatility has supported its medicinal use by humans for at least the past five thousand years. The earliest known record of its medicinal use dates back to 4,700 B.P. in China. The legend of a Chinese emperor Shen-Nung, a father of Chinese agriculture and medicine, tells the story of how he personally tested various drugs and wrote Pen Ts'ao - a kind of herbal Materia Medica that contained 365 natural medicines, including *Cannabis* ('ma'). It was prescribed to treat rheumatism, gout, malaria, beri-beri, constipation, absentmindedness and for menstrual fatigue (Abel, 1980). Ancient texts from India, Persia, Egypt, Greece, and Rome also provide valuable insights into the many medicinal applications of *Cannabis* in the past (e.g., aid to childbirth, as an abortifacient, aphrodisiac, pain treatment, toothaches, earaches; Abel, 1980; Russo, 2002; Clarke & Merlin, 2013; Warf, 2014). Medicinal uses of *Cannabis* were well explored by many cultures, but probably nowhere more than in India. In their traditional medicine, 'bhang' (as *Cannabis* is often called in India) was used as anodyne, hypnotic, analgesic and antispasmodic, as a remedy for external application to piles, in treatment of dysmenorrhoea, rheumatism, chronic

diarrhoea of the sprue type, gonorrhoea, malaria and mental diseases (Chopra & Chopra, 1957). Nowadays, many traditional medicinal applications of *Cannabis* are recognized in modern western medicine, including its use in managing cancer-related pain and chronic pain (Lynch & Ware, 2015; Blake *et al.*, 2017), alleviating spasticity and pain associated with multiple sclerosis (Mecha *et al.*, 2020), and reducing inflammation in bowel disease (Perisetti *et al.*, 2020). However, numerous other uses reported in ethnopharmacological surveys have yet to be studied more extensively to be scientifically validated and developed into effective treatments.

Alimentary use

Beyond its well-known psychoactive and medicinal uses, *Cannabis* has also represented an important part in human and animal diets across the world (Clarke & Merlin, 2013). The fruits of the *Cannabis* plant, often referred to as seeds, were likely among the first parts collected by early humans (Small, 2015). Especially in Asia, *Cannabis* seeds have been a crucial component of the human diet and continue to be consumed in various forms, such as raw, roasted, pickled, ground, parched, or pressed for oil (Figure 4; Clarke & Merlin, 2013). While areal plant parts are generally characterised by the presence of considerable amounts of cannabinoids, terpenoids, flavonoids, and sterols (Jin *et al.*, 2020), *Cannabis* seeds are highly nutritious, containing over 30% oil with an ideal omega-3 to omega-6 fatty acid ratio of 1:3, which is considered optimal for human health. Additionally, they are comprised of 25% of easily digestible protein, are high in dietary fibre, and are a rich source of vitamins and minerals (Callaway, 2004). This nutritional profile has contributed to their growing popularity as a snack and dietary supplement (Clarke & Merlin, 2013). Although some authors have reported the presence of cannabinoids in oil extracted from *Cannabis*

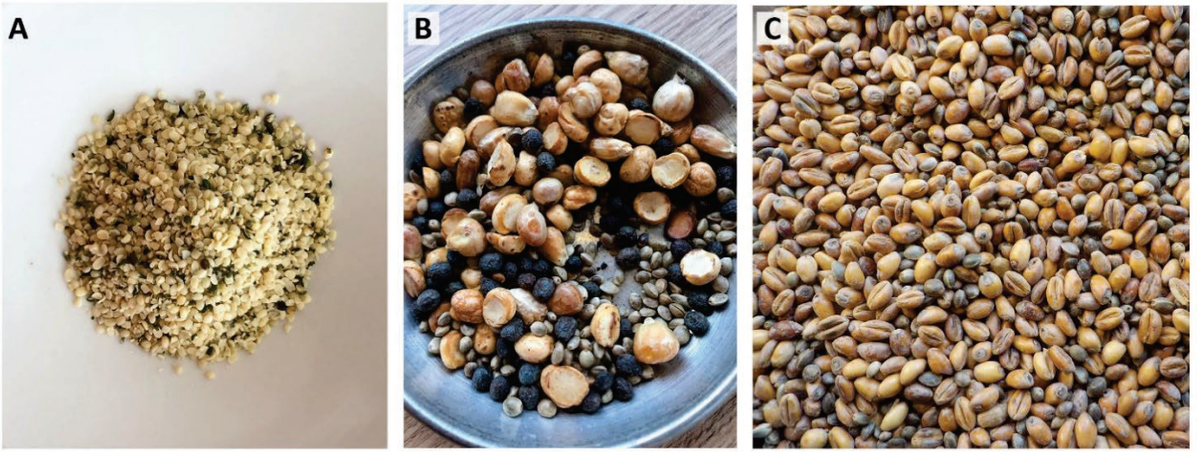


Figure 4. Alimentary uses of *Cannabis* seeds (achenes). *Cannabis* seeds can be peeled to reveal the seed, crushed and eaten raw **(A**, Photo: Mira Balant**)**. The entire seeds can also be roasted and served as appetizers together with a mix of other seeds as those in Turkey (*kavurga*; **B**, Photo: İrem Erdoğan) and Armenia (*aghandz*; **C**, Photo: Joan Vallès).

seeds, this is likely due to contamination (Ross *et al.*, 2000). Cannabinoids are synthesized in glandular trichomes, which are absent on the seeds themselves but are abundant on the surrounding bracts that encase the seeds.

Fibre and other uses

Cannabis is also known for producing one of the strongest and most durable natural fibres, which is why it has been used for centuries in the production of clothing, coarse canvas, sackings, twine, rope, rugs, and paper pulp (Kišgeci, 1994; Clarke, 2010). It was especially important in naval industry, where hemp fibres were used for making anchor ropes, rigging and lashing lines, canvas sail cloth, oakum, fishing nets and many other maritime uses (Clarke, 2023). Today, *Cannabis* fibres are finding new applications in sustainable industries, such as house insulation, hemp fibre interior panels in the automotive sector, animal bedding, nonwoven agricultural fleece, matting, and mulch for weed suppression and erosion control (Clarke & Merlin, 2013). Beyond these modern uses, other *Cannabis* parts have historically been utilized in various other traditional applications worldwide. Stems have served as firewood, while seed oil was employed for lighting, as well as in the production of paints and lacquers. Seed oil has also found its way into cosmetic products, including soaps and hair care items (Shah, 2004; Afzal *et al.*, 2009). Additionally, the aerial parts of the plant have been used for pest control, as insect repellents, and as green manure (Bhardwaj *et al.*, 2011; Ona *et al.*, 2022; Soares *et al.*, 2023)

Contrary to some other plants, substantial knowledge of *Cannabis* traditional uses exists. Various books and review articles dedicated to *Cannabis* have been published (e.g., Li, 1974; Abel, 1980; Kišgeci, 1994; Clarke, 1998; Russo, 2005; Clarke & Merlin, 2013; Pertwee & Pertwee, 2014; Small, 2015; Pisanti & Bifulco, 2019), but much of the relevant knowledge is also scattered across numerous scientific papers that examine ethnobotanical uses in different languages in regions where *Cannabis* grows freely. Comparing results from these sources is complicated by inconsistent terminology. While ethnobotanical research methods are well-developed, variations in how authors describe the plant's effects, targeted ailments, or body systems create challenges for data integration. Therefore, it is crucial to synthesize and standardize the data dispersed across numerous publications. Organizing and consolidating this information can be greatly improved by using a database, which can serve as a valuable tool, facilitating further research.

OBJECTIVES

OBJECTIVES

In recent years, *Cannabis* has become one of the most extensively studied species, primarily due to its vast range of applications in medicine, agriculture, and other industries. However, studies rarely included wild-growing plants and traditional landraces. Most of the studies done so far focused on modern drug strains and cultivars, often concentrating solely on specific research fields. Consequently, these studies have not provided a clear understanding of the taxonomic status and variability within *Cannabis*. To gain a deeper insight into this complex species, it is essential to include both cultivated and wild-growing individuals from across the entire distribution range and to adopt a multidisciplinary approach.

The primary objective of this thesis is to enhance our understanding of the genetic, morphological, and phytochemical diversity of wild-growing *Cannabis* populations and traditional landraces and their traditional uses. We have outlined the main objectives into four key goals, each with specific subgoals:

Objective 1: Study the genetic diversity and clarify the taxonomic status *Cannabis*

a. Investigate the extent of genome size and ploidy level diversity in *Cannabis* accessions across its distribution area.
b. Evaluate the possibility of using flow cytometry as a standard tool to distinguish between male and female *Cannabis* individuals in both wild-growing and cultivated accessions.
c. Investigate the phylogenomic relationships and genetic structure of wild-growing *Cannabis* populations and traditional landraces.
d. Clarify the phylogeographic history of *Cannabis* and its taxonomic status.

Objective 2: Study the morphological diversity of *Cannabis* leaves

a. Develop a methodology that would enable the application of geometric morphometrics techniques to measure leaf shape diversity within *Cannabis.*
b. Evaluate the diversity in the leaf morphology between *Cannabis* accessions and investigate if this variability can be used to differentiate among them.

Objective 3: Study the phytochemical diversity of *Cannabis*

a. Characterize the phytochemical diversity in leaves and inflorescences of wild-growing *Cannabis* populations and traditional landraces.
b. Investigate whether the variability in phytochemical profiles can be used to differentiate between *Cannabis* accessions.

Objective 4: Compile and analyse *Cannabis* traditional uses across the world

a. Conduct a literature review and create an accessible database on the traditional uses of *Cannabis.*
b. Analyse the gathered dataset to obtain a general overview of the most common *Cannabis* traditional uses and their diversity.
c. Carry out a detailed analysis of the human medicinal uses and check whether associations between plant parts and treatments of different body systems and ailments exist.
d. Analyse previously elaborated ethnobotanical surveys on traditional *Cannabis* use in Armenia and compare the data with uses found in existing literature.

institut botànic de Barcelona
Parc de Montjuïc
Passeig del Migdia, s/n
08038 Barcelona
T (+34) 93 289 06 11
F (+34) 93 289 06 14
www.ibb.csic.es
infoibb@ibb.csic.es

REPORT OF THE THESIS SUPERVISORS

INFORME DEL DR. DANIEL VITALES SERRANO I LA DRA. TERESA GARNATJE ROCA, DIRECTORS DE LA TESI DE LA SRA. MANICA BALANT, SOBRE LA PARTICIPACIÓ QUE LA DOCTORANDA HA TINGUT EN CADASCUN DELS ARTICLES QUE PRESENTA A LA SEVA TESI I SOBRE LA CATEGORITZACIÓ DE LES REVISTES EN LES QUALS HAN ESTAT PUBLICATS, SÓN EN REVISIÓ O SERAN ENVIATS EN ELS PROPERS MESOS.

La doctoranda Manica Balant és primera autora i autora per a la correspondència en totes les publicacions incloses en aquesta tesi. En tots els articles, la doctoranda ha estat la principal responsable del disseny i l'execució dels treballs partint del seu projecte de tesi.

Ni els treballs ni els resultats presentats no constitueixen (ni ho faran en el futur) part de cap altra tesi doctoral. A continuació mostrem un llistat dels articles que constitueixen el compendi de publicacions de la tesi doctoral de la Sra. Manica Balant, tot indicant el factor d'impacte de les revistes en l'any de la seva publicació i la posició que ocupa en el seu àmbit, així com la contribució de la doctoranda en cada un dels articles.

1. Balant, M., Quintas, G., Gras, A., Papamichail, A., Vallès, J., De Luna Salvà, N., Montero Molina, S., Garnatje, T., & Vitales, D. Phytochemical diversity of wild-growing and landrace *Cannabis*: insights into cannabinoid composition across tissues, sexes, and geographic origins (en preparació)
   Participació de la doctoranda: disseny experimental, anàlisi i discussió de les dades i redacció de l'article.

2. Balant, M., Vitales, D., Wang, Z., Barina, Z., Fu, L., Gao, T., Garnatje, T., Gras, A., Hayat, M. Q., Oganesian, M., Pellicer, J., Salami, A. S., Seregin, A. P., Stepanyan-Gandilyan, N., Sultana, N., Tsooj, S., Urgamal, M., Vallès, J., van Velzen, R., & Pokorny, L. (2024). Integrating target capture with whole genome sequencing of recent and natural history collections to explain the phylogeography of wild-growing and cultivated *Cannabis*. Enviat a la revista *Plants People Planet*, preprint a bioRxiv i rebudes les revisions.
   DOI: 10.1101/2024.10.17.618884
   Participació de la doctoranda: disseny experimental, anàlisi i discussió de les dades i redacció de l'article.

Categorització de la revista: Inclosa en el *Journal Citation Reports* del *Science Citation Index*. Factor d'impacte: 3,7 (2023). Posició 11 de 75 a *Biodiversity Conservation* i 44 de 197 a *Ecology* (Q1, primer quartil, en ambdós casos). Revista d'accés obert.

3. Balant, M., Garnatje, T., Vitales, D., Oganesian, M., Vallès, J., Stepanyan-Gandilyan, N., & Gras, A. Bridging past and present: Exploring *Cannabis* traditions in Armenia through ethnobotanical interviews and bibliographic prospecting. *Journal of Cannabis Research* (en premsa).
   Participació de la doctoranda: anàlisi i discussió de les dades i redacció de l'article.
   Categorització de la revista: Inclosa en el *Journal Citation Reports* del *Science Citation Index*. Factor d'impacte: 4,1 (2023). Posició 78 de 354 a *Pharmacology & Pharmacy* (Q1, primer quartil). Revista d'accés obert.

4. Balant, M., Garnatje, T., Vitales, D., Hidalgo, O. & Chitwood, D. H. (2024). Intra-leaf modeling of *Cannabis* leaflet shape produces leaf models that predict genetic and developmental identities. *New Phytologist*, 243(2), 781–796.
   DOI: 10.1111/nph.19817
   Participació de la doctoranda: disseny experimental, anàlisi i discussió de les dades i redacció de l'article.
   Categorització de la revista: Inclosa en el *Journal Citation Reports* del *Science Citation Index.* Factor d'impacte: 8,3 (2023). Posició 11 de 265 a *Plant Sciences* (D1, primer decil). Revista d'accés obert.

5. 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.
   DOI: 10.3390/plants11202736
   Participació de la doctoranda: disseny experimental, anàlisi i discussió de les dades i redacció de l'article.
   Categorització de la revista: Inclosa en el *Journal Citation Reports* del *Science Citation Index.* Factor d'impacte: 4,5 (2022). Posició 46 de 265 a *Plant Sciences* (Q1, primer quartil). Revista d'accés obert.

6. Balant, M., Gras, A., Francisco, G., Garnatje, T., Vallès, J., & Vitales, D. (2021). CANNUSE a database of traditional *Cannabis* uses—an opportunity for new research. *Database- The Journal of Biological Databases and Curation*, baab024.
   DOI: 10.1093/database/baab024
   Participació de la doctoranda: disseny experimental, anàlisi i discussió de les dades i redacció de l'article.
   Categorització de la revista: Inclosa en el *Journal Citation Reports* del *Science Citation Index.* Factor d'impacte: 4,462 (2021). Posició 11 de 57 a *Mathematical and Computational Biology* (Q1, primer quartil). Revista d'accés obert.

7. Balant, M., Gras, A., Ruz, M., Vallès, J., Vitales, D., & Garnatje, T. (2021). Traditional uses of *Cannabis*: An analysis of the CANNUSE database. *Journal of Ethnopharmacology,* 279, 114362.
   DOI: 10.1016/j.jep.2021.114362
   Participació de la doctoranda: disseny experimental, anàlisi i discussió de les dades i redacció de l'article.
   Categorització de la revista: Inclosa en el *Journal Citation Reports* del *Science Citation Index.* Factor d'impacte: 5,195 (2021). Posició 4 de 30 a *Integrative & Complementary Medicine* (Q1, primer quartil). Revista d'accés obert.

Firmado por GARNATJE ROCA M.TERESA - DNI ***0208** el día 09/01/2025 con un certificado emitido por AC Sector Público

Firmado por VITALES SERRANO DANIEL - ***3067** el día 09/01/2025 con un certificado emitido por AC FNMT Usuarios

COMPENDIUM OF PUBLICATIONS

CHAPTER 1

CANNABIS GENETIC DIVERSITY

1.1 Novel insights into the nature of intraspecific genome size diversity in *Cannabis sativa* L.

The following chapter is presented in the form of a published article:

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.

DOI: https://doi.org/https://doi.org/10.3390/plants11202736

plants

MDPI

Article

Novel Insights into the Nature of Intraspecific Genome Size Diversity in *Cannabis sativa* L.

Manica Balant [1,*], Roi Rodríguez González [1], Sònia Garcia [1], Teresa Garnatje [1], Jaume Pellicer [1,2], Joan Vallès [3], Daniel Vitales [1,3,†] and Oriane Hidalgo [1,2,†]

1 Institut Botànic de Barcelona (IBB, CSIC-Ajuntament de Barcelona), Passeig del Migdia s.n., 08038 Barcelona, Catalonia, Spain
2 Royal Botanic Gardens, Kew, Kew Green, Richmond TW9 3AE, UK
3 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, Catalonia, Spain
* Correspondence: manica.balant@ibb.csic.es; Tel.: +34-932890611
† These authors contributed equally to this work.

Abstract: *Cannabis sativa* has been used for millennia in traditional medicine for ritual purposes and for the production of food and fibres, thus, providing important and versatile services to humans. The species, which currently has a worldwide distribution, strikes out for displaying a huge morphological and chemical diversity. Differences in *Cannabis* genome size have also been found, suggesting it could be a useful character to differentiate between accessions. We used flow cytometry to investigate the extent of genome size diversity across 483 individuals belonging to 84 accessions, with a wide range of wild/feral, landrace, and cultivated accessions. We also carried out sex determination using the MADC2 marker and investigated the potential of flow cytometry as a method for early sex determination. All individuals were diploid, with genome sizes ranging from 1.810 up to 2.152 pg/2C (1.189-fold variation), apart from a triploid, with 2.884 pg/2C. Our results suggest that the geographical expansion of *Cannabis* and its domestication had little impact on its overall genome size. We found significant differences between the genome size of male and female individuals. Unfortunately, differences were, however, too small to be discriminated using flow cytometry through the direct processing of combined male and female individuals.

Keywords: Cannabaceae; *Cannabis sativa*; genome size; intraspecific genome size variation; population variability; sex chromosomes

Citation: Balant, M.; Rodríguez González, R.; Garcia, S.; Garnatje, T.; Pellicer, J.; Vallès, J.; Vitales, D.; Hidalgo, O. Novel Insights into the Nature of Intraspecific Genome Size Diversity in *Cannabis sativa* L. *Plants* **2022**, *11*, 2736. https://doi.org/10.3390/plants11202736

Academic Editors: Hwan Su Yoon and Kazuo N. Watanabe

Received: 31 August 2022
Accepted: 11 October 2022
Published: 16 October 2022

1. Introduction

Cannabis sativa L. (hereafter referred to as *Cannabis*) is one of the most versatile plants used by humans over millennia. Despite being mostly known for its psychoactive use, *Cannabis* has played an important role in everyday life for hundreds of years. For example, it was extensively used in traditional medicine and became an important source of fibre and food [1]. However, as a consequence of its illegal status, the use of *Cannabis* was abandoned in many parts of the world. Nonetheless, in recent years, the cannabis industry has experienced a rising interest beyond its recreational uses, including more sustainable options in textile, automotive, construction, food, and cosmetic applications [1–4].

The genus most likely originated in the NE Tibetan Plateau more than 25 Mya [5,6], from where it is thought to have spread to North and West Asia and Europe, before continuing to expand eastwards and southwards [5]. Genetic and archaeological evidence suggests that the domestication of *Cannabis* took place approximately 12,000 years ago in East Asia. It was used as a multipurpose crop until c. 4000 years ago, when separate selections for fibre and drug production started [7]. Since then, large-scale cultivation as

a crop has enabled its spread around the world, and today, *Cannabis* has a worldwide distribution [8,9].

The wealth of different applications through centuries resulted in the development of a wide range of cultivars, varieties, and strains adapted to different climates with high morphological and phytochemical diversity [10]. Depending on the cultivation purpose, morphology, and chemical composition, domesticated *Cannabis* can be separated into fibre-type (namely hemp; <0.3% Δ9-tetrahydrocannabinol (THC)) or drug-type (marijuana and medicinal *Cannabis*; >0.3% THC) plants [11]. Within the drug-type plants, different chemotypes are recognised based on their chemical profiles, which are mainly underpinned by the differences in THC/CBD (cannabidiol) ratios. Recently, other secondary metabolites (such as terpenoids and flavonoids) have also gained an important role [12–14]. The morphological and chemical diversity of *Cannabis* has hampered its taxonomic resolution, leading to different taxonomic treatments over the years (see McPartland and Small [15] for a detailed review). Currently, it is considered a monotypic genus, with *C. sativa* as the only accepted species. However, according to a recent evolutionary study based on whole-genome resequencing, fibre-type and drug-type cultivars constitute distinct genetic lineages that diverged from an ancestral gene pool, currently represented by wild or naturalised plants in Central and East Asia, which could have taxonomic implications [7].

Genome size (or C-value) is defined as the amount of DNA in the holoploid genome of an individual [16], and is considered to be relatively constant within a species [17]. Despite reports of intraspecific genome size variation having long been treated with caution, the advent of high-resolution techniques for genome size estimation, such as flow cytometry, has provided strong evidence of intraspecific variability in several taxonomic groups. In general, such variation has been attributed to, e.g., hybridisation and/or polyploidisation events [18,19], B-chromosomes [20], heteromorphic sex chromosomes [21,22], changes in non-coding repetitive DNA [23], presence/absence of specific DNA sequences [24], and illegitimate recombination [25]. In addition to that, intraspecific genome size variation has also been related to extrinsic and/or abiotic factors such as altitude [26–30], latitude [24,31–33], and temperature [31], and to different phenological and morphological characters [27,34].

Cannabis is an annual, wind-pollinated, dioecious plant, although some monoecious cultivars also exist [35]. The diploid genome generally presents 20 chromosomes, 18 autosomal chromosomes and one pair of sex chromosomes. Female and monoecious plants have two X chromosomes (XX), while male plants have heteromorphic X and Y (XY) chromosomes [36,37]. Multiple studies investigating genetic [7,38–40], morphological [41–43], and phytochemical diversity [12,44–47] in *Cannabis* have been published, however, only five of them included genome size measurements [37,48–51]. Most of these studies were carried out on cultivars and centred on either detecting polyploids, or differences in genome size between individuals of different sexes. Certainly, only the study by Lee et al. [50] focused on intraspecific genome size variation in *Cannabis*. These authors detected differences between accessions of different origins, suggesting that genome size could be used as a character to discriminate among accessions. Despite this, intraspecific variability in the genomic content of *Cannabis* has continued to receive little attention. With regard to ploidy levels, natural polyploidisation in *Cannabis* has only been reported once so far, in a wild tetraploid population from India [52]. Small [53] analysed over 200 accessions and found all of them to consistently be diploids (2n = 20). However, artificial polyploids can be induced under laboratory conditions (e.g., chemical treatments), and indeed, triploid, tetraploid, and mixoploid *Cannabis* plants have been produced in plant breeding programs [11,51,54–57].

Many efforts have been made to develop tools to discriminate between male and female *Cannabis* individuals, some of them involving genome size. Although the exact mechanism underpinning sex determination in the species is not yet fully comprehended [8,58], it is thought to be determined by an XY chromosome pair [36,49,59] or by the X to autosome ratio [37,60]. Since the Y chromosome is slightly longer than the X chromosome, male individuals are expected to present a larger genome size. This was corroborated by studies that have found a difference between sexes of Δ = c. 0.05 pg/2C [37,49] or even up to

Δ = 0.15 pg/2C [50]. Early sex determination is usually carried out using male-associated DNA markers [61–67], but the accuracy and reproducibility of some of them have been questioned [67,68]. Based on the above, there is no doubt that developing a method of sex detection through flow cytometry, as previously suggested [50], would be of great interest. However, the reliability and limitations of the method are still to be evaluated for *Cannabis*.

The worldwide distribution of *Cannabis*, its large morphological and phytochemical variability, the existence of heteromorphic sex chromosomes, and the fact that the plant has been a target for selection by humans, could be reflected (to some extent) at the genome size level. So far, most of the studies have focused on a few different (either fibre or drug) *Cannabis* cultivars, but very rarely wild accessions were included. Here, we gathered a large number of wild/feral, landrace and cultivar *Cannabis* accessions, covering a wide distribution area in order to (i) evaluate the extent of genome size and ploidy level diversity in the species; (ii) investigate how this diversity distributes across accessions, geographical ranges, and sexes; and (iii) test whether flow cytometry can be used as a standard tool to distinguish between male and female *Cannabis* individuals in both wild/feral and cultivated accessions.

2. Results and Discussion

2.1. Genome Size in Cannabis: Evidence of Intraspecific Variation

We analysed 483 individuals belonging to 84 accessions (i.e., populations of wild/feral plants, or any landrace and cultivar) from an area spreading over more than 12,000 km and three continents (Figure 1, Table S1). Nuclear DNA content (2C-values) obtained per individual and summarised by accession and geographical region are depicted in Figure 1 and Table S1.

All but one of the individuals analysed were diploid, with genome sizes ranging from 1.810 pg/2C (individual Mongolia 5.14) up to 2.152 pg/2C (individual Armenia 15.3), and an average of 1.956 ± 0.051 pg/2C. One triploid individual was found in a North-Indian wild accession, with a genome size of 2.884 pg/2C. Illustrative flow cytometry histograms for diploid and triploid individuals are presented in Figure 2A,B. The average genome size value for diploid *Cannabis* accessions obtained in our study is slightly higher than average values previously reported (1.720 pg/2C, range = 1.42–1.97 pg/2C; Table S2; [37,48–51]). These differences could be explained by the use of different internal standards, instruments, and stains [69].

The overall genome size difference between diploid individuals spanned over a 1.189-fold range (18.89%). We illustrated for the first time the intraspecific variation in *Cannabis* by processing samples with different genome sizes together and obtaining two peaks (Figure 2C). It is to note that the variation we highlighted through the analyses of 482 diploid individuals is much smaller than the one previously obtained by Lee et al. [50]. Indeed, these authors found a 1.373-fold (37.3%) intraspecific difference through the analysis of 35 individuals.

At the accession's level, we detected significant differences in genome size of diploids across the 84 analysed accessions ($p < 0.001$, Table 1), with average 2C-values ranging from 1.890 ± 0.053 pg/2C (Romania 8) to 2.028 ± 0.022 pg/2C (Armenia 1), which represented a 1.073-fold variation (7.3%). Lee et al. [50] found, however, a much larger variability (1.36-fold range; 35.9%), although they analysed only 14 accessions, with 2C-values ranging from 1.42 to 1.93 pg/2C. In turn, Faux et al. [37] did not find a significant difference among the genome sizes of five *Cannabis* monoecious cultivars. The variation within accessions in our dataset ranged from 1.020-fold (Δ = 0.038 pg/2C, Romania 4) up to 1.123-fold (Δ = 0.236 pg/2C, in Armenia 15), with an average of 1.053-fold (Δ = 0.101 ± 0.032 pg/2C) (Figure 1C, Table S1). Similarly, the study by Lee et al. [50] detected a within-accession variation from 1.006-fold (Δ = 0.01 pg/2C) to 1.127-fold (Δ = 0.22 pg/2C). We found a significant difference in genome size across accessions and distribution areas (Figure 1; ANOVA, $p < 0.001$, Table 1), however, no accession nor area could be clearly separated from the rest through the Tukey HSD post hoc test.

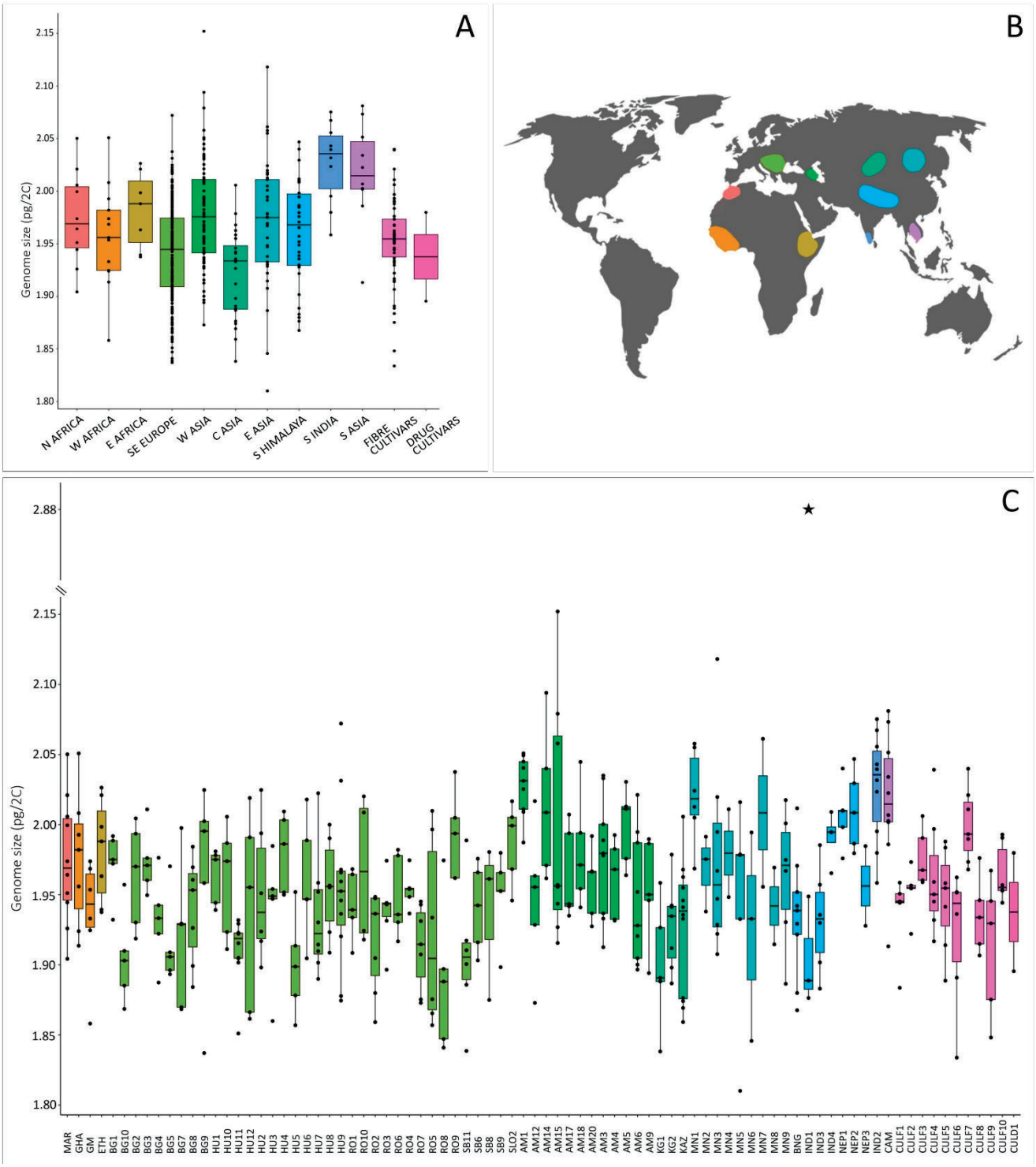


Figure 1. (**A**) Boxplots showing the distribution of genome size in diploid *Cannabis* individuals in different distribution areas. (**B**) Map of the areas of origin of the sampled accessions. (**C**) Boxplots showing the distribution of genome size in *Cannabis* individuals per accessions (the star indicates the genome size of the triploid individual found in the accession IND1—North India).

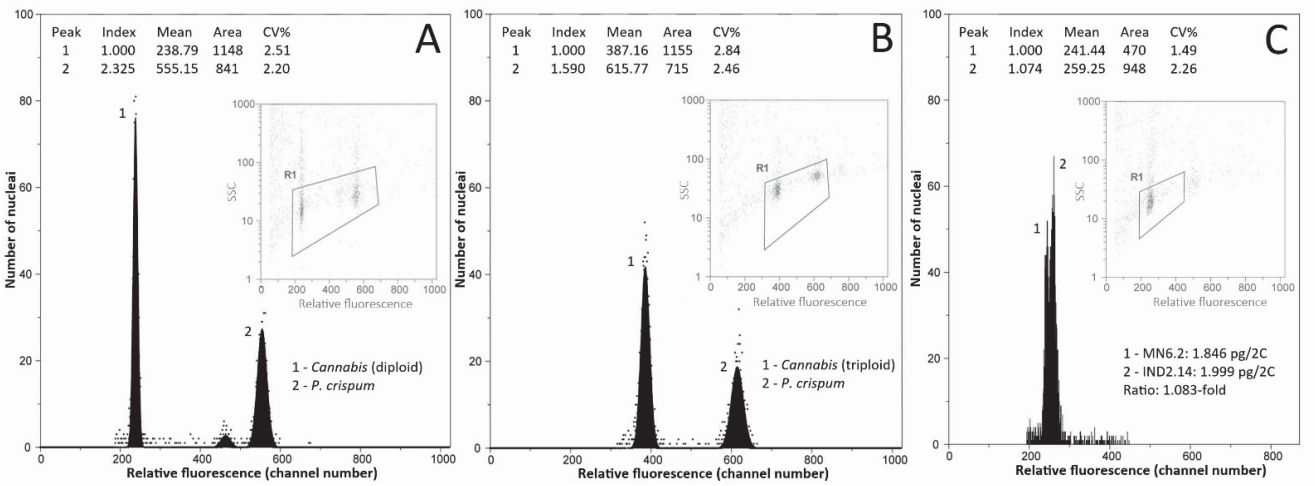


Figure 2. Flow histograms obtained from analysing (**A**) diploid *Cannabis* individual (accession KAZ, Kazakhstan) (peak 1) and (**B**) triploid *Cannabis* individual (accession IND1, North India), using *Petroselinum crispum* (4.5 pg/2C, peak 2) as the internal standard. (**C**) Flow histogram obtained from co-processing diploid individuals from accessions MN6 (Mongolia) and IND2 (South India).

Table 1. Results of ANOVA analysis comparing the effect of accessions, distribution areas, and sex on genome size values of *Cannabis*.

Variable	No. ind.	DF	Sum Sq.	Mean Sq.	F Value	*p* Value
Accessions	482					
Accession		83	0.5206	0.006272	3.386	<0.001
Residuals		398	0.7372	0.001852		
Distribution area	482					
Distribution area		11	0.2185	0.019863	8.983	<0.001
Residuals		470	1.0393	0.002211		
Sex	96					
Sex		1	0.0397	0.03965	11.62	<0.001
Residuals		94	0.3208	0.00341		

Taking together these results, despite the differences in the degree of genome size variation when compared with previous studies, our results provide compelling evidence of genuine intraspecific variation in *Cannabis*.

2.2. Potential Factors Influencing Genome Size Variation in Cannabis

Intraspecific genome size variation of taxa with a large distribution area or isolated populations has been mostly attributed to changes in ploidy level, though, cases of intraspecific variation at the same ploidy level as found in *Cannabis* have also been reported, such as in *Urtica dioica* (2*x* and 4*x* populations with 3.05% within 2*x* accessions and 9.8% variability within 4*x* accessions [33]), *Festuca pallens* (2*x* and 4*x* populations with 16.6% variation in 2*x* and 15% in 4*x* [70]), *Picris hieracioides* (37.6% variability [71]), *Senecio carniolicus* (13.1% variability in 2*x*, 10.2% in 4*x*, 5.4% in 5*x*, and 10.5% in 6*x* populations [72]), *Ranunculus parnassifolius* (2*x* populations with 8.58% and 4*x* with 1.29% variability [73]), and *Euphrasia arctica* (27.4% variability in 2*x* accessions [24]). Intraspecific genome size variation in species with characteristics comparable to *Cannabis*, i.e., a large distribution area and/or the presence of numerous cultivars, has also been reported in *Chenopodium album* (Europe–China; 6.13% [74]), *Chenopodium quinoa* (Americas; 5.9% [75]), *Prunus armeniaca* (Europe–China; 2.3% [76]), and *Cardamine occulta* (Europe–Japan; 8.98% [77]). The intraspecific variation in genome size we found in diploid *Cannabis* at the level of the individuals (18.89%) and accessions (7.3%) is, therefore, similar to that found in other taxa.

Given that no differences in chromosome numbers—except for a few cases—have been found in *Cannabis* (see [50,53] and Table S1), the variation we observed is unlikely to be caused by aneuploidy (i.e., changes in chromosome number). *Cannabis* has heteromorphic sex chromosomes [49], therefore, the sex of individuals could account for some of the variation in genome size. Even though most of our analysed dioecious accessions included both male and female individuals, their frequencies within accessions were not always the same, which could affect the average genome size values per accession. However, according to our results, sex does not fully explain the variation detected between accessions (further discussed below). In the absence of chromosome number variation, another possible explanation for intraspecific genome size variation could be the differences in repetitive DNA sequence content. Pisupati et al. [78] found that 64% of the *Cannabis* genome is made up of repetitive sequences. This is less than in *Zea mays* (c. 85% [79]), but more than in *Arabidopsis thaliana* (c. 21% [80]), where intraspecific genome size variation has also been found [29,81]. Finally, although we have made a great effort to optimise the method for genome size assessment in *Cannabis* by testing a wide range of plant tissues, growing stages, and nuclei extraction buffers (see Section 3. Materials and Methods), we cannot entirely rule out that part of the variation could be due to a technical error. Indeed, all *Cannabis* parts are very rich in secondary metabolites [44], and previous studies have shown that chemical compounds can interfere with DNA binding of the stain, thus, potentially altering the genome size assessments [82–87]. However, we are confident that we have minimised this effect by using only very young leaves from newly germinated seedlings, which provided the best quality measurements in our preliminary tests.

2.3. *Events of Polyploidy in* Cannabis *Are Extremely Rare*

We found one triploid and 482 diploid individuals (Figure 1, Table S1). These results are similar to the previous evidence of Small [53] and Lee et al. [50], showing consistent diploidy (with minimal exceptions) in the species. We confirmed chromosomally that the ploidy levels inferred with flow cytometry by carrying out chromosome counts in 10 individuals from 10 accessions. We found 2n = 20 in diploids and 2n = 30 in the triploid individual (wild North-Indian accession IND1; Figure 3; Table S1). This is the first report of a wild-born triploid individual in *Cannabis*. Records of non-diploid *Cannabis* individuals were indeed so far limited to a tetraploid population in North India [52], or they were otherwise induced by chemical treatment [51]. From the same accession as the triploid individual, the genome size of three other individuals was measured—they were all diploids. The triploid was a male, had a similar morphology than other individuals, and it flowered normally. Unfortunately, we were not able to study this accession further due to the limited number of seeds available, but it would certainly be interesting to investigate whether other ploidy levels could be found in this or more accessions.

Our results confirm that natural polyploidy seems to be extremely rare or even practically non-existent in *Cannabis*, despite its rich domestication background. This contrasts with evidence found in many other species, where genome polyploidisation is preceding or concomitant with their domestication [88,89]. Whole genome multiplication and subsequent diploidization processes provide plants with increased allelic diversity, heterozygosity, and enhanced meiotic recombination, which may increase their adaptive plasticity and evolutionary success [89]. It is, therefore, not surprising that the domestication of some of the most economically important cultivated plants is associated with a polyploidization event, e.g., *Avena sativa* [90], *Triticum* sp. [91], *Ipomoea batatas* [92], *Brassica rapa* [93], and *Musa* sp. [94], among others. In *Cannabis*, artificial polyploids have been obtained by several breeding programs; however, the changes in morphology and phytochemistry of the polyploids have not been extensively investigated so far, thus, requiring more research to be carried out [95].

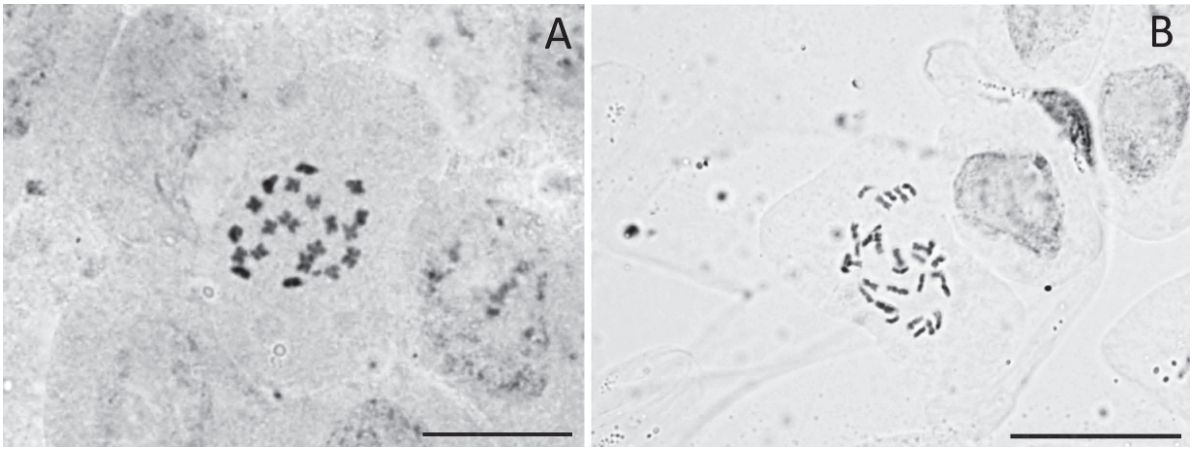


Figure 3. Somatic metaphase plates of a diploid *Cannabis* individual from the accession IND4—North India (2n = 20) (**A**) and a triploid individual from the accession IND1—North India (2n = 30) (**B**). Scale bars = 10 μm.

2.4. *Differences in Genome Size Values of Male and Female* Cannabis *Individuals*

From the 99 individuals with the previously measured genome size selected for sex determination, a MADC2 male-associated band of 390 bp amplified in 46 of them (considered males), while the male-associated band was absent in 49 (considered females). Four individuals showed inconclusive results, with either no PCR bands or two non-indicative bands.

The average female genome size was 1.947 ± 0.065 pg/2C (1.810–2.152 pg/2C), and the average male genome size was 1.987 ± 0.0521 pg/2C (1.920–2.112 pg/2C) (Figure 4 and Figure S1; Table 2). Using ANOVA, we found a significant difference in genome size between male and female plants ($p < 0.001$) (Figure 4, Table 1). The 2C-value of male individuals was in general larger than females for Δ = c. 0.050 pg (0.0009–0.114 pg), which agrees with previous studies [37,49,50]. However, we found few cases where within the same accession, male individuals had a smaller genome size than females. Additionally, the overlap of genome size values of male and female individuals within accessions was, in general, quite high (Figure S1).

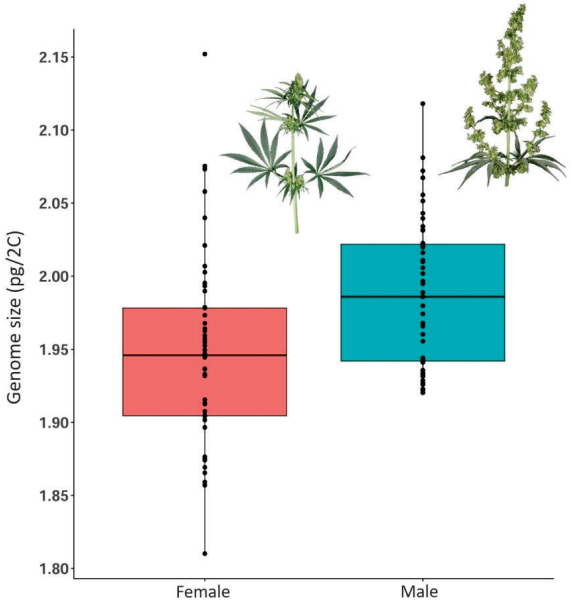


Figure 4. Boxplots showing the genome size distribution of female and male *Cannabis* individuals.

Table 2. Differences in genome size between male and female individuals in the 15 selected *Cannabis* accessions. More details of the accessions can be found in the Supplementary Table (Table S1).

	Female Genome Size (pg/2C)					Male Genome Size (pg/2C)					
Accession	**No. ind.**	**Mean**	**SD [1]**	**Min.**	**Max.**	**No. ind.**	**Mean**	**SD [1]**	**Min.**	**Max.**	**Difference**
AM15	4	2.021	0.106	1.916	2.152	3	1.942	0.014	1.927	1.956	0.079
AM3	3	1.927	0.013	1.913	1.937	1	1.989	/	1.989	1.989	0.061
BG3	1	1.950	/	1.950	1.950	2	1.986	0.036	1.960	2.011	0.036
CAM	4	1.999	0.066	1.913	2.073	6	2.029	0.034	1.986	2.081	0.030
HU11	3	1.913	0.017	1.902	1.932	2	1.923	0.000	1.922	1.923	0.010
HU9	6	1.939	0.034	1.875	1.968	4	1.997	0.068	1.920	2.072	0.058
IND2	2	2.035	0.057	1.995	2.075	6	2.036	0.030	1.980	2.067	0.001
CUL7	5	2.003	0.027	1.973	2.040	2	1.989	0.030	1.968	2.011	0.014
KAZ	6	1.914	0.045	1.869	1.963	5	1.957	0.031	1.933	2.006	0.043
MAR	3	1.938	0.031	1.904	1.964	2	1.973	0.067	1.926	2.021	0.036
MN3	2	1.927	0.028	1.908	1.947	5	1.997	0.080	1.922	2.118	0.069
MN5	4	1.925	0.080	1.810	1.979	1	2.016	/	2.016	2.016	0.091
RO2	2	1.882	0.032	1.859	1.905	0	/	/	/	/	/
RO3	1	1.897	/	1.897	1.897	5	1.947	0.016	1.932	1.974	0.051
RO5	4	2.021	0.106	1.916	2.152	3	1.942	0.014	1.927	1.956	0.079

[1] SD: standard deviation.

Cannabis is showing significant genome size differences between male and female individuals, which is not always the case in dioecious species (e.g., in *Juniperus thurifera* [96]). The presence of a larger genome size in males has been reported in most plant species with heteromorphic sex chromosomes. While some dioecious species have differences in genome size between male and female individuals of similar magnitude to those found in *Cannabis* (2.05%), e.g., 0.45% in *Simmondsia chinensis* [97], 1.97% in *Viscum album* [97], and 2.09–4.19% in *Silene latifolia* [97,98], other species present much larger differences, e.g., 7.14% in *Rumex acetosa* [99], 9.83% in *R. hastatulus* [100], and 10% in *Coccinia grandis* [101]. A larger genome size in male is probably related with Y chromosome degeneration in plants, likely involving the accumulation of repeats in this non-recombining chromosome, as found in *R. acetosa* [102], *Cannabis*, and some *Humulus* species [103].

2.5. Sex Determination in Cannabis Using Flow Cytometry

Peaks of male and female *Cannabis* individuals from the same accession analysed together through flow cytometry overlapped in all cases. This can be explained by the fact that the largest difference between male and female individuals we intended to discriminate was $\Delta = 0.076$ pg/2C (Armenia 3; Table S3), which is well below the smallest genome size difference for which we obtained distinguishable fluorescence double peaks in *Cannabis* (i.e., $\Delta = 0.130$ pg/2C). Our results showed that while differences between the genome size of male and female individuals are significant (according to ANOVA; see part 2.3 for more details), they are simply too small to be discriminated using flow cytometry, by directly processing together male and female individuals. In previous reports, the differences between male individuals on the one hand, and female and monoecious individuals (in both the sex is determined by two X chromosomes) on the other, detected by Faux et al. [37] and Sakamoto et al. [49] ($\Delta = 0.046$ pg/2C and $\Delta = 0.048$ pg/2C, respectively), were also extremely small. Only Lee et al. [50] found larger differences of $\Delta = 0.05$–0.15 pg/2C (2.90–10.56%) between sexes, that could potentially be discriminated in flow cytometry histograms. Unfortunately, the individuals demonstrating these large differences were not processed together to confirm these results. It should be noted, however, that our results were obtained using propidium iodide as the dye in the flow cytometry experiments. Certainly, other methods of flow cytometry, such as the use of other fluorochromes (for example DAPI) or flow sorting, that could offer an improved resolution limit of the technique, should be explored in the future for inexpensive and high-throughput early sex determination in *Cannabis*. Indeed, a previous study has shown the suitability of DAPI flow cytometry for direct sex identification in *Silene latifolia* (formerly *Melandrium album*) and *Silene dioica* (formerly *M. rubrum*) [22], allowing for the discrimination of approximately 1.04-fold genome size difference.

3. Materials and Methods

3.1. Plant Sampling and Cultivation

We analysed 483 *Cannabis* individuals from 84 accessions distributed worldwide, spanning over 12,000 km (Table S4). On average, 5 individuals from each accession were analysed (see Table S1 for details on specific accessions). Seeds from the studied accessions were germinated in Petri dishes and transplanted to pots after the emergence of the first leaves. Plants were cultivated in a growth chamber under controlled conditions (25 °C, 18 h light/6 h dark). Studied individuals were grown for approximately 2–3 weeks until the development of the first or second pair of leaves.

3.2. Flow Cytometry Measurements

Genome size was determined using a CyFlow Space instrument (Sysmex-Partec GmbH, Goerlitz, Germany), fitted with a 100 mW green solid-state laser (Cobolt Samba, Cobolt AB, Solna, Sweden). The internal standard *Petroselinum crispum* 'Champion Moss Curled' (2C = 4.50 pg) [104] was used.

Cannabis plants have many secondary metabolites [44] that could potentially interfere with DNA staining and worsen the quality of the measurements. To overcome such potential issues, different plant tissues and growing stages were tested. The best results were obtained using the first or second pair of leaves of young *Cannabis* plants. Additionally, different flow cytometry buffers (LB01 [105], Ebihara [106], Cystain Ox Protect and PI Absolute buffers (Sysmex-Partec GmbH)) were tested as well, before choosing the general purpose buffer GPB [107] supplemented with 3% PVP-40 [108] as the most appropriate one. Additional measures, such as reducing chopping intensity and working in ice-cold conditions, were taken to reduce the potential effects of secondary metabolites.

We followed the one-step procedure [109] with some modifications. Fresh leaf samples of *Cannabis* and the standard were co-chopped in a Petri dish over ice using 2 mL of the selected nuclei extraction buffer. The sample was then filtered, stained with 40 µL of propidium iodide (PI), and vortexed; samples were left on ice for approximately 30 min before the measurement.

For each sample, the nuclear DNA content was estimated by counting approximately 1000 nuclei per fluorescence peak. Each sample was assessed two times and the results averaged to obtain the final genome size value for the individual. The histograms were analysed using the FlowMax software (v. 2.9, Sysmex-Partec GmbH). Histograms with coefficients of variation (CVs) larger than 5% were discarded.

3.3. Chromosome Counts

Root meristems from each accession were collected for chromosome counts, pre-treated for 2.5 h in 0.05% aqueous colchicine and fixed in fresh absolute ethanol and glacial acetic acid (3:1) for 3 h at room temperature, before being stored in the fixative at 4 °C. They were hydrolysed for 10 min at 60 °C in 1N HCl and stained in 1% aqueous aceto-orcein for at least two hours. Root tips were subsequently squashed in a drop of 45% acetic acid-glycerol (9:1) and observed with a Zeiss Axioplan microscope (Carl Zeiss, Oberkochen, Germany). Metaphases were photographed using a Zeiss AxioCam HRm camera (Carl Zeiss).

3.4. Sex Determination Using Male-Associated Marker and Flow Cytometry

To address the potential differences in genome size between male and female individuals, leaf material from 15 accessions (99 individuals) (Table S5) was collected after genome size measurements and stored in silica gel. DNA was extracted either using the E.Z.N.A. SP Plant DNA Kit (Omega Bio-Tek, Norcross, GA, USA) or the CTAB protocol, following the method by Doyle and Doyle [110] with some modifications.

The sex of individuals was tested using a male-associated DNA marker MADC2, with sequences 5′-GTGACGTAGGTAGAGTTGAA-3′, corresponding to the positions 1–20, and 5′-GTGACGTAGGCTATGAGAG-3′, corresponding to the positions 373–391 [62]. PCR reactions were performed in a 25 µL reaction mixture, containing 1 µL of genomic DNA

(approximately 50 ng), 14.3 µL of sterile water, 2.5 µL of 2 mM MgCl2, 2.5 µL of 10X Gene Taq Universal buffer (Applied Biosystems, Carlsbad, CA, USA), 2.5 µL of 2.5 mM dNTPs mixture, 1 µL of each primer (5 pmol/µL), and 0.2 µL of AmpliTaq DNA polymerase (Applied Biosystems, Carlsbad, CA, USA). The amplification was carried out following the steps: 94 °C for 5 min followed by 37 cycles of 94 °C for 30 s, 58 °C for 1 min, 72 °C for 1 min, and a final step of 72 °C for 5 min. PCR products and ladder (HyperLadder™ 100 bp; Meridian Bioscience, Cincinnati, OH, USA) were separated on 2% agarose gels stained with SYBR Safe-DNA Gel Stain (Thermo Fisher Scientific, MA, USA), and were run at 100 V.

As the reliability of the MADC2 marker used here has been questioned in the past, we first tested the marker on 43 individuals of wild/feral, landrace, and cultivar *Cannabis* accessions with previously known sex (plants grown until the reproductive phase). The marker proved to be a reliable method to assign the correct sex in all but one case, which was inconclusive. No false positives were detected.

To test the suitability of flow cytometry to discriminate between male and female *Cannabis* plants, we selected five accessions (Table S3) displaying a particularly wide range of genome sizes in a preliminary genome size survey (Table S4). New plants from these accessions were cultivated. The first leaf of all individuals was collected and dried in silica gel, and this material was then used to detect the sex-associated marker MADC2 as described above. The genome size was determined by flow cytometry. Samples of each sex from the same or different accessions showing the most divergent genome size values were processed together to test whether genome size differences were large enough to be detected directly by flow cytometry (presence of double peaks).

3.5. Statistical Analyses

To analyse genome sizes across different accessions and distribution areas, we used the dataset composed of all 482 diploid individuals from 84 accessions (Table S4). We analysed the differences using the analysis of variance (ANOVA). The difference in genome size between male and female individuals was also analysed using ANOVA on a dataset of 96 individuals from 15 accessions for which the sex was previously determined with the MADC2 marker (95 individuals); one additional individual where the MADC2 marker showed inconclusive results, but rapidly reached the reproductive phase, was also included (Table S5). Before performing the ANOVA tests, the normality of the datasets was tested on residuals using the Shapiro–Wilk test and Q-Q plots, and homogeneity of variances with Bartlett's test. All the analyses and data visualisations were performed using R version 4.2.1 [111].

4. Conclusions

This study evidenced the extent of intraspecific genome size variation in *Cannabis* and its distribution between and within accessions in an extended sampling covering a wide range of wild/feral, landrace, and cultivated accessions. Our results suggest that the geographical expansion of *Cannabis* and its domestication had little impact on its genome size. In this sense, the pattern observed for genome size is similar to that of other traits in *Cannabis* (e.g., leaf and inflorescence phenotype): a high variability of difficult interpretation, as it does not seem tightly related to its geographical distribution or to infraspecific taxonomic differentiation. Consequently, further studies will be needed to confidently determine whether the observed pattern is a consequence of the history of *Cannabis*, tightly linked to humans, or an intrinsic characteristic of the species.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/plants11202736/s1, Supplementary Figure S1: Genome size of male and female individuals per accession; Table S1: Details of analysed accessions, Table S2: Review of the previously published genome size assessments of *Cannabis*, Table S3: Genome size values of the five selected accessions for sex differentiation, Table S4: Dataset with all *Cannabis* diploid accessions analysed, Table S5: Dataset with 15 selected *Cannabis* diploid accessions analysed.

Author Contributions: Conceptualization, M.B., T.G., O.H., D.V., with the assistance of the remaining authors; fieldwork, management of collection and cultivation permits, M.B., T.G., J.V., D.V.; preliminary tests for the optimisation of genome size assessments in *Cannabis*, M.B., O.H.; genome size assessments, M.B., R.R.G.; sex determination through genetic marker and statistical analyses, M.B., R.R.G., T.G., D.V.; writing—original draft preparation, M.B.; writing—review and editing, M.B., S.G., T.G., J.P., J.V., D.V., O.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministerio de Ciencia, Innovación y Universidades, grant numbers CGL2017-84297-R, PRE2018-083226, Generalitat de Catalunya, grant number 2017SGR1116, S.G. and J.P. benefited from a Ramón y Cajal grant funded by MCIN/AEI/ 10.13039/501100011033 and by "ESF Investing in your future", grant numbers RYC-2014-16608 and RYC-2017-2274.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the Supplementary Materials.

Acknowledgments: This research was supported by projects WECANN (CGL2017-84297-R, Ministerio de Ciencia, Innovación y Universidades), 2017SGR1116 from the Generalitat de Catalunya, and M. Balant FPI predoctoral contract of the Ministerio de Ciencia, Innovación y Universidades (PRE2018-083226). J. Pellicer benefited from a Ramón y Cajal grant ref: RYC-2017-2274, and S. Garcia benefited from a Ramón y Cajal grant ref: RYC-2014-16608 funded by MCIN/AEI/ 10.13039/501100011033 and by "ESF Investing in your future". We would also like to thank Cannaflos—Gesellschaft für medizinisches Cannabis mbH and Joan Uriach Marsal from Uriach Laboratories for additional financial support. We would also like to acknowledge Cannaflos—Gesellschaft für medizinisches Cannabis mbH, and especially Carlos Sáez and Carlos Ribelles for the collaboration in sample collection, and contributions during the original proposal for the WECANN project design. We would also like to thank the colleagues who helped with the sample collection and laboratory work: Zoltán Barina, Ciprian C. Buzna, Marco Calvi, Teodora Dalmacija, Branko Dolinar, İrem Erdoğan, Airy Gras, María Luisa Gutiérrez Merino, Kunigunda Macalik, Marine Oganesian, Alexandra Papamichail, Sonja Siljak-Yakovlev, Nina Stepanyan-Gandilyan, Vladimir Stevanović, Nusrat Sultana, Boštjan Surina, Paul Szatmari, Shagdar Tsooj, Magsar Urgamal, and Vladimir Vladimirov.

Conflicts of Interest: The funders had no role in the design of the study; analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. Cannaflos—Gesellschaft für medizinisches Cannabis mbH contributed to the collection of the samples analysed in this study.

References

1. Clarke, R.; Merlin, M. *Cannabis: Evolution and Ethnobotany*, 1st ed.; University of California Press: Berkley, MI, USA; Los Angeles, CA, USA; London, UK, 2013; ISBN 9780520270480.
2. Gedik, G.; Avinc, O. Hemp fiber as a sustainable raw material source for textile industry: Can we use its potential for more eco-friendly production? In *Sustainability in the Textile and Apparel Industries. Sustainable Textiles: Production, Processing, Manufacturing & Chemistry*; Muthu, S., Gardetti, M., Eds.; Springer: Cham, Switzerland, 2020; pp. 87–109, ISBN 9783030385415.
3. Rupasinghe, H.P.V.; Davis, A.; Kumar, S.K.; Murray, B.; Zheljazkov, V.D. Industrial hemp (*Cannabis sativa* subsp. sativa) as an emerging source for value-added functional food ingredients and nutraceuticals. *Molecules* **2020**, *25*, 4078.
4. Gomez, F.P.; Hu, J.; Clarke, M.A. *Cannabis* as a Feedstock for the Production of Chemicals, Fuels, and Materials: A Review of Relevant Studies to Date. *Energy Fuels* **2021**, *35*, 5538–5557. [CrossRef]
5. McPartland, J.M.; Hegman, W.; Long, T. *Cannabis* in Asia: Its center of origin and early cultivation, based on a synthesis of subfossil pollen and archaeobotanical studies. *Veg. Hist. Archaeobot.* **2019**, *28*, 691–702. [CrossRef]
6. Jin, J.J.; Yang, M.Q.; Fritsch, P.W.; van Velzen, R.; Li, D.Z.; Yi, T.S. Born migrators: Historical biogeography of the cosmopolitan family Cannabaceae. *J. Syst. Evol.* **2020**, *58*, 461–473. [CrossRef]
7. Ren, G.; Zhang, X.; Li, Y.; Ridout, K.; Serrano-Serrano, M.L.; Yang, Y.; Liu, A.; Ravikanth, G.; Nawaz, M.A.; Mumtaz, A.S.; et al. Large-scale whole-genome sequencing unravels the domestication history of *Cannabis sativa*. *Sci. Adv.* **2021**, *7*, eabg2286. [CrossRef]
8. Kovalchuk, I.; Pellino, M.; Rigault, P.; van Velzen, R.; Ebersbach, J.; Ashnest, J.R.; Mau, M.; Schranz, M.E.; Alcorn, J.; Laprairie, R.B.; et al. The genomics of *Cannabis* and its close relatives. *Annu. Rev. Plant Biol.* **2020**, *71*, 20.1–20.27. [CrossRef]
9. Hesami, M.; Pepe, M.; Alizadeh, M.; Rakei, A.; Baiton, A.; Phineas Jones, A.M. Recent advances in cannabis biotechnology. *Ind. Crops Prod.* **2020**, *158*, 113026. [CrossRef]

10. Small, E. Evolution and Classification of *Cannabis sativa* (Marijuana, Hemp) in Relation to Human Utilization. *Bot. Rev.* **2015**, *81*, 189–294. [CrossRef]
11. Hurgobin, B.; Tamiru-Oli, M.; Welling, M.T.; Doblin, M.S.; Bacic, A.; Whelan, J.; Lewsey, M.G. Recent advances in *Cannabis sativa* genomics research. *New Phytol.* **2021**, *230*, 73–89. [CrossRef]
12. Watts, S.; McElroy, M.; Migicovsky, Z.; Maassen, H.; van Velzen, R.; Myles, S. *Cannabis* labelling is associated with genetic variation in terpene synthase genes. *Nat. Plants* **2021**, *7*, 1330–1334. [CrossRef]
13. Hazekamp, A.; Fischedick, J.T. *Cannabis*—From cultivar to chemovar. *Drug Test. Anal.* **2012**, *4*, 660–667. [CrossRef]
14. Bautista, J.L.; Yu, S.; Tian, L. Flavonoids in *Cannabis sativa*: Biosynthesis, Bioactivities, and Biotechnology. *ACS Omega* **2021**, *6*, 5119–5123. [CrossRef]
15. Mcpartland, J.M.; Small, E. A classification of endangered high-THC cannabis (*Cannabis sativa* subsp. indica) domesticates and their wild relatives. *PhytoKeyes* **2020**, *144*, 81–112. [CrossRef]
16. Greilhuber, J.; Doležel, J.; Lysák, M.A.; Bennett, M.D. The origin, evolution and proposed stabilization of the terms "genome size" and "C-value" to describe nuclear DNA contents. *Ann. Bot.* **2005**, *95*, 255–260. [CrossRef]
17. Bennett, M.D.; Leitch, I.J. Genome Size Evolution in Plants. In *The Evolution of the Genome*; Gregory, T.R., Ed.; Elsevier Inc.: Amsterdam, The Netherlands, 2005; pp. 89–162, ISBN 9780123014634.
18. Pellicer, J.; López-Pujol, J.; Aixarch, M.; Garnatje, T.; Vallès, J.; Hidalgo, O. Detecting introgressed populations in the Iberian endemic *Centaurea podospermifolia* through genome size. *Plants* **2021**, *10*, 1492. [CrossRef]
19. Fernández, P.; Hidalgo, O.; Juan, A.; Leitch, I.J.; Leitch, A.R.; Palazzesi, L.; Pegoraro, L.; Viruel, J.; Pellicer, J. Genome Insights into Autopolyploid Evolution: A Case Study in *Senecio doronicum* (Asteraceae) from the Southern Alps. *Plants* **2022**, *11*, 1235. [CrossRef]
20. González, G.E.; Poggio, L. Intragenomic conflict between knob heterochromatin and b chromosomes is the key to understand genome size variation along altitudinal clines in maize. *Plants* **2021**, *10*, 1859. [CrossRef]
21. Costich, D.E.; Meagher, T.R.; Yurkow, J.E. A rapid means of sex identification in *Silene latifolia* by use of flow cytometry. *Plant Mol. Biol. Report.* **1991**, *9*, 359–370. [CrossRef]
22. Doležel, J.; Göhde, W. Sex determination in dioecious plants *Melandrium album* and *M. rubrum* using high-resolution flow cytometry. *Cytometry* **1995**, *19*, 103–106. [CrossRef]
23. Zhang, S.J.; Liu, L.; Yang, R.; Wang, X. Genome Size Evolution Mediated by Gypsy Retrotransposons in Brassicaceae. *Genom. Proteom. Bioinforma* **2020**, *18*, 321–332. [CrossRef]
24. Becher, H.; Powell, R.F.; Brown, M.R.; Metherell, C.; Pellicer, J.; Leitch, I.J.; Twyford, A.D. The nature of intraspecific and interspecific genome size variation in taxonomically complex eyebrights. *Ann. Bot.* **2021**, *128*, 639–651. [CrossRef] [PubMed]
25. Devos, K.M.; Brown, J.K.M.; Bennetzen, J.L. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* **2002**, *12*, 1075–1079. [CrossRef] [PubMed]
26. Achigan-Dako, E.G.; Fuchs, J.; Ahanchede, A.; Blattner, F.R. Flow cytometric analysis in *Lagenaria siceraria* (Cucurbitaceae) indicates correlation of genome size with usage types and growing elevation. *Plant Syst. Evol.* **2008**, *276*, 9–19. [CrossRef]
27. Benor, S.; Fuchs, J.; Blattner, F.R. Genome size variation in *Corchorus olitorius* (Malvaceae s.l.) and its correlation with elevation and phenotypic traits. *Genome* **2011**, *54*, 575–585. [CrossRef]
28. García-Fernández, A.; Iriondo, J.M.; Vallès, J.; Orellana, J.; Escudero, A. Ploidy level and genome size of locally adapted populations of *Silene ciliata* across an altitudinal gradient. *Plant Syst. Evol.* **2012**, *298*, 139–146. [CrossRef]
29. Díez, C.M.; Gaut, B.S.; Meca, E.; Scheinvar, E.; Montes-Hernandez, S.; Eguiarte, L.E.; Tenaillon, M.I. Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytol.* **2013**, *199*, 264–276. [CrossRef]
30. Bilinski, P.; Albert, P.S.; Berg, J.J.; Birchler, J.A.; Grote, M.N.; Lorant, A.; Quezada, J.; Swarts, K.; Yang, J.; Ross-Ibarra, J. Parallel altitudinal clines reveal trends in adaptive evolution of genome size in *Zea mays*. *PLoS Genet.* **2018**, *14*, e1007162. [CrossRef]
31. Walker, D.J.; Moñino, I.; Correal, E. Genome size in *Bituminaria bituminosa* (L.) C.H. Stirton (Fabaceae) populations: Separation of "true" differences from environmental effects on DNA determination. *Environ. Exp. Bot.* **2006**, *55*, 258–265. [CrossRef]
32. Pellicer, J.; Vallès, J.; Garcia, S.; Garnatje, T. Changes in genome size in a fragmented distribution area: The case of *Artemisia crithmifolia* L. (Asteraceae, Anthemideae). *Caryologia* **2009**, *62*, 152–160. [CrossRef]
33. Cronk, Q.; Hidalgo, O.; Pellicer, J.; Percy, D.; Leitch, I.J. *Salix* transect of Europe: Variation in ploidy and genome size in willow-associated common nettle, *Urtica dioica* L. sens. lat., from Greece to arctic Norway. *Biodivers. Data J.* **2016**, *4*, e10003. [CrossRef]
34. Qiu, Y.; Hamernick, S.; Ortiz, J.B.; Watkins, E. DNA content and ploidy estimation of *Festuca ovina* accessions by flow cytometry. *Crop Sci.* **2020**, *60*, 2757–2767. [CrossRef]
35. Small, E.; Cronquist, A. A Practical and Natural Taxonomy for *Cannabis*. *Taxon* **1976**, *25*, 405–435. [CrossRef]
36. Hirata, K. Sex determination in hemp (*Cannabis sativa* L.). *J. Genet.* **1925**, *19*, 65–79. [CrossRef]
37. Faux, A.M.; Berhin, A.; Dauguet, N.; Bertin, P. Sex chromosomes and quantitative sex expression in monoecious hemp (*Cannabis sativa* L.). *Euphytica* **2014**, *196*, 183–197. [CrossRef]
38. Soorni, A.; Fatahi, R.; Haak, D.C.; Salami, S.A.; Bombarely, A. Assessment of Genetic Diversity and Population Structure in Iranian *Cannabis* Germplasm. *Sci. Rep.* **2017**, *7*, 15668. [CrossRef]

39. Bocharkina, Y.V.; Alexandrov, O.S.; Razumova, O.V.; Karlov, G.I. The comparative analysis of the DNA repeat composition among *Cannabis sativa* L., *Humulus lupulus* L. and *Humulus japonicus* Siebold & Zucc. with heteromorphic sex chromosomes. *Preprints* **2021**, 2021120040. [CrossRef]
40. Grassa, C.J.; Weiblen, G.D.; Wenger, J.P.; Dabney, C.; Poplawski, S.G.; Motley, S.T.; Michael, T.P.; Schwartz, C.J. A new *Cannabis* genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytol.* **2021**, *230*, 1665–1679. [CrossRef]
41. Márquez, F.; Lozada, M.; Idaszkin, Y.L.; González-José, R.; Bigatti, G. *Cannabis* Varieties Can Be Distinguished by Achene Shape Using Geometric Morphometrics. *Cannabis Cannabinoid Res.* **2022**, *7*, 409–414. [CrossRef]
42. Petit, J.; Salentijn, E.M.J.; Paulo, M.J.; Thouminot, C.; van Dinter, B.J.; Magagnini, G.; Gusovius, H.J.; Tang, K.; Amaducci, S.; Wang, S.; et al. Genetic Variability of Morphological, Flowering, and Biomass Quality Traits in Hemp (*Cannabis sativa* L.). *Front. Plant Sci.* **2020**, *11*, 102. [CrossRef]
43. Vergara, D.; Feathers, C.; Huscher, E.L.; Holmes, B.; Haas, J.A.; Kane, N.C. Widely assumed phenotypic associations in *Cannabis sativa* lack a shared genetic basis. *PeerJ* **2021**, *9*, e10672. [CrossRef]
44. Jin, D.; Dai, K.; Xie, Z.; Jie, C. Secondary metabolites profiled in *Cannabis* inflorescences, leaves, stem barks, and roots for medicinal purposes. *Sci. Rep.* **2020**, *10*, 3309. [CrossRef]
45. Busta, L.; Dweikat, I.; Sato, S.J.; Qu, H.; Xue, Y.; Zhou, B.; Gan, L.; Yu, B.; Clemente, T.E.; Cahoon, E.B.; et al. Chemical and genetic variation in feral *Cannabis sativa* populations across the Nebraska climate gradient. *Phytochemistry* **2022**, *200*, 113206. [CrossRef] [PubMed]
46. Di Sotto, A.; Gullì, M.; Acquaviva, A.; Tacchini, M.; Di Simone, S.C.; Chiavaroli, A.; Recinella, L.; Leone, S.; Brunetti, L.; Orlando, G.; et al. Phytochemical and pharmacological profiles of the essential oil from the inflorescences of the *Cannabis sativa* L. *Ind. Crop. Prod.* **2022**, *183*, 114980. [CrossRef]
47. Sharma, K.; Rana, R.; Ashawat, M.S. Preliminary Phytochemical Screening of Leaves, Stems and Roots of wild *Cannabis sativa*. *Res. J. Pharmacogn. Phytochem.* **2022**, *14*, 19–22. [CrossRef]
48. Kubešová, M.; Moravcová, L.; Suda, J.; Jarošík, V.; Pyšek, P. Naturalized plants have smaller genomes than their non-invading relatives: A flow cytometric analysis of the Czech alien flora. *Preslia* **2010**, *82*, 81–96.
49. Sakamoto, K.; Akiyama, Y.; Fukui, K.; Kamada, H.; Satoh, S. Characterization; Genome Sizes and Morphology of Sex Chromosomes in Hemp (*Cannabis sativa* L.). *Cytologia (Tokyo)* **1998**, *63*, 459–464. [CrossRef]
50. Lee, I.-J.; Uchiyama, H.; Sekita, S.; Goda, Y.; Satake, M. Analytical studies on *Cannabis sativa* L. introduced into Japan. Part III. The range variation of genome size in *C. sativa*. *Foods Food Ingred. J. Jpn.* **2003**, *208*, 382–386.
51. Parsons, J.L.; Martin, S.L.; James, T.; Golenia, G.; Boudko, E.A.; Hepworth, S.R. Polyploidization for the Genetic Improvement of *Cannabis sativa*. *Front. Plant Sci.* **2019**, *10*, 476. [CrossRef]
52. Sharma, V.; Srivastava, D.K.; Gupta, R.C.; Singh, B. Abnormal Meiosis in Tetraploid (4x) *Cannabis sativa* (L.) from Lahaul-Spiti (Cold Desert Higher Altitude Himalayas)-A Neglected But Important Herb. *J. Biol. Chem. Chron.* **2015**, *2*, 38–42.
53. Small, E. Interfertility and chromosomal uniformity in *Cannabis*. *Can. J. Bot.* **1972**, *50*, 1947–1949. [CrossRef]
54. Bagheri, M.; Mansouri, H. Effect of Induced Polyploidy on Some Biochemical Parameters in *Cannabis sativa* L. *Appl. Biochem. Biotechnol.* **2015**, *175*, 2366–2375. [CrossRef]
55. Mansouri, H.; Bagheri, M. Induction of Polyploidy and Its Effect on *Cannabis sativa* L. In *Cannabis sativa L.—Botany and Biotechnology*; Chandra, S., Lata, H., ElSohly, M.A., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 365–383, ISBN 9783319545646.
56. Galán-Ávila, A.; García-Fortea, E.; Prohens, J.; Herraiz, F.J. Development of a Direct in vitro Plant Regeneration Protocol From *Cannabis sativa* L. Seedling Explants: Developmental Morphology of Shoot Regeneration and Ploidy Level of Regenerated Plants. *Front. Plant Sci.* **2020**, *11*, 645. [CrossRef] [PubMed]
57. Kurtz, L.E.; Brand, M.H.; Lubell-Brand, J.D. Production of tetraploid and triploid hemp. *HortScience* **2020**, *55*, 1703–1707. [CrossRef]
58. Lynch, R.C.; Vergara, D.; Tittes, S.; White, K.; Schwartz, C.J.; Gibbs, M.J.; Ruthenburg, T.C.; DeCesare, K.; Land, D.P.; Kane, N.C. Genomic and Chemical Diversity in *Cannabis*. *CRC. Crit. Rev. Plant Sci.* **2016**, *35*, 349–363. [CrossRef]
59. McPhee, H.C. The genetics of sex in Hemp. *J. Agric. Res.* **1926**, *31*, 935–943.
60. Westergaard, M. The Mechanism of Sex Determination in Dioecious Flowering Plants. *Adv. Genet.* **1958**, *9*, 217–281. [CrossRef] [PubMed]
61. Sakamoto, K.; Abe, T.; Matsuyama, T.; Yoshida, S.; Ohmido, N.; Fukui, K.; Satoh, S. RAPD markers encoding retrotransposable elements are linked to the male sex in *Cannabis sativa* L. *Genome* **2005**, *48*, 931–936. [CrossRef] [PubMed]
62. Mandolino, G.; Carboni, A.; Forapani, S.; Faeti, V.; Ranalli, P. Identification of DNA markers linked to the male sex in dioecious hemp (*Cannabis sativa* L.). *Theor. Appl. Genet.* **1999**, *98*, 86–92. [CrossRef]
63. Flachowsky, H.; Schumann, E.; Weber, W.E.; Peil, A. Application of AFLP for the detection of sex-specific markers in hemp. *Plant Breed.* **2001**, *120*, 305–309. [CrossRef]
64. Törjék, O.; Bucherna, N.; Kiss, E.; Homoki, H.; Finta-Korpelová, Z.; Bócsa, I.; Nagy, I.; Heszky, L.E. Novel male-specific molecular markers (MADC5, MADC6) in hemp. *Euphytica* **2002**, *127*, 209–218. [CrossRef]
65. Peil, A.; Flachowsky, H.; Schumann, E.; Weber, W.E. Sex-linked AFLP markers indicate a pseudoautosomal region in hemp (*Cannabis sativa* L.). *Theor. Appl. Genet.* **2003**, *107*, 102–109. [CrossRef]

66. Sakamoto, K.; Shimomura, K.; Komeda, Y.; Kamada, H.; Satoh, S. A Male-Associated DNA Sequence in a Dioecious Plant, *Cannabis sativa* L. *Plant Cell Physiol.* **1995**, *36*, 1549–1554. [CrossRef]
67. Toth, J.A.; Stack, G.M.; Cala, A.R.; Carlson, C.H.; Wilk, R.L.; Crawford, J.L.; Viands, D.R.; Philippe, G.; Smart, C.D.; Rose, J.K.C.; et al. Development and validation of genetic markers for sex and cannabinoid chemotype in *Cannabis sativa* L. *GCB Bioenergy* **2020**, *12*, 213–222. [CrossRef]
68. Mandolino, G.; Carboni, A. Potential of marker-assisted selection in hemp genetic improvement. *Euphytica* **2004**, *140*, 107–120. [CrossRef]
69. Doležel, J.; Greilhuber, J.; Lucretti, S.; Meister, A.; Lysák, M.A.; Nardi, L.; Obermayer, R. Plant genome size estimation by flow cytometry: Inter-laboratory comparison. *Ann. Bot.* **1998**, *82*, 17–26. [CrossRef]
70. Šmarda, P.; Bureš, P. Intraspecific DNA content variability in *Festuca pallens* on different geographical scales and ploidy levels. *Ann. Bot.* **2006**, *98*, 665–678. [CrossRef]
71. Slovák, M.; Vít, P.; Urfus, T.; Suda, J. Complex pattern of genome size variation in a polymorphic member of the Asteraceae. *J. Biogeogr.* **2009**, *36*, 372–384. [CrossRef]
72. Suda, J.; Weiss-Schneeweiss, H.; Tribsch, A.; Schneeweiss, G.M.; Trávníček, P.; Schönswetter, P. Complex distribution patterns of di-, tetra-, and hexaploid cytotypes in the European high mountain plant *Senecio carniolicus* (Asteraceae). *Am. J. Bot.* **2007**, *94*, 1391–1401. [CrossRef]
73. Cires, E.; Cuesta, C.; Revilla, M.Á.; Fernández Prieto, J.A. Intraspecific genome size variation and morphological differentiation of *Ranunculus parnassifolius* (Ranunculaceae), an Alpine-Pyrenean-Cantabrian polyploid group. *Biol. J. Linn. Soc.* **2010**, *101*, 251–271. [CrossRef]
74. Vít, P.; Krak, K.; Trávníček, P.; Douda, J.; Lomonosova, M.N.; Mandák, B. Genome size stability across Eurasian *Chenopodium* species (Amaranthaceae). *Bot. J. Linn. Soc.* **2016**, *182*, 637–649. [CrossRef]
75. Kolano, B.; Siwinska, D.; Pando, L.G.; Szymanowska-Pulka, J.; Maluszynska, J. Genome size variation in *Chenopodium quinoa* (Chenopodiaceae). *Plant Syst. Evol.* **2012**, *298*, 251–255. [CrossRef]
76. Li, W.; Liu, L.; Wang, Y.; Fan, G.; Zhang, S.; Wang, Y.; Liao, K. Determination of genome size and chromosome ploidy of selected taxa from *Prunus armeniaca* by flow cytometry. *Sci. Hortic. (Amsterdam)* **2020**, *261*, 108987. [CrossRef]
77. Šlenker, M.; Zozomová-Lihová, J.; Mandáková, T.; Kudoh, H.; Zhao, Y.; Soejima, A.; Yahara, T.; Skokanová, K.; Španiel, S.; Fls, K.M. Morphology and genome size of the widespread weed *Cardamine occulta*: How it differs from cleistogamic *C. kokaiensis* and other closely related taxa in Europe and Asia. *Bot. J. Linn. Soc.* **2018**, *187*, 456–482. [CrossRef]
78. Pisupati, R.; Vergara, D.; Kane, N.C. Diversity and evolution of the repetitive genomic content in *Cannabis sativa*. *BMC Genom.* **2018**, *19*, 156. [CrossRef] [PubMed]
79. Schnable, P.S.; Ware, D.; Fulton, R.S.; Stein, J.C.; Wei, F.; Pasternak, S.; Liang, C.; Zhang, J.; Fulton, L.; Graves, T.A.; et al. The B73 maize genome: Complexity, diversity, and dynamics. *Science* **2009**, *326*, 1112–1115. [CrossRef] [PubMed]
80. Quesneville, H. Twenty years of transposable element analysis in the *Arabidopsis thaliana* genome. *Mob. DNA* **2020**, *11*, 28. [CrossRef] [PubMed]
81. Schmuths, H.; Meister, A.; Horres, R.; Bachmann, K. Genome size variation among accessions of *Arabidopsis thaliana*. *Ann. Bot.* **2004**, *93*, 317–321. [CrossRef] [PubMed]
82. Loureiro, J.; Rodriguez, E.; Doležel, J.; Santos, C. Flow cytometric and microscopic analysis of the effect of tannic acid on plant nuclei and estimation of DNA content. *Ann. Bot.* **2006**, *98*, 515–527. [CrossRef] [PubMed]
83. Sharma, S.; Kaushik, S.; Raina, S.N. Estimation of nuclear DNA content and its variation among Indian Tea accessions by flow cytometry. *Physiol. Mol. Biol. Plants* **2019**, *25*, 339–346. [CrossRef]
84. Noirot, M.; Barre, P.; Louarn, J.; Duperray, C.; Hamon, S. Nucleus-cytosol interactions—A source of stoichiometric error in flow cytometric estimation of nuclear DNA content in plants. *Ann. Bot.* **2000**, *86*, 309–316. [CrossRef]
85. Noirot, M.; Barre, P.; Duperray, C.; Louarn, J.; Hamon, S. Effects of caffeine and chlorogenic acid on propidium iodide accessibility to DNA: Consequences on genome size evaluation in coffee tree. *Ann. Bot.* **2003**, *92*, 259–264. [CrossRef]
86. Bennett, M.D.; Price, H.J.; Johnston, J.S. Anthocyanin inhibits propidium iodide DNA fluorescence in *Euphorbia pulcherrima*: Implications for genome size variation and flow cytometry. *Ann. Bot.* **2008**, *101*, 777–790. [CrossRef]
87. Price, H.J.; Hodnett, G.; Johnston, J.S. Sunflower (*Helianthus annuus*) leaves contain compounds that reduce nuclear propidium iodide fluorescence. *Ann. Bot.* **2000**, *86*, 929–934. [CrossRef]
88. Akagi, T.; Jung, K.; Masuda, K.; Shimizu, K.K. Polyploidy before and after domestication of crop species. *Curr. Opin. Plant Biol.* **2022**, *69*, 102255. [CrossRef]
89. Salman-Minkov, A.; Sabath, N.; Mayrose, I. Whole-genome duplication as a key factor in crop domestication. *Nat. Plants* **2016**, *2*, 16115. [CrossRef]
90. Peng, Y.; Yan, H.; Guo, L.; Deng, C.; Wang, C.; Wang, Y.; Kang, L.; Zhou, P.; Yu, K.; Dong, X.; et al. Reference genome assemblies reveal the origin and evolution of allohexaploid oat. *Nat. Genet.* **2022**, *54*, 1248–1258. [CrossRef]
91. Peng, J.H.; Sun, D.; Nevo, E. Domestication evolution, genetics and genomics in wheat. *Mol. Breed.* **2011**, *28*, 281–301. [CrossRef]
92. Yang, J.; Moeinzadeh, M.H.; Kuhl, H.; Helmuth, J.; Xiao, P.; Haas, S.; Liu, G.; Zheng, J.; Sun, Z.; Fan, W.; et al. Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nat. Plants* **2017**, *3*, 696–703. [CrossRef]

93. Qi, X.; An, H.; Hall, T.E.; Di, C.; Blischak, P.D.; McKibben, M.T.W.; Hao, Y.; Conant, G.C.; Pires, J.C.; Barker, M.S. Genes derived from ancient polyploidy have higher genetic diversity and are associated with domestication in *Brassica rapa*. *New Phytol.* **2021**, *230*, 372–386. [CrossRef]
94. Šimoníková, D.; Němečková, A.; Čížková, J.; Brown, A.; Swennen, R.; Doležel, J.; Hřibová, E. Chromosome painting in cultivated bananas and their wild relatives (*Musa* spp.) reveals differences in chromosome structure. *Int. J. Mol. Sci.* **2020**, *21*, 7915. [CrossRef]
95. Hesami, M.; Baiton, A.; Alizadeh, M.; Pepe, M.; Torkamaneh, D.; Phineas Jones, A.M. Advances and Perspectives in Tissue Culture and Genetic Engineering of *Cannabis*. *Int. J. Mol. Sci.* **2021**, *22*, 5671. [CrossRef]
96. Romo, A.; Hidalgo, O.; Boratyński, A.; Sobierajska, K.; Jasińska, A.K.; Vallès, J.; Garnatje, T. Genome size and ploidy levels in highly fragmented habitats: The case of western Mediterranean *Juniperus* (Cupressaceae) with special emphasis on *J. thurifera* L. *Tree Genet. Genomes* **2013**, *9*, 587–599. [CrossRef]
97. Marie, D.; Brown, S.C. A cytometric exercise in plant DNA histograms, with 2C values for 70 species. *Biol. Cell* **1993**, *78*, 41–51. [CrossRef]
98. Vagera, J.; Paulíková, D.; Doležel, J. The Development of Male and Female Regenerants by In Vitro Androgenesis in Dioecious Plant *Melandrium album*. *Ann. Bot.* **1994**, *73*, 455–459. [CrossRef]
99. Błocka-Wandas, M.; Sliwinska, E.; Grabowska-Joachimiak, A.; Musial, K.; Joachimiak, A.J. Male gametophyte development and two different DNA classes of pollen grains in *Rumex acetosa* L., a plant with an XX/XY1Y2 sex chromosome system and a female-biased sex ratio. *Sex. Plant Reprod.* **2007**, *20*, 171–180. [CrossRef]
100. Grabowska-Joachimiak, A.; Kula, A.; Książczyk, T.; Chojnicka, J.; Sliwinska, E.; Joachimiak, A.J. Chromosome landmarks and autosome-sex chromosome translocations in *Rumex hastatulus*, a plant with XX/XY1Y2 sex chromosome system. *Chromosom. Res.* **2015**, *23*, 187–197. [CrossRef]
101. Sousa, A.; Fuchs, J.; Renner, S.S. Molecular cytogenetics (FISH, GISH) of *Coccinia grandis*: A ca. 3 myr-old species of cucurbitaceae with the largest Y/autosome divergence in flowering plants. *Cytogenet. Genome Res.* **2013**, *139*, 107–118. [CrossRef]
102. Charlesworth, D. Plant sex determination and sex chromosomes. *Heredity (Edinb.)* **2002**, *88*, 94–101. [CrossRef]
103. Razumova, O.V.; Divashuk, M.G.; Alexandrov, O.S.; Karlov, G.I. GISH painting of the Y chromosomes suggests advanced phases of sex chromosome evolution in three dioecious Cannabaceae species (*Humulus lupulus*, *H. japonicus*, and *Cannabis sativa*). *Protoplasma* **2022**. [CrossRef]
104. Obermayer, R.; Leitch, I.J.; Hanson, L.; Bennett, M.D. Nuclear DNA C-values in 30 species double the familial representation in pteridophytes. *Ann. Bot.* **2002**, *90*, 209–217. [CrossRef]
105. Doležel, J.; Binarová, P.; Lucretti, S. Analysis of Nuclear DNA content in plant cells by Flow cytometry. *Biol. Plant.* **1989**, *31*, 113–120. [CrossRef]
106. Ebihara, A.; Ishikawa, H.; Matsumoto, S.; Lin, S.J.; Iwatsuki, K.; Takamiya, M.; Watano, Y.; Ito, M. Nuclear DNA, chloroplast DNA, and ploidy analysis clarified biological complexity of the *Vandenboschia radicans* complex (Hymenophyllaceae) in Japan and adjacent areas. *Am. J. Bot.* **2005**, *92*, 1535–1547. [CrossRef]
107. Loureiro, J.; Rodriguez, E.; Doležel, J.; Santos, C. Two new nuclear isolation buffers for plant DNA flow cytometry: A test with 37 species. *Ann. Bot.* **2007**, *100*, 875–888. [CrossRef]
108. Pellicer, J.; Powell, R.F.; Leitch, I.J. The Application of Flow Cytometry for Estimating Genome Size, Ploidy Level Endopolyploidy, and Reproductive Modes in Plants. In *Molecular Plant Taxonomy: Methods and Protocols*; Besse, P., Ed.; Springer: New York, NY, USA, 2021; pp. 325–361, ISBN 9781071609972.
109. Doležel, J.; Greilhuber, J.; Suda, J. Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* **2007**, *2*, 2233–2244. [CrossRef]
110. Doyle, J.; Doyle, J.L. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem. Bull.* **1987**, *19*, 1–15.
111. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.

1.2 Integrating target capture with whole genome sequencing of recent and natural history collections to explain the phylogeography of wild-growing and cultivated *Cannabis*

The following chapter is presented in the form of a manuscript published as a preprint in bioRxiv:

Balant, M., Vitales, D., Wang, Z., Barina, Z., Fu, L., Gao, T., Garnatje, T., Gras, A., Hayat, M. Q., Oganesian, M., Pellicer, J., Salami, A. S., Seregin, A. P., Stepanyan-Gandilyan, N., Sultana, N., Tsooj, S., Urgamal, M., Vallès, J., van Velzen, R., & Pokorny, L. (2024). Integrating target capture with whole genome sequencing of recent and natural history collections to explain the phylogeography of wild-growing and cultivated *Cannabis*.

DOI: https://doi.org/10.1101/2024.10.17.618884

Integrating target capture with whole genome sequencing of recent and natural history collections to explain the phylogeography of wild-growing and cultivated *Cannabis*

Manica Balant[1,2*‡], Daniel Vitales[1*], Zhiqiang Wang[3*], Zoltán Barina[4], Lin Fu[5], Tiangang Gao[6,7], Teresa Garnatje[1,8], Airy Gras[2], Muhammad Q. Hayat[9], Marine Oganesian[10], Jaume Pellicer[1,11], Alireza S. Salami[12,13], Alexey P. Seregin[14], Nina Stepanyan-Gandilyan[10], Nusrat Sultana[15], Shagdar Tsooj[16], Magsar Urgamal[16], Joan Vallès[2,17], Robin van Velzen[18,19†], Lisa Pokorny[20†‡]

[‡]corresponding authors

*equal contribution

[†]co-senior authors

Affiliations

[1]Institut Botànic de Barcelona (IBB, CSIC-CMCNB), Passeig del Migdia sn, 08038, Barcelona, Spain

[2]Laboratori de Botànica—Unitat Associada CSIC, Facultat de Farmàcia i Ciències de l'Alimentació—Institut de Recerca de la Biodiversitat IRBio, Universitat de Barcelona (UB), 08028 Barcelona, Spain

[3]Sichuan Zoige Alpine Wetland Ecosystem National Observation and Research Station, College of Grassland Resources, Southwest Minzu University, Chengdu, 610041, China

[4]WWF Magyarország, Budapest, 1399. Pf. 694/112, Hungary

[5]South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China

[6]State Key Laboratory of Plant Diversity and Specialty Crops, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

[7]University of Chinese Academy of Sciences, Beijing 100049, China

[8]Jardí Botànic Marimurtra - Fundació Carl Faust, pg. Carles Faust 9, 17300 Blanes, Spain

[9]Medicinal Plants Research Laboratory, Department of Plant Biotechnology, Atta-Ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan

[10]Institute of Botany 'A. Takhtajyan', National Academy of Sciences of the Republic of Armenia, Acharyan str. 1, 0040 Yerevan, Armenia

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

[12]Industrial and Medical Cannabis Research Institute (IMCRI), Tehran 14176-14411, Iran

[13]Department of Horticultural Science, Faculty of Agriculture, University of Tehran, Iran

[14]Herbarium, Faculty of Biology, Lomonosov Moscow State University, Moscow, 119991, Russia

[15]Department of Botany, Faculty of Life and Earth Sciences, Jagannath University, Dhaka 1100, Bangladesh

[16]Laboratory of Flora and Plant Systematics, Botanic Garden and Research Institute, Mongolian Academy of Sciences, Ulaanbaatar 13330, Mongolia

[17]Secció de Ciències Biològiques, Institut d'Estudis Catalans, Carrer del Carme 47, 08001 Barcelona, Spain

[18]Biosystematics Group, Department of Plant Sciences, Wageningen University & Research, 6708PB Wageningen, the Netherlands

[19]Bedrocan International, 9640CA, Veendam, the Netherlands

[20]Real Jardín Botánico (RJB-CSIC), Plaza de Murillo 2, 28014 Madrid, Spain

Authors for correspondence:

Manica Balant

Email: manica.balant@ibb.csic.es

Lisa Pokorny

Email: pokorny@rjb.csic.es

Summary

- *Cannabis* has provided important and versatile services to humans for millennia. Domestication and subsequent dispersal have resulted in various landraces and cultivars. Unravelling the phylogeography of this genus poses considerable challenges due to its complex history.
- We relied on a Hyb-Seq approach (combining target capture with shotgun sequencing), with the universal Angiosperms353 enrichment panel, to explore the genetic structure of wild-growing accessions and cultivars by implementing phylogenomic and population genomic workflows on the same Hyb-Seq data.
- Our findings support the treatment of *Cannabis* as a monotypic genus (*C. sativa* L.), structured into three main genetic groups—E Asia, Paleotropis, and Boreal—with clear phylogeographic signal despite significant levels of admixture. The E Asia group was sister to the Paleotropis and the Boreal groups. Individuals within the Paleotropis group could be further structured into three subgroups: Iranian Plateau, C & S China and Himalayas, and Indoafrica. Individuals from the Boreal group split into two subgroups: Eurosiberia and W Mongolia and Caucasus and Mediterranean. Hemp and drug-type landraces and cultivars consistently matched their putative geographic origin.
- These findings enhance our understanding of the genetic patterns in *Cannabis* and provide a framework for future research into its current and past genetic diversity.

Keywords: Angiosperms353, Cannabaceae, Hemp, Herbariomics, Hyb-Seq, Population Genomics, Phylogenomics, Single nucleotide polymorphisms

INTRODUCTION

Cannabis sativa L. (hereafter referred to as *Cannabis*) is one of the oldest multi-purpose crops, utilised by humans worldwide for thousands of years (Clarke & Merlin, 2013). It has been used as fibre (ropes, fabric, paper), medicinally (over 200 recorded uses), as food (nutrient-rich seeds), as well as in various magico-religious rituals (Balant *et al.*, 2021a,b). Despite its long history of use, *Cannabis* was broadly deemed illegal at the beginning of the 20th century, primarily because of its psychoactive properties. Consequently, studies on *Cannabis* became scarce and relied almost completely on hemp cultivars or on plant material confiscated by law enforcement. Nonetheless, spurred by recent legalization efforts, the *Cannabis* research and industry are now experiencing a revival in the agronomic, medicinal, and recreational sectors. Although there are several chromosome-level reference genomes and abundant whole genome sequencing (WGS) data available for *Cannabis*, these data predominantly originate from modern hemp cultivars or drug strains with unknown geographic origins and limited genetic diversity (e.g., van Bakel *et al.*, 2011; Braich *et al.*, 2020; Grassa *et al.*, 2021; but see also Gao *et al.*, 2020; Ren *et al.*, 2021; Chen *et al.*, 2022). Meanwhile, comprehensive studies including wild-growing and landrace individuals remain scant, which is why sampling these individuals across the entire natural distribution of this genus is much needed to better understand *Cannabis* genetic diversity and geographic structure (Kovalchuk *et al.*, 2020).

Cannabis belongs to the Cannabaceae, an angiosperm family with ten genera and over 100 species (WFO, 2024). Within the family, two closely related species stand out for their economic significance: hops (*Humulus lupulus* L.), which plays a key role in the beer industry; and *Cannabis*, which is widely used in both medical and recreational sectors (Fu *et al.*, 2023). *Cannabis* is a dioecious plant (except for some monoecious cultivars; Clarke & Merlin, 2013; Heer *et al.*, 2024), typically a diploid (2n = 20; although natural triploids and tetraploids exist), with an average genome size of ~1 pg/1C (Sharma *et al.*, 2015; Balant *et al.*, 2022; Philbrook *et al.*, 2023).

Different centres of origin of the genus across Eurasia have been proposed, but palaeobotanical studies on subfossil pollen indicate that *Cannabis* most probably originated somewhere close to the NE Tibetan Plateau ~27 million years ago (Mya) (Clarke & Merlin, 2013; McPartland *et al.*, 2018, 2019; Zhang *et al.*, 2018a; McPartland & Small, 2020). From there, it likely first spread west, reaching Europe approximately 6 Mya, and then east, arriving in E China around 1.2 Mya. Despite its current widespread use across India, the oldest subfossil pollen remains indicate that it reached the Indian subcontinent only ~30 thousand years ago (Kya) (McPartland *et al.*, 2019; Rull, 2022).

Similarly, the domestication of *Cannabis* has long been the subject of discussion. Some authors proposed a single C Asian domestication event (Schultes *et al.*, 1974), whereas others suggested several independent ones (Vavilov, 1926; McPartland *et al.*, 2018, 2019; Jin *et al.*, 2021); however, the high concentration of early archaeological remains, together with the latest study by Ren *et al.* (2021), suggest that *Cannabis* was first domesticated in E Asia, approximately 12 Kya. Although it was initially cultivated as a multipurpose crop, selection for specific type-use cultivars might have started ~4 Kya, leading to the development of separate 'Hemp-type' vs. 'Drug-type' plants (Ren *et al.*, 2021). Since then, humans have been instrumental in *Cannabis* dispersal across C and E Asia, Europe, along the Himalayas, and on the Indian subcontinent. Subsequently, with the establishment of numerous trading routes, such as the Silk Road, and the expansion of multiple empires, human-mediated dispersal intensified across Eurasia and towards Africa, reaching the Americas with the European colonization and the Atlantic slave trade. Currently, dispersal in the opposite direction is happening and modern cultivars are being reintroduced into native areas, resulting in admixture with local landraces and wild-growing *Cannabis* populations (Abel, 1980; Clarke & Merlin, 2013).

The taxonomy of *Cannabis* has historically been complex, influenced by cultural biases and legal issues that led to confusion, with numerous synonyms inconsistently applied to taxa across different geographic regions (McPartland & Guy, 2017). The first known differentiation between European and Asian *Cannabis* was recorded by Ibn-al-Baitār ca. 1240 (Lozano Cámara, 2017; McPartland & Guy, 2017); however, it was not until the 18th century that Linnaeus (*C. sativa*; 1753) and Lamarck (*C. indica* Lam.; 1783) scientifically described two distinct species. In the past two centuries, various taxonomic approaches based on genetics, morphology, and phytochemistry have been proposed, with several researchers treating *Cannabis* as a polytypic genus, identifying two or three species with various subspecies or varieties (Janischevsky, 1924; Vavilov, 1935-translated in 1992; Emboden, 1974; Schultes *et al.*, 1974; Anderson, 1980; Clarke & Merlin, 2013; Jin *et al.*, 2021). One of the first comprehensive studies, including a broad range of wild-grown and landrace *Cannabis* accessions with a worldwide distribution, was conducted by Hillig (2005a). Based on allozyme variation, morphological characters, and phytochemical profiles, he recognised two *Cannabis* species with six so-called ‘biotypes’: *C. sativa* for the accessions from the Levant, Europe, and N Asia (with hemp and feral ‘biotypes’) and *C. indica* for accessions from S, W, and E Asia, as well as Africa (with narrow-leaflet drug, wide-leaflet drug, hemp, and feral ‘biotypes’). He suggested a third species, *C. ruderalis* Janisch., might also exist; however, the sampling of individuals potentially belonging to this third putative species was too sparse to confirm its existence (Hillig, 2005b). Based on Hillig’s findings (2005a,b), Clarke and

Merlin (2013) adopted a similar classification, with three species and six subspecies.

In contrast to this polytypic taxonomic concept, others considered *Cannabis* to be a monotypic genus, recognizing only *C. sativa* (Small & Cronquist, 1976; Sawler *et al.*, 2015; Small, 2015; Lynch *et al.*, 2016; McPartland *et al.*, 2018; Ren *et al.*, 2021; Lapierre *et al.*, 2023), albeit with different infraspecific taxonomic divisions. McPartland & Small (2020), who follow the classification proposed by Small & Cronquist (1976) that recognises two subspecies within *C. sativa* (ssp. *sativa* and ssp. *indica*), carried out a large-scale revision of morphological traits, building on past genetic and phytochemical studies. Thus, within ssp. *indica*, they identified two domesticated (D) and two wild type (WT) varieties: var. *indica* (D) and var. *himalayensis* (WT) from S Asia, and var. *afghanica* (D) and var. *asperrima* (WT) from C Asia. The study by Ren *et al.* (2021), which mostly included hemp cultivars and drug strains, along with some wild-growing populations, also indicated that *Cannabis* should be considered as a single species, with individuals clustering into four genetic groups: 'Basal cannabis', 'Hemp-type', 'Drug-type feral', and 'Drug-type'. Other WGS and microsatellite markers studies have also observed differentiation between geographic regions, and between hemp and drug accessions, sometimes with further distinctions within the drug genetic pool, identifying two separate groups (Sawler *et al.*, 2015; Lynch *et al.*, 2016; Schwabe & McGlaughlin, 2019; Woods *et al.*, 2023). However, none of these studies included feral samples from either Mongolia or Africa, and they included few samples from the Caucasus, the Levant, and C & W Asia—areas otherwise reported as potentially very diverse (Soorni *et al.*, 2017; McPartland & Small, 2020; Dehnavi *et al.*, 2024). Moreover, several investigations analysing only within-country genetic diversity, found complex population structure within *Cannabis* in, e.g., China (Zhang *et al.*, 2018a; Chen *et al.*, 2022), USA (Busta *et al.*, 2022), Iran (Soorni *et al.*, 2017; Shams *et al.*, 2020; Dehnavi *et al.*, 2024), Morocco (Benkirane *et al.*, 2024), and India (Pandey *et al.*, 2023).

Based on cultivation purpose, morphology, and chemical composition, *Cannabis* plants can also be described as hemp-type (primarily grown for fibre and seed production) and drug-type, based on Δ9-tetrahydrocannabinol (THC) concentration (Hurgobin *et al.*, 2021) or on the THC and cannabidiol (CBD) ratio (THC-dominant, balanced THC:CBD, and CBD dominant, that is, Type I, Type II, and Type III, respectively; Small & Beckstead, 1973). Outside of academic environments, drug-type plants are typically classified as 'sativa', 'indica', or 'hybrid' (McPartland & Guy, 2017); however, several studies have demonstrated that these informal classifications are not supported by genetic data (Sawler *et al.*, 2015; Schwabe & McGlaughlin, 2019; Watts *et al.*, 2021).

Recent studies have relied on high-throughput sequencing approaches such as genotyping-by-sequencing (GBS) or whole genome sequencing (WGS) to study *Cannabis*; however, no previous study has attempted to use a target-capture sequencing (TCS) approach to investigate the evolution of *Cannabis*. Furthermore, none of these studies has explored the potential of herbarium specimens, which could offer valuable insights into the past distribution of *Cannabis* genotypes. Hyb-Seq (Weitemier *et al.*, 2014; Dodsworth *et al.*, 2019), that is, TCS combined with low-coverage WGS, is an affordable method (Hale *et al.*, 2020) proven very effective for sequencing not only recent and silica-dried tissue, but also historical collections (i.e., herbarium tissue), where DNA template is often highly degraded, which up until recently had thwarted their inclusion in genetic studies (Villaverde *et al.*, 2018; Brewer *et al.*, 2019; Shee *et al.*, 2020). Different probe sets (TCS kits) for specific plant families (e.g., Asteraceae, Mandel *et al.*, 2014; Euphorbiaceae, Villaverde *et al.*, 2018; Dioscoreaceae, Soto Gomez *et al.*, 2019) or larger taxonomic groups (e.g., flagellate land plants, Breinholt *et al.*, 2021) have been developed. The universal Angiosperms353 enrichment panel is a probe set which includes 353 orthologous nuclear protein-coding genes found in single copy across all flowering plants (Johnson *et al.*, 2019). Although originally conceived to study phylogenetic relationships above the species level, it has successfully been used for population-level analyses of various flowering plant groups (Slimp *et al.*, 2021; Wenzell *et al.*, 2021; Beck *et al.*, 2021; Yardeni *et al.*, 2022; Crowl *et al.*, 2022; Phang *et al.*, 2023), as well as domesticated landraces (Van Andel *et al.*, 2019).

To address the taxonomic inconsistencies and to gain a clearer understanding of the genetic structure of *Cannabis*, we conducted a comprehensive sampling (with emphasis on wild-growing populations) focusing on its native distribution range (taking special care to include individuals from previously under-sampled areas). Relying on the same Hyb-Seq dataset, we carried out phylogenomic analyses to clarify the taxonomic status of wild-growing and landrace *Cannabis* accessions, and we implemented population genomics analyses to better understand how populations are structured. In this manner, we linked macro- and microevolutionary scales to shed light on the phylogeography of *Cannabis*.

MATERIAL AND METHODS

Sampling and Molecular Protocols

For the ingroup, we sampled 94 *Cannabis sativa* L. individuals with emphasis on populations across Eurasia (Fig. 1). Fifty-eight samples were obtained from living plants, dried in silica gel,

and 36 samples were secured from herbaria. For the outgroup, three *Humulus scandens* and three *H. lupulus* SRAs, corresponding to WGS and RNA-sequencing data, were downloaded from the NCBI repository (see Table S1 for details).

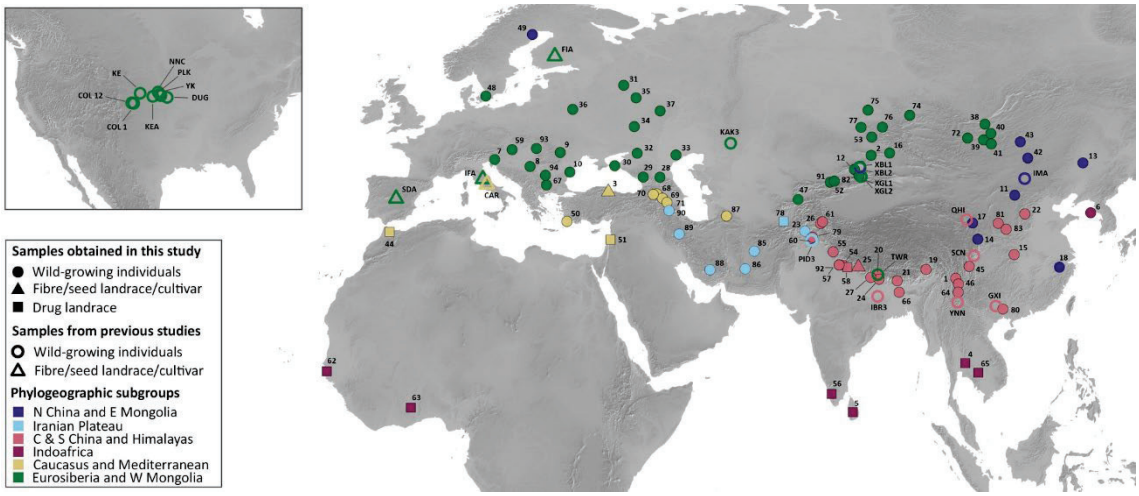


Fig. 1 Geographic distribution of samples included in this study, with individuals coloured according to the subgroups obtained in the phylogenomic analysis (see Fig. 2). The shapes indicate *Cannabis* accession types, them being, wild-growing (circles), fibre/seed (triangles), and drug (squares) types. Additionally, filled shapes are newly analysed Hyb-Seq samples, while empty shapes are NCBI SRAs corresponding to WGS data mined for our Hyb-Seq targets. The inset shows USA wild-growing populations mined from NCBI SRAs. Drug cultivars mined are not shown. For more detailed information see Supplementary Table S1. The map was made with *Natural Earth* (Free vector and raster map data @ naturalearthdata.com).

DNA of 94 *Cannabis* individuals was extracted either using the E.Z.N.A. SP Plant DNA Kit (Omega Bio-Tek, Norcross, GA, USA) or a modified CTAB protocol (Doyle & Doyle, 1987). DNA concentration was measured with a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) using dsDNA BR Qubit assays. The extractions yielded on average 2,000 ng of DNA.

DNA extractions were sent to Daicel Arbor Biosciences (Ann Arbor, MI, USA), who provide target capture sequencing services (myReads®). They carried out DNA quantitation, genomic library preparation (with dual indexing), target enrichment (nine libraries per capture reaction), and Illumina® sequencing. Captures were performed following the myBaits v5.03 protocol, using the myBaits® Expert Angiosperms353 enrichment panel (Johnson *et al.*, 2019), with an overnight hybridization and washes at 65° C. Enriched libraries were then pooled in approximately

equimolar ratios, alongside the original genomic libraries at a ratio of 75% enriched to 25% original genomic libraries. Samples were sequenced on an Illumina® NovaSeq 6000 platform on a partial S4 PE150 lane, resulting in an approximate 108 Gbp total.

Sequencing Data Processing

The de-multiplexed raw sequences were first filtered and trimmed using fastp v0.23.4 (Chen, 2023), removing adapters and low-quality reads (-f 20 -t 5 -F 20 -T 5 -g -x -W 3 -r -M 20 -q 20 -l 40 --detect_adapter_for_pe; for lower quality samples flags -q 15 and -l 30 were used instead), and checked with FastQC (Andrews, 2010) and MultiQC (Ewels *et al.*, 2016) before and after filtering with fastp.

HybPiper v2.1.6 (Johnson *et al.*, 2016) was then used to recover the single-copy nuclear genes from the Angiosperms353 enrichment panel with the target file mega353.fasta (McLay *et al.*, 2021) and the *assemble* flag and the 'bwa' option. We checked for potential paralogues using the *paralog_retriever* flag, calculated statistics with the *stats* flag, and visualised the gene recovery using the *recovery_heatmap* flag. Due either to the presence of paralogues (73_RUS_SB) or because of extremely low coverage (84_CHN_ANH), we eliminated two individuals from further analysis. The *max_overlap* script (Shee *et al.*, 2020) was then used to calculate a coverage score for each of the remaining accessions and sequences. Four more *Cannabis* individuals (12_CHN_XIN, 15_CHN_HUB, 17_CHN_QIN, and 18_CHN_ZHN), three *Humulus* accessions (DRR024392, SRR24774240, and SRR24774242), and eight genes (6514, 6886, 6705, 6893, 6713, 6565, 6557, and 5354) were also eliminated to reduce noise and remove underrepresented, incomplete, and unevenly distributed sequences across accessions from our data matrix. The supercontigs (exons plus flanking regions) of 345 target genes for the 91 remaining individuals (88 *Cannabis* and three *Humulus* individuals) were then retrieved using the *retrieve_sequences* flag selecting the 'supercontig' option.

Nuclear Species Tree Inference

Retrieved sequences were then aligned with MAFFT v7.520 (Katoh & Standley, 2013) (using flag *auto*). Exploratory gene trees were constructed with FastTree 2 v2.1.11 (Price *et al.*, 2010) and TreeShrink v1.2.1 (Mai & Mirarab, 2018) was used to automatically prune outlier branches, using the false positive tolerance rate (α) of 0.05 and the 'per-species' option. The output was then re-aligned using MAFFT (same settings as above) and trimmed with trimAl v1.4.1 (Capella-Gutiérrez *et al.*, 2009) using relaxed settings (gap threshold set to 0.3, while keeping at least 30% of the original alignment).

The gene trees were inferred under maximum likelihood (ML) with IQ-TREE v2.2.6 (Nguyen *et al.*, 2015) using ModelFinder to select the best fit DNA substitution model (Kalyaanamoorthy *et al.*, 2017) and choosing the non-parametric Shimodaira–Hasegawa approximate likelihood ratio tests (SH-aLRT; Guindon *et al.*, 2010) for assessing branch support values with 1,000 replicates. For the resulting ML gene trees, unsupported branches were collapsed using the 'nw_ed' tool from the Newick Utilities v1.6.0 package (Junier & Zdobnov, 2010) with threshold 0% SH-aLRT, as recommended by Simmons & Gatesy (2021). The coalescent species tree was then inferred using ASTRAL-III (Zhang *et al.*, 2018b) and, since branch lengths in the resulting topology come in coalescent units, RAxML-NG v 1.2.1 (Kozlov *et al.*, 2019) was used (with flag *evaluate*) to estimate branch lengths in substitutions per site (pre-requisite for some of our downstream analyses). Gene tree vs. species tree incongruence was visualised with the *AstralPlane* package (Hutter, 2021) in R v4.3.2 (R Core Team, 2022), using the *astralProjection* function to plot quartet scores calculated in ASTRAL-III (using the '-t 2' option) as pie charts. Trees were visualised in FigTree v1.4.4 (Rambaut, 2018).

Phylogenomic Placement of WGS Accessions

We downloaded 64 publicly available *Cannabis sativa* WGS SRAs from the NCBI repository (see Suppl. Table S1 for details), which we then placed in our nuclear species tree. Using fastp (Chen, 2023), these raw sequences were also quality-filtered and trimmed (with flags -f 15 -t 5 -F 15 -T 5 -g -x -W 3 -r -M 20 -q 20 -l 40 --detect_adapter_for_pe; additionally, and to prevent batch effects, alternative quality filters were also used, i.e., -f 20 -t 7 -F 20 -T 7), and quality-checked with FastQC (Andrews, 2010) and MultiQC (Ewels *et al.*, 2016). The four previously eliminated samples (12_CHN_XIN, 15_CHN_HUB, 17_CHN_QIN, and 18_CHN_ZHN) were added to the 64 NCBI SRAs.

To recover the Angiosperms353 target genes from these WGS SRAs, we also used HybPiper (Johnson *et al.*, 2016), following the same steps described above. No paralogues were found in the downloaded dataset; however, due to the diverse approaches (e.g., varying levels of sequencing depth) implemented by the different research teams who produced and shared their *Cannabis* WGS data, many samples had poor target gene recovery (to be expected, given that our 353 targets mostly appear in single-copy in the nucleus). We discarded 32 individuals that had < 200 target genes with sequences with < 50% of the mean target length, as well as the same eight genes flagged by the abovementioned *max_overlap* script (see Table S2 for details). As a result, we were left with 36 accessions, for which we retrieved supercontigs (exons and flanking regions) of 345 target genes using the *retrieve_sequences* flag and the 'supercontig' option in

HybPiper.

These supercontigs were then aligned using the 91-individual alignment above as a constraint in MAFFT (Katoh & Standley, 2013) (with flags *add* and *keeplength*). The resulting alignments were pruned to extract the 36 accessions, which we then placed in the 91-individual ASTRAL species tree (following branch-length recalculation with flag *evaluate* in RAxML-NG, see above) with EPA-ng v0.3.8 (Barbera *et al.*, 2019), using the best.Model file previously obtained with RAxML-NG (also with flag *evaluate*). The placement output was converted with GAPPA (Czech *et al.*, 2020), using the function *guppy tog*, and visualised in FigTree.

SNP Calling and Population Genomics Analyses

For SNP calling, we used the supercontig sequences (exons and flanking regions) of the 88 *Cannabis* individuals that were also included in the phylogenomic analyses. To call the SNPs, we followed the workflow designed by Slimp *et al.* (2021), with minor modifications. In brief, we generated a combined reference sequence from the longest supercontig recovered for each of the Angiosperms353 target genes. The variant detection was carried out with GATK4 v4.5.0.0 (McKenna *et al.*, 2010). We refined the combined SNP data matrix using filters 'QD < 5.0', 'FS > 60.0', 'MQ < 40.0', 'MQRankSum < −12.5', and 'ReadPosRankSum < −8.0', with flag *missing-values-evaluate-as-failing*. Only SNPs that passed all filters above were then processed using BCFtools v1.20 (Danecek *et al.*, 2021) and VCFtools v0.1.16 (Danecek *et al.*, 2011) to eliminate multi-allelic variants, and to only keep SNPs with minimum 30% quality, minimum and maximum mean depth of 10 and 200, respectively, maximum missingness of 10%, and minor allele frequency of at least 10%. All individuals had coverage < 36 and > 40 missingness. Using PLINK v1.9 (Purcell *et al.*, 2007), we additionally filtered the SNPs based on linkage disequilibrium, with settings --indep 50 5 2. On this fully filtered and unlinked SNP data matrix we calculated eigenvalues and eigenvectors for 20 principal component analysis (PCA) axes, also with PLINK.

The analysis of population structure was first carried out with STRUCTURE v.2.3.4 (Pritchard *et al.*, 2000), as implemented in the ipyrad toolkit v0.9.52 (Eaton & Overcast, 2020). The filtered VCF file was first converted into a HDF5 file, with a linkage block size of one. We assigned individuals into six population groups, which matched the subgroups in our ASTRAL species tree, and then ran the analysis with burnin length one million and three million replicates. A range of K values (2–10) was tested in five independent runs and the most likely number of clusters was selected by detecting the highest values of the ΔK statistic (Evanno *et al.*, 2005). The population structure was visualised as an ancestry matrix using the *geom_bar* function from the *ggplot2* package in R, and pie charts were projected onto a map obtained from *Natural Earth* with

MAPMIXTURE package in R (Jenkins, 2024). Heterozygosity and pairwise identity-by-descent were calculated using PLINK. The fixation index (F_{ST}) was calculated for each pair of the previously detected six phylogenetic groups using PLINK v.2.0 (Chang *et al.*, 2015), and following the Hudson method (Bhatia *et al.*, 2013).

RESULTS

Phylogenomic analyses reveal geographically defined groups

We used the Hyb-Seq approach to sequence 94 *Cannabis* individuals from across its entire native distribution (Fig. 1). Target enrichment with the Angiosperms353 universal probe set was successful for both silica-dried and herbarium samples in all but one individual. On average we obtained more than 18 million reads per individual, with ~20% reads on target for silica-dried tissue and ~23% for herbarium samples. Using the HybPiper 'supercontig' option, gene recovery rate was very high (median value for genes with at least 50% targeted gene-length recovered was 336 for silica-dried tissue and 334 for herbarium samples). We detected only six genes with putative paralogues, present in a single individual that was eliminated from downstream analyses. Finally, 88 *Cannabis* individuals (ingroup) and three *Humulus* accessions (outgroup) were included in the final dataset used to infer a species tree under the multispecies coalescent (MSC) theoretical framework (Figs. 2A & S1). Because WGS is not targeted, HybPiper retrieval was less efficient for downloaded WGS data for which, despite the high number of reads per sample (average > 96 million reads), on average only 0.42% reads mapped to the Angiosperms353 targets (with values ranging between 0.1% and 2.4%). Detailed information on target recovery statistics and max_overlap outputs can be found in Supplementary Tables S2 to S7.

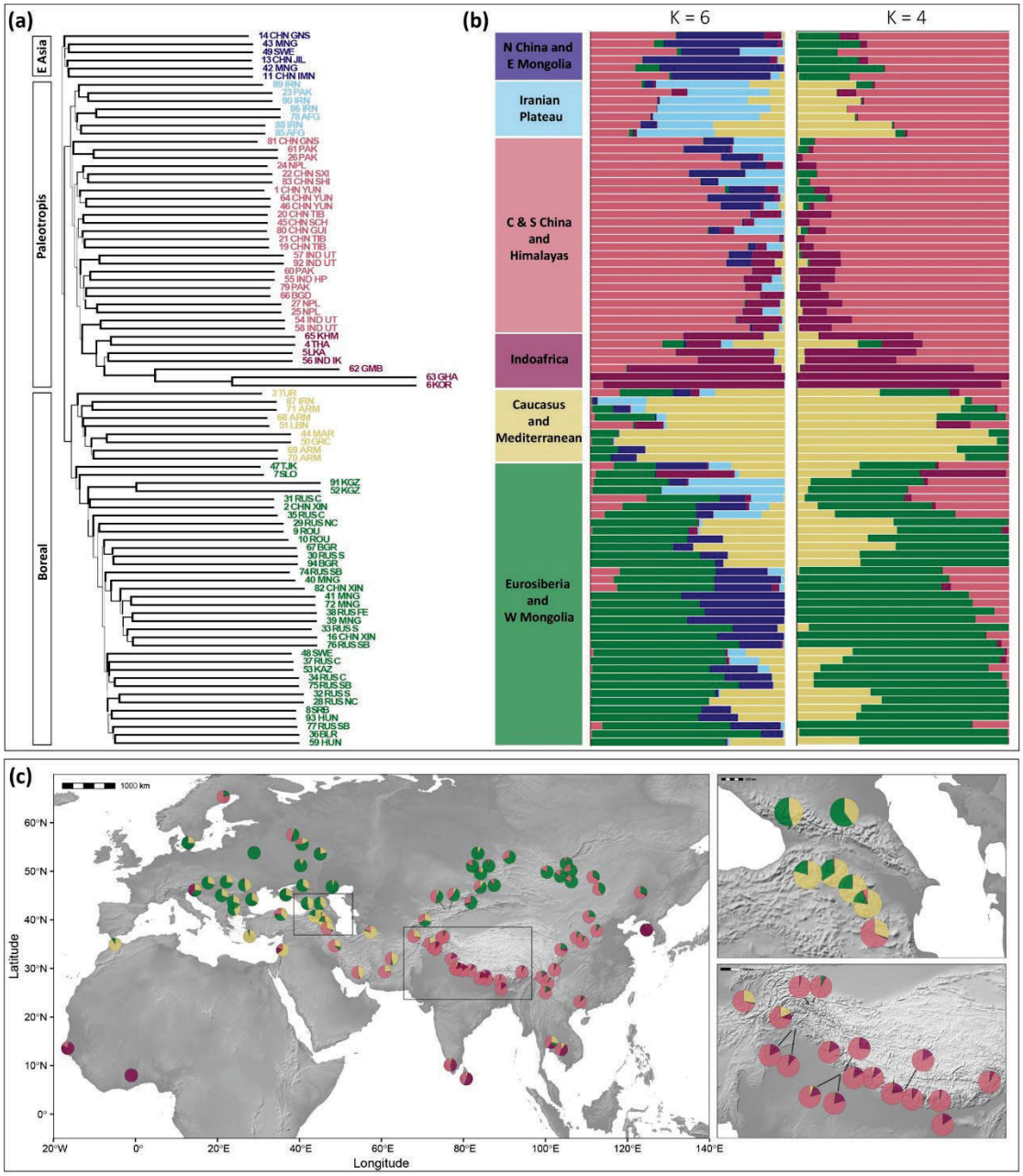


Fig. 2 Phylogenomic and population genomic analyses reveal the complex genetic structure of *Cannabis sativa* (a) ASTRAL-III nuclear species tree (for topology with the outgroup see Supp. Fig. S1) inferred from 345 ML gene trees (estimated with IQ-TREE2 from filtered MAFFT alignments), showing three main groups (E Asia, Paleotropis, and Boreal) subdivided into six subgroups matching the geographic distribution of the samples analysed (only the 88 highest-quality samples shown). Branch thickness in the species tree is proportional to support measured as local posterior probabilities (LPP), and branch length is shown in coalescent units. (b) Admixture plots estimated in STRUCTURE from 2,875 (filtered and unlinked) SNPs called from the same 345 nuclear ortholog targets used to estimate the species tree. We show genetic admixture plots for the two most likely clustering scenarios (K = 4 and K = 6, as per ΔK statistic values). (c) Geographic

distribution of samples coloured for K=4, with two insets zooming into the Caucasus and the Himalayas. The map was made with *Natural Earth* (Free vector and raster map data @ naturalearthdata.com).

The MSC species tree inferred from 345 nuclear gene trees clustered all *Cannabis* individuals together in a clade sister to genus *Humulus* (LPP = 1.0; Fig. S1). Within *Cannabis*, a division into three main genetic groups was observed. While local posterior support for these three main groups was very low, the individuals comprising them clustered into geographically distinctive subgroups (Figs. 2B & S1). The first group (E Asia group), which is sister to the other two main groups, consists of individuals from N China (provinces of Jilin, Gansu, and Inner Mongolia) and E Mongolia. All other *Cannabis* individuals belonged to either of the remaining monophyletic groups. The second group (Boreal group) consists of individuals predominantly present at latitudes above 40° N and the third group (Paleotropis group) of individuals from lower latitudes. The Paleotropis group can be further divided into the Iranian Plateau subgroup (a poorly supported group sister to all other Paleotropis individuals that includes accessions from Pakistan, Afghanistan, and Iran), and the C & S China and Himalayas subgroup (from China, Nepal, Bangladesh, N India and N Pakistan). This latter subgroup further extends into S India, Sri Lanka, W Africa, and SE Asia, forming a distinctive, highly supported, Indoafrica subgroup that is well-nested within the C & S China and Himalayas subgroup. On the other hand, the Boreal group is divided into two subgroups, them being the Caucasus and Mediterranean subgroup (from Armenia, Turkey, Greece, Lebanon, and Morocco) and the Eurosiberia and W Mongolia subgroup (from Europe, Russia, Kazakhstan, NW China, Kyrgyzstan, Tajikistan, and W Mongolia), which are reciprocally monophyletic, albeit with low support.

No batch effects were observed with regards to the placement of the 32 downloaded WGS SRAs and 4 newly sequenced individuals (with lower target capture success) into the existing MSC species tree. Instead, their placement matched the geographic origin of the samples, and not their use (i.e., hemp-type vs. drug-type; Figs. 1, 3 & S1, S3). The phylogenetically placed samples display longer branches (measured in substitutions per site) than those already present in the MSC species tree, which we attribute to an artefact resulting from their higher proportion of missing data.

Including the newly sequenced individuals and the downloaded WGS SRAs, most of these samples were collected from wild-growing plants (101), but we also incorporated seven hemp cultivars, one high CBD cultivar, and 16 other drug strains. As previously stated, rather than by

their use type, samples matched their geographic origin. Thus, the Carmagnola (CAR) hemp cultivar and an unnamed hemp cultivar from Turkey (3_TUR) were nested within wild-growing plants of the Caucasus and Mediterranean subgroup, while the Finola (FIA), Fibranova (IFA), Delta Llosa (SDA), and Fedora (FED) hemp cultivars fell in the Eurosiberia and W Mongolia subgroup, both within the Boreal group. A multipurpose landrace from Nepal (25_NPL) primarily used for fibre production was nested in the C & S China & Himalayas subgroup, within the Paleotropis group. As for the drug types, many were placed in the Paleotropis group; Haze drug strain (HAE) and landraces from Thailand (4_THA), Cambodia (65_KHM), Sri Lanka (7_LKA), S India (56_IND_IK), and W Africa (62_GMB & 63_GHA) all belonged to the Indoafrica subgroup, while N India (54_IND_UT & 58_IND_UT) drug landraces were nested in the C & S China and Himalayas subgroup. Drug strains Ruderalis indica (RIA), Hindu Kush (HKH), Purple Kush (PPK), Top 44 (TOP), and Afghanistan landrace (78_AFG) were all nested in the Iranian Plateau subgroup. However, drug landraces from Morocco (44_MAR) and Lebanon (51_LBN) belonged to the Caucasus and Mediterranean subgroup, nested within the Boreal group. Lastly, the high CBD cultivar (CBDRx) was placed within the Iranian Plateau subgroup, in the Paleotropis group.

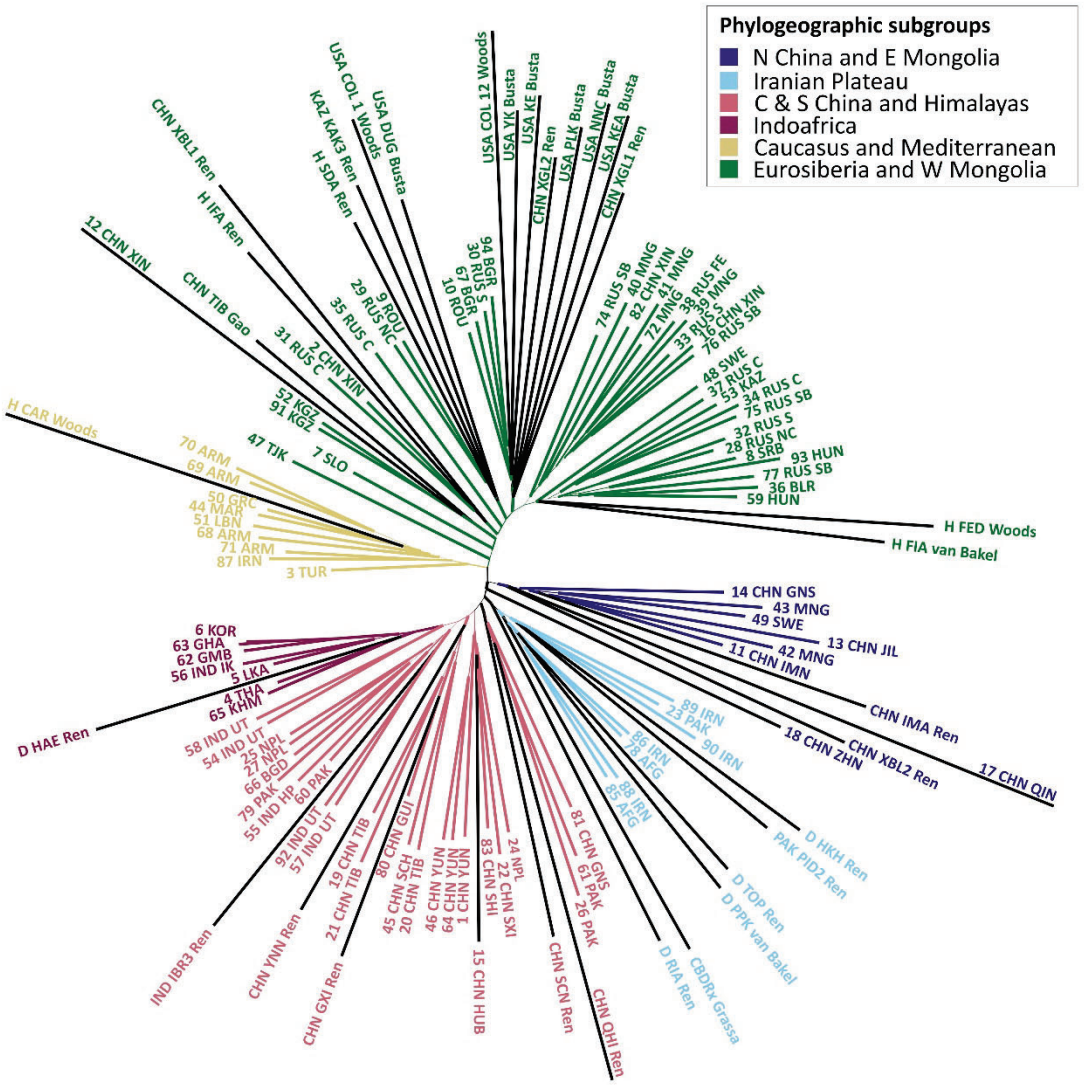


Fig. 3 Unrooted topology depicting the phylogenomic placement (done with EPA-ng) of the four lower-quality Hyb-Seq samples and the 32 WGS samples downloaded from the NCBI SRA database (black terminal branches) into the nuclear species tree inferred from 345 nuclear targets for *Cannabis* (for rooted topology with the outgroup see Supp. Fig. S2).

Population genomic analyses reveal extensive admixture across the native range

Using the longest supercontig sequences (exons and flanking regions) per target gene for the variant mapping, we were able to recover a total of 68,212 single nucleotide polymorphisms (SNPs), from the 88 *Cannabis* accessions also included in our phylogenomic workflow. Of these SNPs, 2,875 passed our robust filtering settings and were used for downstream population structure analyses.

The PCA of the filtered and unlinked SNP dataset confirmed the geographical signal revealed by

the phylogenomic analyses (Fig. 4). The two first PCA axes confirm the separation of the three main groups recovered in the MSC nuclear tree. There is also clear clustering of individuals following the subgroups these main groups are divided into in the nuclear species tree (Fig. 1), which is consistent throughout different PCA axes (Fig. 4). Together, the first two axes explain ~30% of the total variation (PC1: 20.46%; PC2: 9.56%). *Cannabis* individuals are thus structured into a Eurosiberia and E Mongolia subgroup (top-centre left), a Caucasus and Mediterranean subgroup (bottom centre), a N China and E Mongolia subgroup (top centre), an Iranian Plateau subgroup (middle), a C & S China and Himalayas subgroup (top right), and a Indoafrica subgroup (bottom right); with some exceptions (i.e., 3_TUR, 4_THA, 7_LKA, 51_LBN; Fig. 4A). The third axis (PC3) explains 8.32% of the variance and places the Indoafrica subgroup and the Caucasus and Mediterranean subgroup at opposite ends of a continuum, with most other individuals clustering in the middle (Fig. 4B & 4C).

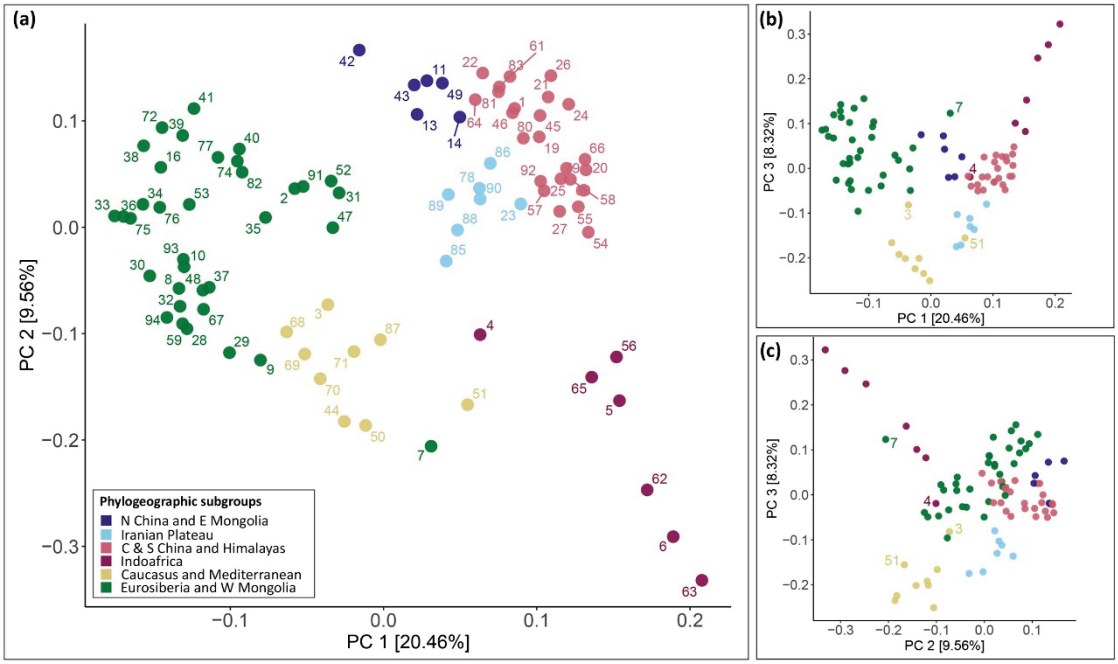


Fig. 4 Principal component analysis (PCA; done with PLINK) of *Cannabis sativa* individuals for the 2,875 (filtered and unlinked) single nucleotide polymorphisms (SNPs) called from the same 345 nuclear ortholog targets (comprising exons and their flanking regions) used to estimate the nuclear species tree (supercontig data matrix). Colours correspond to the six phylogeographic subgroups identified in the phylogenomic analysis (see Fig. 2). (a) First and second PCA axes. (b) First and third PCA axes. (c) Second and third PCA axes.

The STRUCTURE analyses (Fig. 2B) indicated that the most optimal number of clusters is four (Fig. S3). This clustering scheme is apparently inconsistent with the three main groups divided into six subgroups we observe in our nuclear species tree (Fig. 2A). Interestingly, some of these six subgroups are characterised by specific admixture patterns, also evident when we map the geographic distribution of the individuals' ancestry (Fig. 2C). Indeed, when increasing the cluster number to K=6 (second most likely clustering scheme), the specific admixture pattern characterising these six subgroups comes to the foreground (Fig. 2B). For instance, the N China and E Mongolia subgroup is a mixture of two clusters (K=4, salmon and green; K=6, salmon and dark blue). The same could be said of the Iranian Plateau subgroup (K=4, salmon and mustard; K=6, salmon and mustard with light blue); however, for this latter subgroup, the predominant cluster when K=6 (light blue) is barely found elsewhere (except for the Kyrgyzstan individuals, 52_KGZ & 91_KGZ). Meanwhile, the Indoafrica subgroup is mostly composed of a single cluster (maroon), with barely any hints in most of its individuals of the predominant cluster in the Paleotropis group (salmon), where this subgroup is otherwise nested. Similarly, the Caucasus and Mediterranean subgroup is mostly composed of a single cluster (mustard, Fig. 4), with barely a touch of the predominant cluster in the Boreal group (green), where this subgroup belongs. However, this latter predominant cluster does characterize the Eurosiberia and W Mongolia subgroup (green), regardless of the clustering scheme.

There are some individuals (i.e., 3_TUR, 4_THA, 7_ SLO and 47_TJK) with noticeably admixed ancestry profiles for either clustering scheme (measured as inbreeding coefficient, F). The genetic structure of the Turkish hemp cultivar for example is showing a highly admixed profile with high outbreeding (F_{03_TUR} = -0.522), and points to recent admixture with genetically distant individuals from different genetic backgrounds. The Slovenian sample, found growing wild near a field, likely resulted from a recent unintentional crossing between a nearby drug strain and a hemp cultivar, as evidenced by its high outbreeding (F_{7_SLO} = -0.483). The herbarium sample 47_TJK (collected from a wheat field in Tajikistan, back in 1969) shows mixed genetic ancestry, but it does not present too high outbreeding (F_{47_TJK} = 0.071). On the opposite end, we can find some individuals (i.e., 6_KOR, 19_CHN_TIB, 21_CHN_TIB, 26_PAK, 33_RUS_S, 36_BLR, 44_MAR, 50_GRC, 61_PAK, and 63_GHA) barely showing any admixture at all. While the mean inbreeding coefficient for the highly admixed samples was generally low (F = -0.358), the latter unmixed samples show relatively high mean inbreeding coefficient (F = 0.276), compared to the mean F value of 0.083 for all the samples (Table S8). As expected, the Lebanese (F_{51_LBN} = 0.2677) and Moroccan (F_{44_MAR} = 0.2717) drug landraces exhibit high inbreeding.

With regards to the inbreeding coefficient of the six subgroups identified in the nuclear species tree, the Indoafrica subgroup had the highest mean value (F = 0.133), while the Iranian Plateau had the lowest mean value (F = -0.005; Table S8). To check if the phylogenetic subgroups existing in close geographic proximity also shared the most genetic diversity, we calculated the fixation index (F_{ST}) between them. The lowest genetic distance was found between the N China and E Mongolia subgroup and the C & S China and Himalayas subgroup (F_{ST} = 0.036), and the highest between the Eurosiberia and W Mongolia subgroup and the Indoafrica subgroup (F_{ST} = 0.155; Table 1). Samples from the Indoafrica subgroup in general exhibit the highest genetic divergence from other phylogeographic subgroups and show high levels of inbreeding (F > 0.2). Only two samples in this subgroup show high outbreeding (5_LKA and 4_THA; F < -0.2), which may be due to recent hybridization for landrace improvement, as suggested by their genetic admixture profiles. Notably, the sample with the highest outbreeding (F_{4_THA} = −0.498) appeared roughly in the middle of our PCA (PC1 through PC3 axes), most distant to all other Indoafrican samples (label 4, Fig. 4). The highest proportion of shared alleles was found between individuals 6_KOR and 63_GHA (0.791), while the average value was 0.023 (Table S9).

Table 1. Pairwise fixation index (Hudson F_{ST}) values between phylogeographic subgroups.

Phylogeographic subgroup pairs		Hudson F_{ST}
Eurosiberia and W Mongolia	Indoafrica	0.155
Caucasus and Mediterranean	Indoafrica	0.136
N China and E Mongolia	Indoafrica	0.126
Indoafrica	Iranian Plateau	0.120
Indoafrica	C & S China and Himalayas	0.090
Caucasus and Mediterranean	N China and E Mongolia	0.086
Eurosiberia and W Mongolia	C & S China and Himalayas	0.080
Eurosiberia and W Mongolia	Iranian Plateau	0.077
Caucasus and Mediterranean	C & S China and Himalayas	0.076
Caucasus and Mediterranean	Eurosiberia and W Mongolia	0.061
N China and E Mongolia	Iranian Plateau	0.060
N China and E Mongolia	Eurosiberia and W Mongolia	0.058
Caucasus and Mediterranean	Iranian Plateau	0.048
Iranian Plateau	C & S China and Himalayas	0.039
N China and E Mongolia	C & S China and Himalayas	0.036

DISCUSSION

The classification of genus *Cannabis* has historically been subject to numerous interpretations, ranging from multiple species to just one. To shed light on the taxonomic status of *Cannabis*, we analysed a comprehensive set of 88 *Cannabis* wild-growing and landrace individuals across its natural distribution area (Figs. 1, S1 & S2), filling in previous sampling gaps (Levant, Caucasus, C & W Asia, and Mongolia). We relied both on recently collected silica-dried tissue samples and on historical herbarium materials. Consistent with previous work (Ren *et al.*, 2021), our MSC species tree (inferred from 345 single-copy nuclear orthologs; Fig. 2A), as well as the results from our population genomics analyses, fully support *Cannabis* as a monotypic genus (*C. sativa*, LPP = 1.0; Fig. S1) sister to the hops genus (*Humulus*). We detected admixture and low genetic differentiation even among the most distantly related populations.

Cannabis sativa Phylogeographic Structure

Within *C. sativa*, we observe three geographically well-defined groups (Figs. 2A & S1), where the E Asia group is sister to the Paleotropis and the Boreal groups. These geographic groups broadly agree with the findings of previous genetic studies (Hillig, 2005b; Ren *et al.*, 2021; Fig. S4). However, contrary to what Ren *et al.* (2021) found, our groups match the geographic distribution of individuals and not the use type, even when we place their WGS data into our nuclear species tree (Figs. 3 & S3). Additionally, we further subdivide the Paleotropis and Boreal groups into three and two subgroups, respectively, albeit with low support. A weakly supported backbone topology has also been inferred for other Angiosperms353 population-level studies (e.g., *Castilleja*, Orobanchaceae; Wenzell *et al.*, 2021), where considerable conflict among gene trees was observed. Indeed, we also detect extensive gene-tree conflict in *Cannabis* (see quartet scores for a selection of branches in Fig. S1). This conflict could stem from introgression or deep coalescence (shared ancestral alleles), as hinted in our population admixture plots (Fig. 2B). To generate these admixture plots in STRUCTURE, we used 2,875 filtered and unlinked SNPs obtained from 345 nuclear orthologs (comprising exons and their flanking regions), from both fresh silica-dried material and herbarium tissue samples. Although exons from the Angiosperms353 targets are relatively conserved across land plants, the non-coding flanking regions provided sufficient variability to uncover consistent phylogeographic patterns within *Cannabis* populations. Granted that WGS data (Ren *et al.*, 2021; Woods *et al.*, 2023) does result in abundant SNPs that may capture greater variability, it does not necessarily resolve problematic nodes, as support values in Ren *et al.* (2021) indicate.

The E Asia Group

The E Asia group comprises samples from N China and E Mongolia (hence the subgroup name). Interestingly, an individual from N Sweden also clustered within this subgroup. Given it was collected nearby a church, it is plausible this plant represents a lineage introduced by Swedish missionaries working in China in the 19th and 20th centuries (Gregersen, 2023). One could argue that the most recent common ancestor of *C. sativa* might have originated close to this N China and E Mongolia region (which would be in agreement with previous palaeobotanical findings; McPartland *et al.*, 2019) and, from there, spread westwards into the Altai Mountains (giving rise to the Boreal group), and south-westward into the Hengduan Shan and the Himalayas (resulting in the Paleotropis group), to later intersect in the axis formed by the C Asian Tien-Shan, Pamir-Alay, and Hindu-Kush mountain ranges.

Out of the six 'Basal cannabis' accessions from Ren *et al.* (2021) that we could include in our study, only two fell in our E Asia group (Figs. 3, S3 & S5). This discrepancy may result from varying levels of missing data in their dataset versus ours or it could be attributed to molecular methodological differences, i.e., WGS is anonymous (orthology needs to be assessed) and Hyb-Seq is targeted (orthology is known), where anonymous sequencing could introduce excessive noise, while targeted sequencing could lack sufficient signal (Fuentes-Pardo & Ruzzante, 2017; Dodsworth *et al.*, 2019). Additionally, as suggested by Halpin-McCormick *et al.* (2024), sampling biases could be driving the results instead. Our study focused on wild-growing individuals and landraces, with special attention to previously underrepresented areas across the entire distribution (e.g., Levant, Caucasus, and Mongolia), whereas the study by Ren *et al.* (2021) predominantly sampled commercial hemp cultivars and drug strains.

The Paleotropis Group

Within the Paleotropis group, we identified three distinct subgroups. The first one, the Iranian Plateau subgroup, includes wild-growing and drug-type plants from Iran, Afghanistan, and Pakistan. These samples exhibit a specific admixture pattern characterising its two adjacent subgroups (Fig. 2C). Prior to our study, Iranian samples had not been extensively included in *Cannabis* phylogeographic analyses. Soorni *et al.* (2017) identified two distinct groups within Iranian *Cannabis* populations: one comprised of plants from W Iran and another from E Iran. This differentiation may be influenced by admixture between W Iranian accessions and the Caucasus and Mediterranean subgroup, as evidenced by the distinct patterns observed in our results (Fig. 2).

The second subgroup in the Paleotropis group, the C & S China and Himalayas subgroup, includes

Cannabis plants with a distribution ranging from Pakistan to C China, spreading across the Himalayas, the Hengduan, and even the Qinling mountains. It encompasses both hemp- and drug- type plants, as well as wild-growing plants.

Nested within C & S China and Himalayas subgroup is a third one we denominate the Indoafrica subgroup, which is primarily composed of drug-type plants typically found at latitude ca. 10° N. A comparable subgroup was also found in a study by Lynch *et al.* (2016). While Hillig (2005b) did not find distinctions between individuals from Iran, Afghanistan, and Pakistan compared to other Asian samples, he did note that samples from sub-Saharan Africa and SE Asia clustered together within what he called the '*indica* gene pool', which has a distribution like that of our Paleotropis group (Fig. S4). Interestingly, a sample from an island off the coast of S Korea (6_KOR) also belongs to our Indoafrica subgroup; however, as it was collected from a ruderal environment, it is possible that it represents an escaped individual with a genotype closely related to those found in W Africa.

Numerous other studies have consistently identified two subgroups within drug-type *Cannabis* accessions, commonly referred to as 'sativa' and 'indica' (Sawler *et al.*, 2015; Lynch *et al.*, 2016; Schwabe & McGlaughlin, 2019; Vergara *et al.*, 2021). In our phylogenomic placement analyses (Figs. 3 & S3), we included five drug strains (van Bakel *et al.*, 2011; Ren *et al.*, 2021) and one high CBD cultivar (Grassa *et al.*, 2021). The strain named Haze grouped with samples from the Indoafrica subgroup, likely reflecting its development from landraces primarily originating from Thailand and S India (Clarke & Merlin, 2013). The remaining samples (CBDRx, HKH, PPK, TOP, and RIA) were placed within the Iranian Plateau subgroup and are generally believed to stem from landraces from Pakistan and Afghanistan, which subsequently have been extensively crossbred (Clarke & Merlin, 2013). Put simply, drug-type accessions (be them cultivated or wild-growing) are most closely related to the populations where they might have originated. It is therefore likely that these subgroups (Indoafrican versus Iranian Plateau) represent separate ancestral gene pools that eventually gave rise to the so-called 'indica' and 'sativa' drug types. Nowadays, these distinctions have been blurred due to crossing and inconsistent labelling practices, resulting in low reliability in the naming of *Cannabis* drug strains (Sawler *et al.*, 2015; Schwabe & McGlaughlin, 2019; Watts *et al.*, 2021).

The Boreal Group

The Boreal group spans from NW Africa and Europe, across Russia into C Asia and even W Mongolia (Figs. 1 & 2). Within this group, we identified two subgroups: the Caucasus and Mediterranean subgroup and the Eurosiberia and W Mongolia subgroup. Previous studies did

not detect the former subgroup, likely due to limited sampling in the region. The few studied samples until now were usually classified under what Hillig (2005b) denominated the '*sativa* gene pool' (polytypic framework).

The Caucasus and Mediterranean subgroup encompasses hemp cultivars from Turkey and Italy, drug landraces from Morocco and Lebanon, and wild-growing samples from Iran, Armenia, and Greece. The distinctive admixture pattern the Lebanese drug landrace exhibits seems to corroborate Clarke's (1998) assertion regarding the introduction of germplasm from India. Clarke & Merlin (2013) proposed a S Asian origin for Moroccan landraces, but our study supports a Caucasus/Levantine origin. Onofri *et al.* (2015) identified shared SNP mutations between Moroccan and certain Afghan landraces, maybe indicating a putative shared genetic ancestry with the Iranian Plateau subgroup through the Caucasus and Mediterranean subgroup (see mustard for K = 6 in Fig. 2B). 'Sieved hashish' production (resin that is collected from dried *Cannabis* plants and filtered through several sieves, to separate and collect trichomes rich in cannabinoids and terpenes) has been documented in Pakistan, Afghanistan, and Iran (Iranian Plateau subgroup). Additionally, Morocco, Lebanon, Turkey, and Greece, all areas where the genotype of the Caucasus and Mediterranean subgroup predominates, are also known for their 'sieved hashish' production (Abel, 1980; Clarke, 1998). It is therefore plausible that both *Cannabis* and the knowledge of 'sieved hashish' production spread across the Mediterranean region through traders from the Caucasus/Levant, who themselves could have acquired that knowledge from C Asian peoples.

The second subgroup in the Boreal group, Eurosiberia and W Mongolia, covers an extensive area and aligns well with Hillig's (2005b) '*sativa* gene pool' (polytypic framework). It primarily consists of wild-growing plants from Europe, Russia, C Asia, NE China, and W Mongolia, but it also includes all North American individuals phylogenomically placed in our species tree (Figs. 1 inset, 4 & S3), partially in agreement with previous studies (Ren *et al.*, 2021; Busta *et al.*, 2022). The presence of C Asian plants in this subgroup is surprising, since they are morphologically different from typical *Cannabis* plants growing in Europe (Vavilov, 1926; Clarke & Merlin, 2013). Nonetheless, Ren *et al.* (2021) found that a sample from Uzbekistan aligned with hemp-type plants, which is consistent with our findings. Samples from NW China, Tajikistan, and Kyrgyzstan exhibit unique admixture patterns, often blending the Iranian Plateau subgroup with the Eurosiberia and W Mongolia subgroup. The samples from Kyrgyzstan show specific genetic structure not observed in other samples (Fig. 2B), with a possible introgression from Iranian Plateau populations. Historically, this region was known for high-quality 'sieved hashish' production by Muslim Uyghurs, although its production in the region was banned by the Soviets

and Chinese in the 19th and 20th centuries, who instead introduced hemp cultivars promoting fibre production (McPartland & Small, 2020 Supp. Mat. 1).

This C Asian region is a melting pot of mountain ranges (Tien-Shan, Pamir-Alay, Hindu-Kush), biogeographic regions (Palearctic and Indomalaya), and cultures that seems to harbour genetically and morphologically diverse *Cannabis* plants in close proximity, which is surprising for a wind pollinated plant with no obvious reproductive barriers (McPartland & Small, 2020; Halpin-McCormick *et al.*, 2024). Our results are in agreement with McPartland & Small (2020), that made a comprehensive revision of over one thousand herbarium specimens and identified the Pamir-Alay and Hindu-Kush mountains as a contact zone for different *Cannabis* genetic pools. Isolation of the Iranian Plateau subgroup, C & S China and Himalayas subgroup, and the Eurosiberia and W Mongolia subgroup was possibly maintained by inaccessible mountainous terrain and/or different cultural practices. In the northern areas of Eurasia drug use was secondary and plants were selected primarily for fibre use and seed production (with some exceptions, e.g., extensive hashish production in the Turpan region in NW China). In southern Eurasian regions *Cannabis* was more frequently consumed for its psychoactive properties (Clarke & Merlin, 2013). While 'sieved hashish' is favoured in Arab countries (Iranian Plateau and Caucasus and Mediterranean subgroups), 'charas' (hand-rubbed resin from living *Cannabis* plants) and 'ganja' (smoking of dried female inflorescences) are preferred in Hindu countries (C & S China and Himalayas subgroup). Clarke & Merlin (2013) proposed that plants in dry climates were selected for 'sieved hashish' production, favouring trichomes that fall off easily. Conversely, in the humid and rainy Himalayas loose trichomes would be disadvantageous, leading to the selection of traits that allow trichomes to withstand the rain and the higher humidity levels. 'Sieved hashish' production in these areas was therefore replaced by production of hand-rubbed *charas* or by directly smoking the dried female inflorescences ('ganja'). These different traits reflect distinct genetic subgroups that are well-suited to their respective environments; the human-driven selection for different drug production styles may have further aided genetic isolation despite their geographic proximity.

Concluding Remarks

Using both phylogenomic and population genomic approaches, we have gained deeper insights into the genetic patterns in *Cannabis*. Our results support the taxonomic treatment of *Cannabis* as a single species, *Cannabis sativa*, with three main groups (E Asia, Paleotropis, and Boreal) and six subgroups (see above). Unlike some previous studies, our findings show that individuals group according to their geographical distribution rather than their use type (e.g., hemp vs. drug-type).

Hyb-Seq has shown success in dealing with older and degraded herbarium specimens (Dodsworth *et al.*, 2019), with some studies including herbarium specimens dating back to the 19th century (Villaverde *et al.*, 2018; Brewer *et al.*, 2019; Shee *et al.*, 2020; Gardner *et al.*, 2022; Moreyra *et al.*, 2023). Our study demonstrated that *Cannabis* herbarium specimens can be successfully analysed and integrated in population genomics studies. This opens the possibility of using historical herbarium specimens, collected before modern germplasm exchange and widespread crossing obscured the population structure of wild *Cannabis* populations, to illuminate the past distribution of this species and potentially detect some unquestionably wild specimens.

Since the Angiosperms353 probe set targets relatively conserved genes across land plants, it may lack the resolution needed for more detailed population genomics analyses in *Cannabis*. While WGS offers higher resolution and would allow for better integration with existing datasets, it does not perform well with older herbarium specimens. The lack of resolution could be addressed by designing a Cannabaceae specific probe set which could integrate the Angiosperms353 targets with other low-to-single-copy genes (e.g., those shared across Rosids that have successfully been used to infer phylogenomic relationships in Cannabaceae; Fu *et al.*, 2023) and further refined by incorporating, functional genes of agronomic interest, such as those involved in cannabinoid biosynthesis or in fibre development (as previously done for the yam family (Dioscoreaceae); Soto Gomez *et al.*, 2019). This combined strategy would offer a more comprehensive tool for future *Cannabis* research. Additionally, morphometric analysis (e.g., using leaves, as proposed by Balant *et al.*, 2024) and phytochemical characterization could further clarify group delimitation within *Cannabis*.

Acknowledgements

We would like to thank Joan Uriach Marsal (Uriach Laboratories), CIJA Preservations SL, and Cannaflos – Gesellschaft für medizinisches Cannabis mbH for additional financial support, and CIJA Preservations SL for providing valuable seed accessions. We would also like to thank Madison Bullock, Ciprian C. Buzna, Marco Calvi, David Criado, Teodora Dalmacija, Branko Dolinar, İrem Erdoğan, Pol Fernandez, Nikolai Friesen, Sònia Garcia, Dmitry German, Peter Glasnović, María Luisa Gutiérrez Merino, Oriane Hidalgo, Neus Ibáñez, Matthew G. Johnson, Bing Liu, Jordi Lopez, Kunigunda Macalik, Marina Olonova, Alexandra Papamichail, Sonja Siljak-Yakovlev, Vladimir Stevanović, Boštjan Surina, Paul Szatmari, Vladimir Vladimirov, Yushi Ye, and several herbaria (ALTB, CL, IBSC, KP, LD, MW, PE, and SNUA) and their curators. This research was

supported by projects WECANN (CGL2017- 84297-R, Ministerio de Ciencia, Innovacion y Universidades), Generalitat de Catalunya (CLT051, 2017SGR1116, and 2021SGR00315), Institut d'Estudis Catalans (PRO2020-2024-S02-VALLES). MB benefited from FPI predoctoral contract funded by the Ministerio de Ciencia, Innovacion y Universidades (PRE2018-083226). LP benefited from a Ramón y Cajal grant (RYC2021-034942-I) funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU"/PRTR. AG benefited from a postdoctoral grant of the Universitat de Barcelona funded by NextGeneration EU funds (Margarita Salas 2022–2024). We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI). This research project was made possible through the access granted by the Galician Supercomputing Center (CESGA) to its supercomputing infrastructure. The supercomputer FinisTerrae III and its permanent data storage system have been funded by the Spanish Ministry of Science and Innovation, the Galician Government, and the European Regional Development Fund (ERDF).

Competing interests

None of the authors had financial or non-financial conflicts of interests regarding this study.

Author contributions

MB, TGar, DV, and LP conceptualized the study, with contributions from RvV regarding the population genomics analyses. TGar, JV, and JP secured the funding. MB, ZB, BD, LF, TGao, AG, MQH, MO, ASS, APS, NS-G, NS, ST, MU, JV, and ZW conducted fieldwork or provided key herbarium specimens. MB did molecular lab work. MB, DV, LP, and RvV designed the analytical workflow, which MB and LP used to process the data. DV, TGar, LP, ZW, and RvV supervised the work. MB and LP wrote the first draft of the paper, which all authors reviewed, commented on, and edited.

Data availability

The raw Hyb-Seq data are available on NCBI Sequence Read Archive (SRA) under BioProject PRJNA1162815. The additional publicly available sequences used in the study are listed in Supplementary Table S1. The intermediate files (trimmed alignments, gene and species trees, filtered VCF file and structure outputs) and workflow are available from Zenodo (10.5281/zenodo.14567235).

ORCID

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

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

Zhiqiang Wang https://orcid.org/0000-0002-2627-7062

Lin Fu https://orcid.org/0000-0003-4933-221X

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

Airy Gras https://orcid.org/0000-0002-2256-3256

Marine Oganesian https://orcid.org/0000-0002-3297-8519

Jaume Pellicer https://orcid.org/0000-0001-7632-9775

Alexey P. Seregin https://orcid.org/0000-0002-1824-7453

Nina Stepanyan-Gandilyan https://orcid.org/0000-0002-3651-4666

Nusrat Sultana https://orcid.org/0000-0003-1373-1286

Joan Vallès https://orcid.org/0000-0002-1309-3942

Robin van Velzen https://orcid.org/0000-0002-6444-7608

Lisa Pokorny https://orcid.org/0000-0002-2478-8555

REFERENCES

Abel EL. **1980**. *Marihuana: The First Twelve Thousand Years*. New York, USA: Springer US.

Van Andel T, Veltman MA, Bertin A, Maat H, Polime T, Hille Ris Lambers D, Tjoe Awie J, De Boer H, Manzanilla V. **2019**. Hidden rice diversity in the Guianas. *Frontiers in Plant Science* **10**: 1161.

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

Andrews S. **2010**. FastQC: a quality control tool for high throughput sequence data. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE. **2011**. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biology* **12**: R102.

Balant M, Garnatje T, Vitales D, Hidalgo O, Chitwood DH. **2024**. Intra-leaf modeling of *Cannabis* leaflet shape produces leaf models that predict genetic and developmental identities. *New Phytologist* **243**: 781–796.

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.

Balant M, Gras A, Francisco G, Garnatje T, Vallès J, Vitales D. **2021a**. CANNUSE, a database of traditional *Cannabis* uses—an opportunity for new research. *Database* **2021**: baab024.

Balant M, Gras A, Ruz M, Vallès J, Vitales D, Garnatje T. **2021b**. Traditional uses of *Cannabis*: An analysis of the CANNUSE database. *Journal of Ethnopharmacology* **279**: 114362.

Barbera P, Kozlov AM, Czech L, Morel B, Darriba D, Flouri T, Stamatakis A. **2019**. EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Systematic Biology* **68**: 365–369.

Beck JB, Markley ML, Zielke MG, Thomas JR, Hale HJ, Williams LD, Johnson MG. **2021**. Are Palmer's Elm-Leaf Goldenrod and the Smooth Elm-Leaf Goldenrod Real? The Angiosperms353 Kit Provides Within-Species Signal in *Solidago ulmifolia* s. l. *Systematic Botany* **46**: 1107–1113.

Benkirane C, Charif M, Müller CM, Taaifi Y, Mansouri F, Addi M, Bellaoui M, Serghini-Caid H, Elamrani A, Abid M. **2024**. Population structure and genetic diversity of Moroccan cannabis (*Cannabis sativa* L.) germplasm through simple sequence repeat (SSR) analysis. *Genetic Resources and Crop Evolution* **71**: 2037–2051.

Bhatia G, Patterson N, Sankararaman S, Price AL. **2013**. Estimating and interpreting F_{ST} : The impact of rare variants. *Genome Research* **23**: 1514–1521.

Braich S, Baillie RC, Spangenberg GC, Cogan NOI. **2020**. A new and improved genome sequence of *Cannabis sativa*. *Gigabyte* **2020**: 1–13.

Breinholt JW, Carey SB, Tiley GP, Davis EC, Endara L, McDaniel SF, Neves LG, Sessa EB, von Konrat M, Chantanaorrapint S, *et al.* 2021. A target enrichment probe set for resolving the flagellate land plant tree of life. *Applications in Plant Sciences* **9**: e11406.

Brewer GE, Clarkson JJ, Maurin O, Zuntini AR, Barber V, Bellot S, Biggs N, Cowan RS, Davies NMJ, Dodsworth S, *et al.* 2019. Factors Affecting Targeted Sequencing of 353 Nuclear Genes From Herbarium Specimens Spanning the Diversity of Angiosperms. *Frontiers in Plant Science* **10**: 1102.

Busta L, Dweikat I, Sato SJ, Qu H, Xue Y, Zhou B, Gan L, Yu B, Clemente TE, Cahoon EB, *et al.* 2022. Chemical and genetic variation in feral *Cannabis sativa* populations across the Nebraska climate gradient. *Phytochemistry* **200**: 113206.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. **2009**. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. **2015**. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**: 7.

Chen S. **2023**. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**: e107.

Chen X, Guo H, Zhang Q, Wang L, Guo R, Zhan Y, Lv P, Xu Y, Guo M, Zhang Y, *et al.* 2022. Whole-genome resequencing of wild and cultivated cannabis reveals the genetic structure and adaptive selection of important traits. *BMC Plant Biology* **22**: 371.

Clarke RC. **1998**. *Hashish!* Los Angeles, CA: Red Eye Press.

Clarke RC, Merlin MD. **2013**. Cannabis*: Evolution and ethnobotany*. Berkley, Los Angeles, London: University of California Press.

Crowl AA, Fritsch PW, Tiley GP, Lynch NP, Ranney TG, Ashrafi H, Manos PS. **2022**. A first complete phylogenomic hypothesis for diploid blueberries (*Vaccinium* section *Cyanococcus*). *American Journal of Botany* **109**: 1596–1606.

Czech L, Barbera P, Stamatakis A. **2020**. Genesis and Gappa: Processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics* **36**: 3263–3265.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al.* 2021. Twelve years of SAMtools and BCFtools. *GigaScience* **10**: giab008.

Dehnavi MM, Damerum A, Taheri S, Ebadi A, Panahi S, Hodgin G, Brandley B, Salami SA, Taylor G. **2024**. Population genomics of a natural *Cannabis sativa* L. collection from Iran identifies novel genetic loci for flowering time, morphology, sex and chemotyping. *bioRxiv*: https://doi.org/10.1101/2024.05.07.593022.

Dodsworth S, Pokorny L, Johnson MG, Kim JT, Maurin O, Wickett NJ, Forest F, Baker WJ. **2019**. Hyb-Seq for Flowering Plant Systematics. *Trends in Plant Science* **24**: 887–891.

Doyle J, Doyle JL. **1987**. Genomic plant DNA preparation from fresh tissue-CTAB method.

Phytochemical bulletin **19**: 11–15.

Eaton DAR, Overcast I. **2020**. ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* **36**: 2592–2594.

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

Evanno G, Regnaut S, Goudet J. **2005**. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620.

Ewels P, Magnusson M, Lundin S, Käller M. **2016**. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**: 3047–3048.

Fu X, Liu S, van Velzen R, Stull GW, Tian Q, Li Y, Folk RA, Guralnick RP, Kates HR, Jin J, *et al.* 2023. Phylogenomic analysis of the hemp family (Cannabaceae) reveals deep cyto-nuclear discordance and provides new insights into generic relationships. *Journal of Systematics and Evolution* **61**: 806–826.

Fuentes-Pardo AP, Ruzzante DE. **2017**. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology* **26**: 5369–5406.

Gao S, Wang B, Xie S, Xu X, Zhang J, Pei L, Yu Y, Yang W, Zhang Y. **2020**. A high-quality reference genome of wild *Cannabis sativa*. *Horticulture Research* **7**: 73.

Gardner EM, Ahmad Puad AS, Pereira JT, Tagi J anak, Nyegang S anak, Miun P, Jumian J, Pokorny L, Zerega NJC. **2022**. Engagement with indigenous people preserves local knowledge and biodiversity alike. *Current Biology* **32**: R511–R512.

Grassa CJ, Weiblen GD, Wenger JP, Dabney C, Poplawski SG, Motley ST, Michael TP, Schwartz CJ. **2021**. A new *Cannabis* genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytologist* **230**: 1665–1679.

Gregersen M. **2023**. Nordic Connections: Norwegian and Swedish Missionaries in Changsha in the 1920s. *Journal of Imperial and Commonwealth History* **51**: 442–463.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. **2010**. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**: 307–321.

Hale H, Gardner EM, Viruel J, Pokorny L, Johnson MG. **2020**. Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in

plants. *Applications in Plant Sciences* **8**: e11337.

Halpin-McCormick A, Heyduk K, Kantar MB, Batora NL, Masalia RR, Law KB, Kuntz EJ. **2024**. Examining population structure across multiple collections of Cannabis. *Genetic Resources and Crop Evolution*: https://doi.org/10.1007/s10722-024-01928-1.

Heer LG, Mieog J, Burn A, Kretzschmar T. **2024**. Why not XY? Male monoecious sexual phenotypes challenge the female monoecious paradigm in *Cannabis sativa* L. *Frontiers in Plant Science* **15**: 1412079.

Hillig KW. **2005a**. *A Systematic Investigation of* Cannabis. Bloomington, IN, USA: Indiana University.

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

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.

Hutter CR. **2021**. AstralPlane. Available online at: https://github.com/chutter/AstralPlane.

Janischevsky DE. **1924**. *Cannabis ruderalis*. *Proceedings Saratov* **2**: 14–15.

Jenkins TL. **2024**. MAPMIXTURE: An R package and web app for spatial visualisation of admixture and population structure. *Molecular Ecology Resources* **24**: e13943.

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

Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJC, Wickett NJ. **2016**. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* **4**: 1600016.

Johnson MG, Pokorny L, Dodsworth S, Botigué LR, Cowan RS, Devault A, Eiserhardt WL, Epitawalage N, Forest F, Kim JT, *et al.* 2019. A Universal Probe Set for Targeted Sequencing of 353 Nuclear Genes from Any Flowering Plant Designed Using k-Medoids Clustering (S Renner, Ed.). *Systematic Biology* **68**: 594–606.

Junier T, Zdobnov EM. **2010**. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* **26**: 1669–1670.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. **2017**. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587–589.

Katoh K, Standley DM. **2013**. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**: 772–780.

Kovalchuk I, Pellino M, Rigault P, van Velzen R, Ebersbach J, Ashnest JR, Mau M, Schranz ME, Alcorn J, Laprairie RB, *et al.* 2020. The Genomics of *Cannabis* and Its Close Relatives. *Annual Review of Plant Biology* **71**: 713–739.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. **2019**. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**: 4453–4455.

Lamarck J-B, Poiret J-L-M. **1783**. *Encyclopédie méthodique. Botanique*. Paris, France: Panckoucke; Liège, Belgium: Plomteux.

Lapierre É, Monthony AS, Torkamaneh D. **2023**. Genomics-based taxonomy to clarify cannabis classification. *Genome* **66**: 202–211.

Linnaeus C. **1753**. *Species plantarum*. Stockholm: Laurentius Salvius.

Lozano Cámara I. **2017**. Cultivo y usos etnobotánicos del cañamo (*Cannabis sativa* L.) en la ciencia árabe (siglos VIII-XVII). *Asclepio. Revista de Historia de la Medicina y de la Ciencia* **69**: 197–209.

Lynch RC, Vergara D, Tittes S, White K, Schwartz CJ, Gibbs MJ, Ruthenburg TC, DeCesare K, Land DP, Kane NC. **2016**. Genomic and Chemical Diversity in *Cannabis*. *Critical Reviews in Plant Sciences* **35**: 349–363.

Mai U, Mirarab S. **2018**. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* **19**: 272.

Mandel JR, Dikow RB, Funk VA, Masalia RR, Staton SE, Kozik A, Michelmore RW, Rieseberg LH, Burke JM. **2014**. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Applications in Plant Sciences* **2**: 1300085.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, *et al.* 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**: 1297–1303.

McLay TGB, Birch JL, Gunn BF, Ning W, Tate JA, Nauheimer L, Joyce EM, Simpson L, Schmidt-

Lebuhn AN, Baker WJ, *et al.* 2021. New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Sciences* **9**: e11420.

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, Guy GW, Hegman W. **2018**. *Cannabis* is indigenous to Europe and cultivation began during the Copper or Bronze age: a probabilistic synthesis of fossil pollen studies. *Vegetation History and Archaeobotany* **27**: 635–648.

McPartland JM, Hegman W, Long T. **2019**. *Cannabis* in Asia: its center of origin and early cultivation, based on a synthesis of subfossil pollen and archaeobotanical studies. *Vegetation History and Archaeobotany* **28**: 691–702.

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

Moreyra LD, Garcia-Jacas N, Roquet C, Ackerfield JR, Arabacı T, Blanco-Gavaldà C, Brochmann C, Calleja JA, Dirmenci T, Fujikawa K, *et al.* 2023. African Mountain Thistles: Three New Genera in the *Carduus-Cirsium* Group. *Plants* **12**: 3083.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. **2015**. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**: 268–274.

Onofri C, de Meijer EPM, Mandolino G. **2015**. Sequence heterogeneity of cannabidiolic- and tetrahydrocannabinolic acid-synthase in *Cannabis sativa* L. and its relationship with chemical phenotype. *Phytochemistry* **116**: 57–68.

Pandey Y, Chaturvedi T, Swaroop H, Gupta AK, Shanker K, Tiwari G. **2023**. Phytochemical and genetic marker (SCoT and CBDP) based study of genetic diversity and population structure in natural populations of *Cannabis sativa* L.: A high-value sustainable biodiversity of North-Indian Himalaya. *Industrial Crops and Products* **200**: 116892.

Phang A, Pezzini FF, Burslem DFRP, Khew GS, Middleton DJ, Ruhsam M, Wilkie P. **2023**. Target capture sequencing for phylogenomic and population studies in the Southeast Asian genus *Palaquium* (Sapotaceae). *Botanical Journal of the Linnean Society* **203**: 134–147.

Philbrook R, Jafari M, Gerstenberg S, Say KL, Warren J, Jones AMP. **2023**. Naturally Occurring Triploidy in *Cannabis*. *Plants* **12**: 3927.

Price MN, Dehal PS, Arkin AP. **2010**. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **5**: e9490.

Pritchard JK, Stephens M, Donnelly P. **2000**. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**: 945–959.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, *et al.* 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**: 559–575.

R Core Team.2022. *R: A Language and Environment for Statistical Computing. Available online at: https://www.R-project.org.*

Rambaut A. **2018**. FigTree. Available online at: http://tree.bio.ed.ac.uk/software/figtree/ and https://github.com/rambaut/figtree.

Ren G, Zhang X, Li Y, Ridout K, Serrano-Serrano ML, Yang Y, Liu A, Ravikanth G, Nawaz MA, Mumtaz AS, *et al.* 2021. Large-scale whole-genome resequencing unravels the domestication history of *Cannabis sativa*. *Science Advances* **7**: eabg2286.

Rull V. **2022**. Origin, early expansion, domestication and anthropogenic diffusion of *Cannabis*, with emphasis on Europe and the Iberian Peninsula. *Perspectives in Plant Ecology, Evolution and Systematics* **55**: 125670.

Sawler J, Stout JM, Gardner KM, Hudson D, Vidmar J, Butler L, Page JE, Myles S. **2015**. The Genetic Structure of Marijuana and Hemp. *PLoS ONE* **10**: e0133292.

Schultes RE, Klein WM, Plowman T, Lockwood TE. **1974**. *Cannabis*: An example of taxonomic neglect. *Botanical Museum Leaflets, Harvard University* **23**: 337–367.

Schwabe AL, McGlaughlin ME. **2019**. Genetic tools weed out misconceptions of strain reliability in *Cannabis sativa*: implications for a budding industry. *Journal of Cannabis Research* **1**: 3.

Shams R, Azizi A, Hamzei J, Noroozisharaf A, Moghadam S, Kordrostami M. **2020**. Genetic structure and diversity of Iranian Cannabis populations based on phytochemical, agro-morphological and molecular markers. *Industrial Crops and Products* **158**: 112950.

Sharma V, Srivastava DK, Gupta RC, Singh B. **2015**. Abnormal Meiosis in Tetraploid (4x) *Cannabis sativa* (L.) from Lahaul-Spiti (Cold Desert Higher Altitude Himalayas)-A Neglected But Important Herb. *Journal of Biological and chemical Chronicles* **2**: 38–42.

Shee ZQ, Frodin DG, Cámara-Leret R, Pokorny L. **2020**. Reconstructing the Complex Evolutionary History of the Papuasian *Schefflera* Radiation Through Herbariomics. *Frontiers in Plant Science* **11**: 258.

Simmons MP, Gatesy J. **2021**. Collapsing dubiously resolved gene-tree branches in phylogenomic coalescent analyses. *Molecular Phylogenetics and Evolution* **158**: 107092.

Slimp M, Williams LD, Hale H, Johnson MG. **2021**. On the potential of Angiosperms353 for population genomic studies. *Applications in Plant Sciences* **9**: e11419.

Small E. **2015**. Evolution and Classification of *Cannabis sativa* (Marijuana, Hemp) in Relation to Human Utilization. *Botanical Review* **81**: 189–294.

Small E, Beckstead HD. **1973**. Cannabinoid Phenotypes in *Cannabis sativa*. *Nature* **245**: 147–148.

Small E, Cronquist A. **1976**. A Practical and Natural Taxonomy for *Cannabis*. *Taxon* **25**: 405–435.

Soorni A, Fatahi R, Haak DC, Salami SA, Bombarely A. **2017**. Assessment of Genetic Diversity and Population Structure in Iranian Cannabis Germplasm. *Scientific Reports* **7**: 15668.

Soto Gomez M, Pokorny L, Kantar MB, Forest F, Leitch IJ, Gravendeel B, Wilkin P, Graham SW, Viruel J. **2019**. A customized nuclear target enrichment approach for developing a phylogenomic baseline for *Dioscorea* yams (Dioscoreaceae). *Applications in Plant Sciences* **7**: e11254.

Vavilov NI. **1926**. The origin of the cultivation of 'Primary' crops, in particular of cultivated hemp. In: Studies on the origin of cultivated plants. Leningrad: Institute of Applied Botany and Plant Breeding, 221–233.

Vavilov NI. **1935**. The phyto-geographical basis for plant breeding. In: Origin and geography of cultivated plants (translated by Löve Doris in 1992). Cambridge: Cambridge University Press, 316–366.

Vergara D, Huscher EL, Keepers KG, Pisupati R, Schwabe AL, McGlaughlin ME, Kane NC. **2021**. Genomic Evidence That Governmentally Produced *Cannabis sativa* Poorly Represents Genetic Variation Available in State Markets. *Frontiers in Plant Science* **12**: 668315.

Villaverde T, Pokorny L, Olsson S, Rincón-Barrado M, Johnson MG, Gardner EM, Wickett NJ, Molero J, Riina R, Sanmartín I. **2018**. Bridging the micro- and macroevolutionary levels in

phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist* **220**: 636–650.

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.

Weitemier K, Straub SCK, Cronn RC, Fishbein M, Schmickl R, McDonnell A, Liston A. **2014**. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* **2**: 1400042.

Wenzell KE, McDonnell AJ, Wickett NJ, Fant JB, Skogen KA. **2021**. Incomplete reproductive isolation and low genetic differentiation despite floral divergence across varying geographic scales in *Castilleja*. *American Journal of Botany* **108**: 1270–1288.

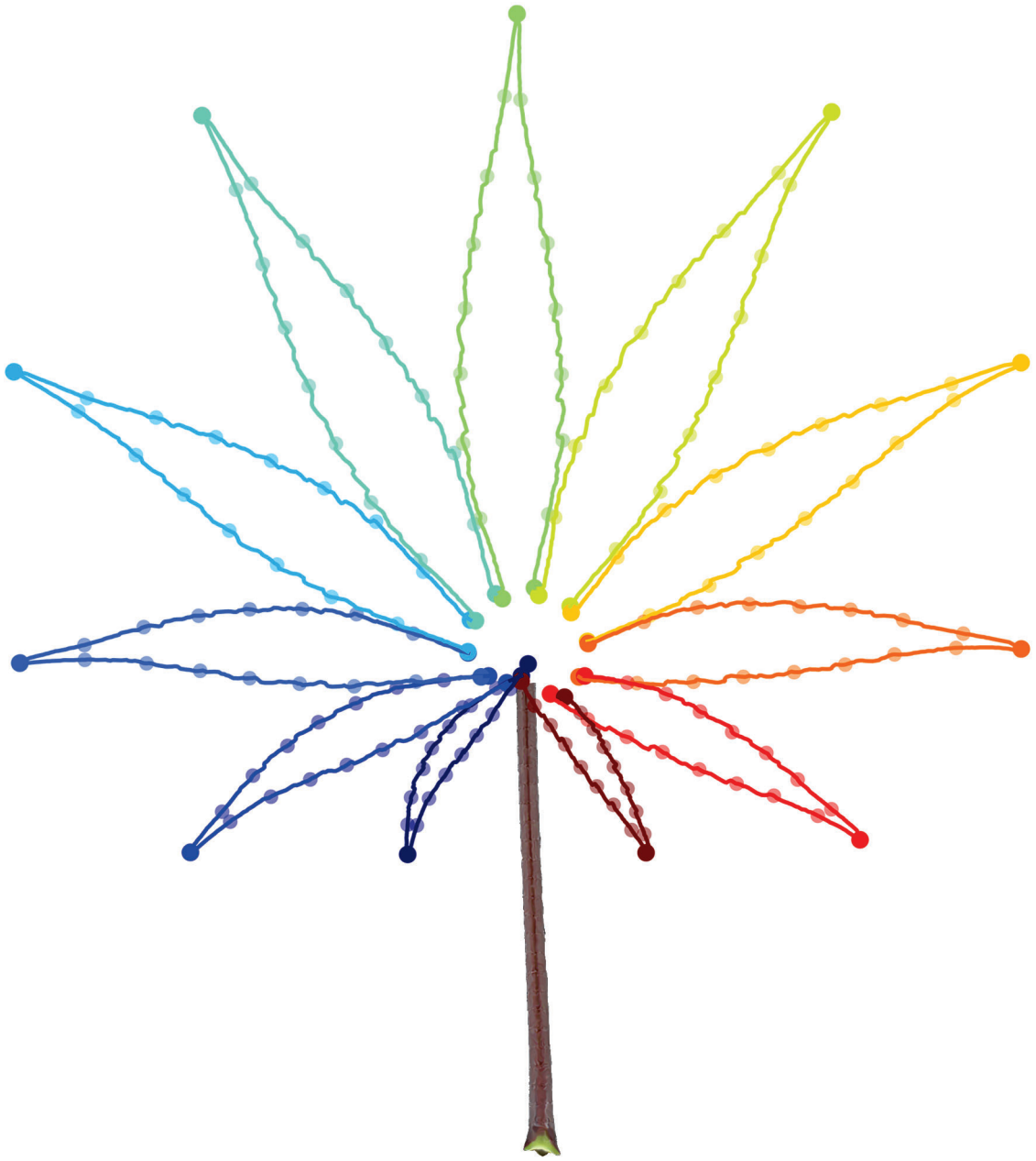
WFO. **2024**. Cannabaceae Martinov. World Flora Online, http://www.worldfloraonline.org/, Accessed on 22/09/2024.

Woods P, Price N, Matthews P, McKay JK. **2023**. Genome-wide polymorphism and genic selection in feral and domesticated lineages of *Cannabis sativa*. *G3: Genes, Genomes, Genetics* **13**: jkac209.

Yardeni G, Viruel J, Paris M, Hess J, Groot Crego C, de La Harpe M, Rivera N, Barfuss MHJ, Till W, Guzmán-Jacob V, *et al.* 2022. Taxon-specific or universal? Using target capture to study the evolutionary history of rapid radiations. *Molecular Ecology Resources* **22**: 927–945.

Zhang Q, Chen X, Guo H, Trindade LM, Salentijn EMJ, Guo R, Guo M, Xu Y, Yang M. **2018a**. Latitudinal Adaptation and Genetic Insights Into the Origins of *Cannabis sativa* L. *Frontiers in Plant Science* **9**: 1876.

Zhang C, Rabiee M, Sayyari E, Mirarab S. **2018b**. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**: 153.



CHAPTER 2

CANNABIS MORPHOLOGICAL DIVERSITY

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

The following chapter is presented in the form of a published article:

Balant, M., Garnatje, T., Vitales, D., Hidalgo, O., & Chitwood, D. H. (2024). Intra-leaf modeling of *Cannabis* leaflet shape produces leaf models that predict genetic and developmental identities. *New Phytologist*, 243(2), 781–796.

DOI: https://doi.org/10.1111/nph.19817

New Phytologist

Research

Methods Article

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

Manica Balant[1,2,3], Teresa Garnatje[1,4], Daniel Vitales[1], Oriane Hidalgo[1,5] and Daniel H. Chitwood[3,6]

[1]Institut Botànic de Barcelona, IBB (CSIC-CMCNB), Passeig del Migdia s.n., 08038, Barcelona, Spain; [2]Laboratori de Botànica, Unitat Associada al CSIC, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona (UB), Av. Joan XXIII 27–31, 08028, Barcelona, Spain; [3]Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA; [4]Jardí Botànic Marimurtra – Fundació Carl Faust, pg. Carles Faust, 9, 17300, Blanes, Spain; [5]Royal Botanic Gardens, Kew, Richmond, TW9 3AE, UK; [6]Department of Computational Mathematics, Science & Engineering, Michigan State University, East Lansing, MI 48824, USA

Authors for correspondence:
Daniel H. Chitwood
Email: dhchitwood@gmail.com

Manica Balant
Email: manica.balant@ibb.csic.es

Received: *2 February 2024*
Accepted: *28 April 2024*

New Phytologist (2024) **243:** 781–796
doi: 10.1111/nph.19817

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

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 within a single plant during development, which has so far prevented the proper use of common morphometric techniques.
- Here, we present a new method that overcomes the challenge of nonhomologous landmarks in palmate, pinnate, and lobed leaves, using *Cannabis* as an example. We model corresponding pseudo-landmarks for each leaflet as angle-radius coordinates and model them as a function of leaflet to create continuous polynomial models, bypassing the problems associated with variable number of leaflets between leaves.
- We analyze 341 leaves from 24 individuals from nine *Cannabis* accessions. Using 3591 pseudo-landmarks in modeled leaves, we 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.

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 archeological evidence suggests that *Cannabis* was domesticated *c.* 12 000 yr ago in East Asia, initially serving as a multipurpose crop before separate selections for fiber and drug production emerged *c.* 4000 yr 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 world-wide has led to a surge in the cannabis industry (Prohibition Partners, 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, *C. sativa* (Small & Cronquist, 1976; Sawler *et al.*, 2015; Small, 2015; McPartland, 2018; McPartland & Small, 2020; Ren *et al.*, 2021). Hillig (2005a) 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, wide-leaflet drug, hemp and feral biotype, and *C. sativa* as hemp and feral biotype. Recently, Lapierre *et al.* (2023)

New Phytologist

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 < 0.3% of the psychoactive compound Δ9-tetrahydrocannabinol (THC), 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 cannabidiol (CBD) 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*.

Cannabis arguably possesses one of the most iconic leaves among all plants. Its palmately compound leaves with a varying number of leaflets are a popular culture symbol. *Cannabis* exhibits a remarkable degree of phenotypic plasticity, further accentuated by selection pressure during the domestication process (Small, 2015). Extensive variability in leaf morphology has already been described by Quimby *et al.* (1973) and later Anderson (1980), who was the first to quantify the width, length, and ratio of the central leaflet. This or similar methods were then commonly used in studies investigating the morphological characteristics of *Cannabis* species, subspecies, cultivars, biotypes, and chemotypes (Small *et al.*, 1976; de Meijer *et al.*, 1992; de Meijer & Keizer, 1996; Hillig, 2005a; Clarke & Merlin, 2013; Lynch *et al.*, 2016; Karlov *et al.*, 2017; Parsons *et al.*, 2019; McPartland & Small, 2020; Carlson *et al.*, 2021; Islam *et al.*, 2021; Jin *et al.*, 2021a; Vergara *et al.*, 2021; Buzna & Sala, 2022; Chen *et al.*, 2022; Murovec *et al.*, 2022), often with contradictory results. Leaf shape has therefore played an important and sometimes controversial role in *Cannabis* taxonomy. While researchers in previous *Cannabis* studies were aware of enormous plasticity and the effect the environment has on leaf shape (Vergara *et al.*, 2021; Murovec *et al.*, 2022), they very rarely paid attention to the effects of developmental processes, even though heteroblastic changes (differences in leaf shape arising from juvenile-to-adult phase transitions in the meristem) profoundly affect the arrangement and shape of *Cannabis* leaves along the shoot. While some studies briefly mention the developmental changes in leaves (Hillig, 2005a; Carlson *et al.*, 2021; Jin *et al.*, 2021b; Spitzer-Rimon *et al.*, 2022), the only two studies focusing on heteroblastic phase changes in leaves along the plant axis were done by Heslop-Harrison & Heslop-Harrison (1958) and Hesami *et al.* (2023). In the lower part of the shoot, *Cannabis* leaves exhibit opposite phyllotaxy and one to three leaflets, transitioning to alternate phyllotaxy and leaves with up to 11 or 13 leaflets in the upper section (Hillig, 2005a; Clarke & Merlin, 2013; Small, 2015). Additionally, the changes in leaflet number are not uniform between different *Cannabis* accessions (Hillig, 2005a). These changes during development not only complicate categorization of plant accessions based on leaf shape but also prevent the use of morphometric techniques.

Morphometrics is the quantitative analysis of shape. It includes a wide range of methods, from measuring allometric differences in dimensions such as lengths, widths, and angles in relation to size (Niklas, 1994) to geometric techniques that measure shape comprehensively, such as 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. 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* (and several other species with lobed, pinnate, and palmate compound leaves) precludes analysis methods that rely on homologous, comparable points to measure shape comprehensively. Methods to automatically isolate individual leaflets (Failmezger *et al.*, 2018) or to model developmental trajectories, such as heteroblastic series (Biot *et al.*, 2016), were proposed previously for morphometrical analysis in such cases. In *Cannabis*, 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 seek to build on these works and conceptually extend our framework of continuously modeling leaflets within a palmate leaf. We model corresponding pseudo-landmarks for each leaflet as angle-radius coordinates relative to the petiolar junction and model angle and radius as a function of leaflet number to create continuous polynomial models that bypass the problems associated with variable numbers of leaflets between leaves. This enabled us to compare leaves with different numbers of leaflets within a plant and to discern differences between genotypes rather than the heteroblastic series. Analyzing over 300 *Cannabis* leaves, we model theoretical leaves with nine leaflets and 3591 comparable pseudo-landmarks. 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

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

Accession ID	Accession type	Location/cultivar name	No. of individuals	No. of leaves collected	No. of leaves analyzed
AM15	Wild/feral	Armenia, Sjunik marz, Goris town	5	90	74
BNG	Wild/feral	Bangladesh, Rangpur, Carmichael College Campus	1	14	10
FUT75	Cultivar	Futura 75	2	45	30
HU1	Wild/feral	Hungary, Nyírvasvári	4	83	68
IK	Landrace	India, Kerala	4	92	53
IKL	Landrace	India, Kullu	4	69	47
MAR	Landrace	Morocco, North Morocco	1	18	15
MN9	Wild/feral	Mongolia, Selenge aimag, Baruunburen sum	1	14	10
RO1	Wild/feral	Romania, Mangalija	2	36	34

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.

Materials and Methods

Plant material and growing conditions

This study includes 24 individuals from nine accessions of *C. sativa* L. (Table 1; Fig. 1), encompassing both wild/feral accessions and cultivated varieties with a wide distribution area. The plants were grown from seeds in a growth chamber (D1200PLL; Fitoclima, Aralab, Portugal) to minimize the influence of the environment. Before sowing, the seeds were sterilized overnight in a 5% of H_2O_2 solution with the addition of Inex-A solution (Cosmocel, Zaragoza, Spain) at room temperature. Sterilized seeds were then transferred to Petri dishes and placed in the growth chamber for germination. Once the first leaves emerged, the seedlings were transferred to small peat pots with a pre-fertilized soil substrate (Kilomix Atami, Oldbury, UK). During this phase, the environmental conditions were set to 25°C, with an 18 h : 6 h, day : night photoperiod, and a light intensity of 50 μmol m^{-2} s^{-1} (Master PL-L 55W; Philips, Amsterdam, the Netherlands). After 2 wk, the surviving plants were transplanted to 3.5-l pots with the same soil substrate. The light intensity was gradually increased to 300 μmol m^{-2} s^{-1} over the following week, without changing the photoperiod and temperature. The onset of flowering in some *Cannabis* accessions is photoperiod-dependent; therefore, after 4 wk, the photoperiod was changed to 12 h of daylight and 12 h of darkness, and the light intensity was gradually increased to 700 μmol m^{-2} s^{-1} over the following week, while keeping the temperature at 25°C. The plants remained in these environmental conditions until the flowering stage. Plants received daily irrigation with tap water, without any application of nutrient or phytosanitary control.

Leaf sampling and imaging

A total of 461 leaves were sampled during the flowering stage, with the exception of individuals from the accession IK, which did not begin to flower during the 2-month cultivation period. Leaves along the main axis of the plants were collected and immediately scanned using a flatbed photograph scanner (Epson Perfection V370, Suwa, Japan) at 1200 dpi resolution. A piece of velvet fabric was placed between the leaf and the scanner cover to avoid any shadows. No adjustments to the angle of individual leaflets were made before scanning. Each leaf was scanned with a scale and a label indicating the node it originated from, followed by a sequential lowercase letter, since typically two leaves are present per node. Starting at the base of the plant, the first two leaves were labeled as leaves 'a' and 'b' from node number 1, and so on, until the shoot apex.

Cannabis leaves display a marked heteroblastic, or juvenile-to-adult, leaf shape progression. Mature, juvenile leaves located on the first node at the base of the plant usually have a simple, serrated leaf. As node number increases so does the leaflet number, reaching a maximum of 9–13 leaflets in young, adult leaves at the growing tip. Eventually, leaves transition into an inflorescence type. During this transition, the number of leaflets per leaf starts to decrease again until the top of the inflorescence. Leaves at the shoot base have opposite phyllotaxy and transition to alternate phyllotaxy in the upper section on the stem and inflorescence (Heslop-Harrison & Heslop-Harrison, 1958; Hillig, 2004; Potter, 2009; Spitzer-Rimon *et al.*, 2022). To ensure that only stem leaves were included in our analysis, we separated the two types (i.e. stem and inflorescence leaves) based on the point where the decrease in the number of leaflets appeared. This point determined the 'total node number', the number of nodes per plant used for further analysis. Total node number varied among individuals. To compare node positions, a relative node number was calculated, which was defined by the node position divided by the total node number for the individual plant, where zero is at the plant base and one at the last node included in the analysis (Fig. 1). Because of the nature of plant growth, the leaves at the base of the plant were frequently too senesced to be incorporated into the analysis or were entirely lost. Nevertheless, the nodes could still be identified, which allowed them to be taken into account in the calculation of relative node number.

Image analysis and landmarking

After eliminating damaged and deformed leaves (39), simple leaves (4), leaves with even leaflet numbers (3), and leaves with

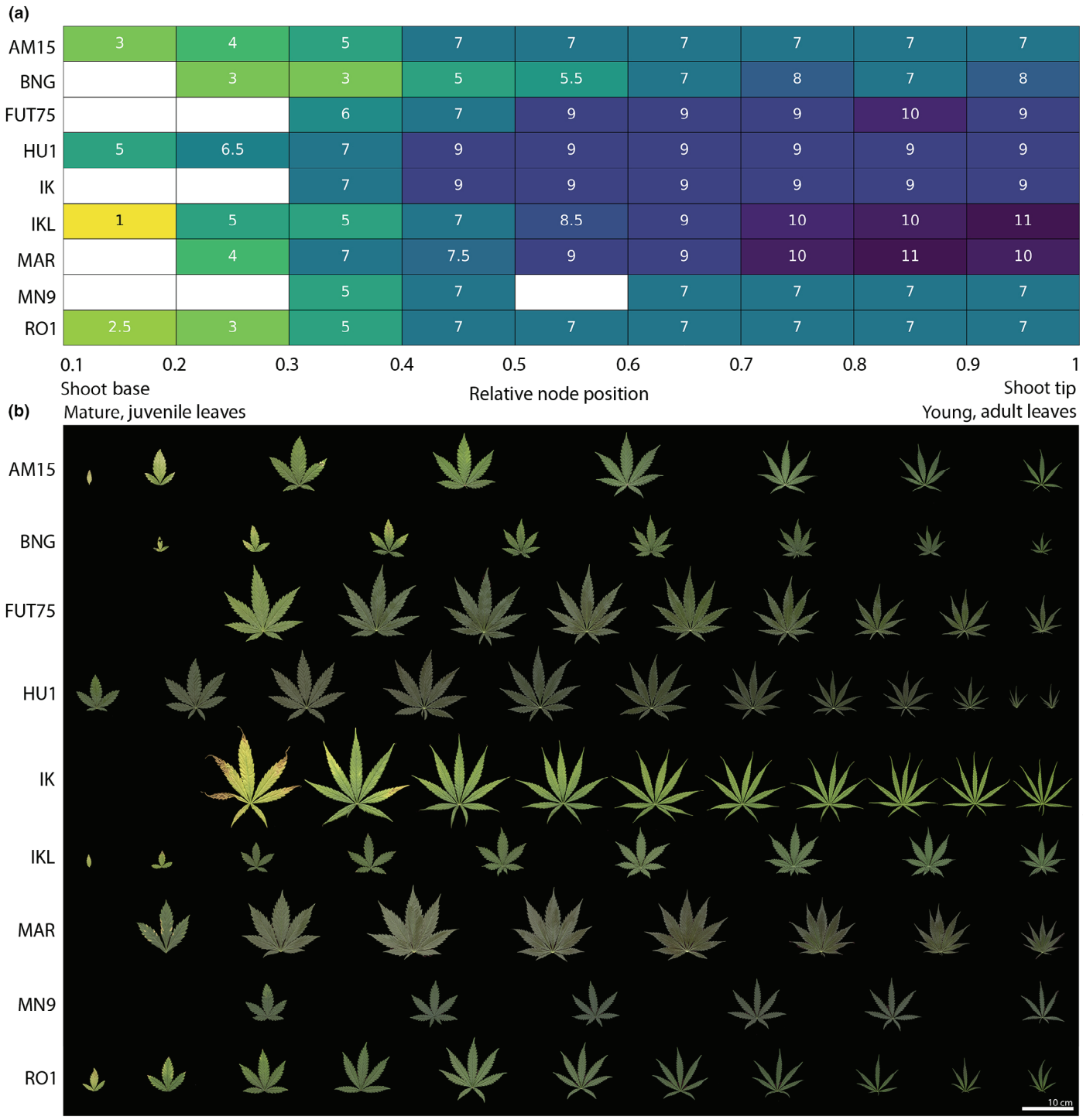


Fig. 1 Changes in the leaf shape and leaflet number during the development in nine *Cannabis* accessions. (a) Median values for all available leaflet number for each relative node number for the nine *Cannabis* accessions. (b) Changes in leaf shape between different developmental stages in different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

relative node values above one (57), a total of 358 *Cannabis* leaves were used for image analysis and landmarking. 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_updated). The *x* and *y* coordinates of blade outlines and landmarks were extracted using IMAGEJ (Abràmoff *et al.*, 2004). The

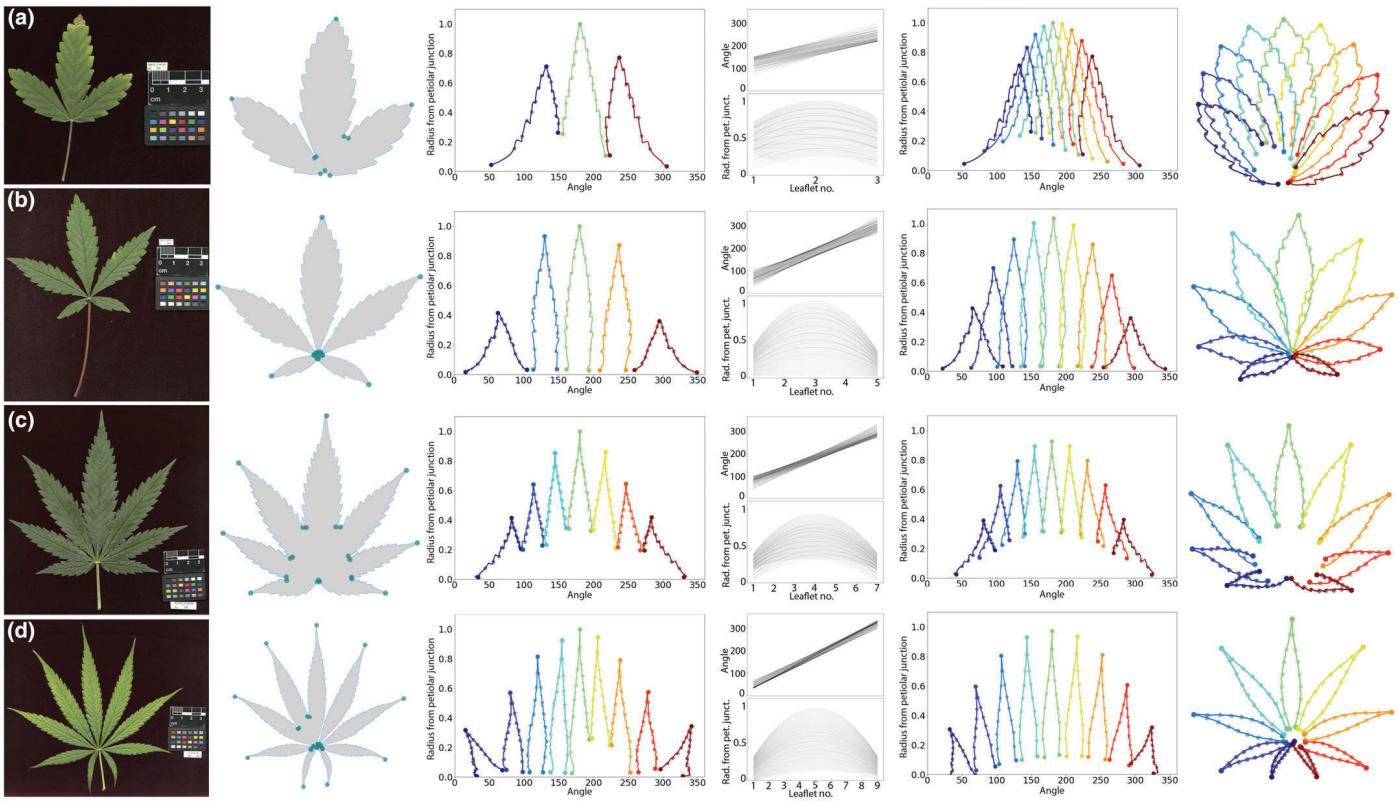


Fig. 2 Process of modeling theoretical *Cannabis* leaves for a leaf with (a) three leaflets from accession AM15, (b) five leaflets from accession IKL, (c) seven leaflets from accession FUT75, and (d) nine leaflets from accession IK. The first column shows the scans of the leaves, which we use to extract the outline and place the landmarks on the tip, start, and end of each leaflet and on the petiolar junction (second column). These coordinates are used to generate 200 equidistant pseudo-landmarks on each side of each leaflet, sharing the landmark on the tip of the leaflet for a total of 399 pseudo-landmarks. These coordinates are then converted into polar coordinates. Each transformed leaflet is defined with 399 equidistant pseudo-landmarks, with three landmarks, two at the base and one at the tip. Large points are placed every 25 pseudo-landmarks to emphasize that leaflet outlines are defined by points (third column). Second-degree polynomials for angles and for radius from petiolar junction are then fitted through these 399 pseudo-landmarks (fourth column). A modeled theoretical leaf with nine leaflets defined by 3591 pseudo-landmarks can then be modeled using the collection of 798 polynomial models for each leaf (399 polynomial models for angles and 399 for radius from petiolar junction) (fifth column) and visualized in the Cartesian coordinate system (sixth column).

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, second column). 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_updated).

Reconstruction of the new modeled leaves

To analyze leaves with different numbers of leaflets, pseudo-landmarks of each leaflet were modeled as second-degree polynomial models of angles and radius as functions of leaflet number within a leaf, in order to use the models to construct a modeled theoretical leaf with a desired number of leaflets. The PYTHON code, presented as a Jupyter notebook with detailed description, is available on GitHub (https://github.com/BalantM/Cannabis_leaf_morpho_updated). The x and y coordinates of the leaf outline were first interpolated to create an arbitrarily high number of coordinates to increase resolution of the leaf outline. The coordinates of manually selected landmarks were then compared against the high-resolution coordinates of the leaf outline, and the nearest neighboring point of the high-resolution coordinates to each original landmark was identified and specified as the new landmark point. Next, the outline and new landmark coordinates were rotated, translated, and scaled so that the central leaflet had a length of one and pointed in the same direction. The transformed points were then interpolated to generate 200 pseudo-landmarks on each side of each leaflet (from the landmark at the bottom until the tip of the leaflet), sharing the landmark on the tip of the leaflet (i.e. a total of 399 pseudo-landmarks per leaflet). These pseudo-landmarks were then converted to polar coordinates, where each point was defined by a radius and angle relative to

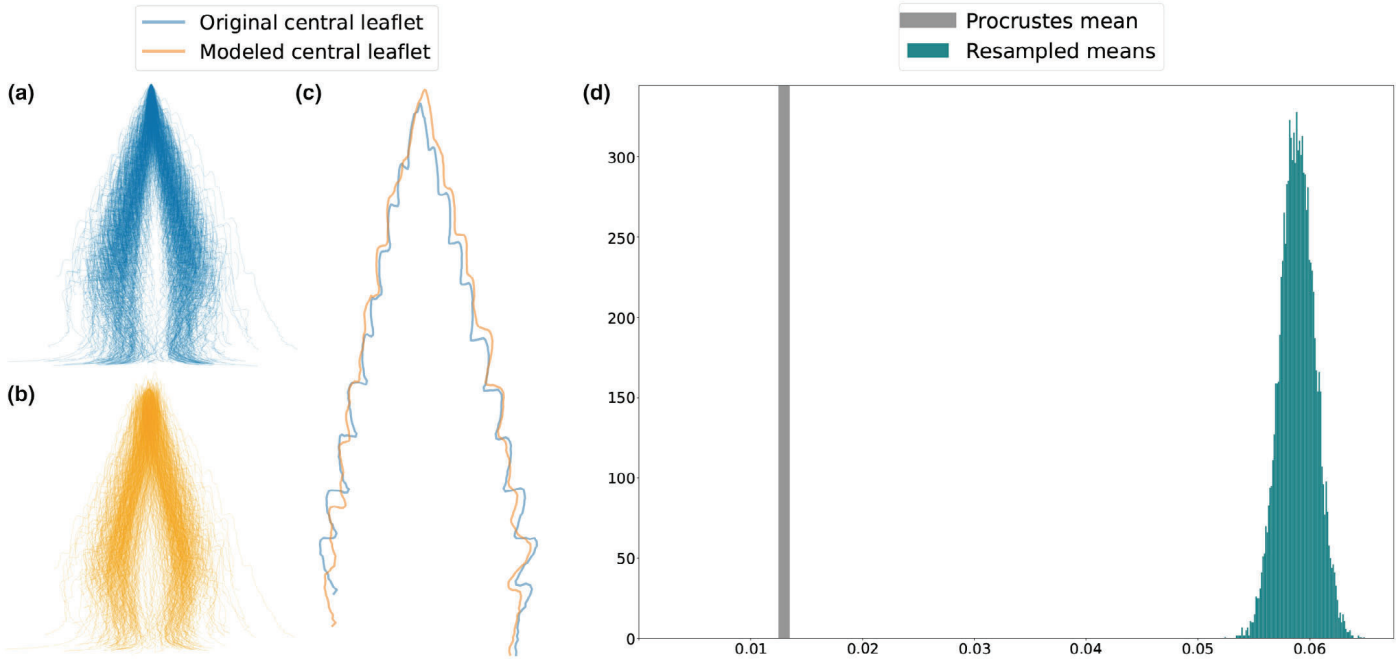


Fig. 3 Modeling approach validation using Procrustes analysis and bootstrap resampling. The (a) original and (b) modeled central leaflets in polar coordinate system were superimposed (c) and Procrustes distances calculated. (d) The resampled mean was plotted as a distribution (green histogram) against the actual Procrustes mean (gray vertical line).

the landmark of the petiolar junction and tip of the central leaflet (Fig. 2, third column).

Using the polar coordinates of each leaflet, second-degree polynomial models for x (angle) and y (radius from petiolar junction) values were fit through each of the 399 corresponding pseudo-landmarks for each leaflet using the PYTHON *scipy.optimize.curve_fit* function (Virtanen *et al.*, 2020), modeling angle and radius as a function of leaflet number (Fig. 2, fourth column). Using the coefficients for second-degree polynomial models, we then model each pseudo-landmark as a function of leaflet number to reconstruct the new theoretical leaf with an arbitrary number of leaflets. Meaning that for each leaflet, each of the 399 x and y pseudo-landmarks (i.e. angle and radius from petiolar junction coordinates) was calculated using the second-degree polynomial function, with coefficients obtained from the previous step, and the newly defined leaflet number (9 in this case). The optimal number of reconstructed leaflets was tested for the best prediction accuracy in LDA modeling, and the highest accuracy was achieved by reconstructing nine leaflets (Supporting Information Table S1). It is important to note that the reconstructions start with the first real leaflet and end with the last real leaflet. These nine reconstructed leaflets are then equally divided between these two points.

Nine leaflets were reconstructed using the collection of coefficients of 798 second-degree polynomial models for each leaf; the 399 models for angle were used to model theoretical x (i.e. angle) and 399 models for radius were used to model theoretical y (i.e. radius from petiolar junction) pseudo-landmarks as a function of nine leaflets.

The coordinates defining the 3591 pseudo-landmarks for each of the modeled leaves (399 pseudo-landmarks for each of the nine reconstructed leaflets) were then plotted and visually inspected. We detected 17 inaccurately modeled leaves, most likely caused by the position of the petiole landmark compared with the landmark marking the start and end landmarks of the leaflet. A total of 341 *Cannabis* leaves were then used in the analysis.

Validation of the leaf modeling approach

To validate our modeling approach, we extracted the polar coordinates of the original central leaflets (Fig. 3a) and central leaflets of the modeled leaves (Fig. 3b) and used them in Procrustes analysis using *Procrustes* function from scipy.spatial module (Virtanen *et al.*, 2020). Procrustes analysis minimizes the distance between all points for a set of landmarks/pseudo-landmarks between two samples through translation, rotation, and scaling, and returns new points of the two sets, superimposed to each other (Fig. 3c). We then calculated the Procrustes distance between the original central leaflet (angle and radius coordinates) to its corresponding modeled reconstruction, a measure of their similarity. The mean distance was calculated and compared with that of simulated bootstrapped mean values by resampling (10 000 resamples) through randomly sorting original leaflet coordinates against coordinates of reconstructed leaflets.

Morphometric analysis of the central leaflet shape using previously established methodologies

The width : length ratio (W : L ratio), first described by Anderson (1980), was frequently used to describe the shape of *Cannabis* leaves or even differentiate between different *Cannabis* taxa. With

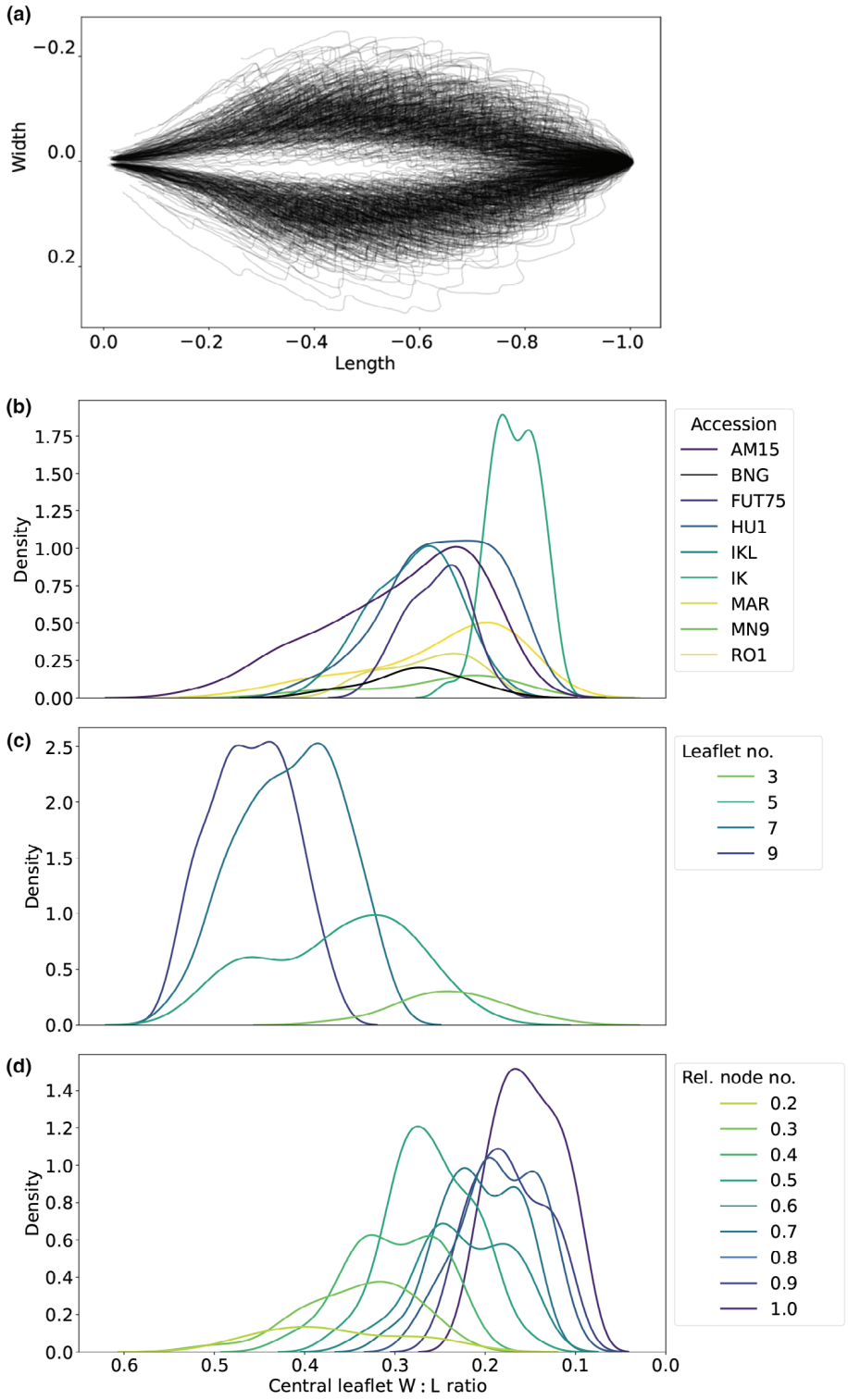


Fig. 4 Analysis of *Cannabis* leaf shape using the approach adapted from Anderson (1980). (a) Visualization of the 341 central leaflets used in the analysis. Width : length ratios plotted by (b) accession, (c) leaflet number, and (d) relative node number.

previously established morphometric methods, the shape analysis of central leaflets (that all leaves share) would also be possible, using EFDs or pseudo-landmark approach. To evaluate the effectiveness of these two previous methods for the shape analysis of *Cannabis* leaves, we first extracted the Cartesian coordinates of central leaflets (Fig. 4a), which were previously scaled, rotated, and translated so that they were all pointing in the same direction and had the length of one. We then interpolated 200 pseudo-landmarks on each side of each leaflet, sharing the landmark on the tip of the leaflet (i.e. a total of 399 pseudo-landmarks per leaflet).

To measure the W : L ratio, we calculated width of the leaf (as the leaves were already normalized to length of one), calculating the minimum bounding rectangle. The distribution of widths

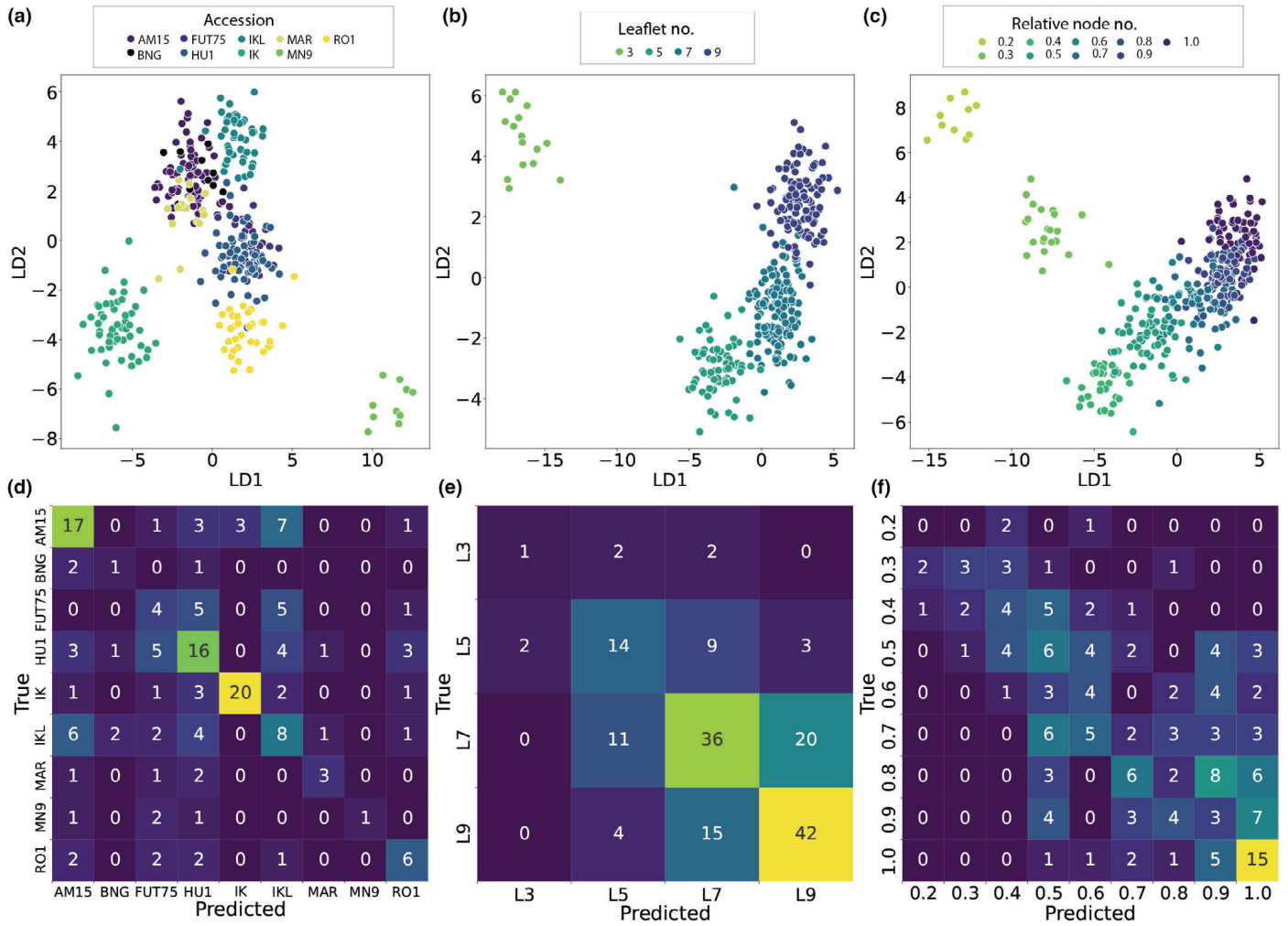


Fig. 5 Accession, leaflet number, and relative node numbers prediction of *Cannabis* leaves using the outline of central leaflets. Linear discriminant analysis (LDA) plots for (a) accession, (b) leaflet number, and (c) relative node number. In the lower row, the confusion matrices show the true and predicted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

was then plotted using PYTHON package *seaborn.kdeplot*. To see whether the analyzed accessions differed significantly in their W : L ratios, Kruskal–Wallis test was calculated using *stats.kruskal* function from the scipy.stats module. To see which of the accessions differ in W : L ratio, we calculated Dunn's multiple comparison test with SCIKIT_POSTHOCS package in PYTHON (Terpilowski, 2019), using the *posthoc_dunn* function.

Linear discriminant analysis was applied to model accession, leaflet number, and relative node number as the function of central leaflet coordinate values, using the *LinearDiscriminantAnalysis* function from the scikit-learn module in PYTHON (Pedregosa *et al.*, 2011). To test the performance of the LDA model, the dataset was divided into two parts. Since most of the analyzed leaves exhibit opposite phyllotaxy, wherein the nodes were represented by two leaves (a and b) in the same developmental phase with the same number of leaflets, the dataset was split into a training dataset (leaf a) comprising 180 leaves and a test dataset (leaf b) containing 161 leaves. The *predict* function from *LinearDiscriminantAnalysis* in the scikit-learn module was used to predict the accession identity, leaflet number, and relative node number, based on the central leaflet coordinate values. The accuracy of the LDA model was calculated and visualized using the function *confusion_matrix* from scikit-learn. Spearman's rank correlation was calculated for true and predicted results for relative node number with *spearmanr* function from the scipy.stats module.

Data analysis of modeled leaves

A principal component analysis (PCA) was performed on the coordinates of the modeled leaves using scikit-learn module in PYTHON and proportions of explained variance for each principal component and the cumulative variance was calculated. Points representing the leaves were colored by the accession identity, leaflet number, or relative node number (Fig. 6). To see which of the first two PCs explains most of the leaf shape variation for accessions, leaflet number, and relative node number, Kruskal–Wallis test was calculated using *stats.kruskal* function from the

Table 2 Predictive power of genetic and developmental identities using the LDA model on the central leaflet shape of *Cannabis* leaves.

	Correct prediction (*n*)	False prediction (*n*)	Prediction accuracy (%)	Correlation coefficient (rho)	*P* value
Accession	76	85	47.20	NA	NA
Leaflet number	93	68	57.76	NA	NA
Relative node number	NA	NA	NA	0.629	< 0.0001

NA, data not analyzed.

scipy.stats module. To visualize an average leaf for each accession, leaflet number, and relative node number, the average coordinate values of modeled leaves were calculated for each of the categories and plotted using the Matplotlib module in Python (Fig. 6).

To see whether the modeled leaves can be used to model accession, leaflet number, and relative node number, we followed the same steps as before for shape analysis of central leaflet. Linear discriminant analysis was applied to model accession, leaflet number, and relative node number. The dataset was again split into a training and test dataset to see whether we were able to predict accession, leaflet number, and relative node number identity, based on the coordinates of modeled leaves. The same was done on a combined dataset with 3990 coordinates, created by concatenating coordinates of modeled leaves and the coordinates of the original central leaflets.

Results

Heteroblastic changes in leaflet number along the main axis

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

Validation of the leaf modeling approach

The modeling approach was validated by calculating the mean Procrustes distance of modeled central leaflet coordinates to original central leaflet coordinates using 10 000 bootstrap replicas, assessing resampled means against the actual Procrustes mean value. None of the 10 000 resamples yielded a mean lower than the observed Procrustes value, confirming the robustness of the novel modeling approach (Fig. 3d).

Width : length ratio and central leaflet shape analysis

Our results indicate that the W : L ratio of central leaflets is not able to differentiate well between different *Cannabis* leaf accessions based on this information alone (Fig. 4). While the Kruskal–Wallis test did show overall significance between accessions (Table S2), Dunn's *post hoc* test indicated significance in leaf morphology for just one accession (Table S3). The W : L ratio significantly differs from the rest only for the IK accession, characterized by particularly narrow leaves (Table S3). The Kruskal–Wallis test was also significant for leaflet numbers and relative node numbers (Table S2). Dunn's *post hoc* test revealed that while we can differentiate between leaflet numbers based on the W : L ratio of central leaflet, we can only separate the lower and higher relative nodes (Table S3).

To test whether the outline of the central leaflet can better predict the genetic and developmental identity of *Cannabis* leaves, we used LDA to model each factor as a function of 399 pseudo-landmark points defining the shape of central leaflet (Fig. 5a–c). To evaluate model accuracy, accession was treated as a categorical variable, as was leaflet number, as it not only has a small number of levels (3, 5, 7, and 9 leaflets) but each level is well-separated from the others. To evaluate the accuracy of relative node number, we treated it as a continuous variable, due to a high number of levels (9) that continuously overlap with each other. Models revealed low accuracy, as the accession was correctly determined only in 47.20% (Table 2). The LDA model for the shape of central leaflet showed no overlap for the accessions IK and MN9, but the remaining accessions showed significant overlap (Fig. 5a). The confusion matrix revealed that only two accessions were correctly identified more than half the time (AM15 – 53.13% and IK – 71.43% prediction accuracy) (Fig. 5d). The LDA model showed better success when identifying the leaflet number (57.76% overall accuracy) and relative node number, where the true and predicted values show significant, but moderate correlation (rho = 0.629, $P < 0.0001$) (Fig. 5b,c,e,f; Table 2).

Principal component analysis on modeled leaves

Using the outline and landmark coordinates of 341 leaves, we modeled new theoretical leaves, all with nine leaflets. Each leaf is defined by 3591 pseudo-landmarks, which overcomes the problems associated with variable leaflet numbers and permits dimension reduction using PCA (Fig. 6a–c) and the visualization of average *Cannabis* leaves (Fig. 6d–f). The first and second PCs account for 85.85% and 7.25% of the shape variation, respectively (Fig. 6a–c). Examining the PC1 and PC2 with Kruskal–Wallis test reveals that accession, leaflet number, and relative node number all vary significantly along the first PC axis. The variation along the PC2 for accession and leaflet number is less pronounced, however still significant, while PC2 values for

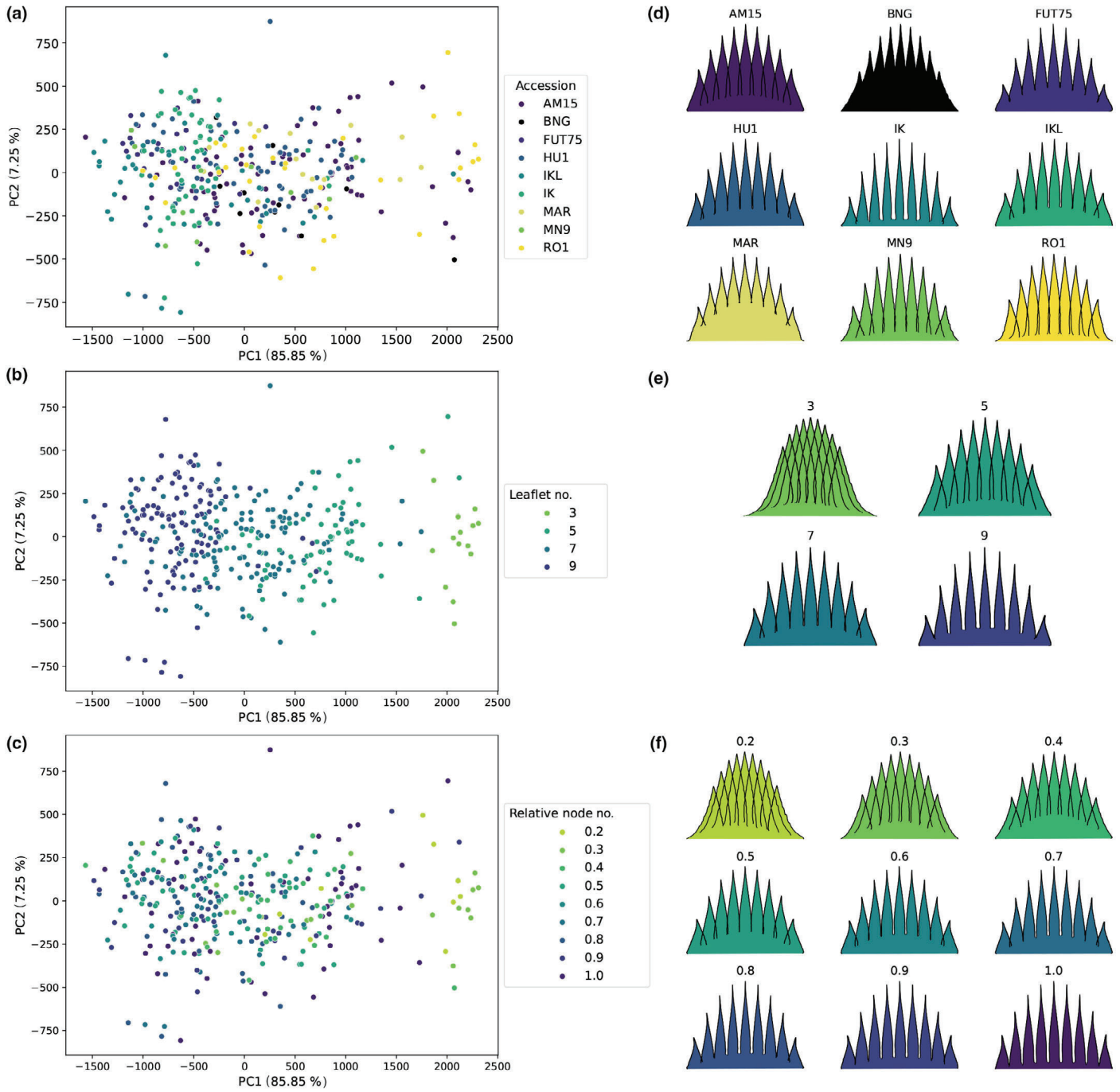


Fig. 6 Principal component analysis (PCA) of the *Cannabis* accessions performed on modeled leaves using the 3591 pseudo-landmarks (a–c). The first PC explains 85.85% and the second 7.25% of variation. The images on the right show the average modeled leaf shapes for each of the (d) nine analyzed accessions, (e) leaflet number, and (f) relative node number.

relative node numbers do not vary significantly (Fig. 6; Table 3). This indicates that the changes in leaf shape between accessions are not independent from developmental variation. That is, a facet of variation in accession leaf shape covaries with developmental variation across the shoot in leaflet, and relative node number suggests a heterochronic mechanism by which accession differences in leaf shape arise from changes in developmental timing and contrasts with the historical focus on changes in timing arising from plasticity (Goebel, 1908; Ashby, 1948).

The average modeled leaf shapes show that the most pronounced change in leaf shape between the accessions and during the development corresponds to narrow vs wide leaflets that are stereotypical descriptions of *sativa* vs *indica* or *wide-* vs *narrow-leaflet* drug varieties. Furthermore, the leaves with the lower

Table 3 Kruskal–Wallis test was used to test the *Cannabis* leaf shape variation along PC1 and PC2 for accessions, leaflet number, and relative node number.

	PC1		PC2	
	H	*P* value	*H*	*P* value
Accession	112.64	< **0.0001**	18.57	< **0.05**
Leaflet number	204.36	< **0.0001**	10.75	< **0.05**
Relative node number	49.73	< **0.0001**	2.98	> 0.05

The bold font indicates the values that are statistically significant.

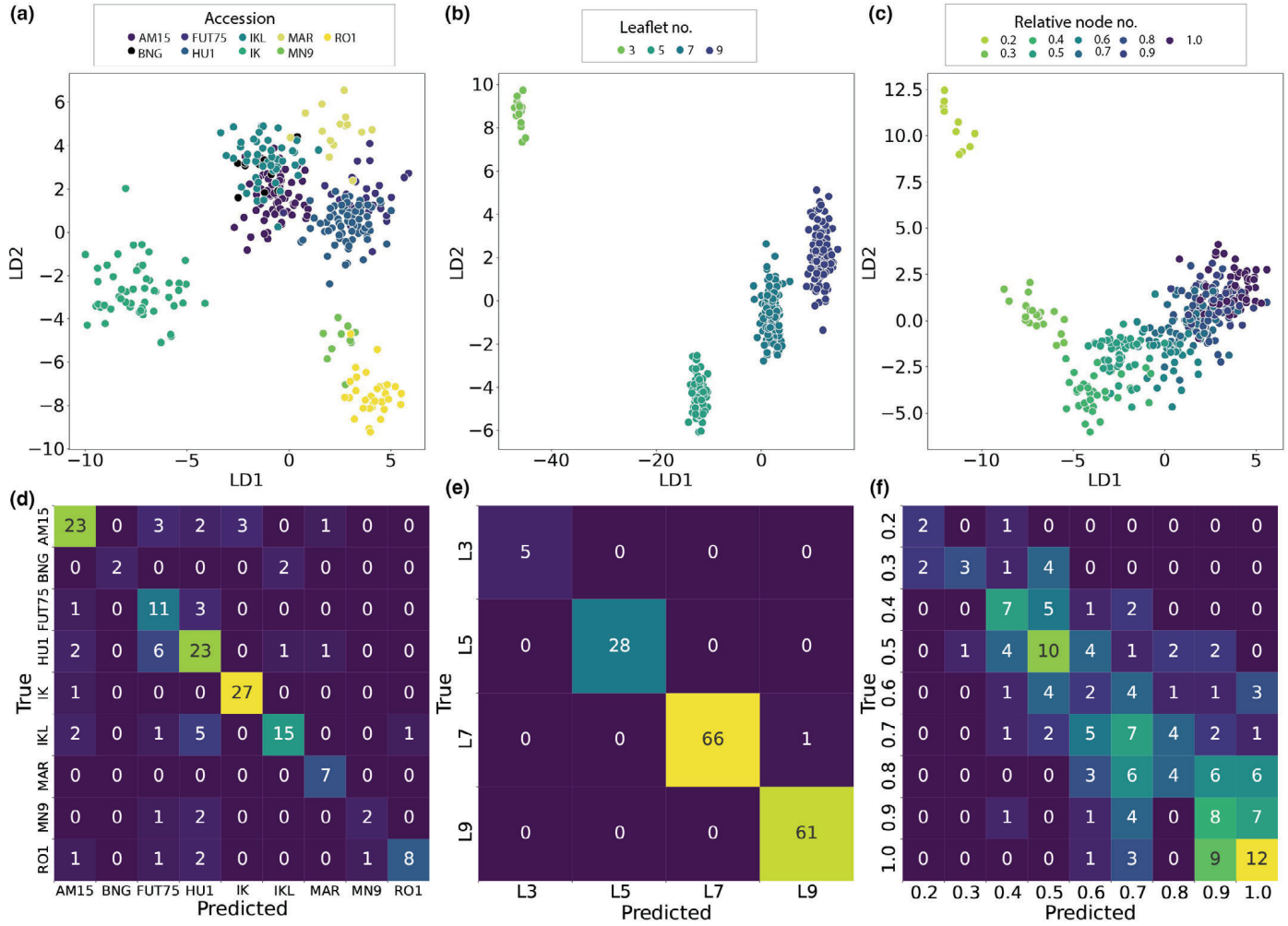


Fig. 7 Accession, leaflet number, and relative node numbers of *Cannabis* leaves can be predicted independently of each other using modeled leaves. Linear discriminant analysis (LDA) plots for (a) accession, (b) leaflet number, and (c) relative node number. In the lower row, the confusion matrices show 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.

number of leaflets have more acute leaflet tips, which slowly transition into acuminate. Additionally, the outer leaflets in the leaves from lower nodes (and in certain accessions) are longer, than the central leaflet, and become shorter higher up (Fig. 6d,e).

Linear discriminant analysis and prediction of genetic and developmental identities on modeled leaves

As in the analysis of central leaflet shape before, we used LDA to model accession, leaflet number, and relative node number as a function of all 3591 pseudo-landmark points defining the complete modeled leaves (Fig. 7). Accuracy of the model was calculated on the split dataset, treating accession and leaflet number as categorical and relative node number as continuous variable. Linear discriminant analysis models for both accession and leaflet number were highly accurate (73.29% and 99.38%, respectively) (Table 4), significantly improving the results obtained by analyzing solely the outline of the central leaflet (Table 2). The model for relative node number is highly accurate as well, as inferred by a highly significant Spearman's rank correlation coefficient value

New Phytologist

Table 4 Predictive power of genetic and developmental identities using the LDA model on the modeled *Cannabis* leaves.

	Correct prediction (*n*)	False prediction (*n*)	Prediction accuracy (%)	Correlation coefficient (rho)	*P* value
Accession	118	43	73.29	NA	NA
Leaflet number	160	1	99.38	NA	NA
Relative node number	NA	NA	NA	0.747	< 0.0001

NA, data not analyzed.

Table 5 Predictive power of genetic and developmental identities using the LDA model on a combined dataset created by concatenating coordinates of modeled *Cannabis* leaves and the coordinates of the original central leaflets.

	Correct prediction (*n*)	False prediction (*n*)	Prediction accuracy (%)	Correlation coefficient (rho)	*P* value
Accession	115	46	71.43	NA	NA
Leaflet number	161	0	100	NA	NA
Relative node number	NA	NA	NA	0.787	< 0.0001

NA, data not analyzed.

between actual and predicted values (rho = 0.747, $P < 0.0001$) (Table 4).

A confusion matrix reveals that the LDA model in most cases had a high accuracy for predicting accession identity (Fig. 7d; Table 4), much higher, as compared to the accuracy achieved by using only the outline of the central leaflet (Fig. 5d; Table 2). Accessions IK, RO1, and MN9 show practically no overlap in LDA space, while AM15, BNG, FUT75, HU1, IKL, and MAR show more overlap (Fig. 7a). The model showed an almost 100% success rate in determining leaflet number, again, much higher than before.

Results of both methods revealed that leaves with only three leaflets are markedly different from the rest, and the prediction model on theoretical leaves consistently classified them correctly (Fig. 7e). Leaves with five to nine leaflets showed less pronounced differences in shape, resulting in a slightly lower accuracy of the prediction model for these cases. However, an examination of the confusion matrix revealed that misclassifications only occurred once between leaves with neighboring leaflet numbers (7 and 9 leaflets) (Fig. 7e). The marked difference in shape of leaves with three leaflets from the rest may suggest that this developmental mechanism is biased toward variation at the base of the shoot. Similar to leaflet number, the confusion matrix for the relative node model reveals high rates of misclassification between the neighboring relative node numbers, as is expected, and leaves from lower nodes were very rarely classified as those from higher nodes (Fig. 5f). A pronounced change in leaf shape occurs between the relative nodes 0.3 and 0.4, while the shape changes in later relative nodes are more gradual (Fig. 7c).

Compared with only using the modeled leaves, the accuracy of the LDA model did not improve significantly when using a combined dataset. A confusion matrix revealed that the LDA model (Fig. S1) was slightly less successful in accession identity classification (71.43%) but was higher for leaflet number (100%). The Spearman's rank correlation coefficient was slightly higher and highly significant (rho = 0.787, $P < 0.0001$) (Table 5).

Discussion

Like grapevines, striking variation in leaf shape (Fig. 1) has historically played a significant role in taxonomic classification of *Cannabis*. Leaf shape and differences in phyllotaxy were among the characters Lamarck used to describe a new *Cannabis* species (Lamarck & Poiret, 1783). Anderson (1980) introduced a quantitative approach by quantifying the length : width ratio of the central leaflet. Further studies using different characters – including plant height, stem diameter, achene shape, and phytochemical profiles – to characterize accessions have only confirmed the importance of leaf characteristics (Small *et al.*, 1976; Hillig, 2005a). The central leaflet W : L ratio has been adopted by researchers as one of the main characters for determining species, subspecies, biotypes, and chemotypes of *Cannabis* (Hillig, 2005a; Clarke & Merlin, 2013; McPartland & Small, 2020). However, this method is only able to capture a limited aspect of leaf shape variation, neglecting other important characteristics that we measure in this study, such as leaflet outlines, serrations, angles, and relative changes in leaflet shape across the leaf. By modeling leaflet shape as a function of leaflet number, we model theoretical leaves with the same number of leaflets for which high densities of corresponding pseudo-landmarks capture high-resolution shape features (Fig. 2). To validate the modeling approach, we have compared the outline of the original central leaflet and the outline of the modeled theoretical central leaflet. The Procrustes analysis showed that the two leaflets are very similar in shape and that the modeling is even able to preserve the serration pattern to some degree (Fig. 3c). The modeling approach validated using 10 000 bootstrap replicas confirmed the robustness of the novel modeling approach (Fig. 3d). This method can be applied not only on palmately composed leaves as in *Cannabis* but also on pinnate and lobed leaves. To demonstrate the proof of concept, we applied the method to a pinnate leaf of *Cardamine flexuosa* With. and lobate leaf of *Quercus macrocarpa* Michx. (Fig. 8), showing the method could be applied in other leaf types. However, the method needs to be improved

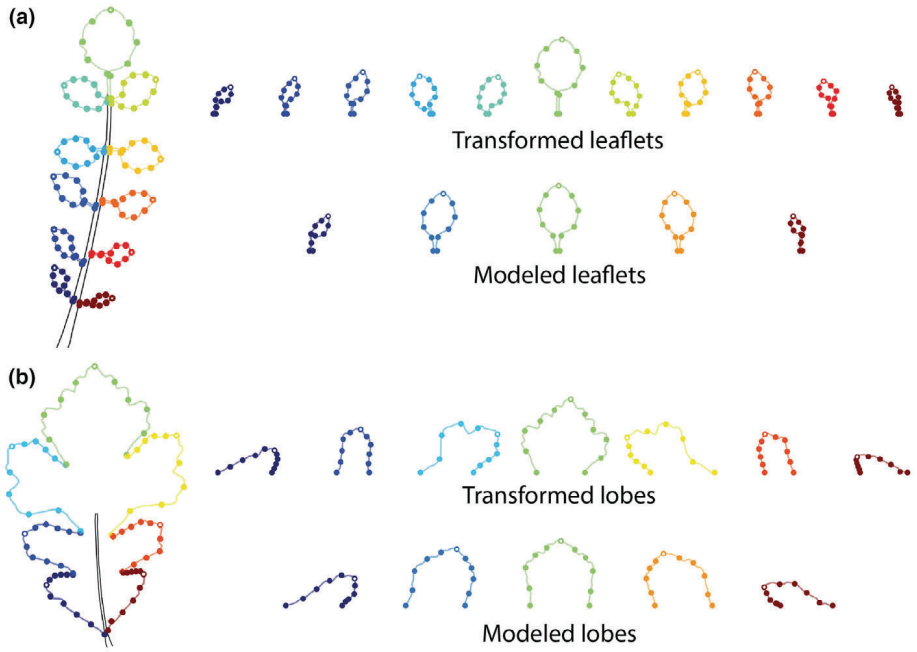


Fig. 8 Intra-leaf modeling of leaflets and lobes extended to pinnate leaves: Leaves from (a) *Cardamine flexuosa* and (b) *Quercus macrocarpa*. Leaflets and lobes are defined by 100 equidistant pseudo-landmarks on each side, each defined by three landmarks, two at the base and one at the tip. Large points are placed every 20 pseudo-landmarks to emphasize that leaflet outlines are defined by points. The landmarks defining the base of each leaflet or lobe are aligned to the rachis or midvein and the transformed leaflets and lobes have been oriented parallel to the rachis, as defined by the landmarks at their base. The modeled leaflets and lobes are created from second-degree polynomial models for each *x* and *y* coordinate value for each pseudo-landmark as a function of leaflet or lobe number. From these models, an equivalent number of modeled leaflets or lobes can be reconstructed (in this case, five), permitting morphometric analysis.

before being applied to other species but shows the possible utility of intra-leaf modeling.

The method presented in this study can accurately determine accession based on leaf shape, regardless of its developmental stage (Fig. 7a,d). The method works effectively not only on stabilized or cloned cultivar accessions but also on wild or feral accessions cultivated from seed that can exhibit distinct plant phenotypes (Table 1), indicating its robustness and potential value in future germplasm classification. Compared with the low accuracy and prediction ability of the previously known methods (W : L ratio and shape analysis of central leaflets), the newly proposed method demonstrates significantly improved results (Tables 2, 4, S2, S3). The combined dataset of both, data for modeled leaves and outline of the central leaflet, did not return significantly better results, further confirming the effectiveness of the new modeling approach (Table 5).

When observing the shape changes between averaged leaves for accessions and between developmental stages, the most obvious are changes in leaflet widths, similar to stereotypical classifications of *sativa* and *indica* plants or *wide-* vs *narrow-leaflet* drug varieties. However, other important changes in shape occur, such as transition from acute to acuminate leaflet tip and changes in the relative length of outer most leaflets compared with the central leaflet that previous methods could not successfully capture (Fig. 6d–f). The reliance on the non-quantitative leaf shape descriptors in previous methods has led to numerous cultivars with unreliable names, inconsistent genetic origins, and phytochemical profiles (Sawler *et al.*, 2015; Schwabe & McGlaughlin, 2019; Jin *et al.*, 2021a; Watts *et al.*, 2021). For example, Jin *et al.* (2021b) conducted a study on clones of 21 cultivars and found a strong negative correlation between the width and length ratios of central leaflets and CBD, and a positive correlation with THC; however, Vergara *et al.* (2021) and Murovec *et al.* (2022) were unable to confirm these findings. All three studies used low-resolution morphometric approaches. Sex of the plants also plays a crucial role in the cannabis industry, where the presence of male plants and inevitable pollination leads to decreases in cannabinoid production as plants shift the use of energy into seed development. Several methods have been employed to differentiate between male and female plants at early stages, but only genetic methods were successful so far (Prentout *et al.*, 2020; Toth *et al.*, 2020; Campbell *et al.*, 2021; Balant *et al.*, 2022; Torres *et al.*, 2022). Our results quantify the variation in leaf shape between accessions that can potentially be used to classify accessions and predict chemical profiles and plant sex faster and more accurately.

Unlike grapevine, where developmental variance is orthogonal and separate from genetic variance, in *Cannabis*, these two factors are correlated. That is, the developmental source of variation is colinear with accession identity suggests that part of the differences between accession leaf shape is explained by shifts in developmental timing, or heterochrony.

Cannabis plants demonstrate extreme phenotypic plasticity depending on the environmental conditions in which they grows (Small, 2015). Some *Cannabis* accessions are photoperiod-dependent and can remain in vegetative phase for longer periods of time under long-day conditions (typically 18 h : 6 h, darkness : light), until the transition to short-day (12 h : 12 h, darkness : light) induces the formation of the apical inflorescence. Previous investigations showed that other morphological changes, such as decrease in leaf area, number of leaflets per leaf, and serration number, occur after the change in the environmental conditions one or two nodes after (Heslop-Harrison & Heslop-Harrison, 1958; Hesami *et al.*, 2023). However, differences, especially in flowering time and growth rates between cultivars, have been observed before (de Meijer & Keizer, 1996; Hillig, 2005a; Spitzer-Rimon *et al.*, 2019; Carlson *et al.*, 2021;

Naim-Feil *et al.*, 2021; Stack *et al.*, 2021; Chen *et al.*, 2022) and differences in cannabinoid profiles, leaflet index, and phenological development were proposed as characteristics to discriminate between them (de Meijer & Keizer, 1996). Heterochronic shifts are apparent in the differential rates in which accessions increase leaflet number across nodes, as well as maximum and average leaflet counts across accessions (Fig. 1). Remarkably, stages in developmental timing are conserved despite being shifted. For example, a significant shape change exhibited between the leaves with three and leaves with five leaflets, with leaflets becoming more acuminate and narrower. By contrast, changes in shape between leaves with a higher number of leaflets were more gradual. Additionally, we observed a similar shift in leaf shape between the Nodes 0.3 and 0.4, potentially indicating a transition between the juvenile and adult phases of leaf development. Similar results were obtained in previous research. Spitzer-Rimon *et al.* (2022) demonstrated that flowering buds were initiated at Node 7, while Moliterni *et al.* (2004) analyzing a different cultivar found developing flower buds in the fourth node, suggesting that transitions in growth phases are conserved but not synchronized across cultivars. Due to the differences in developmental timing between accessions, the use of continuous models along the shoot could further improve the success predicting accession identity, as was the case in grapevine (Bryson *et al.*, 2020).

Conclusions

In grapevine, leaf shape has long been utilized for variety identification. However, in the case of *Cannabis*, previous attempts were hindered by the variability in leaflet numbers. In this study, we present a pioneering method that successfully addresses this issue. By generating theoretical leaves with customizable leaflet counts, we can now employ high-resolution morphometric techniques to accurately classify different wild/feral and cultivated *Cannabis* accessions. Through the use of 3591 densely placed pseudo-landmarks, we were able to predict the accession identity with almost 74% accuracy. The method works well not only on stabilized cultivars but also on phenotypically more variable wild/feral accessions grown from seed. Unifying the number of leaflets allowed us, for the first time, to make comparisons among several leaves along the main axis, enabling us to investigate developmental changes in leaf shape and detect heterochronic mechanisms influencing the leaf shape in *Cannabis*. The implications of this new high-resolution method in both the cannabis industry and research extend beyond its role in determining *Cannabis* accessions. It also offers a promising tool for developmental studies, and for studying the correlation between leaf shape and phytochemical profiles and the sex of the plants, where lower 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 palmate, pinnate, and lobate leaves with varying numbers of lobes and leaflets where the use of geometric morphometrics methods was not previously possible to this extent.

Acknowledgements

This research was supported by projects WECANN (CGL2017-84297-R, Ministerio de Ciencia, Innovación y Universidades), Generalitat de Catalunya (grant no. 2021SGR00315), 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, Nusrat Sultana, 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 wrote the first draft of the paper that all authors read, commented and edited. MB analyzed the data.

ORCID

Manica Balant https://orcid.org/0000-0002-4316-8511
Daniel H. Chitwood https://orcid.org/0000-0003-4875-1447
Teresa Garnatje https://orcid.org/0000-0001-6295-6217
Oriane Hidalgo https://orcid.org/0000-0002-1547-8627
Daniel Vitales https://orcid.org/0000-0002-3109-5028

Data availability

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

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.

Ashby E. 1948. Studies in the morphogenesis of leaves. I. An essay on leaf shape. *New Phytologist* **47**: 153–176.

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.

Biot E, Cortizo M, Burguet J, Kiss A, Oughou M, Maugarny-Calès A, Gonçalves B, Adroher B, Andrey P, Boudaoud A *et al.* 2016. Multiscale quantification of morphodynamics: MorphoLeaf, software for 2-D shape analysis. *Development* **143**: 3417–3428.

Bookstein FL. 1997. *Morphometric tools for landmark data: geometry and biology.* Cambrige, UK: Cambridge University Press.

Buzna C, Sala F. 2022. Non-destructive method to determining the leaf area in hemp, *Cannabis sativa* L. *Life Science and Sustainable Development* **3**: 25–32.

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**: e11404.

Campbell LG, Peach K, Wizenberg SB. 2021. Dioecious hemp (*Cannabis sativa* L.) plants do not express significant sexually dimorphic morphology in the seedling stage. *Scientific Reports* **11**: 16825.

Carlson CH, Stack GM, Jiang Y, Taşklran B, Cala AR, Toth JA, Philippe G, Rose JKC, Smart CD, Smart LB. 2021. Morphometric relationships and their contribution to biomass and cannabinoid yield in hybrids of hemp (*Cannabis sativa*). *Journal of Experimental Botany* **72**: 7694–7709.

Chen X, Guo H, Zhang Q, Wang L, Guo R, Zhan Y, Lv P, Xu Y, Guo M, Zhang Y *et al.* 2022. Whole-genome resequencing of wild and cultivated cannabis reveals the genetic structure and adaptive selection of important traits. *BMC Plant Biology* **22**: 371.

Chitwood DH. 2021. The shapes of wine and table grape leaves: an ampelometric study inspired by the methods of Pierre Galet. *Plants, People, Planet* **3**: 155–170.

Chitwood DH, Klein LL, O'Hanlon R, Chacko S, Greg M, Kitchen C, Miller AJ, Londo JP. 2016a. Latent developmental and evolutionary shapes embedded within the grapevine leaf. *New Phytologist* **210**: 343–355.

Chitwood DH, Mullins J, Migicovsky Z, Frank M, VanBuren R, Londo JP. 2021. Vein-to-blade ratio is an allometric indicator of leaf size and plasticity. *American Journal of Botany* **108**: 571–579.

Chitwood DH, Ranjan A, Martinez CC, Headland LR, Thiem T, Kumar R, Covington MF, Hatcher T, Naylor DT, Zimmerman S *et al.* 2014. A modern ampelography: a genetic basis for leaf shape and venation patterning in grape. *Plant Physiology* **164**: 259–272.

Chitwood DH, Rundell SM, Li DY, Woodford QL, Yu TT, Lopez JR, Greenblatt D, Kang J, Londo JP. 2016b. Climate and developmental plasticity: interannual plasticity in grapevine leaf morphology. *Plant Physiology* **170**: 1480–1491.

Chitwood DH, Sinha NR. 2016. Evolutionary and environmental forces sculpting leaf development. *Current Biology* **26**: R297–R306.

Clarke R, Merlin M. 2013. *Cannabis: evolution and ethnobotany.* Berkley, CA, USA; Los Angeles, CA, USA; London, UK: University of California Press.

Demmings EM, Williams BR, Lee C-R, Barba P, Yang S, Hwang C-F, Reisch BI, Chitwood DH, Londo JP. 2019. Quantitative trait locus analysis of leaf morphology indicates conserved shape loci in grapevine. *Frontiers in Plant Science* **10**: 1373.

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

Failmezger H, Lempe J, Khadem N, Cartolano M, Tsiantis M, Tresch A. 2018. MowJoe: a method for automated-high throughput dissected leaf phenotyping. *Plant Methods* **14**: 27.

Galet P. 1952. *Précis d'ampélographie pratique.* Imprimerie P. Déhan. Montpellier, France: Lavoisier.

Galet P, trans. Morton L. 1979. *A practical ampelography: grapevine identification.* Ithaca, NY, USA: Cornell University Press (English translation).

Goebel K. 1908. *Einleitung in die experimentelle Morphologie der Pflanzen.* Leipzig, Germany: B.G. Teubner.

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.

Hesami M, Pepe M, Jones AMP. 2023. Morphological characterization of *Cannabis sativa* L. throughout its complete life cycle. *Plants* **12**: 3646.

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. Bloomington, IN, USA: 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.

Islam MJ, Ryu BR, Azad MOK, Rahman MH, Rana MS, Kang C-W, Lim J-D, Lim Y-S. 2021. Comparative growth, photosynthetic pigments, and osmolytes analysis of hemp (*Cannabis sativa* L.) seedlings under an aeroponics system with different LED light sources. *Horticulturae* **7**: 239.

Jin D, Henry P, Shan J, Chen J. 2021a. Classification of cannabis strains in the Canadian market with discriminant analysis of principal components using genome-wide single nucleotide polymorphisms. *PLoS ONE* **16**: e0253387.

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

Karlov GI, Razumova OV, Alexandrov OS, Divashuk MG, Kroupin PY. 2017. Classical and molecular cytogenetics of *Cannabis sativa* L. In: Chandra S, Lata H, ElSohly M, eds. *Cannabis sativa L. - Botany and biotechnology.* Cham, Switzerland: Springer, 385–394.

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

Lamarck J-B, Poiret J-L-M. 1783. *Encyclopédie méthodique. Botanique.* Paris, France: Panckoucke; Liège, Belgium: Plomteux.

Lapierre É, Monthony AS, Torkamaneh D. 2023. Genomics-based taxonomy to clarify *Cannabis* classification. *Genome* **66**: 202–211.

Lynch RC, Vergara D, Tittes S, White K, Schwartz CJ, Gibbs MJ, Ruthenburg TC, deCesare K, Land DP, Kane NC. 2016. Genomic and chemical diversity in *Cannabis*. *Critical Reviews in Plant Sciences* **35**: 349–363.

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. *PhytoKeys* **144**: 81–112.

de Meijer EPM, van der Kamp HJ, van Eeuwijk FA. 1992. Characterisation of *Cannabis* accessions with regard to cannabinoid content in relation to other plant characters. *Euphytica* **62**: 187–200.

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.

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**: 294.

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

Parsons JL, Martin SL, James T, Golenia G, Boudko EA, Hepworth SR. 2019. Polyploidization for the genetic improvement of *Cannabis sativa*. *Frontiers in Plant Science* **10**: 476.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V *et al.* 2011. scikit-learn: machine learning in Python. *Journal of Machine Learning Research* **12**: 2825–2830.

Potter DJ. 2009. *Optimisation of* Cannabis sativa *L. as a phytopharmaceutical*. PhD thesis, Department of Pharmaceutical Science Research, King's College London, London, UK.

Prentout D, Razumova O, Rhoné B, Badouin H, Henri H, Feng C, Käfer J, Karlov G, Marais GAB. 2020. An efficient RNA-seq-based segregation analysis identifies the sex chromosomes of *Cannabis sativa*. *Genome Research* **30**: 164–172.

Prohibition Partners. 2022. *The global Cannabis report, 3rd edn*. London, UK; Berlin, Germany; Barcelona, Spain: Prohibition Partners.

Quimby MW, Doorenbos NJ, Turner CE, Masoud A. 1973. Mississippi-Grown Marihuana – *Cannabis sativa* cultivation and observed morphological variations. *Economic Botany* **27**: 117–127.

Ravaz L. 1902. *Les vignes américaines: porte-greffes et producteurs directs*. Goulet, Montpellier, Paris, France: Digitized by Google Books from Cornell University.

Ren G, Zhang X, Li Y, Ridout K, Serrano-Serrano ML, Yang Y, Liu A, Ravikanth G, Nawaz MA, Mumtaz AS *et al.* 2021. Large-scale whole-genome resequencing unravels the domestication history of *Cannabis sativa*. *Science Advances* **7**: eabg2286.

Sawler J, Stout JM, Gardner KM, Hudson D, Vidmar J, Butler L, Page JE, Myles S. 2015. The genetic structure of marijuana and hemp. *PLoS ONE* **10**: e0133292.

Schultes RE, Klein WM, Plowman T, Lockwood TE. 1974. *Cannabis*: an example of taxonomic neglect. *Botanical Museum Leaflets, Harvard University* **23**: 337–367.

Schwabe AL, McGlaughlin ME. 2019. Genetic tools weed out misconceptions of strain reliability in *Cannabis sativa*: implications for a budding industry. *Journal of Cannabis Research* **1**: 3.

Small E. 2015. Evolution and classification of *Cannabis sativa* (Marijuana, Hemp) in relation to human utilization. *Botanical Review* **81**: 189–294.

Small E, Beckstead HD. 1973. Cannabinoid phenotypes in *Cannabis sativa*. *Nature* **245**: 147–148.

Small E, Cronquist A. 1976. A practical and natural taxonomy for *Cannabis*. *Taxon* **25**: 405–435.

Small E, Jui PY, Lefkovitch LP. 1976. A numerical taxonomic analysis of *Cannabis* with special reference to species delimitation. *Systematic Botany* **1**: 67–84.

Spitzer-Rimon B, Duchin S, Bernstein N, Kamenetsky R. 2019. Architecture and florogenesis in female *Cannabis sativa* plants. *Frontiers in Plant Science* **10**: 350.

Spitzer-Rimon B, Shafran-Tomer H, Gottlieb GH, Doron-Faigenboim A, Zemach H, Kamenetsky-Goldstein R, Flaishman M. 2022. Non-photoperiodic transition of female 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.

Terpilowski M. 2019. scikit-posthocs: pairwise multiple comparison tests in Python. *Journal of Open Source Software* **4**: 1169.

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.

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.

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

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Accession, leaflet number, and relative node numbers prediction of *Cannabis* leaves using the combined dataset.

Table S1 Prediction accuracy in Linear discriminant analysis modeling for accession, leaflet, and relative node number using 3–15 leaflets.

Table S2 Results of Kruskal–Wallis test for accession, leaflet, and relative node number, using width : length ratios.

Table S3 Results of Dunn's *post hoc* test for accession, leaflet, and relative node number, using width : length ratios.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

CHAPTER 3

CANNABIS PHYTOCHEMICAL DIVERSITY

3.1 Phytochemical diversity of wild-growing and landrace *Cannabis*: insights into cannabinoid composition across tissues, sexes, and geographic origins

Manuscript in preparation:

Balant, M., Quintas, G., Gras., A, Papamichail, A., Vallès, J., De Luna Salvà, N., Montero Molina, S., Garnatje, T., Vitales, D.

Phytochemical diversity of wild-growing and landrace *Cannabis*: insights into cannabinoid composition across tissues, sexes, and geographic origins

Manica Balant[1,2,*], Guillermo Quintas[3], Airy Gras[2], Alexandra Papamichail[1], Joan Vallès[2,4], Noemi De Luna Salvà[3], Sonia Montero Molina[3], Teresa Garnatje[1,5,**], Daniel Vitales[1,**]

[*] corresponding author
[**] equal contribution

Affiliations
[1] Institut Botànic de Barcelona (IBB, CSIC-CMCNB), Passeig del Migdia sn, 08038, Barcelona, Spain
[2] Laboratori de Botànica—Unitat Associada CSIC, Facultat de Farmàcia i Ciències de l'Alimentació—Institut de Recerca de la Biodiversitat IRBio, Universitat de Barcelona (UB), Barcelona, Spain
[3] Health & Biomedicine Department of Leitat Technological Center, 08225 Terrassa, Spain
[4] Secció de Ciències Biològiques, Institut d'Estudis Catalans, Carrer del Carme 47, 08001, Barcelona, Catalonia, Spain
[5] Jardí Botànic Marimurtra - Fundació Carl Faust, pg. Carles Faust, 9, 17300 Blanes, Spain

Keywords: Cannabaceae, cannabinoids, chemovars, crop wild relatives, hemp, phytochemical composition, UPLC-MS

ABSTRACT

This study examined the phytochemical diversity of *Cannabis*, emphasizing wild-growing and landrace populations, along with underutilized plant parts like leaves and male plants. We analysed cannabinoid composition across plant tissues, sexes, and phylogeographic groups to identify distinguishing patterns. Significant differences in total cannabinoid concentrations and compositions were found between tissues, with inflorescences generally having higher cannabinoid levels than leaves, particularly in female plants. While leaves showed no statistically significant differences in cannabinoid composition between male and female plants, male and female inflorescences exhibited notable variations. Geographic origins could not be precisely determined by cannabinoid composition alone, but a global trend emerged: tCBD-dominant plants were predominantly located north of 30°N, and tTHC-dominant plants south of 30°N, with exceptions. Better differentiation was observed between cultivated drug-type landraces and wild-growing plants, reflecting the impact of centuries of selective breeding. These findings underscore the importance of conserving traditional landraces and wild accessions as valuable resources for breeding, conservation, and broader *Cannabis* applications. Further research with expanded datasets is needed to deepen our understanding of *Cannabis* phytochemical diversity and evolutionary patterns.

INTRODUCTION

Cannabis sativa L. (hereafter referred to as *Cannabis*) is one of humanity's earliest domesticated plants and has been widely utilized for thousands of years across a variety of purposes (Abel, 1980). Though the initial plant parts utilized by humans remain unknown, traditional knowledge records the extensive use of all parts of the plant. Stems were processed to obtain durable, high-quality fibre, valued for producing textiles and cordage, and seeds were pressed to produce oil used in dietary and other applications (Clarke & Merlin, 2013). However, the most diverse and widespread is probably the use of *Cannabis* in traditional medicine, where leaves and resin-rich female inflorescences, seeds, and roots have been extensively employed and over 200 traditional medicinal uses treating diverse ailments have been recorded (Balant *et al.*, 2021a,b). This versatility was in part enabled by the extraordinary phytochemical profile of *Cannabis*, containing over 500 secondary metabolites (ElSohly *et al.*, 2017). Among the most well-known is the Δ^9-tetrahydrocannabinol (THC), responsible for its psychoactive effects (Hanuš *et al.*, 2016). Given its importance, the cannabinoid composition has been closely studied and even considered an important factor in *Cannabis* taxonomic classification (Small & Beckstead, 1973; Hillig & Mahlberg, 2004; Hazekamp & Fischedick, 2012; Hazekamp *et al.*, 2016; Herwig *et al.*, 2024). Despite substantial research, recent studies have primarily focused on cultivated *Cannabis* plants, while limited attention was given to wild-growing plants that may possess unique phytochemical profiles. Furthermore, although leaves have long been used in traditional medicine, most research still focuses solely on cannabinoids present in female inflorescences, neglecting plant leaves and entire male plants, which are frequently considered as waste in the medicinal cannabis industry.

The phytochemical profile of *Cannabis* is remarkably diverse, comprising hundreds of secondary metabolites, including cannabinoids, terpenoids, flavonoids, sterols, fatty acids and vitamins, each contributing to the plant's unique pharmacological characteristics (Turner *et al.*, 1979). Cannabinoids are of particular interest due to their diverse physiological effects, which arise from their interactions with the endocannabinoid system (ECS) in humans and animals. The ECS is regulated by endocannabinoids, which are naturally occurring molecules produced within their bodies. Both endocannabinoids and phytocannabinoids, synthesised in plants, share similar structures, enabling them to bind to the same cannabinoid receptors found within the ECS (Hazekamp *et al.*, 2010; Stasiłowicz *et al.*, 2021). These receptors are distributed throughout various body systems, including the central nervous, immune, and digestive systems, where they play essential roles in regulating key physiological functions related to appetite, inflammation, pain perception, and mood, among others (Mechoulam & Hanuš, 2001; Hazekamp *et al.*, 2010; Osafo *et al.*, 2021). Alongside cannabinoids, *Cannabis* also contains over 120 terpenoids, contributing to its distinct aroma and therapeutic

properties (Hazekamp *et al.*, 2010; Liktor-Busa *et al.*, 2021), often modulating the effects of cannabinoids through a phenomenon known as the “entourage effect” (Russo, 2011; Koltai & Namdar, 2020).

Cannabinoids were first identified in the *Cannabis* plant, and to date, over 170 distinct cannabinoids have been isolated, some of which are naturally occurring degradation (Hanuš *et al.*, 2016). Cannabinoids have also been discovered in various other plants, such as in genera *Trema*, *Helichrysum*, *Amorpha*, and *Glycyrrhiza*, as well as in certain liverworts like *Radula* and even in the fungal genus *Cylindrocarpon* (Quaghebeur *et al.*, 1994; ElSohly & Slade, 2005; Hanuš *et al.*, 2016; Ribeiro *et al.*, 2024). The precise function of cannabinoids in plants remains unclear; however, it has been hypothesized that they may function in protective roles against ultraviolet (UV) radiation and herbivore attacks (Clarke & Merlin, 2013). Research has indicated that certain cannabinoids can induce apoptosis in various organisms, which may act as a deterrent for herbivores (Sirikantaramas *et al.*, 2005).

The synthesis of cannabinoids in *Cannabis* occurs in glandular trichomes. They are located on the plant's aerial parts and are especially abundant on the bracts of female flowers (Livingston *et al.*, 2020). The synthesis begins with the precursor compounds olivetolic acid and geranyl pyrophosphate (GPP), which are converted to cannabigerolic acid (CBGA) through the enzyme CBGA synthase. The CBGA is subsequently secreted into the extracellular storage cavity of glandular trichomes, where CBGA is further transformed into tetrahydrocannabinolic acid (THCA), cannabidiolic acid (CBDA), and cannabichromenic acid (CBCA) by their respective synthases (THCA synthase, CBDA synthase, and CBCA synthase; van Velzen & Schranz, 2021). With exposure to heat, light, and atmospheric oxygen, the acidic forms of cannabinoids undergo non-enzymatic decarboxylation. This decarboxylation process begins early in the vegetative stage of the plant and continues throughout its growth and even more extensively during storage (Kajima & Piraux, 1982). Notably, THCA is non-psychoactive in its acidic form. To experience the psychoactive effects associated with THC, THCA is typically heated via smoking, vaping, or baking before the consumption, which converts it to THC through decarboxylation. Additionally, THC can degrade into cannabinol (CBN) when stored for extended periods of time (Hazekamp *et al.*, 2010). While THC and CBD are the most well-known and researched cannabinoids, *Cannabis* also contains a variety of minor cannabinoids, including cannabigerol (CBG), cannabichromene (CBC), cannabielsoin (CBE), cannabidivarin (CBDV), tetrahydrocannabivarin (THCV), cannabigerol monomethylether (CBGM), and their acidic forms (Hanuš *et al.*, 2016). Initially it was believed that THC and CBD ratios in *Cannabis* are regulated by two alleles of a single gene (de Meijer *et al.*, 2003; Hillig & Mahlberg, 2004), however it was later demonstrated that specific cannabinoid composition of *Cannabis* plants is defined by the presence or absence of single-copy genes within the cannabinoid oxidocyclase family (van Velzen & Schranz, 2021).

Phytochemical composition has often been considered valuable criteria for *Cannabis* classification in taxonomic studies (Small & Cronquist, 1976; Clark & Bohm, 1979; Hillig & Mahlberg, 2004; Hazekamp & Fischedick, 2012; Herwig *et al.*, 2024), as only a specific subset of the hundreds of identified secondary metabolites is typically expressed in any given plant. Fetterman (1971) proposed a classification system based on the ratio of the two primary cannabinoids: plants with large amounts of THC on one hand, and plants with large amounts of CBD on the other. This classification is similar to what is used nowadays to separate *Cannabis* plants cultivated for fibre production (i.e., hemp) that generally contain less than 0.3% THC and *Cannabis* plants cultivated for medicinal and recreational purposes, with higher levels of THC and/or CBD (Hurgobin *et al.*, 2021). Small and Beckstead (1973) extended this classification into three categories, based on the THC/CBD ratio to type I (THC-dominant), type II (balanced THC and CBD levels), and type III plants (CBD-dominant). Later, they incorporated an additional group (Type IV), based on the presence or absence of the minor cannabinoid CBGM (Small *et al.*, 1975). Mandolino and Carboni (2004), who worked primarily with cultivated plants, established a similar classification system, describing five chemotypes based on the percentage of cannabinoids in dry weight (d.w.): chemotype I with prevalent THC (THC > 0.3% d.w., CBD < 0.5% d.w.), intermediate chemotype II (THC ≥ 0.3% d.w., CBD > 0.5% d.w.), chemotype III with prevalent CBD (THC < 0.3% d.w., CBD > 0.5% d.w.), chemotype IV with prevalent CBG (CBG > 0.3% d.w., CBD < 0.5% d.w.) and chemotype V, that contains almost no cannabinoids (total cannabinoid content < 0.2% d.w.). Some studies suggest that the phytochemical composition of *Cannabis* plants may be influenced by their geographic origins (Fetterman *et al.*, 1971; Small & Beckstead, 1973; Baker *et al.*, 1980; Hillig & Mahlberg, 2004; McPartland & Small, 2020). Their results indicate that *Cannabis* plants from regions below 30°N tend to exhibit higher levels of THC compared to those from higher latitudes. Additionally, both the ratios of the two primary cannabinoids and the presence and absence of minor cannabinoids have been associated to geographic origin.

While *Cannabis* phytochemical composition is primarily determined by genetics, many studies have demonstrated that environmental conditions (Small *et al.*, 1975; de Meijer *et al.*, 1992; Saloner & Bernstein, 2021; De Prato *et al.*, 2022; Park *et al.*, 2022; Zandkarimi *et al.*, 2023; Reichel *et al.*, 2024) and the developmental phase (Turner *et al.*, 1979; Aizpurua-Olaizola *et al.*, 2016; Stack *et al.*, 2021; Tremlová *et al.*, 2021) also play significant roles. These factors contribute to the variability in phytochemical traits, complicating taxonomic classifications based solely on phytochemistry (Booth & Bohlmann, 2019). Significant differences in phytochemical composition have also been found when comparing different plant parts or the same plant tissue along the plant (e.g., terminal and lateral inflorescences; Hemphill *et al.*, 1980; Bernstein *et al.*, 2019; Jin *et al.*, 2020). Differences have also been found between the male and female plants (Fetterman *et al.*, 1971; Turner *et al.*, 1979; Nagy *et al.*,

2019; Busta *et al.*, 2022), as well as between the plants of the same sex within a population (Busta *et al.*, 2022). However, studies show that the variability is mainly found in the amount of each secondary compound (quantitative analysis), and when using the content ratios, or scoring for presence and absence of the compounds (qualitative analysis), the variability is significantly decreased (Barni-Comparini *et al.*, 1984; Pacifico *et al.*, 2008; Aizpurua-Olaizola *et al.*, 2016).

Cannabis exhibits a remarkable range of phytochemical diversity, yet substantial gaps remain in our understanding of its phytochemical composition. Further research focusing on the phytochemical diversity of wild-growing *Cannabis* populations and underutilized plant parts, such as leaves and entire male plants, is needed to advance this understanding. The objectives of this study were to i) investigate the cannabinoid diversity in wild-growing and landrace *Cannabis* populations, ii) compare the cannabinoid composition of different plant parts, specifically leaves and inflorescences, across male and female plants and iii) compare the cannabinoid composition to assess whether it could be used to differentiate between phylogeographic groups and iv) between wild-growing and cultivated *Cannabis* plants.

MATERIAL AND METHODS

Plant material and growing conditions

In this study 34 individuals from 15 *Cannabis* accessions from diverse geographical regions were cultivated, encompassing wild-growing plants, and cultivated drug-type and multipurpose landraces. The accessions were classified into five of the six phylogeographic groups as defined by Balant *et al.* (2024). List of the accessions used is provided in Table 1, and additional details are summarized in Table S1.

The plants were propagated from seed and cultivated in controlled conditions to reduce external environmental variability. Prior to sowing, the seeds were sterilized in a 5% hydrogen peroxide (H_2O_2) solution supplemented with Inex-A (Cosmocel, Spain) at ambient temperature overnight. Following sterilization, seeds were scarified and placed in Petri dishes for initial germination at 25°C. When the seeds germinated, they were transplanted into small peat pots filled with a nutrient-enriched soil substrate (Kilomix Atami, Spain). These peat pots were placed into larger 7-liter pots filled with Cocopeat substrate (Projar, Spain) to support further growth.

Throughout the initial phase of cultivation, the temperature was maintained at 25°C with an 18-hour light/6-hour dark cycle. To account for potential photoperiod sensitivity in certain *Cannabis* accessions, the photoperiod was adjusted to 12 hours of light and 12 hours of darkness after three weeks of growth. The light intensity was gradually increased from 150 to 700 µmol/m²/s (BX120c2; Valoya

Finland) over the course of cultivation. These controlled conditions were maintained until the onset of flowering. Plants were irrigated daily with a nutrient solution (see Table S2), with no pest control treatments administered. Plant positioning within the growth room was periodically randomized to minimize microenvironmental variation. A comprehensive description of growth conditions across different stages is provided in Table S2.

Table 1: Accession details and number of *Cannabis* plants, their sex, and number of leaf and inflorescence samples collected and analysed in the study. The accessions were assigned to phylogeographic groups based on those defined in Balant *et al.* (2024).

Population ID	Phylogeographic group	Accession type	No. of plants	No. of male plants	No. of female plants	No. of leaf samples	No. of inflorescence samples
MN1	N China and E Mongolia	W	2	1	1	2	2
MN4	N China and E Mongolia	W	2	2	0	2	2
BNG	C & S China and Himalayas	W	1	1	0	1	1
IKL	C & S China and Himalayas	W	3	1	2	3	3
ISAT	C & S China and Himalayas	D	3	1	2	3	3
NO	C & S China and Himalayas	M	2	1	1	2	1
CAM	Indoafrica	D	3	1*	2	2	4
ETH3	Indoafrica	D	3	2	1	3	3
IK2	Indoafrica	D	3	1	2	3	1
MAR	Caucasus and Mediterranean	D	2	2	0	2	2
AM15	Caucasus and Mediterranean	W	2	1	1	1	2
AM18	Caucasus and Mediterranean	W	2	1	1	2	1
AM20	Caucasus and Mediterranean	W	2	2	0	2	2
BG3	Eurosiberia and W Mongolia	W	1	1	0	1	1
KAZ	Eurosiberia and W Mongolia	W	3	2	1	3	3

* The analysed individual was a hermaphrodite; W – Wild-growing accessions, D – Drug landrace, M – Multipurpose landrace

Sampling was performed shortly after the onset of flowering. Male inflorescences were collected once at least one-third of the flowers had opened, while female inflorescences were sampled upon the appearance of stigmas. However, due to variability in developmental timing among individuals, not all samples were collected at the at precisely the same stage. Despite these challenges, efforts were made to standardize the sampling time as much as possible. For inflorescence samples, the terminal section of the inflorescence was collected; however, due to significant variation in size and shape, the material obtained was sometimes insufficient for analysis, particularly in certain wild-growing plants. In such cases, additional material was taken from higher lateral inflorescences and subsequently combined with the material from the terminal inflorescence. Both male and female inflorescences were dissected to eliminate the leaves and stem parts, with as little handling as possible. Since variation in

cannabinoids along the plant was observed in previous studies (Bernstein *et al.*, 2019), leaf samples were always collected from the middle section of the main stem, typically between nodes 6 and 7 (exceptionally from node 4 to node 12 if leaves in nodes 6 and 7 were not in suitable condition). All samples were flash-frozen on dry ice and stored at -20°C until analysis. For two individuals from population IK2 that did not reach flowering by the end of the cultivation period, only leaf samples were collected. In total, 19 male plants, 14 female plants, and one hermaphrodite were sampled, for a total of 32 leaf samples and 31 inflorescence samples.

Sample preparation

For each extraction, a total of 2.1 grams of flash-frozen plant material was prepared, homogenised, and divided into three replicates of 0.7 grams each. The extraction was conducted using a Precellys homogenizer (Precellys evolution, Bertin Technologies, Montigny-le-Bretonneux, France). Each tube contained 0.7 grams of the prepared sample mixed with 5 mL of methanol. The homogenization process was performed for 30 seconds at 10,000 rpm, using a Precellys Lysing Kit (Tissue Grinding). Given that 5 mL of methanol per replicate might not be sufficient for all future experiments, an additional 2 mL of methanol was added to the homogenized extract, which was then mixed thoroughly. The mixture was allowed to stand for an additional hour at room temperature to ensure maximum extraction of phytochemical constituents from the plant material. In cases where the available material was limited, the extraction was done in fewer replicas and with smaller quantities of material, but the µM ratio was maintained to allow for comparative analysis.

Cannabinoid analysis

Sample extracts were diluted in 1:1 H_2O (0.1% v/v HCOOH) and analysed using an Acquity HSS T3 C18 (100 × 2.1 mm, 1.7 µm) column in an Agilent 6460 (Agilent) system. Mobile phases were H_2O (0.1% v/v HCOOH) (A) and (0.1% v/v HCOOH) CH_3CN (B). The gradient program was as follows: starting at 45% A and 55% B at 0 min, % A decreased down to 40 at 6 min, and at 20% at 9 min. The % A was further reduced to 1% A at 10 min, and this composition was held until 13 min. At 13.1 min, the gradient returned to the initial condition of 45% A and 55% B for column re-equilibration. Injection volume, flow rate and column temperature were set at 1 µL, 6000 µL/min and 30°C, respectively. Autosampler temperature was set at 6°C during sample analysis. Electrospray ionization was carried out using the following conditions: capillary 3.5 kV, cone 25 V, source temperature 100°C, desolvation temperature 400°C, N2 cone and desolvation gas flow rates were 50 and 800 L/h, respectively. The following multiple reaction monitoring parameters were selected: CBDVA 329.2 > 285.2 (RT = 2.91 min), CBDV 287.2 > 123.1 (RT = 3.14 min), CBDA 357.2 > 245.2 (RT = 4.36 min), CBD 315.3 > 193.2 (RT = 4.69 min),

CBGA 359.2 > 341.2 (RT = 4.63 min), CBG 317.2 > 193.0 (RT = 4.68 min), THCVA 329.2 > 285.2 (RT = 5.37 min), THCV 287.2 > 123.1 (RT = 4.71 min), THCA 357.2 > 245.2 (RT = 6.34 min), D9D8THC 315.3 > 193.2 (RT = 5.85 min), CBC 315.3 > 193.2 (RT = 6.27 min), and CBN 311.2 > 223.0 (RT = 5.50 min). Peak area values were interpolated in external linear calibration curves for the estimation of the cannabinoid concentrations in the extracts.

Statistical analyses

Samples were prepared in replicates, and the cannabinoid content was analyzed by UPLC-MSM for each replicate. The median value of the replicates was then calculated and used for further statistical analysis. Correlations between the acidic and non-acidic forms of cannabinoids were assessed using the Pearson's correlation test with the *cor()* function from the R package 'stats' (R Core Team, 2022), visualizing the results as a correlation matrix built using the 'ggcorrplot' package (Wickham, 2016). Since statistically significant correlations were found, and to minimize variations due to the decarboxylation of acidic cannabinoids, downstream analysis was conducted using the combined values of acidic and non-acidic forms for each cannabinoid. These combinations included CBDVA + CBDV = tCBDV, CBDA + CBD = tCBD, THCVA + THCV = tTHCV, THCA + Δ^9-THC + Δ^8-THC = tTHC, and CBGA + CBG = tCBG. For CBC, only the non-acidic form was analysed. The correlation between the sums of acidic and non-acidic forms was analysed using the same methodology as before.

Our cannabinoid analyses employed a different pre-processing method compared to previous studies with similar sampling (i.e., freezing samples on dry ice vs. drying the samples). Therefore, as the percentage of cannabinoids per dry weight was unavailable in our dataset, we opted to calculate the tTHC/tCBD ratios instead. This approach provided a comparable metric to assess the cannabinoid profiles across different types of samples and between different studies. The ratio provides a more valid comparison of many studies that grew plants under different conditions and sampled plants at different developmental points, as the ratio is stable throughout the life cycle, while the concentration of individual cannabinoids changes (Fetterman *et al.*, 1971; Barni-Comparini *et al.*, 1984; Aizpurua-Olaizola *et al.*, 2016; Grassi & McPartland, 2017). To classify the samples into chemotypes (sometimes also called phenotypes or types), we followed the method used by Hillig and Mahlberg (2004), which utilized the $\log_{10}$-transformed THC/CBD ratio, as the dry weight percentage data necessary for the classification systems of Small and Beckstead (1973) or Mandolino and Carboni (2004) was unavailable in our dataset.

The Shapiro-Wilk Normality Test revealed that our data did not follow a normal distribution; therefore, non-parametric tests were applied. To analyse differences between inflorescences and leaves within

individual plants, we used a subset of 27 individuals where measurements for both inflorescence and leaf samples were available and conducted a paired Wilcoxon Signed-Rank Test using the R package 'stats' (R Core Team, 2022). To test for differences in cannabinoid concentrations between leaves from male and female plants, as well as male and female inflorescences, we first excluded the single hermaphroditic leaf sample (CAM-11_Lf_HE). We then performed the Kruskal-Wallis Rank Sum Test to compare cannabinoid concentrations between the two sexes for individual cannabinoids, using the R package 'stats' (R Core Team, 2022).

For the analysis of differences between phylogeographic groups, we divided the dataset into three subsets: leaves (32 samples), male inflorescences (20 samples), and female inflorescences (11 samples). Samples were *a priori* assigned to five of the six phylogeographic subgroups defined in Balant *et al*. (2024) based on the geographic origin of the analysed accessions. Differences in concentrations of individual cannabinoids were visualised with boxplots with 'ggplot2' package in R (Wickham, 2016). We performed the Kruskal-Wallis Rank Sum Test for each subset (leaves; male inflorescences; and female inflorescences) to compare among phylogeographic groups for differences in individual cannabinoid concentration. If significant differences were found, Pairwise Wilcoxon Rank Sum Test with FDR adjusted p-values was conducted using the R package 'stats' (R Core Team, 2022).

Differences in concentrations of individual cannabinoids across wild-growing and cultivated plants (further divided between drug-type and multipurpose landraces) for leaves and male inflorescences were visualized with boxplots using 'ggplot2' package in R. However, statistical significance was not calculated since the multipurpose landrace group only contained three samples. Female inflorescences were not visualized due to insufficient sample size (n = 11). Differences in the tTHC/tCBD ratios between the three groups were also visualised and the differences were analysed with Kruskal-Wallis Rank Sum Test and Pairwise Wilcoxon Rank Sum Tests with FDR adjustment p-values using the R package 'stats' (R Core Team, 2022).

Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA) were performed using PLS Toolbox 9.5 (Eigenvector Research Inc., Wenatchee, WA, USA) and MATLAB 2021a (MathWorks Inc., Natick, MA, USA), with scripts developed by the authors.

RESULTS

Cannabinoid concentration and composition of the analysed samples

Our results revealed substantial variability in both total cannabinoid concentration and composition among the analysed samples (Fig. 1). The sample with the highest total cannabinoid concentration was a female inflorescence from a wild-growing population in Armenia (AM15-1; 2143.22 µM). In contrast,

the lowest total cannabinoid concentration was observed in a leaf sample from a hermaphroditic plant of drug-type landrace from Cambodia (CAM-11_Lf_HE; 78.11 µM; Table S3).

Among the quantified cannabinoids, CBDA was the most abundant, followed by THCA and CBGA, with median concentrations of 265.65 µM, 57.23 µM, and 16.35 µM, respectively. The least abundant cannabinoids detected were CBDV and THCV, present in almost negligible amounts (median value of < 0.001 µM), while CBN below the limit of quantification in all samples. As CBN is a degradation product of THC, its absence confirms that no significant degradation occurred between sample collection and analysis.

The analysis revealed the predominance of acidic forms of cannabinoids, with CBDA as most abundant followed by THCA and CBGA. This is attributable to the immediate flash-freezing of samples on dry ice post-collection and their storage at -20°C, which effectively prevented degradation and preserved cannabinoids in the acidic forms. Consequently, the presence of non-acidic forms was minimal. For example, the median CBDA concentration was 256.65 µM, compared to just 0.02 µM for CBD (see Table S4 for remaining cannabinoids). A strong correlation was observed between acidic and non-acidic forms of cannabinoids (average $r > 0.70$; Fig. S1), allowing their paired sums to be used in the later analysis, except for CBC. In this case, CBCA was not measured, so only the non-acidic form (i.e., CBC) was included. In the continuation we will refer to the sums of the acidic and non-acidic form as the total non-acidic forms (i.e., tCBG = CBG + CBGA, tCBD = CBD + CBDA, tCBDV = CBDV + CBDVA, tTHC = THC + THCA, tTHCV = THCV + THCVA). Previous studies recorded the presence of CBGM and CBGMA in *Cannabis* samples (Small *et al.*, 1975; Hillig & Mahlberg, 2004; de Meijer *et al.*, 2009). Due to the lack of CBGM and CBGMA analytical standards, sample extracts were analysed by UPLC-TOF-MS in full scan mode. Nonetheless, the chromatographic peaks corresponding to the protonated (ESI+) or deprotonated (ESI-) CBGM and CBGMA were not detected using a *m/z* accuracy of 20 ppm.

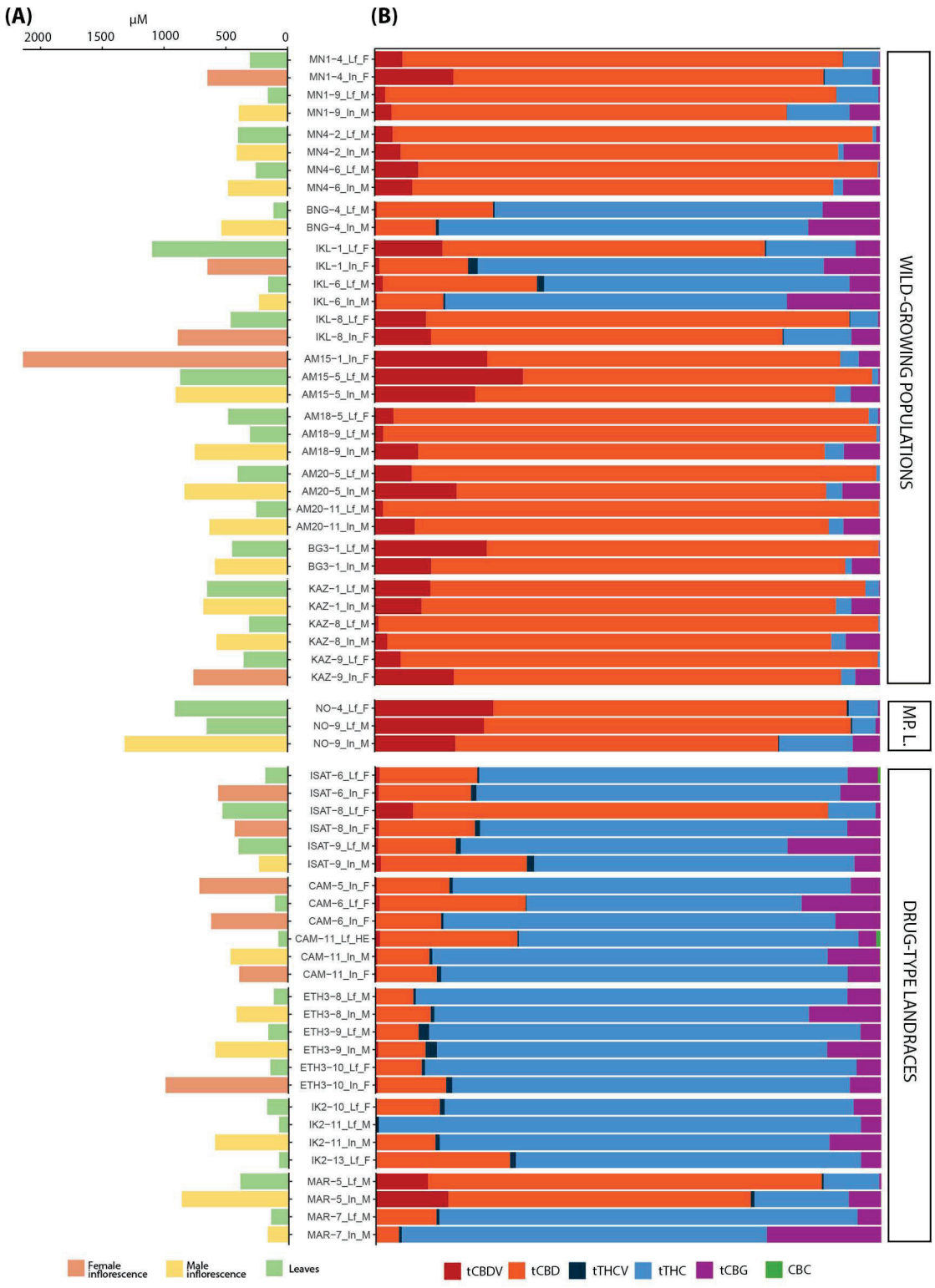


Figure 1. The total cannabinoid concentration **(A)** and cannabinoid composition **(B)** in the analysed samples. Concentrations of the acidic and non-acidic forms of cannabinoids are summarised and shown together, except for CBC (i.e., CBDVA + CBDV = tCBDV, CBDA + CBD = tCBD, THCVA + THCV = tTHCV, THCA + Δ^9-THC + Δ^8-THC = tTHC, and CBGA + CBG = tCBG).

Abbreviations: MP. L. – Multipurpose landrace; In – Inflorescence; M – Male; F – Female.

Based on the cannabinoid composition, the samples could be broadly categorized into two major groups: tCBD-dominant and tTHC-dominant (Fig. 1). While in some samples minor cannabinoids represented considerable proportions of total cannabinoids (i.e., tCBDV – 29.35% in a leaf sample in wild-growing population from Armenia and tCBG – 22.63% in a male inflorescence from drug-type landrace from Morocco), we did not find any sample, where the minor cannabinoids would dominate (Fig. 1, Table S3). The least abundant minor cannabinoid was CBC, with its highest detected proportion being 0.84% in a leaf sample from a hermaphroditic plant of a drug-type cultivar from Cambodia. However, given the relatively similar median values of CBC, THC, and CBD (0.01%, 0.51%, and 0.02%, respectively; Table S4), it is possible that including the acidic form of CBC (i.e., CBCA) in the analysis, would result in a higher combined concentration for this cannabinoid.

The tCBG was positively correlated with all the analysed cannabinoids (Fig. 2). We observed a strong positive correlation between tCBD and tCBDV (r = 0.87), as well as between tTHC and tTHCV (r = 0.83; Fig. 2). Both THC and tTHC also showed positive correlations with CBC (r = 0.70 and r = 0.38, respectively; Fig. 2, S1). We observed a negative correlation between tTHC and tCBD (r = -0.38), and less so for tTHCV and tCBDV (r = -0.09).

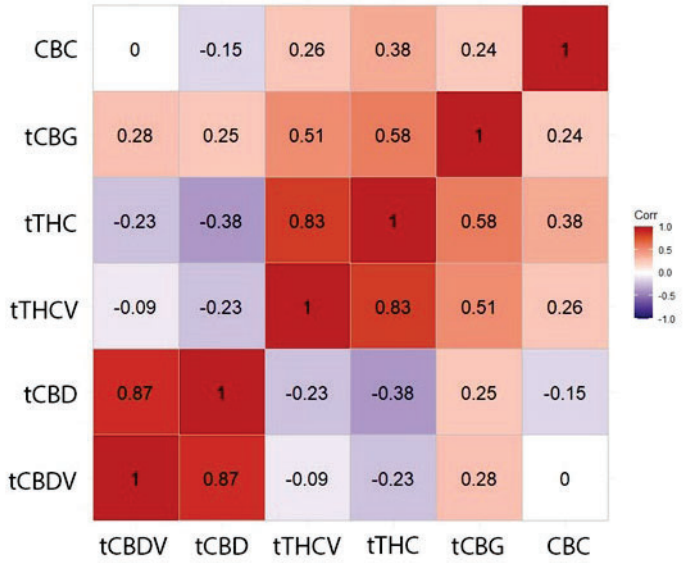


Figure 2. The correlation coefficient (r) between the concentration of the analysed cannabinoids. Concentrations of the acidic and non-acidic forms of cannabinoids are summarised and shown together, except for CBC (i.e., tCBG = CBG + CBGA, tCBD = CBD + CBDA, tCBDV = CBDV + CBDVA, tTHC = THC + THCA, tTHCV = THCV + THCVA).

The tTHC/tCBD ratio

As the percentage of cannabinoids per dry weight was unavailable in our dataset, we opted to calculate the tTHC/tCBD ratios instead. The calculated ratios ranged from 0.002 to 16.15 (median = 0.19), with one extreme outlier of 53215.5 in leaf sample from individual IK2-11 (Table S3).

Most samples in our study were categorized as CBD-dominant (chemotype III; n = 32) or intermediate, with a more balanced THC and CBD concentration (chemotype II; n = 27). Only four samples—leaves from two male plants from Ethiopia, one male plant from India Kerala, and one male inflorescence from Morocco—were classified as THC-dominant (chemotype I). The median value for the chemotype I was 13.83 (min 10.28; max 53215.47; Table S3), chemotype II 4.20 (min 0.23; max 9.45; Table S3) and chemotype III 0.04 (min 0.002; max 0.19; Table S3). No samples matched type IV (CBGM-dominant) as defined by Small *et al.* (1975) or chemotype IV (CBG-dominant) and chemotype V (cannabinoid-free) as defined Mandolino and Carboni (2004). Leaves and inflorescences generally belonged to the same chemotype, except for five male individuals, where either inflorescences (NO-9, MAR-5, and MAR-7) or leaves (ETH3-8 and IK2-11) displayed a higher tTHC/tCBD ratio.

Differences in cannabinoid composition of leaves and inflorescences in male and female plants

Significant differences in both the total cannabinoid concentration and composition were observed between leaves and inflorescences within a plant (Table S5). The PLS-DA model based on the cannabinoid composition was able to correctly classify the samples identity as leaves or inflorescences (Fig. 3a), showing high area under the curve (AUC) value of 0.94 for calibration and 0.93 for cross-validation (p-value < 0.05). The cannabinoid contributing most to the observed discrimination between leaves and inflorescences was tCBG (Fig. S2).

On average, inflorescences presented 2.52 higher total cannabinoid concentrations than leaves within the same plant, with values ranging from 0.59 to 7.54 times higher (Table S6). In female plants, inflorescences averaged 2.92 times higher cannabinoid concentrations than leaves (range: 0.59–6.79), while in male plants, inflorescences contained on average 2.34 times higher concentrations than leaves (range: 0.59–7.54). Leaves and inflorescences show statistically significant differences in the concentrations of all total cannabinoids, except CBC, which did not exhibit such variation (Table S5). However, despite differences in cannabinoid composition and concentration, no significant variation was observed in the tTHC/tCBD ratios between these tissue types, as confirmed by the paired Wilcoxon signed-rank test (Table S5).

We also examined whether the cannabinoid composition varies based on the plant's sex within specific tissue types. To identify potential differences, the dataset was divided into leaves and inflorescences, and each group was analysed separately for significant variations between sexes. Kruskal–Wallis test revealed no significant differences between leaves from male and female plants for the analysed cannabinoids (Table S7, Fig. 4a). Additionally, the discriminant analysis revealed insufficient discrimination between these two leaf categories (Fig. 3b). A PLS-DA showed a non-statistically significant discrimination between the two leaf groups (AUROC cv < 0.5, p-value>0.05), indicating a strong overlap between the cannabinoid profiles of male and female leaves (Fig. S3a-c). In contrast, significant differences between male and female inflorescences for tTHC and tTHCV concentration were observed (Table S8, Fig. 3c), enabling the development of a statistically significant PLS-DA model for the discrimination of male and female inflorescences (AUCcv = 0.81 (p-value < 0.05); see Fig. S3d-e).

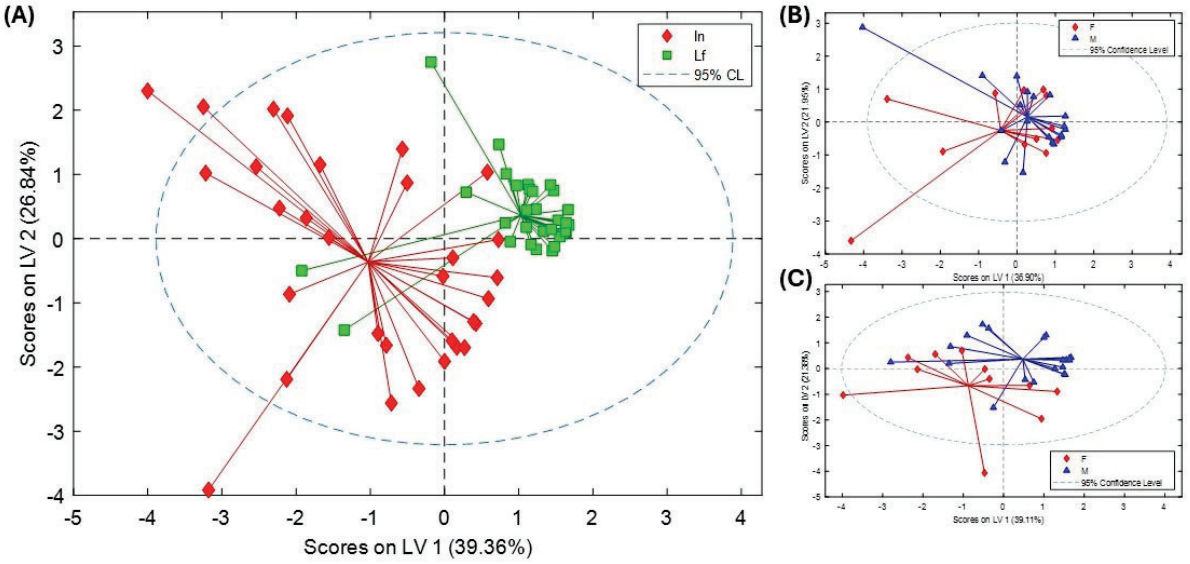


Figure 3. PLS-DA scores plot showing the differences in cannabinoid concentrations between various sample groups. **(A)** Comparison between leaves (Lf) and inflorescences (In) based on cannabinoid concentrations. **(B)** Comparison between leaves from male (M) and female (F) plants. **(C)** Comparison between male (M) and female inflorescences (F).

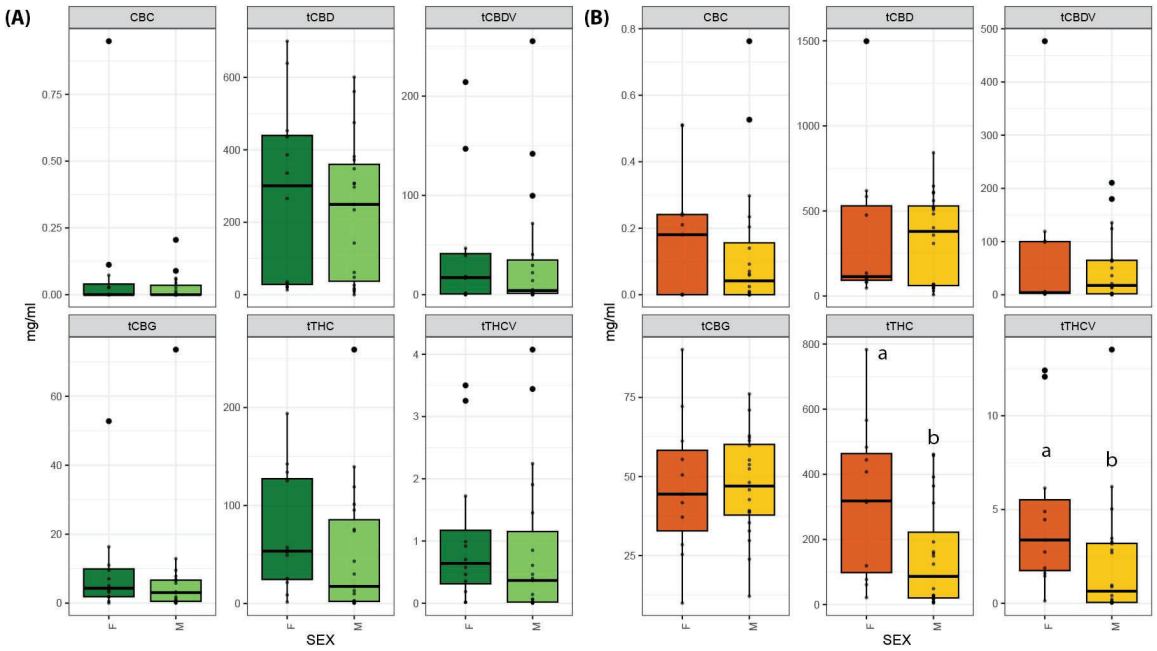


Figure 4. The boxplots showing the variability of cannabinoid concentrations in leaves of male (light green) and female (dark green) plants **(A)** and male (yellow) and female (orange) inflorescences **(B)**. The letters indicate the significant differences between the groups as calculated by Kruskal–Wallis test (see also Table S6).

Differences in cannabinoid composition between phylogeographic groups

As previous research indicated significant variability in cannabinoid composition among *Cannabis* accessions from different geographic areas, we tested if the tTHC/tCBD ratios could provide sufficient information for separating the samples into corresponding phylogeographic groups (as defined by Balant et al. 2024). As the tTHC/tCBD ratios of leaves and inflorescences did not differ significantly, we analysed them together.

Significant differences in ratios among the phylogeographic groups were identified. The Wilcoxon rank-sum exact test showed that the Indoafrica and C & S China and Himalayas groups had significantly higher tTHC/tCBD ratios compared to the other groups, including Caucasus and Mediterranean, Eurosiberia and W Mongolia, and N China and E Mongolia, which in turn did not exhibit significant differences (Fig. 5b, Table S9). We also detected significant differences between the Indoafrica group and C & S China and Himalayas group, the first showing higher tTHC/tCBD ratios.

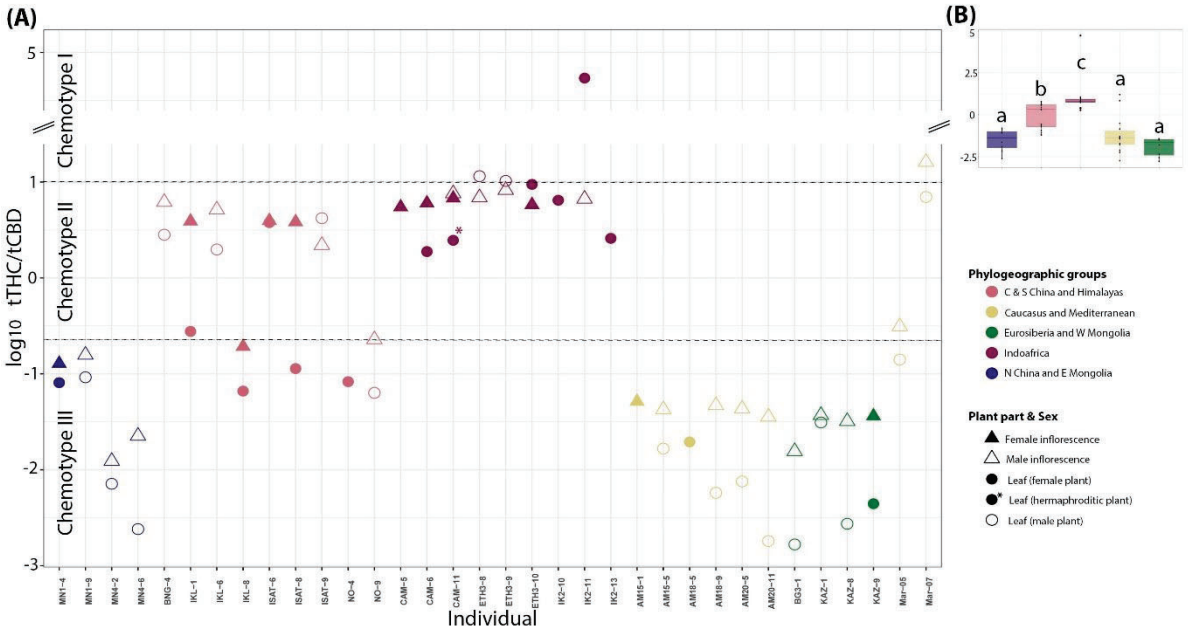


Figure 5. The $\log_{10}$ values of the tTHC/tCBD ratios of analysed samples and the division into the three chemotypes following Hillig and Mahlberg (2004) criteria **(A)**. The boxplots for analysed samples separated per phylogeographic groups based on Balant *et al.* (2024). The letters mark significant differences between the groups **(B)**.

Since the tTHC/tCBD ratio provided only limited separation between phylogeographic groups, we also performed a multivariate analysis using the complete cannabinoid composition. The first two PC axis including all analysed samples explained 77.02% of variation and revealed partial separation of the groups (Fig. S4). The Indoafrica group was separated from the Caucasus and Mediterranean group, Eurosiberia and W Mongolia group, and N China and E Mongolia group, while the samples from C & S China and Himalayas group were positioned between both clusters. This differentiation was primarily driven by variations in the levels of tCBD and tCBDV, and tTHC and tTHCV. However, there was still significant overlapping between the groups (Fig. S4).

Given that we found significant differences in cannabinoid composition among leaves and inflorescences and between male and female inflorescences, we decided to divide the dataset into three subsets and assess whether analysing each tissue type separately would improve the resolution of phylogeographic group differentiation. The PCA of the leaf subset showed only slightly improved separation, with the first two principal components explaining 75.39% of the variance (Fig. 6). The primary separation of clusters was based on the division into the CBD-dominant from THC-dominant plants, forming three clusters. Like before, the first cluster included the Caucasus and Mediterranean group, Eurosiberia and W Mongolia group, and N China and E Mongolia group, predominantly

characterised by higher concentration of tCBD and tCBDV in leaves. The second cluster comprised the Indoafrica group, characterised CBC and tTHC, while the third, the most diverse, was formed by the C & S China and Himalayas group (Fig. 6). While the CBC did not differ significantly between any of the phylogeographic groups (Table S10), the PCA suggests an association between the samples from Indoafrica group and the presence of tTHC and CBC (Fig. 6). The leaves from plants belonging to the Indoafrica group displayed a relatively uniform cannabinoid composition and differed from all the remaining groups by significantly lower concentration of tCBD and tCBDV (Fig. 6, S5, Table S10). The Eurosiberia and W Mongolia, and N China and E Mongolia groups also showed uniformity in cannabinoid composition of leaves, resulting in tighter clustering within the PCA (Fig. 6). They were both characterised by very low tCBG, tTHC and tTHCV concentrations in the leaves (Fig. 6, S5, Table S10). The PLS-DA model based on the analysed cannabinoids concentrations in leaves revealed weak predictive power for most of the phylogeographic groups. Only the C & S China and Himalayas group demonstrated strong model performance, with an AUC of 0.88 for calibration and 0.83 for cross-validation ($p = 0.02$). In contrast, the Caucasus and Mediterranean group showed worse performance, with a low AUC (0.74 for calibration and 0.45 for cross-validation; $p = 0.7$), indicating weak predictive power. Eurosiberia and W Mongolia group, Indoafrica group and N China and E Mongolia group showed moderate to high AUC values for calibration (0.91, 0.92, and 0.83, respectively), but lower values for cross-validation (0.61, 0.80, and 0.61, respectively), indicating variability in model accuracy (Fig. S6).

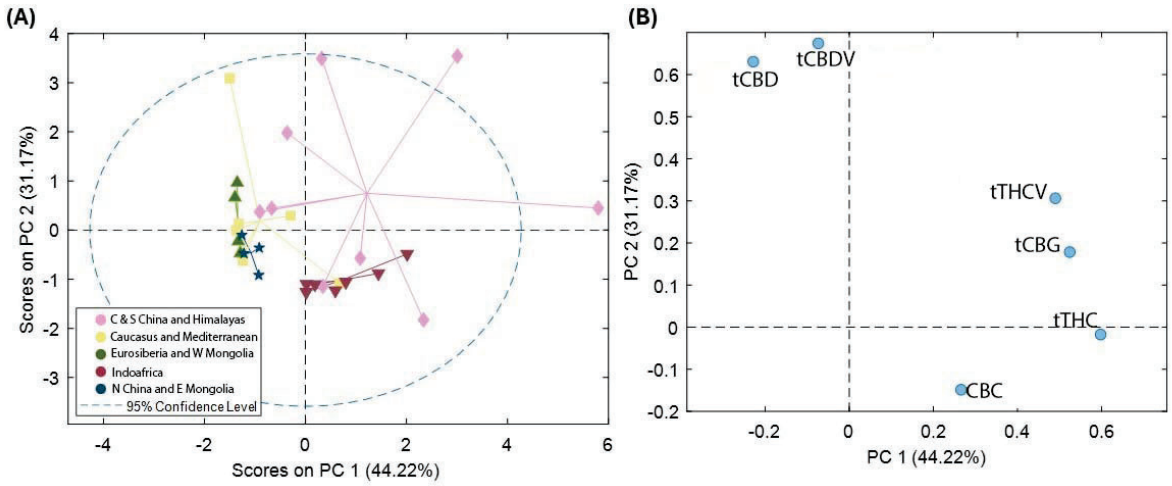


Figure 6. PCA scores **(A)** and loadings **(B)** plots summarizing 75% of the initial data variance in the data set including cannabinoid concentrations from the male and female leaf sample set. Samples **(A)** were coloured according to the phylogeographic groups (as defined by Balant *et al.* 2024)**.** The six cannabinoids are presented in the loadings plot as the sum of their acidic and non-acidic forms, except

for CBC (e.g., CBDVA + CBD = tCBDV, CBDA + CBD = tCBD, THCVA + THCV = tTHCV, THCA + Δ9-THC + Δ8-THC = tTHC, and CBGA + CBG = tCBG).

Similar patterns were observed when analysing male inflorescences, with the first two principal components explaining 78.27% of the variance (Fig. S7). Male inflorescences from plants belonging to Indoafrica group clustered close together and were characterised by higher tTHC, tTHCV and CBC concentrations (Fig. S7, S8). On the opposite side of the PCA scores plot, samples of male inflorescences from Eurosiberia and W Mongolia group, and N China and E Mongolia group, characterised by lower tTHC and tTHCV and higher tCBD and tCBDV concentrations were clustered. The remaining samples from Caucasus and Mediterranean group and C & S China and Himalayas group showed the most diverse composition, with samples clustering together with either samples showing higher concentrations of tTHC, tTHCV and CBC or higher concentrations of tCBD and tCBDV (Fig. S7, S8). Due to the relatively low number of samples (n = 20), significance for differences in concentrations for individual cannabinoids were not calculated. The PLS-DA model again revealed weak predictive power for most of the phylogeographic groups. The best model performance was found for the Indoafrica group and Caucasus and Mediterranean group (AUC of 0.92 and 0.86 for calibration, and 0.87 and 0.81 for cross-validation, respectively). On the other hand, C & S China and Himalayas group, Eurosiberia and W Mongolia group and N China and E Mongolia group showed worse performance, with lower AUC values for both calibration and cross-validation and a non-significant p-value (Fig. S9). Female inflorescences subset was too limited in size (n = 11) to draw meaningful conclusions and was therefore not analysed.

Cannabinoid composition in wild-growing and cultivated plants

Since our dataset included both wild-growing plants, drug-type landraces and a multipurpose landrace, we also investigated differences in cannabinoid composition between those groups of *Cannabis* plants.

Kruskal–Wallis test showed significant differences in tTHC/tCBD ratios between the three groups of plants (p-value < 0.001), and the pairwise Wilcoxon rank sum exact test revealed that significant difference exists only between drug-type landraces and the wild-growing and the multipurpose landrace, while the ratios of the last two groups did not differ significantly (Fig. 7, Table S11).

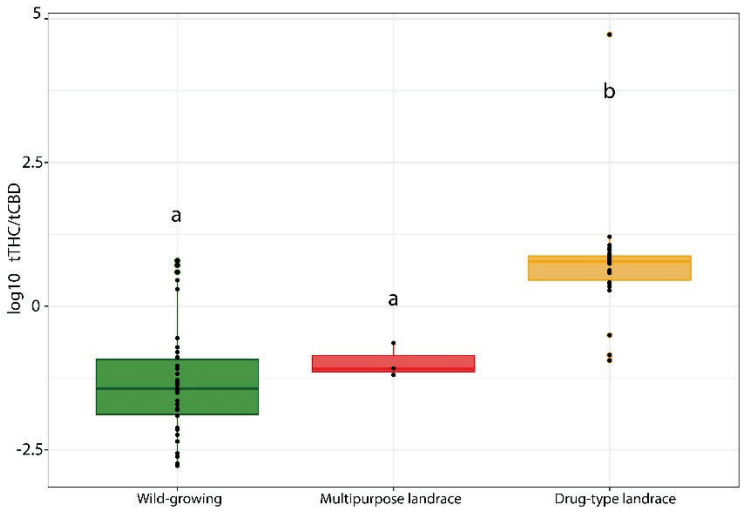


Figure 7. The $\log_{10}$ values of the tTHC/tCBD ratios of analysed samples and the division into the three groups (i.e., wild-growing plants (green) and multipurpose (red) and drug-type (yellow) landrace plants). The letters mark significant differences between the groups calculated using pairwise Wilcoxon rank sum exact test.

When considering complete cannabinoid composition, the PCA showing the first two PC axis revealed relatively good separation of samples among wild-growing plants and drug-type plants, while the three samples of the multipurpose landrace clustered together with wild-growing samples (Fig. 8). The boxplots visualising concentrations of individual cannabinoids in leaves of male and female plants show that the drug-type landraces are characterised by higher CBC, tCBG, tTHC and tTHCV concentrations, while the opposite is true for most of the wild-growing samples. Leaves of the multipurpose landrace from Nepal (NO), revealed the most diverse profile, with high concentrations of CBC, tCBD, tCBDV, and tTHCV (Fig. S10a). Concentrations of cannabinoids in male inflorescences revealed a similar pattern, with drug-type landraces containing higher concentrations of CBC, tTHC, and tTHCV, and male inflorescences of wild-growing plants showing higher concentrations of tCBD and tCBDV (Fig. S8b). The one male inflorescence from the multipurpose landrace revealed high concentrations of tCBD, tCBDV and tCBG (Fig. S10b).

Since differences in cannabinoid composition were found, we also tested if the PLS-DA model would perform better when classifying samples into different use types than phylogeographic groups. Due to low number of multipurpose landrace and female inflorescences samples we only tested it on the subset of leaf samples from wild-growing and drug-type plants, and on the subset of male inflorescences from wild-growing and drug-type plants. The PLS-DA model based on the analysed

cannabinoids revealed strong predictive power for both leaves and male inflorescence, with the calibration AUC values of 0.96 and 0.93 for cross-validation in both cases (Fig. S11).

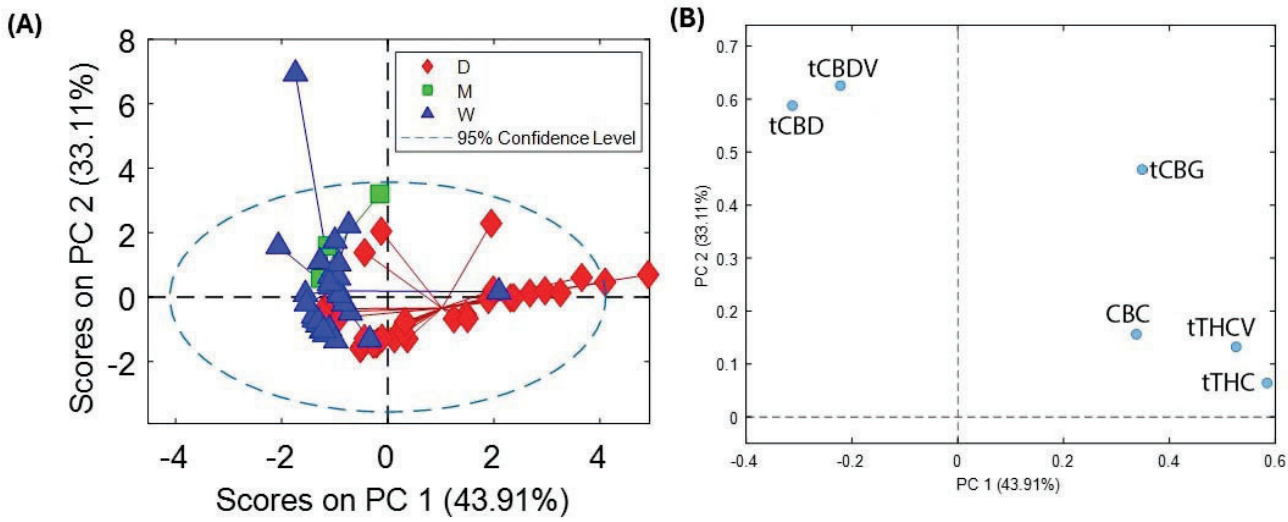


Figure 8. The first two PC axis, showing samples of wild-growing (W), drug-type landrace (D) and multipurpose landrace (M) cannabis plants **(A)** and the loading vectors of the cannabinoids analysed **(B)**. The six cannabinoids are presented in the loadings plot as the sum of their acidic and non-acidic forms, except for CBC (e.g., CBDVA + CBD = tCBDV, CBDA + CBD = tCBD, THCVA + THCV = tTHCV, THCA + Δ9-THC + Δ8-THC = tTHC, and CBGA + CBG = tCBG).

DISCUSSION

Cannabinoid composition of the analysed samples

This study aimed to investigate the phytochemical composition of male and female inflorescences and leaves from wild-growing and cultivated landrace *Cannabis* plants, with a focus on cannabinoid content. To minimize environmental influences on variability between populations, plants were cultivated under controlled conditions. Despite this, our results revealed substantial variability in both total cannabinoid concentration and composition among the analysed samples (Fig. 1), suggesting high heterogeneity in wild-growing and landrace *Cannabis* accessions. Similar findings were reported in a study examining wild-growing *Cannabis* populations in Nebraska (Busta *et al.*, 2022).

We also observed different correlation patterns among the concentration of individual cannabinoids. The tCBG was positively correlated with all analysed cannabinoids, which aligns with its role as a primary precursor in cannabinoid biosynthesis (Fig. 2). A strong positive correlation was also observed

between tCBD and tCBDV, as well as between tTHC and tTHCV (Fig. 2). CBDV and THCV are analogues of CBD and THC, respectively, differing by the presence of an n-propyl side chain instead of an n-pentyl side chain. The biosynthetic pathway of CBDVA and THCVA diverges early from that of CBDA and THCA since they are synthesised from CBGVA instead of CBGA (de Meijer *et al.*, 2009). However, THCA and CBDA synthases are believed to lack substrate selectivity and can convert CBGVA to CBDVA and THCVA, respectively (de Meijer *et al.*, 2009). This could explain the high correlations observed in this study. Both THC and tTHC also showed positive correlations with CBC (Fig. 2, S1). This contrasts with earlier findings by Hillig and Mahlberg (2004), who reported an insignificant correlation between THC and CBC. Conversely, they observed a strong correlation between CBC and CBD, a relationship that we did not detect (Fig. 2, S1). Mudge (2019) reported yet another pattern, identifying slight negative correlations between CBC and THCA, a slight positive correlation with CBDA, and considerable positive correlations between CBC, THC, and CBD. Variations in CBC correlation patterns across studies may stem from differences in the selection of accessions analysed. Small (2015) suggested that CBC is commonly associated with narcotic, high-THC strains originating in Africa. The positive correlation between CBC and tTHC observed in our study could be influenced by the fact that three of the five drug-type accessions analysed are of African origin or closely related to them, belonging to the same Indoafrica phylogeographic group (Balant *et al.*, 2024). We also observed a negative correlation between tTHC and tCBD. Previous studies with similar results suggested that this negative correlation might indicate a competitive biosynthesis pathway where the metabolic energy is directed toward either THCA or CBDA production, but not in both simultaneously (Ren *et al.*, 2021; Yoosefzadeh Najafabadi & Torkamaneh, 2024).

The tTHC/tCBD ratio

Because the dataset did not include the percentage of cannabinoids per dry weight, we calculated tTHC/tCBD ratios and classified the samples into chemotypes using the $\log_{10}$-transformed THC/CBD ratio method of Hillig and Mahlberg (2004). We found only 4 samples that were classified as chemotype I, while the most were classified as chemotype III, followed by chemotype II. Apart from the few samples that were classified as chemotype I, the ratio values we obtained for this group were also relatively low, compared to some other studies. For example, Hillig and Mahlberg (2004) reported ratio values around 50 for chemotype I plants, whereas the values in our study were mostly around 12 (Fig. 5). The lack of samples in our study classified as THC-dominant chemotype I - either for leaves or inflorescences - highlights the singularity of our sampling compared to modern *Cannabis* cultivars, which are often selectively bred for high THC content. This distinctiveness likely reflects the inclusion

of wild-growing plants and traditional landraces in our study, which have undergone less selective breeding for high THC content. Part of the difference may stem from the methodologies used: Hillig and Mahlberg (2004) analysed dried inflorescences, where non-acidic cannabinoids were more prominent, while we used freshly frozen material, which had very low levels of non-acidic cannabinoids. Additionally, they analysed the samples using gas chromatography, where the samples are vaporized, and the acidic forms of cannabinoids cannot be reliably analysed. We did not find any samples with the presence of cannabinoid CBGM, and therefore no samples could be classified as type IV as defined by Small *et al.* (1975). Similarly, no samples displayed CBG dominance or negligible cannabinoid levels, corresponding to chemotypes IV and V, as defined by Mandolino and Carboni (2004), respectively. These findings suggest that such plants are likely products of modern cultivar selection and are uncommon in wild-growing plants and traditional landraces.

Differences in cannabinoid composition of leaves and inflorescences in male and female plants

Significant differences were observed in the cannabinoid composition of leaves and inflorescences within individual plants. On average, inflorescences contained 2.52 times more cannabinoids than leaves, with these differences generally being more pronounced in female plants compared to male plants. This variation provided sufficient discriminatory power for the PLS-DA model to successfully predict sample identity as either leaves or inflorescences. However, the observed differences were smaller than those reported in previous studies, where inflorescences were found to contain 10- to 20-fold higher cannabinoid levels than leaves (Bernstein *et al.*, 2019; Jin *et al.*, 2020). This discrepancy may reflect the characteristics of the accessions analysed, which, in our case, were not modern cultivars selectively bred for high cannabinoid production. Additionally, the smaller differences between leaves and inflorescences in our study could be affected by sampling inflorescences early in the flowering stage, as similar lower values have been reported in studies examining immature or young inflorescences and leaves (i.e., 2.2-fold, Nagy *et al.*, 2019; 2.5-fold, Park *et al.*, 2022). Sampling later might have yielded higher cannabinoid concentrations in the inflorescences, which could result in larger overall differences. While total and most individual cannabinoid concentrations differed significantly between leaves and inflorescences, CBC levels showed minimal variation (Table S5), consistent with findings by Bernstein *et al.* (2019), who reported smaller differences in CBC levels between leaves and inflorescences compared to other cannabinoids.

While the composition and cannabinoid concentration showed significant differences between the inflorescences and leaves, this was not the case for tTHC/tCBD ratios. Previous studies have shown that the THC/CBD ratio in *Cannabis* plants is determined early and remains stable throughout the life

cycle, from young leaves to the inflorescences (Pacifico *et al.*, 2008; Aizpurua-Olaizola *et al.*, 2016; Jin *et al.*, 2020). In line with this, we found that leaves and inflorescences of individual plants generally clustered into the same chemotype, except for five male individuals in which either the leaves or the inflorescences exhibited a higher tTHC/tCBD ratio than the other.

We also investigated the differences in cannabinoid composition of leaves from male and female plants. Using the PLS-DA model, we were not able to differentiate between leaves coming from male and female plants (Fig. 3b, S3a-c), which is consistent with previous studies (Pacifico *et al.*, 2008; Li *et al.*, 2022). In contrast, the differentiation was possible for male and female inflorescences (Fig. 3c, S3d-f). Busta *et al.* (2022) found that the male inflorescences produced on average 40% of the total cannabinoids compared to female ones. The differences observed in our study were less pronounced, as male inflorescences produced, on average, 73% of the cannabinoids found in female inflorescences. Some other studies have even reported no significant differences between male and female inflorescences (Ghosh *et al.*, 2024). Diverse results were also recorded by Small *et al*. (1975), who observed that the differences varied depending on the phenotype. In their study, THC-dominant plants showed minimal differences between male and female inflorescences, whereas other phenotypes exhibited larger differences. Since most of the accessions analysed in our study were not THC-dominant, our findings align with these later results. This suggests that male plants could also serve as a notable source of cannabinoids.

Differences in cannabinoid composition between phylogeographic groups, and wild-growing and cultivated plants

Consistent with previous research (Small *et al.*, 1975; Hillig & Mahlberg, 2004; McPartland & Small, 2020), we found significant differences in tTHC/tCBD ratios among phylogeographic groups. The Indoafrica and C & S China and Himalayas groups had significantly higher tTHC/tCBD ratios compared to the other groups, generally found in more northern regions (Fig. 5). These findings align with those of Hillig and Mahlberg (2004), who reported that *Cannabis* populations originating from S and SE Asia and Africa more frequently exhibited a high THC/CBD ratio, while populations from more northern regions of Eurasia contained lower proportion of plants with high THC/CBD ratio. Significant differences were also observed between the C & S China and Himalayas group and the Indoafrica group. This divergence likely reflects the selection of plants analysed in each group: the C & S China and Himalayas group comprises a mix of wild-growing plants, multipurpose, and drug-type landraces, whereas the Indoafrica group consists exclusively of drug-type landraces, which would explain its elevated tTHC/tCBD ratios.

Multivariate analysis of the complete cannabinoid composition yielded similar results (Fig. 6, S4). Samples from Indoafrica group (characterised by CBC, and tTHC) were distinct from those of the Caucasus and Mediterranean group, Eurosiberia and W Mongolia group, and N China and E Mongolia group (characterized mainly by higher concentrations of tCBD and tCBDV), while the samples from C & S China and Himalayas group were positioned between these clusters. Dividing the dataset into three subsets—leaves, male inflorescences, and female inflorescences—to reduce variation introduced by combining different tissue types, produced similar results. The diversity in the C & S China and Himalayas group is driven, at least in part, by its varied composition of wild-growing, multipurpose, and drug-type landraces. The accessions analysed from this group included both either tTHC-dominant and tCBD-dominant plants. Notably, such variability can be observed even within the same population of wild-growing plants (e.g., IKL) or within landraces (e.g., ISAT; Fig. 1). These findings align with those of McPartland and Small (2020), who reported significant variability in THC/CBD ratios among wild-type *Cannabis* plants from S Asia. Ghosh *et al.* (2024) reported that northern Indian *Cannabis* populations are dominated by intermediate chemotypes (plants with a balanced concentration of CBD and THC), and that CBD-rich fibre-type accessions are rarely found in nature. Most of the wild-growing samples from North India analysed in this study also belong to intermediate chemotype II. However, we did identify some populations (i.e., ISAT and IKL) with particular individuals in which tCBD was the dominant cannabinoid. These populations, with such contrasting cannabinoid compositions, could represent an ancient escape of multipurpose landraces common in the region, where plants were not selected for high concentrations of a single cannabinoid (e.g., THC). Alternatively, they may reflect the characteristics of a truly wild *Cannabis* population. Ren *et al.* (2021) suggested that the ancestral state of *Cannabis* involved both CBDAs and THCAs genes in a functional state, with the loss of one gene occurring during domestication based on selection for specific uses. However, since most of the other wild-growing populations from Bangladesh, Mongolia, Armenia, and Kazakhstan (Fig. 1) did not exhibit such extreme variability, the former explanation seems more likely.

The Caucasus and Mediterranean group, while less diverse than C & S China and Himalayas group, similarly includes both wild-growing plants and drug-type landraces, which exhibit contrasting cannabinoid profiles. Notably, significant variability was observed within the Moroccan drug-type landrace, where both a tCBD-dominant and a tTHC-dominant plant was identified (Fig. 1). In contrast, plants from the Indoafrica group (also drug-type landraces) displayed a far more uniform cannabinoid composition and differed from all the remaining groups by significantly lower concentrations of tCBD and tCBDV in leaves. This differences in the intra-population variability of cannabinoid composition between drug-type landraces from different phylogeographic groups may be linked to the form of drug product that is traditionally produced from these plants. Indoafrica plants are typically consumed as

individual dried inflorescences, known as 'ganja', allowing for individual selection of THC-dominant plants with low CBD concentrations. Conversely, Moroccan landraces are predominantly used for 'hashish' production, which involves mixing and processing the resin from multiple plants. The latter method does not allow the individual selection of THC-dominant plants and results in mixed population of THC- and CBD-dominant plants (Clarke, 1998; Clarke & Merlin, 2013).

Previous studies have also reported differences in cannabinoid composition between populations from different geographic regions. Small *et al.* (1975) observed that plants from latitudes south of 30°N generally had a THC-dominant cannabinoid composition, while those from northern regions were characterized by higher CBD levels, leading them to propose the classification of two subspecies (*Cannabis sativa* subsp. *sativa* and *C. sativa* subsp. *indica*) based on this criterion. Hillig and Mahlberg (2004) observed a similar pattern, but proposed two separate species (*C. sativa* and *C. indica*) based on the frequency of the dominant cannabinoid in *Cannabis* populations. For *C. sativa*, less than 25% of individuals in a population were classified as THC-dominant chemotype I, while in *C. indica* populations, more than 25% of individuals were chemotype I. They also noted that while CBD-dominant plants (*C. sativa*) are not found below 35°N, THC-dominant plants (*C. indica*) can occur further north, in regions like Afghanistan and Pakistan. Our findings generally align with their results, as most populations where the dominant cannabinoid was tTHC originated from areas south of 30°N, while in northern regions tCBD was generally the more dominant cannabinoid. However, our results show that it is difficult to determine a clear separation line. For example, the population from Nepal (NO), which originates from approximately 29°N, is tCBD-dominant, while populations from northern India (IKL) and Morocco (MAR), both originating north of 30°N, contain both tTHC- and tCBD- dominant individuals.

Minor cannabinoids have also previously been linked to specific geographic distributions. Small (2015) found that CBC is frequently found in high-THC strains of *C. sativa* from Africa. While in our case the CBC did not differ significantly between leaves of any of the phylogeographic groups (Table S10), the PCA suggests a possible association between the tTHC and CBC (Fig. S4). Previous studies have shown that THCV could be characteristic for plants from South and Southeast Asia, Afghanistan, and Africa, while CBDV is more common in plants from Central Asia (Hillig & Mahlberg, 2004; Small, 2015; McPartland & Small, 2020). In our analysis, we observed significant differences in the concentrations of tTHCV in leaves from African and Asian samples. The samples with the highest proportions of tTHCV—ranging from 1-2% of total cannabinoids—were from India (IK, IKL, ISAT) and Ethiopia (ETH3; Table S3). However, since tTHCV was also found in an individual from Morocco and showed a strong correlation with tTHC, it may not be a reliable distinguishing characteristic. The tCBDV concentration was highest in samples from Nepal (NO), Armenia (AM15) and Bulgaria (BG3), with values of

approximately 20-30% of total cannabinoids, that also contained high proportions of tCBD (Table S3). The Eurosiberia and W Mongolia group, and N China and E Mongolia groups showed uniformity in cannabinoid profiles, both presenting very low tCBG, tTHC and tTHCV concentrations in leaves and male inflorescences (Fig. S5, S8). Small and Beckstead (1975) and Hillig and Mahlberg (2004) reported that plants from Northeast Asia (NE China, Japan, South Korea) and hemp landraces of *Cannabis* were characterized by the presence of the minor cannabinoid CBGM, suggesting it could serve as a useful chemotaxonomic marker. Although we included samples from Eastern Mongolia, we did not detect CBGM or CBGMA in our analysis. Unfortunately, we did not analyse samples from NE Asian hemp landraces, which could have revealed the presence of these compounds.

Despite significant differences in cannabinoid composition, the variability and limited number of samples hindered the PLS-DA model from reliably distinguishing between phylogeographic groups. Previous studies have noted that cannabinoid composition is not a dependable indicator of geographic origin, particularly for populations selected and cultivated for specific purposes (Hillig & Mahlberg, 2004; Busta *et al.*, 2022). This study analysed a highly diverse sample set, including wild-growing plants, drug-type, and multipurpose landraces, which likely contributed to the variability in cannabinoid content across high- and low-THC samples. In contrast, the PLS-DA model was much more effective at classifying samples as either wild-growing or drug-type landrace, whether analysing leaves or male inflorescences. While the primary variation observed among the samples was in the tTHC concentration, the domesticated varieties did not necessarily show higher total cannabinoid concentrations compared to wild-growing plants; in fact, the highest concentration was found in a wild-growing accession from Armenia. This could be due to the early sampling, as many drug-type accessions came from lower latitudes where plants have longer life cycles and inflorescences take more time to mature, potentially resulting in lower cannabinoid concentrations at collection time. Alternatively, it could simply be a result of the inclusion of traditional landraces in our study, rather than modern cultivars subjected to intense selection for high THC or CBD content. The PCA analysis revealed that wild-growing plants were primarily characterized by the presence of tCBD and tCBDV, with only one sample from Bangladesh (BNG) clustering with drug-type landraces. The multipurpose landrace from Nepal showed a phytochemical composition more similar to wild-growing plants, being predominantly characterized by tCBD and tCBDV.

CONCLUSIONS

This study underscores the variability in cannabinoid composition across wild-growing *Cannabis* plants, and multipurpose and drug-type landraces from diverse phylogeographic regions. While cannabinoid composition was not sufficient to precisely distinguish geographic origins, our results

revealed a global trend: tCBD-dominant plants were predominantly found in regions north of 30°N, while tTHC-dominant plants were mostly distributed south of 30°N. However, some individuals and populations did not follow this pattern, pointing to the limitations of previous taxonomic classifications of *Cannabis* based mainly on phytochemical composition. In contrast, we observed clearer differentiation between cultivated drug-type landraces and wild-growing plants, reflecting the influence of centuries of cultivation practices and selective breeding in *Cannabis*. Wild-growing populations, often overlooked, exhibited some of the highest total cannabinoid concentrations, highlighting their potential as valuable genetic resources. However, the small sample size in this study limits the scope of our conclusions and emphasizes the need for further research with larger datasets to enable more accurate interpretation of the observed patterns. Nevertheless, our findings highlight the importance of preserving traditional landraces and wild accessions, which offer unique phytochemical diversity that holds promise for advancing breeding programs, conservation strategies, and the broader utilization of *Cannabis*.

Acknowledgements

We would like to thank Joan Uriach Marsal † (Uriach Laboratories), CIJA Preservations SL, and Cannaflos – Gesellschaft für medizinisches Cannabis mbH for additional financial support, and CIJA Preservations SL for providing valuable seed accessions. We would also like to thank Carlos Sáez, Carlos Ribelles, David Benavente López, Magsar Urgamal, Shagdar Tsooj, Nusrat Sultana, Marine Oganesian, and Nina Stepanyan-Gandilyan for help with the sample collection and the cultivation process. This research was supported by projects WECANN (CGL2017-84297-R, Ministerio de Ciencia, Innovacion y Universidades), Generalitat de Catalunya (CLT051, 2017SGR1116, and 2021SGR00315), Institut d'Estudis Catalans (PRO2020-2024-S02-VALLES). MB benefited from FPI predoctoral contract funded by the Ministerio de Ciencia, Innovacion y Universidades (PRE2018-083226). AG benefited from a postdoctoral grant of the Universitat de Barcelona funded by NextGeneration EU funds (Margarita Salas 2022–2024).

Competing interests

None declared.

Funding

This research was funded by Project WECANN (grant number CGL2017-84297-R, AEI/FEDER, UE) from the Spanish government, Generalitat de Catalunya (grant numbers CLT051, 2017SGR1116,

2021SGR00315), Institut d'Estudis Catalans (grant number PRO2020-2024-S02-VALLES), FPI predoctoral contract of the Ministerio de Ciencia, Innovación y Universidades (grant number PRE2018-083226) and postdoctoral grant of the Universitat de Barcelona funded by NextGeneration EU funds (Margarita Salas 2022–2024).

Author contributions

MB, TG and DV conceived the study, with inputs from JV and AG in a preliminary design phase of the project. MB, JV and AG carried out collection of samples in the field. MB, AP and TG were involved with the cultivation of plants and the sample collection. NLS, SMM carried out the extraction of samples and GQ carried out the cannabinoid analysis. MB and GQ carried out the statistical analysis. MB wrote the first draft of the paper, which all authors reviewed, commented on, and edited.

Data availability

The dataset and supplementary material are available at Zenodo (https://doi.org/10.5281/zenodo.14604544).

REFERENCES

Abel EL. **1980**. *Marihuana: The First Twelve Thousand Years*. New York, USA: Springer US.

Aizpurua-Olaizola O, Soydaner U, Öztürk E, Schibano D, Simsir Y, Navarro P, Etxebarria N, Usobiaga A. **2016**. Evolution of the cannabinoid and terpene content during the growth of *Cannabis sativa* plants from different chemotypes. *Journal of Natural Products* **79**: 324–331.

Baker PB, Gough TA, Taylor BJ. **1980**. Illicitly imported Cannabis products: Some physical and chemical features indicative of their origin. *Bulletin on Narcotics* **32**: 31–40.

Balant M, Gras A, Gálvez F, Garnatje T, Vallès J, Vitales D. **2021a**. CANNUSE, a database of traditional *Cannabis uses*—an opportunity for new research. *Database* **2021**: baab024.

Balant M, Gras A, Ruz M, Vallès J, Vitales D, Garnatje T. **2021b**. Traditional uses of *Cannabis*: An analysis of the CANNUSE database. *Journal of Ethnopharmacology* **279**: 114362.

Balant M, Vitales D, Wang Z, Barina Z, Fu L, Gao T, Garnatje T, Gras A, Hayat MQ, Oganesian M, *et al.* 2024. Integrating target capture with whole genome sequencing of recent and natural history collections to explain the phylogeography of wild-growing and cultivated *Cannabis*. *bioRxiv*.

Barni-Comparini I, Ferri S, Centini F. **1984**. Cannabinoid level in the leaves as a tool for the early discrimination of cannabis chemiovariants. *Forensic Science International* **24**: 37–42.

Bernstein N, Gorelick J, Koch S. **2019**. Interplay between chemistry and morphology in medical cannabis (*Cannabis sativa* L.). *Industrial Crops and Products* **129**: 185–194.

Booth JK, Bohlmann J. **2019**. Terpenes in *Cannabis sativa* – From plant genome to humans. *Plant Science* **284**: 67–72.

Busta L, Dweikat I, Sato SJ, Qu H, Xue Y, Zhou B, Gan L, Yu B, Clemente TE, Cahoon EB, *et al.* 2022. Chemical and genetic variation in feral *Cannabis sativa* populations across the Nebraska climate gradient. *Phytochemistry* **200**: 113206.

Clark MN, Bohm BA. **1979**. Flavonoid variation in *Cannabis* L. *Botanical Journal of the Linnean Society* **79**: 249–257.

Clarke RC. **1998**. *Hashish!* Los Angeles, CA: Red Eye Press.

Clarke RC, Merlin MD. **2013**. Cannabis*: Evolution and ethnobotany*. Berkley, Los Angeles, London: University of California Press.

ElSohly MA, Radwan MM, Gul W, Chandra S, Galal A. **2017**. Phytochemistry of *Cannabis sativa* L. In: Kinghorn A, Falk H, Gibbons S, Kobayashi J, eds. Phytocannabinoids. Progress in the chemistry of organic natural products, vol 103. 1–36.

ElSohly MA, Slade D. **2005**. Chemical constituents of marijuana: The complex mixture of natural cannabinoids. *Life Sciences* **78**: 539–548.

Fetterman PS, Keith ES, Waller CW, Guerrero O, Doorenbos NJ, Quimby MW. **1971**. Mississippi-grown *Cannabis sativa* L.: Preliminary observation on chemical definition of phenotype and variations in tetrahydrocannabinol content *versus* age, sex, and plant part. *Journal of Pharmaceutical Sciences* **60**: 1246–1249.

Ghosh D, Kundu A, Chaudhary N, Singh S, Shanker K, Kumar B, Kumar N. **2024**. Morpho-anatomical, phytochemical, and molecular diversity of North-Indian *Cannabis sativa* L. germplasms for CBD-rich genotypes. *The Nucleus*.

Grassi G, McPartland JM. **2017**. Chemical and morphological phenotypes in breeding of *Cannabis sativa* L. In: *Cannabis sativa* L. - Botany and Biotechnology. Cham: Springer International Publishing, 137–160.

Hanuš LO, Meyer SM, Muñoz E, Taglialatela-Scafati O, Appendino G. **2016**. Phytocannabinoids: a

unified critical inventory. *Natural Product Reports* **33**: 1357–1392.

Hazekamp A, Fischedick JT. **2012**. Cannabis - from cultivar to chemovar. *Drug testing and Analysis* **4**: 660–667.

Hazekamp A, Fischedick JT, Llano MD, Lubbe A, Ruhaak RL. **2010**. Chemistry of Cannabis. In: Liu H-W (Ben), Mander L, eds. Comprehensive Natural Products Chemistry II. Oxford: Elsevier Ldt., 1033–1084.

Hazekamp A, Tejkalová K, Papadimitriou S. **2016**. Cannabis: From cultivar to chemovar II - A metabolomics approach to cannabis classification. *Cannabis and Cannabinoid Research* **1**: 202–215.

Hemphill JK, Turner JC, Mahlberg PG. **1980**. Cannabinoid content of individual plant organs from different geographical strains of *Cannabis sativa* L. *Journal of Natural Products* **43**: 112–122.

Herwig N, Utgenannt S, Nickl F, Möbius P, Nowak L, Schulz O, Fischer M. **2024**. Classification of cannabis strains based on their chemical fingerprint—A broad analysis of chemovars in the German market. *Cannabis and Cannabinoid Research* **00**: 000–000.

Hillig KW, Mahlberg PG. **2004**. A chemotaxonomic analysis of cannabinoid variation in *Cannabis* (Cannabaceae). *American Journal of Botany* **91**: 966–975.

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, Dai K, Xie Z, Jie C. **2020**. Secondary metabolites profiled in cannabis inflorescences, leaves, stem barks, and roots for medicinal purposes. *Scientific Reports* **10**: 3309.

Kajima M, Piraux M. **1982**. The biogenesis of cannabinoids in *Cannabis sativa*. *Phytochemistry* **21**: 67–69.

Koltai H, Namdar D. **2020**. Cannabis phytomolecule 'entourage': From domestication to medical use. *Trends in Plant Science* **25**: 976–984.

Li C-R, Yang L-X, Guo Z-F, Yang H, Zhang Y, Wang Y-M, Zhang G-Z, Li P, Gao W. **2022**. LC-MS-based untargeted metabolomics reveals chemical differences of *Cannabis* leaves from different regions of China. *Industrial Crops and Products* **176**: 114411.

Liktor-Busa E, Keresztes A, Lavigne J, Streicher JM, Largent-milnes TM. **2021**. Analgesic potential of terpenes derived from *Cannabis sativa*. *Pharmacological Reviews* **73**: 98–126.

Livingston SJ, Quilichini TD, Booth JK, Wong DCJ, Rensing KH, Laflamme-Yonkman J, Castellarin SD,

Bohlmann J, Page JE, Samuels AL. **2020**. Cannabis glandular trichomes alter morphology and metabolite content during flower maturation. *Plant Journal* **101**: 37–56.

Mandolino G, Carboni A. **2004**. Potential of marker-assisted selection in hemp genetic improvement. *Euphytica* **140**: 107–120.

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

Mechoulam R, Hanuš L. **2001**. The cannabinoids: An overview. Therapeutic implications in vomiting and nausea after cancer chemotherapy, in appetite promotion, in multiple sclerosis and in neuroprotection. *Pain Research and Management* **6**: 67–73.

de Meijer EPM, Bagatta M, Carboni A, Crucitti P, Moliterni VMC, Ranalli P, Mandolino G. **2003**. The inheritance of chemical phenotype in *Cannabis sativa* L. *Genetics* **163**: 335–346.

de Meijer EPM, Hammond KM, Sutton A. **2009**. The inheritance of chemical phenotype in *Cannabis sativa* L. (IV): cannabinoid-free plants. *Euphytica* **168**: 95–112.

de Meijer EPM, van der Kamp HJ, van Eeuwijk FA. **1992**. Characterisation of *Cannabis* accessions with regard to cannabinoid content in relation to other plant characters. *Euphytica* **62**: 187–200.

Mudge EM. **2019**. *Chemometrics and metabolomics of* Cannabis sativa *L. PhD thesis.* Okanagan: The University of British Columbia.

Nagy DU, Cianfaglione K, Maggi F, Sut S, Dall'Acqua S. **2019**. Chemical characterization of leaves, male and female flowers from spontaneous *Cannabis* (*Cannabis sativa* L.) growing in Hungary. *Chemistry and Biodiversity* **16**: e1800562.

Osafo N, Yeboah OK, Antwi AO. **2021**. Endocannabinoid system and its modulation of brain, gut, joint and skin inflammation. *Molecular Biology Reports* **48**: 3665–3680.

Pacifico D, Miselli F, Carboni A, Moschella A, Mandolino G. **2008**. Time course of cannabinoid accumulation and chemotype development during the growth of *Cannabis sativa* L. *Euphytica* **160**: 231–240.

Park S-H, Pauli CS, Gostin EL, Staples SK, Seifried D, Kinney C, Vanden Heuvel BD. **2022**. Effects of short-term environmental stresses on the onset of cannabinoid production in young immature flowers of industrial hemp (*Cannabis sativa* L.). *Journal of Cannabis Research* **4**: 1.

De Prato L, Ansari O, Hardy GESJ, Howieson J, O'Hara G, Ruthrof KX. **2022**. Morpho-physiology and cannabinoid concentrations of hemp (*Cannabis sativa* L.) are affected by potassium fertilisers and

microbes under tropical conditions. *Industrial Crops and Products* **182**: 114907.

Quaghebeur K, Coosemans J, Toppet S, Compernolle F. **1994**. Cannabiorci- and 8-chlorocannabiorcichromenic acid as fungal antagonists from *Cylindrocarpon olidum*. *Phytochemistry* **37**: 159–161.

R Core Team. R: A Language and Environment for Statistical Computing. 2022. *Available online at: https://www.R-project.org*.

Reichel P, Munz S, Hartung J, Graeff-Hönninger S. **2024**. Harvesting Light: The interrelation of spectrum, plant density, secondary metabolites, and *Cannabis sativa* L. yield. *Agronomy* **14**: 2565.

Ren G, Zhang X, Li Y, Ridout K, Serrano-Serrano ML, Yang Y, Liu A, Ravikanth G, Nawaz MA, Mumtaz AS, *et al.* 2021. Large-scale whole-genome resequencing unravels the domestication history of *Cannabis sativa*. *Science Advances* **7**: eabg2286.

Ribeiro R, Silva YC, Finotti R, Carneiro GRA, Santos GRC dos, Pereira HMG, Padilha MC, Veiga-Junior VF. **2024**. Can Trema micranthum (L.) Blume (Cannabaceae) become a new source of cannabinoids? A UHPLC- HRMS/MS analysis. *Research Square; Preprint https://doi.org/10.21203/rs.3.rs-5125664/v1*.

Russo EB. **2011**. Taming THC: Potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. *British Journal of Pharmacology* **163**: 1344–1364.

Saloner A, Bernstein N. **2021**. Nitrogen supply affects cannabinoid and terpenoid profile in medical cannabis (*Cannabis sativa* L.). *Industrial Crops and Products* **167**: 113516.

Sirikantaramas S, Taura F, Tanaka Y, Ishikawa Y, Morimoto S, Shoyama Y. **2005**. Tetrahydrocannabinolic acid synthase, the enzyme controlling marijuana psychoactivity, is secreted into the storage cavity of the glandular trichomes. *Plant and Cell Physiology* **46**: 1578–1582.

Small E. **2015**. Evolution and classification of *Cannabis sativa* (marijuana, hemp) in relation to human utilization. *Botanical Review* **81**: 189–294.

Small E, Beckstead HD. **1973**. Cannabinoid phenotypes in *Cannabis sativa*. *Nature* **245**: 147–148.

Small E, Beckstead HD, Chan A. **1975**. The evolution of cannabinoid phenotypes in *Cannabis*. *Economic Botany* **29**: 219–232.

Small E, Cronquist A. **1976**. A practical and natural taxonomy for *Cannabis*. *Taxon* **25**: 405–435.

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.

Stasiłowicz A, Tomala A, Podolak I, Cielecka-Piontek J. **2021**. *Cannabis sativa* L. as a natural drug meeting the criteria of a multitarget approach to treatment. *International Journal of Molecular Sciences* **22**: 778.

Tremlová B, Mikulášková HK, Hajduchová K, Jancikova S, Kaczorová D, Zeljković SĆ, Dordevic D. **2021**. Influence of Technological Maturity on the Secondary Metabolites of Hemp Concentrate (*Cannabis sativa* L .). *Foods* **10**: 1418.

Turner CE, Elsohly MA, Cheng PC, Lewis G. **1979**. Constituents of *Cannabis sativa L*., XIV: Intrinsic problems in classifying cannabis based on a single cannabinoid analysis. *Journal of Natural Products* **42**: 317–319.

van Velzen R, Schranz ME. **2021**. Origin and evolution of the cannabinoid oxidocyclase gene family. *Genome Biology and Evolution* **13**: evab130.

Wickham H. **2016**. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag New York.

Yoosefzadeh Najafabadi M, Torkamaneh D. **2024**. Machine learning-enhanced multi-trait genomic prediction for optimizing cannabinoid profiles in cannabis. *The Plant Journal*: 1–12.

Zandkarimi F, Decatur J, Casali J, Gordon T, Skibola C, Nuckolls C. **2023**. Comparison of the cannabinoid and terpene profiles in commercial cannabis from natural and artificial cultivation. *Molecules* **28**: 833.

CHAPTER 4

TRADITIONAL USES OF *CANNABIS*

4.1 CANNUSE, a database of traditional *Cannabis* uses—an opportunity for new research

The following chapter is presented in the form of a published article:

Balant, M., Gras, A., Francisco, G., Garnatje, T., Vallès, J., & Vitales, D. (2021). CANNUSE, a database of traditional *Cannabis* uses—an opportunity for new research. *Database*, baab024.

DOI: https://doi.org/10.1093/database/baab024

Database, 2021, 1–9
doi:10.1093/database/baab024
Original article

Original article

CANNUSE, a database of traditional *Cannabis* uses—an opportunity for new research

Manica Balant[1,*,†], Airy Gras[1,†], Francisco Gálvez[2], Teresa Garnatje[1], Joan Vallès[3,4] and Daniel Vitales[1,3]

[1]Institut Botànic de Barcelona (IBB, CSIC-Ajuntament de Barcelona), Passeig del Migdia s/n, Barcelona, Catalonia 08038, Spain, [2]Centro de Investigación y Desarrollo de Recursos Científicos BioScripts, 31, Av. de la Reina Mercedes, Sevilla, Andalusia 41012, Spain, [3]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, Avinguda de Joan XXIII, 27-31, Barcelona, Catalonia 08028, Spain and [4]Institut d'Estudis Catalans (IEC), Carrer del Carme, 47, Barcelona, Catalonia 08001, Spain

*Corresponding author: Tel: (+34) 93 289 06 11; Fax: (+34) 93 289 06 14; Email: manica.balant@ibb.csic.es

†These authors have contributed equally to this work and share first authorship.

Citation details: Balant, M., Gras, A., Gálvez, F. *et al.* CANNUSE, a database of traditional *Cannabis* uses—an opportunity for new research. *Database* (2021) Vol. 2021: article ID baab024; doi:10.1093/database/baab024

Received 17 February 2021; Revised 9 April 2021; Accepted 19 April 2021

Abstract

Cannabis is one of the most versatile genera in terms of plant uses and has been exploited by humans for millennia due to its medicinal properties, strong fibres, nutritious seeds and psychoactive resin. Nowadays, *Cannabis* is the centre of many scientific studies, which mainly focus on its chemical composition and medicinal properties. Unfortunately, while new applications of this plant are continuously being developed, some of its traditional uses are becoming rare and even disappearing altogether. Information on traditional uses of *Cannabis* is vast, but it is scattered across many publication sources in different formats, so synthesis and standardization of these data are increasingly important. The CANNUSE database provides an organized information source for scientists and general public interested in different aspects of *Cannabis* use. It contains over 2300 entries from 649 publications related to medicinal, alimentary, fibre and other uses from different geographical areas and cultures around the world. We believe this database will serve as a starting point for new research and development strategies based on the traditional knowledge.

Database URL: http://cannusedb.csic.es

Introduction

Medicinal plants have almost limitless applications and have traditionally been used to treat several illnesses (1). One of the most commonly used plants is *Cannabis*, being known to humans for thousands of years and showing myriad traditional uses globally. The oldest known record of its medicinal use dates back to 4700 B.P. in China. Many other ancient texts from India, Persia, Egypt, Greece and Rome also contain valuable information about a plethora of other *Cannabis* medicinal uses (2). In the beginning of the 20th century, *Cannabis* became widely regarded as an illegal drug with negative effects, resulting in a general reduction in its use as a medicine (3). As a consequence—despite the long-standing recognition of its positive effects—the scientific interest in this plant steeply declined for several decades, and most of the information on *Cannabis* use was limited to the domain of local popular knowledge. However, in the last 20 years, interest in *Cannabis* research has grown, and several medicinal uses originally discovered by traditional knowledge have been tested and developed for commercial medicine production (4), modern fibre applications (5) and food production (6–8). The latest boost to scientific and technological interest in *Cannabis* has been the recent decriminalization or legalization of its medicinal and recreational use in many countries, which has boomed into a billion-dollar industry in just a few years (9).

Databases are one of the tools that enable gathering information in an organized repository, which facilitates further research. Several *Cannabis* databases have been created in recent years to collect and organize information related to its genomic resources (10–12), clinical applications (13) or commercial strains (14). However, despite copious information available for this plant, no database on the traditional uses of *Cannabis* has been established so far. The studies on new and traditional uses of *Cannabis* are numerous and are increasing daily; research papers on this topic are being published in many journals from various scientific fields and circulations (i.e., both local and international publications). Therefore, much important information about its uses can stay unnoticed by the majority of people interested in the topic. Another problem that makes the comparison of results difficult is the terminology used. While methods for ethnobotanical studies are well developed, it is up to the authors to decide whether to state the effect of the plant, the target ailment or merely the body system being treated. The lack of data integration and standardization makes it difficult to use this information in research, so synthesis and standardization of all these data are becoming more and more important.

Inventorying traditional knowledge on biodiversity is a way of ensuring its conservation—especially urgent in many zones, where it is being eroded—and its possible further uses for human well-being (15–18). Public databases are a powerful instrument of such traditional knowledge preservation and represent an excellent tool to accomplish one of the ethical exigencies of ethnobotanical prospection: to return the knowledge to the society, where it came from (18). Nevertheless, some concerns have been formulated regarding the role of the current forms of conservation of traditional knowledge, one of them being the scarce implication of its holders in the processes and their difficulties to access the information (17, 19). In this respect, it is important that databases provide open access to, among other users, local communities (15). Here, we present the CANNUSE database (available at http://cannusedb.csic.es), which we have elaborated in order to provide a thorough information source that will be useful for the whole society, including the scientific community, legislators and non-professionals interested in any aspect of *Cannabis* use. We have undertaken a comprehensive data collection that enabled us to construct the first global database on *Cannabis* medicinal, alimentary, fibre, psychoactive and other ethnobotanical uses. The main aim of the CANNUSE database is to gather and organize this abundant information on traditional *Cannabis* use in a simple manner. Therefore, we believe that the user-friendly web interface constructed for the database will enable easy access to this reserve of information to any type of public and return the knowledge back to society. We hope this resource will facilitate research and development strategies for drug, food or other *Cannabis* products based on traditional knowledge and enable legislators to take decisions on relevant legal dispositions. The database will also serve to bring to light lesser-known traditional *Cannabis* uses (medicinal and others) that have been thus far overlooked and may be disappearing, enhancing new ethnobotanical studies and perhaps promoting beneficial applications of some of these rare uses. In addition, this organized repository of *Cannabis* data may help detect less obvious connections between specific plant parts and illnesses, which could open novel treatment lines based on *Cannabis* products.

Methods

Publication search

Our publication search was carried out in four major online databases—Scopus, Web of Science, PubMed and Google Scholar, using the following set of keywords and exact terms: *Cannabis* AND ('folk medicine' OR 'traditional medicine' OR 'ethnobotany' OR 'traditional knowledge'). Our search returned over 10000 results (Figure 1).

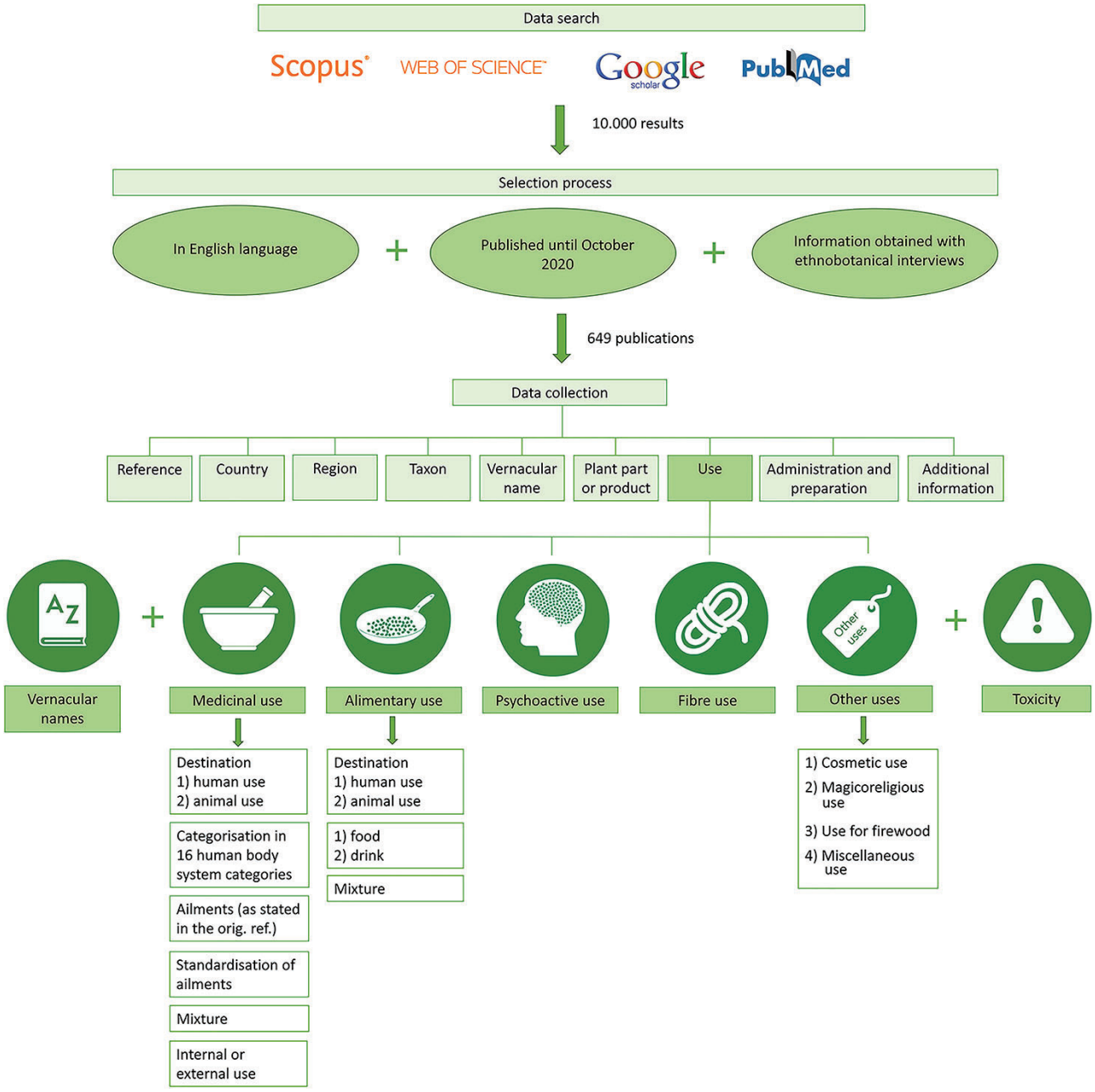


Figure 1. Workflow of the data search, selection process and data collection for the CANNUSE database construction.

During the screening process, we eliminated references that (i) were not published in English language, (ii) were not published by the end of October 2020 and (iii) did not obtain the information included through ethnobotanical interviews. To avoid duplication, information obtained from review papers and books was only used when original research papers could not be found. In further steps, papers containing inconsistencies (e.g. incorrect citations, unclear uses and uses in review papers not matching with the original research papers) were also eliminated. Additional references cited in relevant ethnobotanical papers were added using the snowball method (20).

After filtering and excluding the papers that did not fit our criteria, we obtained a final reference list with 649 publications. Most of these were research papers (607 references, including 6 conference proceedings), but 38 review papers, 2 doctoral theses and 2 master theses were also included. We registered a total of 2330 data entries on traditional uses of *Cannabis*. One data entry is represented by one use quoted in a publication.

Data collection

For each reference, the following information was recorded (Figure 1): (i) type and year of publication, (ii) country, (iii) region, (iv) taxon, (v) vernacular names and (vi) part of the plant (inflorescence, leaf, whole plant, seed, aerial part, stem, bark, root, twig and branch, and other part) or plant product (resin, fibre and other product) used. In this database, the term 'seed' actually refers to the *Cannabis* fruit—a nut (also called achene) (21). In the reviewed literature, this part was referred to with several terms (fruit, young fruit, nut, achene and seed); because the term 'seed' was the most commonly used and generally accepted, the part was referred to in the database under the single term. We also recorded (vii) use categories (medicinal, alimentary, fibre, psychoactive or other), whether *Cannabis* had (viii) animal or human use, if the plant was considered (ix) toxic or noxious (toxicity) or included (x) modes of preparation and administration, whenever they were provided by the authors. When other ingredients (plant, animal or other substance) were added to the alimentary and medicinal preparations, they were categorized as a (xi) mixture. For medicinal uses, (xii) the way of administration (external and internal) was also recorded when possible. Any additional information available (xiii) was also recorded. When vernacular names related to the use of the plant were provided (disease names, names of the recipes or products, etc.), these were included within square brackets.

Structure of the CANNUSE database

The database is structured in five use categories: medicinal (which includes also veterinary), alimentary, fibre, psychoactive and other use. We also added a category for toxicity reports and one for vernacular names (Figure 1). Each database entry was attributed to one or more appropriate categories. For instance, in the case of a reference stating 'traditional drink "thandai" which has a sedative effect and is narcotic', the entry was included in three categories: medicinal use (sedative), alimentary use [drink (*thandai*)] and psychoactive use (narcotic).

Detailed information on CANNUSE categories is described below:

Medicinal use

Medicinal use was divided into human and animal medicinal use. Depending on the reference, the uses were originally formulated in many different ways—sometimes via a name of the relevant disease or condition treated with *Cannabis* (e.g. diabetes) and other times by its putative effect (e.g. antidiabetic). To simplify the search and access to the data, we standardized all the use reports, so they refer the plant's effect, and renamed them according to the Oxford Concise Medical Dictionary (22), but the original use (as stated in the paper) was also retained for easy verification. To make searching faster, human medicinal uses were classified into 16 human body system categories, according to Cook (23) with minor modifications. Categories used in the database are circulatory system and blood disorders, digestive system and nutritional disorders, endocrine system and metabolic disorders, genitourinary system disorders, immune system disorders and neoplasia, infections and infestations, musculoskeletal system disorders and traumas, nervous system and mental disorders, pain and inflammations, poisoning, pregnancy, birth and puerperal disorders, respiratory system disorders, sensory system disorders, skin and subcutaneous tissue disorders, tonic and restorative, and unclassified. Sometimes one use could belong to two system categories (e.g. bladder inflammation is placed under two system categories—pain and inflammations and genitourinary system disorders). Veterinary uses were not further divided into system categories.

Alimentary use

Alimentary use was divided into human and animal use, and again into food and drink categories. Traditional drinks containing *Cannabis*, which had medicinal, psychoactive or religious uses, were automatically added into the alimentary use category even if this use was not additionally specified.

Fibre use

This category contains information on *Cannabis* use for the production of fabric, rope, sack and other products and was not divided further. It only contains information on human uses.

Psychoactive use

The category psychoactive use includes reports related to 'narcotic', 'intoxicating' and other effects altering perception, mood or consciousness. The term 'narcotic' can be defined as 'a drug or other substance that affects mood or behaviour and is consumed for non-medical purposes, especially one sold illegally' or as 'a medical drug that relieves pain and induces drowsiness, stupor, or insensibility' (24). A precise interpretation of the term in publications was not always possible, hence all 'narcotic' references were classified into the category psychoactive use. The category only contains information on human uses.

Other uses

The remaining, less-numerous uses were placed in category other uses, which was further separated into four lower

categories: cosmetic, magicoreligious, firewood and miscellaneous use. The category only contains information on human uses.

Toxicity

Even though in most regions plants from the genus *Cannabis* are considered valuable medicinal plants, in certain regions of the world they are considered toxic, with their consumption (or prolonged consumption and abuse) causing several side effects (e.g. diarrhoea, nausea, poisoning, etc.). All these reports were assigned to this category.

Vernacular names

Wild, cultivated or commercialized, *Cannabis* is widely distributed around the world, and many of its parts are used for a variety of purposes. For these reasons, it has not only been popularly named in many languages, but often with several terms in each one, depending on the part, product or use (cf., just in English, hemp, cannabis and marijuana). For the majority of references, the authors provided vernacular names for *Cannabis* and these are included here.

Quick overview of the data records

The database contains 2330 data entries on medicinal, alimentary, fibre, psychoactive and other ethnobotanical uses of *Cannabis* from different geographical areas and cultures worldwide. It contains information on the intended purpose of plant use, the taxonomic and vernacular name, the country and region of use, bibliographic reference type, plant part used, intended use destination (human or animal use), details on preparation and administration and any other additional information we considered important. Each entry is connected to the original source, which can be accessed easily from the website.

Information was gathered from 649 references from 41 countries worldwide (Figure 2A). The majority of them (71.98%) were published in the last 10 years (Figure 3). Reports from India (41.76%) and Pakistan (25.89%), where the use of *Cannabis* in folk medicine has a long cultural tradition (25–27), represent the greatest proportion of entries in the database. Most of the reported uses were medicinal (75.41%), followed by psychoactive (8.35%), alimentary (7.29%), other uses (5.13%) and fibre use (3.82%) (Figure 2B). The most frequently used plant parts are leaf (50.51%), seed (15.38%) and inflorescence (11.35%), while other plant parts represent a smaller proportion (Figure 2C). We identified *Cannabis* treatments for 210 human and 53 animal ailments. Reports on its toxicity only represent 3.24% of data entries.

User guide and potential data applications

The CANNUSE database is openly accessible at http://cannusedb.csic.es. Besides the web interface, we also provide the data via the DIGITAL.CSIC repository (https://digital.csic.es/handle/10261/226973?mode=full; Table 1), where CANNUSE database can be downloaded as a Microsoft Excel file, under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License.

Search through the database is facilitated by a user-friendly graphical interface. The clean design used is visible from any type of device (e.g. smartphones, laptops and tablets), easy to use and without page reloads so the visitor can use the search quickly and efficiently.

The web version of the database offers two search options (Figure 4). The first one is a general search, based on key words, while the second is an advanced search where additional filter options are available (plant parts and products, country, region, year of publication, etc.) depending on the general category being selected. Due to the limited space available in the graphical interface, abbreviations are used for more than one field, but their explanations can be quickly located in the 'Abbreviations and explanations' section of the website. Where additional information is available, the movement of the cursor over the sign '+' reveals additional text. Original references are connected to each entry and are linked to the tab 'Publications', where full reference information and links to the original publications can be found. Movement of the cursor over the short version of the citation reveals the full reference information. Search results are obtained in a table below the menus.

The CANNUSE database offers an organized and structured dataset that can be used as a basis for research and development strategies in many different scientific and technological fields. For example, at the present time, the majority of medical studies are focused on the application of *Cannabis* inflorescences for new treatments, but traditionally, many other parts of the plant were used for the treatment of different conditions and ailments. The CANNUSE database enables us to filter down to specific plant parts and identify the corresponding ailments for which they have been traditionally used. Furthermore, new applications could be developed in food and nutraceutical, cosmetic or recreational use industries. Analysis of *Cannabis* medicinal (and other) uses in different regions of the world could indicate local variability in *Cannabis* landraces, which would make them more suitable for further development into specific medicines. Furthermore, ethnobotanical records in the CANNUSE database could be considered as relevant additional information (besides genetic diversity and archeological findings) that could

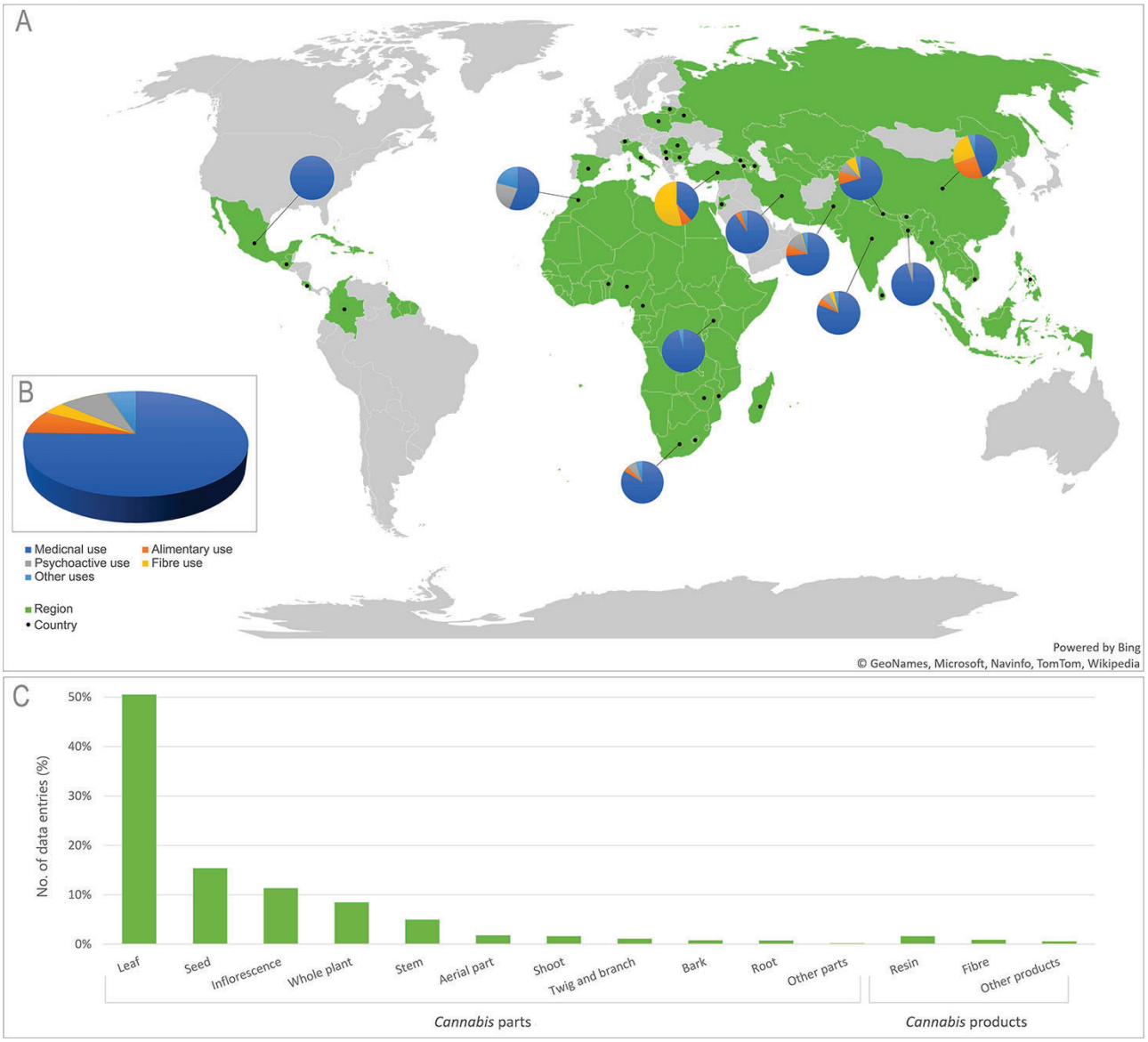


Figure 2. (A) Map of world regions (green) and countries (●) represented in the CANNUSE database, with pie charts showing distribution of uses in the countries with over 50 records. The background map was produced using the Excel Office. (B) Distribution of *Cannabis* uses presented in the database. (C) Distribution of *Cannabis* parts and products presented in the CANNUSE database (in %).

help determine the origin of species and its dispersion history (28).

The CANNUSE database contributes to the conservation and dissemination of many traditional uses in many parts of the world of an emblematic plant in ethnobotany and economic botany. It protects the traditional knowledge holders from misappropriation of their knowledge, using a CC BY-NC-SA 4.0 International License, which allows for sharing and adaptation of the database, with appropriate crediting, but does not permit its use for commercial purposes. Much of the ethnobotanical reports (databases and academic papers) are published in English, which enables a bigger distribution of the knowledge but diminishes their usability for original traditional knowledge holders (16). The CANNUSE database is, indeed, in English, but now powerful translation tools are easily available for many languages. In addition, we have already started to look for publications in other languages for further updates of the database. In this step, we are planning to include local scientists, which would enable us the contact with traditional knowledge holders missing so far.

Future development

The CANNUSE database already contains a comprehensive and globally distributed dataset on traditional *Cannabis*

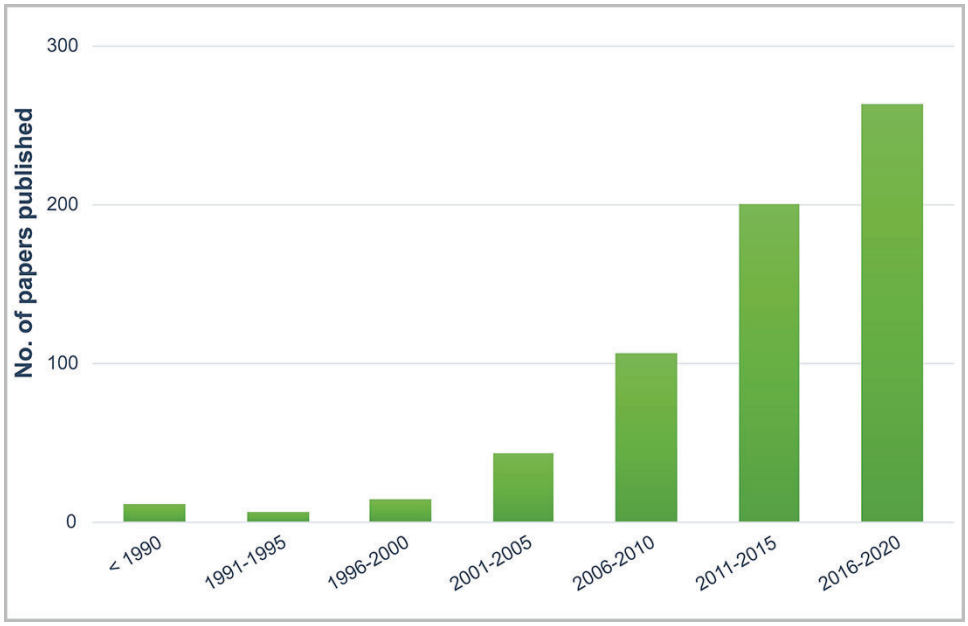


Figure 3. Number of publications in CANNUSE database containing information about *Cannabis* ethnobotanical uses published over the years.

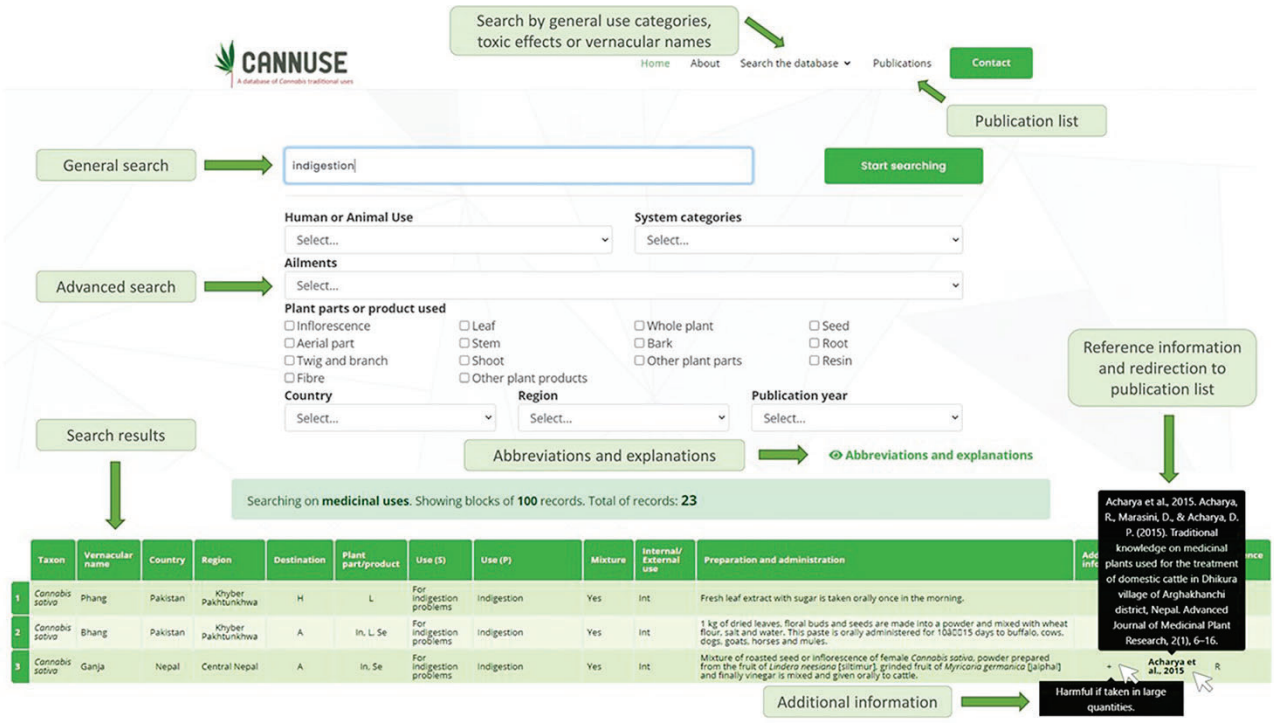


Figure 4. User interface of the Search the database function. The data can be filtered out using general search (search by key words) or by advanced search, where user is selecting the filters for the search.

uses. However, several potential areas for upgrade have already been identified. The data included in this version were obtained only from references written in English, while publications in other languages were up to this point excluded. Ethnobotanical research is often published in lesser-known, local journals, which are not written in English, so many additional uses remain to be included in future updates planned. To improve the protection of the traditional knowledge holders' rights and facilitate their benefit-sharing claims, the information about the ethnic group where the use comes from will be added. The database will be updated annually with new literature and

subsequently information on *Cannabis* uses from historic sources, books, review papers and other secondary sources, and sources in languages other than English. Additional information gathered with our own ethnobotanical interviews will also be included. Researchers are encouraged to submit any additional data they wish to share through the contact form on the website, to facilitate the improvement of the growing dataset. Regular extensions of the database will ensure that updated information on traditional *Cannabis* uses is thoroughly available for basic and applied research purposes. We ask users to cite this paper when data are used in publications or other activities (e.g. teaching and industrial applications) and to also cite the latest version of the database used.

Acknowledgements

We would like to acknowledge the indigenous peoples and local communities that have safeguarded the knowledge about *Cannabis* use during millennia despite the pressures from markets, states and religions. We would also like to thank Dr Joan Uriach from Uriach laboratories and CIJA Preservation S.L. for their support in the project. We further thank Inés Fuentes Garcia for her help with the design, Mario Ruz for collaborating with the publication search and Jonathan Ashworth for help with improving the manuscript.

Funding

Project WECANN (grant number CGL2017-84297-R, AEI/FEDER, UE) from the Spanish government (grant number 2017SGR1116); Generalitat de Catalunya (grant number PRO2020-S02-VALLES); Institut d'Estudis Catalans; FPI predoctoral contract of the Ministerio de Ciencia, Innovación y Universidades (grant number PRE2018-083226).

Conflict of interest. None declared.

References

1. Petrovska,B.B. (2012) Historical review of medicinal plants' usage. *Pharmacogn. Rev.*, **6**, 1–5.
2. Abel,E.L. (1980) *Marihuana: The First Twelve Thousand Years*. Springer US, New York.
3. Pisanti,S. and Bifulco,M. (2019) Medical *Cannabis*: a plurimillennial history of an evergreen. *J. Cell. Physiol.*, **234**, 8342–8351.
4. Abuhasira,R., Shbiro,L. and Landschaft,Y. (2018) Medical use of cannabis and cannabinoids containing products – regulations in Europe and North America. *Eur. J. Intern. Med.*, **49**, 2–6.
5. Vandepitte,K., Vasile,S., Vermeire,S. *et al.* (2020) Hemp (*Cannabis sativa* L.) for high-value textile applications: the effective long fiber yield and quality of different hemp varieties, processed using industrial flax equipment. *Ind. Crops Prod.*, **158**, 112969.
6. Babiker,E.E., Uslu,N., Al Juhaimi,F. *et al.* (2021) Effect of roasting on antioxidative properties, polyphenol profile and fatty acids composition of hemp (*Cannabis sativa* L.) seeds. *LWT – Food Sci. Technol.*, **139**, 110537.
7. Gönce,F., Ersöz,E., Kara,M. *et al.* (2020) Safety assessment of the innovative functional food ingredient from *Cannabis sativa* L. wastes. *EuroBiotech J*, **4**, 134–143.
8. Callaway,J.C. (2004) Hempseed as a nutritional resource: an overview. *Euphytica*, **140**, 65–72.
9. Kang,S.K., O'Leary,J. and Miller,J. (2016) From forbidden fruit to the goose that lays golden eggs: marijuana tourism in Colorado. *SAGE Open*, **6**, 1–12.
10. CannabisDraftMap.Org. (2021) https://www.cannabisdraftmap.org/ (20 December 2020, date last accessed).
11. *C. sativa* (Cannabis) Genome Browser Gateway. (2020) http://genome.ccbr.utoronto.ca/cgi-bin/hgGateway (20 December 2020, date last accessed).
12. Cai,S., Zhang,Z., Huang,S. *et al.* (2021) CannabisGDB: a comprehensive genomic database for *Cannabis Sativa* L. *Plant Biotechnol J*. 10.1111/pbi.13548.
13. Clinical Studies and Case Reports. (2018) http://www.cannabis-med.org/studies/study.php (20 December 2020, date last accessed).
14. Marijuana Strain Library: Effects, Find, History & More – Wikileaf. (2021) https://www.wikileaf.com/strains/ (20 December 2020, date last accessed).
15. Alexander,M., Chamundeeswari,K., Kambu,A. *et al.* (2003) The role of registers and databases in the protection of traditional knowledge: a comparative analysis. UNU-IAS Report.
16. Gaikwad,J. *et al.* (2010) Combining ethnobotany and informatics to discover knowledge from data. In: Rai MK, Acharya D, Rios JL (eds). *Ethnomedicinal Plants: Revitalization of Traditional Knowledge of Herbs*. Science Publishers, Enfield, pp. 444–457.
17. Benyei,P., Arreola,G. and Reyes-García,V. (2020) Storing and sharing: a review of indigenous and local knowledge conservation initiatives. *Ambio*, **49**, 218–230.
18. Vallès,J. (2019) *Etnobotànica: persones, plantes, cultura i benestar. Aspectes generals, i situació i perspectives als Països Catalans*. Institut d'Estudis Catalans, Barcelona.
19. McCarter,J., Gavin,M.C., Baereleo,S. *et al.* (2014) The challenges of maintaining indigenous ecological knowledge. *Ecol. Soc.*, **19**, 39.
20. Goodman,L.A. (1961) Snowball sampling. *Ann. Math. Stat.*, **32**, 148–170.
21. Naraine,S.G.U., Small,E., Laursen,A.E. *et al.* (2020) A multivariate analysis of morphological divergence of 'seeds' (achenes) among ruderal, fibre, oilseed, dioecious/monoecious and marijuana variants of *Cannabis sativa* L. *Genet. Resour. Crop Evol.*, **67**, 703–714.
22. Law,J. and Martin,E. (2020) *Concise Medical Dictionary*. 10th edn. Oxford University Press, Oxford.
23. Cook,F.E.M. (1995) *Economic Botany Data Collection Standard*. Royal Botanic Gardens, Kew.
24. Stevenson,A. and Waite,M. (eds). (2011) *Concise Oxford English Dictionary*. 12th edn. Oxford University Press, Oxford.
25. Dymock,W., Warden,C.J.H. and Hooper,D. (1893) *Pharmacographia Indica: A History of the Principal Drugs*

of Vegetable Origin, Met with in British India. K. Paul, Trench, Trübner, London. https://archive.org/details/b2129737x_0003/page/324/mode/2up?q=Cannabis (20 December 2020, date last accessed).
26. Chopra,I.C. and Chopra,R.N. (1957) The use of Cannabis drugs in India. *Bull. Narc.*, **9**, 4–29. http://www.unodc.org/unodc/en/data-and-analysis/bulletin/bulletin_1957-01-01_1_page003.html (20 December 2020, date last accessed).
27. Russo,E. (2005) Cannabis in India: ancient lore and modern medicine. In Mechoulam R (ed). *Cannabinoids as Therapeutics*. Birkhäuser Basel, Basel, pp. 1–22.
28. Harshberger,J.W. (1896) The purposes of ethno-botany. *Bot. Gaz.*, **21**, 146–154.

4.2 Traditional uses of *Cannabis*: An analysis of the CANNUSE database

The following chapter is presented in the form of a published article:

Balant, M., Gras, A., Ruz, M., Vallès, J., Vitales, D., & Garnatje, T. (2021). Traditional uses of *Cannabis*: An analysis of the CANNUSE database. *Journal of Ethnopharmacology*, 279, 114362.

DOI: https://doi.org/10.1016/j.jep.2021.114362

Journal of Ethnopharmacology 279 (2021) 114362

Contents lists available at ScienceDirect

Journal of Ethnopharmacology

journal homepage: www.elsevier.com/locate/jethpharm

Traditional uses of *Cannabis*: An analysis of the CANNUSE database

Manica Balant [a,*,1], Airy Gras [a,1], Mario Ruz [b], Joan Vallès [b,c], Daniel Vitales [a,b], Teresa Garnatje [a]

[a] Institut Botànic de Barcelona (IBB, CSIC - Ajuntament de Barcelona), Passeig del Migdia s/n, 08038, Barcelona, Catalonia, Spain
[b] 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, Catalonia, Spain
[c] Institut d'Estudis Catalans (IEC), Carrer del Carme, 47, 08001, Barcelona, Catalonia, Spain

ARTICLE INFO

Keywords:
Cannabis
Database
Ethnobotany
Medicinal uses
Traditional uses

ABSTRACT

Ethnopharmacological relevance: Cannabis is one of the most versatile genera in terms of plant use and has been exploited by humans for millennia. Nowadays, *Cannabis* is the centre of many scientific studies, most of them focusing on chemical composition and medicinal values. While new and varied applications are continuously being developed, the knowledge surrounding less common uses of the plant is slowly disappearing.

Aim of the review: We have analysed diversity of global data of *Cannabis* traditional uses, to investigate if certain plant parts are significantly associated with particular *Cannabis* use. We wanted to uncover potential associations between the plant parts used for the treatment of different body systems and ailments.

Materials and methods: We have analysed the extensive database of *Cannabis* traditional uses (CANNUSE). This database contains 2330 data entries of *Cannabis* ethnobotanical uses from over 40 countries across the world. The dataset was divided into five general groups based on the type of use: medicinal, alimentary, psychoactive, fibre and other uses. Given the abundance of human medicinal uses, detailed analysis was done on the subset of 1167 data entries. We analysed the relationship between 16 body system categories and ailments treated with *Cannabis* plant parts. We used a Pearson's chi-square and Fisher's exact test, to determine which *Cannabis* parts are characteristic of treatment for specific ailments.

Results: In this dataset, the majority of reports were represented by medicinal (75.41%), followed by psychoactive (8.35%) and alimentary (7.29%) use. The most commonly used plant parts were leaf (50.51%), seed (15.38%) and inflorescence (11.35%). We found that different *Cannabis* plant parts were significantly associated with different uses; the leaf was typically used for medicinal, seed for alimentary and inflorescence for psychoactive use. Regarding the human medicinal uses, most common were reports for treatments of the digestive system and nutritional disorders (17.66%), nervous system and mental disorders (16.24%), followed by pain and inflammations (12.21%). We found a significant relationship between the use of certain *Cannabis* parts and treatment of ailments and body systems categories; leaf was significantly associated with treatment of two categories: skin and subcutaneous tissue disorders and circulatory system and blood disorders; seed use was associated with musculoskeletal system disorders and traumas; while inflorescence use shows a statistical support for treatment of nervous system and mental disorders.

Conclusion: Several pharmaceutical companies are intensely working on developing new drugs with isolated chemical compounds or crude extracts, almost exclusively from *Cannabis* inflorescences. However, our review revealed that use of leaf or seed in traditional medicine is often more important than use of inflorescence for the treatment of certain ailments. A review of traditional medicine provides a body of knowledge and an initial pathway to identify landraces and plant parts that could have an important role in future medicinal research. We are confident that traditional medicine still has a large potential for modern medicine. As more information on *Cannabis* diversity (genetics, biochemistry, and clinical studies) becomes available, ethnobotanical data are poised to be of much greater significance.

* Corresponding author.
E-mail addresses: manica.balant@ibb.csic.es (M. Balant), airy.gras@ibb.csic.es (A. Gras), mruz995@gmail.com (M. Ruz), joanvalles@ub.edu (J. Vallès), daniel.vitales@ibb.csic.es (D. Vitales), tgarnatje@ibb.csic.es (T. Garnatje).
[1] M. B. and A. G. contributed equally to this work.

https://doi.org/10.1016/j.jep.2021.114362
Received 22 March 2021; Received in revised form 31 May 2021; Accepted 20 June 2021
Available online 24 June 2021

1. Introduction

Cannabis sativa L. (hereafter *Cannabis*) is one of the most versatile plants known to man and has traditional roots among many cultures around the world. Because of its exceptional phenotypic plasticity, *Cannabis* has played an important role in various aspects of human life.

Even though people have used it for thousands of years, details about *Cannabis* origin are still not well known. Latest studies place its origin in Central Asia, in the NE part of the Tibetan plateau (Kovalchuk et al., 2020; McPartland et al., 2019), however theories of South Asian origin have also been proposed (Linné et al., 1737; Zhang et al., 2018). In addition, more research is needed to determine the possible domestication area of *Cannabis*. The oldest archaeological remains are the seeds discovered in Japan about 10,000 years ago (Kudo et al., 2009), but the exact centre of *Cannabis* domestication is still unknown. Domestication most likely started somewhere in Central Asia (Clarke and Merlin, 2013), but theories of multiregional domestication have also been suggested (Long et al., 2017; McPartland et al., 2019; Vavilov, 1992; Zhang et al., 2018).

A long coexistence of *Cannabis* and people managing it has resulted in its worldwide distribution, alongside a high genetic, morphological, and chemical diversity. This variability has impeded the taxonomic resolution within *Cannabis* genus (Clarke and Merlin, 2013). Two hundred years of attempts have produced numerous interpretations, the genus being composed of either: three (*C. sativa* L., *C. indica* Lam., *C. ruderalis* Janisch.; Hillig, 2005; Clarke and Merlin, 2013; Sawler et al., 2015), two (*C. indica* and *C. sativa*; Clarke and Merlin, 2016) or one species (*C. sativa*; Small and Cronquist, 1976; Small, 2015; McPartland, 2018; Zhang et al., 2018). Currently the most widely accepted theory is that the genus consists of a single species, *C. sativa*, with several subspecies and varieties. Depending on the purpose and chemical composition it is mostly divided into fibre-type (hemp; $<$ 0.3% Δ9-tetrahydrocannabinol (THC)) or drug-type (marijuana or medicinal cannabis; $>$ 0.3% THC) plants (Hurgobin et al., 2021). Drug-type plants are known in the vernacular nomenclature as ''*sativa*'' and ''*indica*'' plants based on their CBD/THC ratio. However, this does not always coincide with the taxonomical nomenclature of *C. sativa* and *C. indica* and does not necessarily reflect the common genetic ancestry. For a more detailed review of the taxonomic and popular classification of *Cannabis*, see McPartland and Small (2020) and Small (2015). Since the taxonomy within the genus is still not well resolved and our study does not focus only on one type of plant, we will consider *Cannabis* at the genus level.

Cannabis has a long tradition of use in many cultures around the world. It was traditionally used for medicinal purposes, production of fibres, ropes, textile, and paper, it served as a valuable source of food, and it was an important element in many shamanic rituals (Clarke and Merlin, 2013). Traditional knowledge is the result of centuries of experience and innovations. Practices of indigenous and local communities around the world were passed down from generation to generation and adapted to local culture and environment (Convention on Biological Diversity, 2011). Despite its many uses for millennia, now *Cannabis* is most famous for its psychoactive recreational use. The cannabinoid responsible for its mind-altering effects is Δ-9-tetrahydrocannabinol, better known as THC. It is not clear when and how people first discovered the psychoactive effects of *Cannabis*, but it has probably been used in different ritualistic and religious contexts since the early Palaeolithic period (Clarke and Merlin, 2013). One of the first conclusive evidence of its use in ritual smoking comes from Pamir mountains and dates back 2500 years. The findings of charred seeds, wooden braziers and stones in the Pamirs revealed that *Cannabis* plants were burned intentionally, and chemical analysis suggested high levels of psychoactive chemicals (Ren et al., 2019). Remains from a prehistoric site in China from about the same age also suggest *Cannabis* was used for ritual purposes (Jiang et al., 2016). Recently, ritual use of *Cannabis* was also confirmed at the Judahite Shrine of Arad in Israel, dating to the 8th century BCE (Arie et al., 2020). All these finds reveal that in the past *Cannabis* had an important role in religious rituals. Today, *Cannabis* is the most used recreational drug in the world - an estimated 183 million people were using it in 2014 (UNODC, 2016). In the early 20th century, *Cannabis* became regarded as an illegal drug and its use started to decrease (Pisanti and Bifulco, 2019). However, in the past twenty years, research on *Cannabis* increased and several traditional uses (especially medicinal) have started to gain more attention.

The pharmacological industry's growing interest in *Cannabis* has made it a valuable plant in medical research. Up until now, over 150 cannabinoids and hundreds of other compounds like terpenoids, flavonoids, and alkaloids (with valuable anti-inflammatory, antimicrobial, neuroprotective properties) have been discovered in *Cannabis* (Bonini et al., 2018; Hanuš et al., 2016; Jin et al., 2020). Many traditional medicinal uses of *Cannabis* were already proven and are now medically recognised treatments. It is used for cancer pain and chronic pain management (Blake et al., 2017; Lynch and Ware, 2015), spasticity and pain management associated with multiple sclerosis (Mecha et al., 2020), and inflammation reduction (Perisetti et al., 2020). However, many other uses have been reported in ethnopharmacological surveys but remain to be studied in a broader framework. Several pharmaceutical companies are intensely working on developing new drugs with isolated natural *Cannabis* products, while others are focusing on studying effects of crude extracts from *Cannabis* inflorescence, recently proven superior to the single molecule use in medical treatment (i.e., the entourage effect; Koltai and Namdar, 2020). Despite the deep pharmaceutical inroads, the diversity of *Cannabis* continues to make the research on this plant challenging. Many studies already confirmed differences in chemical profiles between different *Cannabis* landraces and cultivars (Abdollahi et al., 2020; Bueno and Greenbaum, 2021; Eržen et al., 2021; Kornpointner et al., 2021; Li et al., 2020; Nagy et al., 2019; Namdar et al., 2019; Nissen et al., 2010; Novak et al., 2001; Stack et al., 2021). Some of this variation is attributed to the genetic background (Vergara et al., 2019), but the differences are also caused by different growing conditions (Burgel et al., 2020; Saloner and Bernstein, 2021; Wei et al., 2021), and collection period (Kornpointner et al., 2021; Stack et al., 2021). Diversity of material used in clinical trials makes it very difficult to compare the results, because plants with different chemical composition, could be more or less effective for treatment of certain illnesses (Mudge et al., 2019; Namdar et al., 2019). This variation within the plants and plant parts makes standardisation and reproducibility of medicinal products very difficult (Bernstein et al., 2019; Gorelick and Bernstein, 2014). An additional problem in clinical research is the lack of randomized double-blind placebo controlled clinical trials, which are particularly hard to secure when the tested drug is psychoactive, or is considered "miraculous" (Gertsch, 2018; Russo, 2016).

Inflorescences are best-known and almost exclusively used part of *Cannabis* in pharmaceutical industry, even though in the past all plant parts had an important role in traditional medicine (Clarke and Merlin, 2013; Stuart and Smith, 1911). Specific plant parts contain different types and amounts of chemical compounds, and depending on the illness different plant parts and preparations were used (Chopra and Chopra, 1957; Stuart and Smith, 1911). Inflorescences contain the highest density of glandular trichomes, particularly rich in cannabinoids (Livingston et al., 2020), and therefore are the focus of most medicinal studies. Only recently have other plant parts started to gain more attention. In the latest study by Jin et al. (2020), they screened different parts of *Cannabis* and found that inflorescence and leaves are the most abundant source of cannabinoids, mono- and sesquiterpenoids, and flavonoids. However, pharmacologically relevant quantities of triterpenoids and sterols can also be found in roots, stems, and bark. The identification of biochemically active compounds in different plant parts is the basis for development of new medicinal uses (Jin et al., 2020). Nevertheless, a comprehensive review of traditional medicine can also help us identify plant parts and preparations that could potentially be more useful for

treatment of specific illnesses. This traditional knowledge could be the basis for further pharmacological investigation determining key active compounds responsible for the desired medicinal effects.

Apart from well-known psychoactive and medicinal uses, *Cannabis* has played an important role in many other aspects of human life. *Cannabis* fruits (usually referred to as 'seeds'), were probably the first parts of this plant people collected. Throughout Asia, *Cannabis* seeds have represented an important part of human diet and are still consumed in several ways (e.g., raw, roasted, pickled, grinded, parched or pressed for oil) (Clarke and Merlin, 2013). Seeds of non-psychoactive *Cannabis* varieties, commonly known as hemp, contain over 30% of oil, 25% of easily digested protein and are high in dietary fibres, vitamins, minerals, with an optimal ratio of omega-6 to omega-3 fatty acids for human health (Callaway, 2004). *Cannabis* is also known for having one of the strongest and most durable natural fibres, which is why it has long been used in production of clothing, coarse canvas, sackings, twine, rope, fishing nets, rugs, and pulp for paper (Clarke, 2010a, 2010b). *Cannabis* fibres are gaining new uses in sustainable industry as house insulation material, hemp fibre interior panels in automotive industry, animal bedding, nonwoven agricultural fleece, matting, mulch for weed suppression, and erosion control. Furthermore, seeds rich in polyunsaturated fatty acids and proteins have started to gain the popularity as snacks, as well as in oil production (Clarke and Merlin, 2013).

Information on *Cannabis* traditional knowledge is substantial, however there is a strong need to synthesise and standardise these data, since it is scattered among many publication sources. Recently, an online source - the CANNUSE database (http://cannusedb.csic.es) (Balant et al., 2021b) – was released, containing information on *Cannabis* traditional knowledge related to medicinal, alimentary, fibre and other uses from different geographical areas.

In the present study, we analysed the data on traditional *Cannabis* uses included in the CANNUSE database to obtain a general overview of the most common *Cannabis* traditional uses and their diversity. We further investigated if certain plant parts are significantly associated to a particular *Cannabis* use or even treatment of different body systems and ailments.

2. Methodology

2.1. The CANNUSE database content

We have analysed the dataset gathered in the database of *Cannabis* traditional uses – CANNUSE (Balant et al., 2021; https://digital.csic.es/handle/10261/226973?locale=en). The CANNUSE database contains information on literature published in the English language from 1960 until the end of October 2020 comprising of first-hand information obtained through any type of ethnobotanical interviews. The publication search for the database construction was carried out in four major online databases—Scopus, Web of Science, PubMed and Google Scholar, using the following set of keywords and exact terms: *Cannabis* AND ('folk medicine' OR 'traditional medicine' OR 'ethnobotany' OR 'traditional knowledge'). Information obtained from review papers and books was only used when original research papers could not be found. It consists of 2330 entries from 649 publications related to medicinal, alimentary, fibre and other uses from different geographical areas.

For each reference, the following information is provided: type and year of publication, country and region, taxon, vernacular name, and part of the plant (inflorescence, leaf, whole plant, seed, aerial parts, stem, bark, root, twigs and branches, other parts) or plant product (resin, fibre, other products) used. In the database, the term 'seed' refers to the monosperm *Cannabis* fruit, a nut (also called achene) (Naraine et al., 2020). It contains information of the type of use, whether *Cannabis* had animal or human use, and includes modes of preparation and administration, whenever they were provided by the authors. For medicinal use, type of administration (external, internal) is also recorded whenever possible.

The database is divided into five main use categories: medicinal, alimentary, fibre, psychoactive and other uses. Since *Cannabis* is sometimes considered poisonous, with several side effects, an additional category named toxicity is included. The majority of authors also provided vernacular names of *Cannabis,* which can be found next to each use. For more details on data collection and database structure, see Balant et al. (2021b) or the CANNUSE database website (http://cannusedb.csic.es).

2.2. Data analysis

The information included in the CANNUSE database (i.e., 2330 data entries) (Balant et al., 2021a; Table 1 in https://doi.org/10.20350/digitalCSIC/13686) was analysed to obtain a general overview of the most common *Cannabis* uses and their diversity. To investigate the relationship between different uses (i.e., medicinal, alimentary, psychoactive, fibre and other uses) and the plant parts utilised, we analysed the data with Pearson's chi-square test of independence — and Fisher's exact test to calculate the p-values — in XLSTAT 2020.3.1 (Addinsoft, New York, USA). In some references included in the database, plant parts used were not unambiguously specified. Therefore, the analysis of the relationship between plant parts and their uses were performed on a subset of 1725 (74.03%) data entries, where the plant part used was well specified (Balant et al., 2021a; Table 2 in https://doi.org/10.20350/digitalCSIC/13686). Because of low frequencies, reports using whole plant and aerial plant parts were grouped, and less commonly used plant parts (i.e., bark, fibre, root, resin, stem, shoot, twig and branch and other plant parts and products) were grouped under 'other plant parts and products'.

Because of the numerous data entries, a specific analysis of human traditional medicinal uses was done on the subset of 1167 (50.09%) data entries where the plant part used was well specified. Medicinal uses were classified into 16 human body system categories, according to Cook (1995) with minor modifications (Supplementary data Table S1). We tested the relationship between body system categories treated with *Cannabis* and plant parts used (grouped as in the previous step). If the ailment was classified into two categories, it has been considered as two use reports. System categories with less than 30 data entries (i.e., poisoning, pregnancy, birth and puerperal disorders, and sensory system disorders) were grouped together with unclassified ailments under 'other categories and unclassified'. The same analysis was additionally carried out by sub-setting data from two individual countries with over 200 entries (India and Pakistan), to test if the plant parts employed for medicinal use differ between the countries. We also analysed the relationship between specific ailments treated with *Cannabis* and plant parts used. Because only five ailments had over 30 data entries, we chose only those for further statistical analyses. Pearson's chi-square test of independence with 2000 Monte Carlo replicates was performed and p-value was calculated with Fisher's exact test in XLSTAT 2020.3.1. The Pearson's chi-square results were visualized using the function "corrplot" of 'corrplot' package (Wei and Simko, 2017) in R software system 4.0.1 (R Core Team, 2020).

3. Results and discussion

3.1. General overview of the information presented in the CANNUSE database

Traditional uses of *Cannabis* from 41 countries worldwide are represented in the database. The majority of reports come from India (41.76%) and Pakistan (25.89%), two of the countries where the use of *Cannabis* in folk medicine has one of the longest traditions (Chopra and Chopra, 1957; Dymock et al., 1893; Russo, 2005). Unexpectedly, even though there are many documented records of ancient *Cannabis* use in China (Jiang et al., 2006, 2016; Liu et al., 2017; Stuart and Smith, 1911), we found only 12 ethnobotanical papers from this country mentioning

the use of this plant. Cultural changes in China could have a major influence on its use. After the rise of Confucianism, around 200 BCE, ritual, psychoactive and medicinal uses of *Cannabis* started to decline (Touw, 1981), and are nowadays only used for fibre production and consumption of seeds - as snacks and pressed for oil (Clarke and Merlin, 2013). Another possible explanation for the limited reports found from China could be the search strategy carried out to construct the CANNUSE database. Ethnobotanical research is often published in lesser known, often local journals, which are not written in the English language, and the bibliographic search missed those references.

Due to the ambiguous taxonomic status within the genus *Cannabis*, we found eight scientific names within the references included in CANNUSE database. The most frequently employed taxonomic entity was *C. sativa* L. (96.92%), but we also recovered other taxonomic names (in decreasing order) *C. sativa* var. *sativa*, *Cannabis* sp., *C. sativa* var. *indica* (Lam.) E.Small & Cronquist, *C. sativa* subsp. *indica* (Lam.) E.Small & Cronquist, *C. indica* Lam., *C. ruderalis* Janisch., *C. sativa* f. *ruderalis* (Janisch.) Chu, and *Cannabis* spp.

Because *Cannabis* has been used by humans worldwide for thousands of years and for a variety of purposes, we can also find many popular or vernacular names for it. Often it is named differently depending on the use and the plant part used; for example, we can find over 40 names for *Cannabis* in Sanskrit language (Russo, 2005). It is therefore not surprising that the database contains 211 vernacular names. The highest diversity of names was found in references from India (56 vernacular names), South Africa (34) and Pakistan (31). The overall most frequent vernacular name was *bhang* (in 46.22% of references), a prevalent name for *Cannabis* in India. As mentioned before, vernacular names of *Cannabis* do not only change depending on the different countries (or regions within them) but may also depend on the plant part or plant use. In India, for example, the three most common preparations are: *bhang* - dried matured leaves and flowering shoots of female and male plants, *ganja* - dried flowering tops of the cultivated female *Cannabis* plant, and *charas* - the resinous matter collected from the leaves and flowering tops (Chopra and Chopra, 1957). All of them are recorded in the CANNUSE database as vernacular names for *Cannabis*.

The majority of the 2330 entries of the database refer to medicinal use (75.41%), followed by psychoactive (8.35%) and alimentary use (7.29%). Most commonly used plant parts are leaf (50.51%), seed (15.38%) and inflorescence (11.35%). The results of Pearson's chi-square test show, that there is a non-random association between *Cannabis* use categories and plant parts employed ($X^2 = 684.618$; df = 16; $p < 0.0001$) (Supplementary data Table S3). Medicinal reports are significantly associated with the use of leaves, psychoactive reports with inflorescence use and reports of alimentary and other uses with the use of seeds (Fig. 1 and Supplementary data Table S2).

3.2. Medicinal use

Cannabis has been a valuable plant in traditional medicine for thousands of years, so it is not surprising that medicinal use represents the majority of data entries (Fig. 1). According to our analysis, all plant parts have been used for medicinal purposes, but leaf use was reported in over half of data entries (55.76%). The results of Pearson's chi-square and Fisher's exact test show us that different plant parts are not randomly used for medicinal purposes. In fact they show that leaf is strongly associated with medicinal use ($p < 0.0001$) (Supplementary data Table S3). The majority of medicinal uses belong to human medicine, while only 8.54% of them was represented by veterinary use. We recorded 152 entries of 53 ailments treated in animals with the most common being antidiarrhoeal use (9.87%), treatment of dysentery (6.58%), appetite stimulant (4.61%) and treatment of coccidiosis (4.61%).

3.2.1. Human medicinal uses

We analysed 1627 data entries for human medicinal use, which were divided in 16 system categories. The most common ailments belong to digestive system and nutritional disorders (17.66%), nervous system and mental disorders (16.24%), followed by pain and inflammations (12.21%). We recorded *Cannabis* treatments for 210 ailments. The most common uses were sedative (6.02%), analgesic (5.84%), antidiarrhoeal (3.01%), antihaemorrhoidal (2.52%), followed by the use for dysentery (2.27%), wound treatment (2.21%) and as a tonic (2.40%). Some of these uses have already been confirmed by human and/or animal clinical studies, albeit sometimes with contradictory or non-conclusive findings (e.g., Buggy et al., 2003; Maharajan et al., 2020), but many others (i.e., antihaemorrhoidal and wound healing) still need to be verified.

The list of *Cannabis* human medicinal uses is very long, however not all plant parts were similarly used for all treatments. Since different plant parts have different chemical profiles (Burgel et al., 2020; Jin et al., 2020; Nagy et al., 2019; Namdar et al., 2018), they could be more or less effective for the treatment of different illnesses. We analysed the relationship between plant parts used for the treatment of different body systems and ailments, to see if the plant parts are randomly used or an association between them exists. Pearson's chi-square test showed that there is a significant relationship between the two variables ($X^2 = 110.36$, $p = 0.0005$) (Fig. 2 and Supplementary data Table S4).

The leaf was significantly associated with treatment of skin and subcutaneous tissue disorders and circulatory system and blood

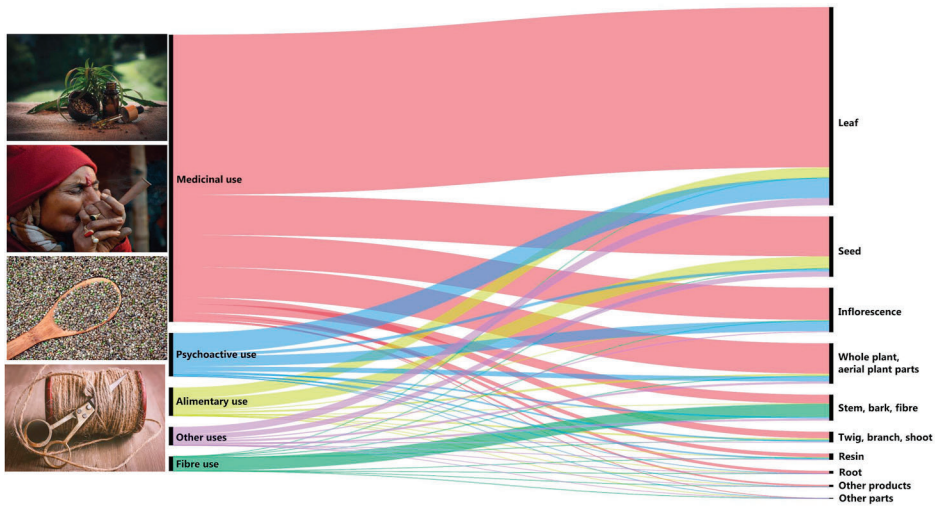


Fig. 1. Different *Cannabis* uses and the plant parts used for each use category. For the medicinal purpose leaves were used in most cases (55.76%), followed by seeds (13.92%) and inflorescences (11.20%). For the psychoactive use leaf use represented majority of reports (44.46%), but inflorescence use is also common (23.85%). In alimentary use, seeds were mostly used (43.59%) and for fibre use, other plant parts (particularly stem, bark, and fibre) were almost exclusively used (93.83%).

disorders. Seed use was associated with musculoskeletal system disorders and traumas, inflorescence with nervous system and mental disorders, while whole plant and aerial plant parts are significantly associated with treatment of pain and inflammation and was often used as tonic and restorative. We also found a significant association between the plant parts used for treatment of different specific ailments (X^2 = 59.447, p = 0.0005) (Fig. 3 and Supplementary data Table S5).

Regarding the analyses of plant part use between countries, most of the results show that reports from India (600 data entries) and Pakistan (548 data entries) yielded similar results as the dataset as a whole. However, we found differences in the plants parts employed for certain body system categories among countries. In the data from India, we found a strong association of leaf use with the treatment of body systems grouped in the category 'other categories and unclassified'. Because this is a very diverse group (poisoning, pregnancy, birth and puerperal disorders, sensory system disorders and unclassified), we cannot assign this relationship to any particular use. Whole plant and aerial plant parts in India were only significantly associated with the use of pain and inflammation treatment and not as a tonic and restorative (Supplementary data Table S6). In Pakistan, seeds were significantly associated with the use of respiratory system disorders and not with musculoskeletal system disorders and traumas, as in other countries. Use of inflorescence in Pakistan was positively associated with treatment of nervous system and mental disorders, but here the relationship was not significant (p = 0.106). The use of other plant parts (especially use of shoots, branches and twigs) was significantly associated with the use as tonic and restorative – this association was not found in the analysis of the rest of the data (Supplementary data Table S7).

These differences between countries could be explained by several factors. Many studies have proven that different landraces and chemovars contain different chemical profiles (Abdollahi et al., 2020; Bueno and Greenbaum, 2021; Eržen et al., 2021; Kornpointner et al., 2021; Li et al., 2020; Nagy et al., 2019; Namdar et al., 2019; Nissen et al., 2010; Novak et al., 2001; Stack et al., 2021), which could be one of the causes for this variation. Detailed analysis of individual countries or regions could help us identify landraces with specific chemical profiles. Currently the dataset obtained from the CANNUSE database only allowed us a detailed analysis of the reports from India and Pakistan, since other countries are still underrepresented. However, the database is currently being updated (Balant et al., 2021b) and can become an important resource for such analysis in the future. Differences in uses between countries can also be caused by other reasons such as local customs, cultural differences and availability of other medicinal plants in the region (Kunwar et al., 2019). Therefore, different traditional uses between countries should be further investigated, through a series of pharmacological and phytochemical studies on local *Cannabis* landraces.

Fig. 2. Frequencies (circle size) and values of adjusted Pearson's chi-square residuals (colour shades) of the plant part use for each body system category. The size of each circle indicates the number of reports for treatment of each system category depending on the plant part used, while the colour shades indicate values of adjusted Pearson's chi-square residuals. The red colour indicates a positive and the blue a negative association between the plant part used and the body system treated. Asterisk indicates a significant positive association between the body system and the plant part used, as calculated with Fischer's exact test (*p < 0.05 and **p < 0.001); LE – leaf, SD – seed, INFL – inflorescence, WP & AP – whole plant and aerial plant parts, OP & P – other plant parts and products (root, resin, twig, branch and shoot, fibre, stem, bark, and other parts). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Fig. 3. Frequencies (circle size) and values of adjusted Pearson's chi-square residuals (colour shades) of the plant use for ailments with over 30 data entries. The size of each circle indicates the number of reports for treatment of each ailment depending on the plant part used, while the colour shades indicate values of adjusted Pearson's chi-square residuals. The red colour indicates a positive and the blue a negative association between the plant part used and the ailment treated. Asterisk indicates a significant positive association between the ailments and the plant part used, as calculated with Fischer's exact test (*$p < 0.05$ and **$p < 0.001$); LE – leaf, SD – seed, INFL – inflorescence, WP & AP – whole plant and aerial plant parts, OP & P – other plant parts and products (root, resin, twig, branch and shoot, fibre, stem, bark and other parts). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3.2.1.1. Leaves. The leaf was the most used plant part for treatments in all system categories (used in 54.69% of all data entries), but the most numerous records correspond to treatments of the digestive system and nutritional disorders (157 data entries; Fig. 2 and Supplementary data Table S8). *Cannabis* leaves contain a considerable amount of cannabinoids that can interact with cannabinoid receptors in the gastrointestinal tract. Many clinical studies already confirmed their effectiveness for treatment of different gastrointestinal disorders, e.g., inflammatory bowel disease (Goyal et al., 2017; Kienzl et al., 2020; Pellesi et al., 2019; Perisetti et al., 2020; Picardo et al., 2019). Other system categories frequently related with *Cannabis* leaves are nervous system and mental disorders (131 entries), skin and subcutaneous tissue disorders (108), infections and infestations (105) and pain and inflammations (101). Among them, Fischer's exact test showed that the use of leaf was significantly associated with the treatment of skin and subcutaneous tissue disorders ($p < 0.0001$) (Fig. 2 and Supplementary data Table S4). According to our data, typical ailments treated with *Cannabis* leaves are wounds, cuts, skin diseases and sores. In these treatments, leaves are either grinded or crushed and applied externally in a form of juice, paste or powder. The analysis of the relationship between plant parts and specific ailments showed that leaf use is significantly associated with the treatment of wounds ($p = 0.001$) (Fig. 3 and Supplementary data Table S5). A recent study by Jin et al. (2020) found that leaves are rich in cannabinoids, terpenes and sesquiterpenoids, but also contain significant quantities of flavonoids and sterols. They all have anti-inflammatory, anti-bacterial and anti-fungal properties that can promote wound healing and can help with different skin problems (Andre et al., 2016; Kupczyk et al., 2009; Wilkinson and Williamson, 2007; Wright et al., 2005). Another system category with which leaf use is significantly associated was circulatory system and blood disorders ($p = 0.005$) (Fig. 2 and Supplementary data Table S4). According to our analysis, this relationship is almost exclusively related to haemorrhoids treatment (also highly associated with leaf use; $p < 0.0001$) (Fig. 3 and Supplementary data Table S5), where leaves are usually applied externally in a form of paste. No clinical tests have been done so far to verify the antihaemorrhoidal effects of *Cannabis* leaves, but the positive effects could be due to the presence of cannabinoids, terpenes, sesquiterpenoids, flavonoids and sterols in leaves, that have anti-inflammatory and analgesic effects (Gallily et al., 2018; Rabgay et al., 2020).

3.2.1.2. Seeds. Seeds are the second most used *Cannabis* part in human medicine, and they represent 14.46% of data entries. Reports of seed use were most frequent for the treatment of digestive system and nutritional disorders, nervous system and mental disorders, followed by pain and inflammation (51, 32, 30 data entries, respectively) (Supplementary data Table S8). Seeds have been used for treatments of these ailments since the early ages. In Arab medicine they were used (among other) for their antiepileptic, antiemetic, and carminative properties and for soothing neurological pain (Lozano, 2003). In traditional Chinese medicine *Cannabis* seeds were used for constipation and obstinate vomiting (Stuart and Smith, 1911), and still today traditional medicinal practitioners prescribe them for digestive and genitourinary problems (Shou-zhong, 1998). A recent study by Xue et al. (2020) found that they have protective effects on intestinal oxidative damage in mice. Indeed, *Cannabis* seeds are commonly used for a large diversity of ailments, but our analysis showed their use is significantly associated with the treatment of musculoskeletal system disorders and traumas ($p = 0.007$) (Fig. 2 and Supplementary data Table S4). In most cases, seed oil is massaged on the affected part, due to the supposed analgesic, antiarthritic, and antirheumatic effects. Clinical studies have already proved that cannabinoids are useful for treating rheumatic pain (Blake et al., 2006; Malfait et al., 2000), however the specific effect of *Cannabis* seeds – or products derived from seeds – still needs to be tested.

3.2.1.3. Inflorescences. In modern medicine, *Cannabis* female inflorescence is the most used part of the plant (Minghetti et al., 2019) and the main focus of many clinical trials. However, in our data, inflorescence use represented only 11.17% of human medicinal reports. Most numerous reports of the inflorescence use correspond to the treatment of nervous system and mental disorders (Supplementary data Table S8), a relationship showing statistical support ($p = 0.002$) (Fig. 2 and Supplementary data Table S4). Many of these data entries represent the use of inflorescence as a sedative, which also showed a statistically significant association ($p = 0.009$) (Fig. 3 and Supplementary data Table S5). The form of administration for the sedative use was given in less than half of the reports, but when specified, it was either smoked or drunk.

Female *Cannabis* inflorescences contain the highest concentration of different cannabinoids, terpenes and sesquiterpenoids that have proven sedative effects, which many studies already confirmed (Choi et al., 2020; Hazekamp et al., 2010; Mondino et al., 2019; Nuutinen, 2018). Treatment of digestive system and nutritional disorders are the second most common use of *Cannabis* inflorescence (Supplementary data Table S8). The three most common digestive ailments treated with inflorescences are dysentery, diarrhoea, and appetite loss, which were also previously confirmed with clinical trials (Mechoulam and Hanuš, 2001; Pellesi et al., 2019).

Analgesic effects of cannabinoids have also been clinically proven and are effectively used for alleviating chronic pain (Aviram et al., 2020; Blake et al., 2017; Cameron and Hemingway, 2020; Lynch and Ware, 2015). However, our results show that traditionally, inflorescences are less frequently used for treatment of pain and inflammation (24 data entries) and we did not find statistical support for such use (Fig. 2 and Supplementary data Table S4).

3.2.1.4. Whole plant and aerial plant parts. The use of whole *Cannabis* plant or its aerial parts is not very frequent and was recorded only in 11.24% of data entries. Whole plant or its aerial parts were most commonly used for treatment of ailments connected to nervous system and mental disorders. The two most common uses were sedative and stimulant use. We found that these preparations were administered in various ways: in the form of decoction or other types of drinks, by bathing in them, smoking or eating them, or they were externally applied. The association between the use of whole plant and its aerial parts for specific treatments is statistically significant for system categories pain and inflammations ($p = 0.015$; used for its analgesic effects, which is statistically significant ($p = 0.008$)) and tonic and restorative ($p = 0.029$; used for tonic preparations) (Figs. 2 and 3 and Supplementary data Tables S4 and S5). Different *Cannabis* parts have been used in tonic preparations for centuries; seeds were used in Chinese traditional medicine (Stuart and Smith, 1911), and other parts were used in indigenous medicine in India (Chopra and Chopra, 1957), Japan (Olson, 1997), and Jamaica (Comitas, 2011).

3.2.1.5. Other plant parts. The versatility of *Cannabis* for human medicine is well reflected in our results, as we found examples of ailments treated with every part of the plant. Although the uses of leaf, seed, inflorescence, whole plant, and aerial plant parts are prevalent, we also found reports of medicinal uses of roots, twigs, branches, shoots, stems, and bark, as well as plant products such as resin and fibre (8.44% of data entries; grouped in the category 'other plant parts'). Most of these data entries fall to the system categories pain and inflammation, digestive system and nutritional disorders and nervous system and mental disorders (21, 20 and 20 data entries, respectively). Even though uses of these parts are not numerously represented, they should not be overlooked. Fibre, stem, and bark were mostly used for their antirheumatic effects and treatment of skin diseases (both 5 data entries). Twigs, branches, and shoots were used for their analgesic effects (4 data entries). Since resin is most abundant on inflorescences, the use was similar for both plant parts - it was mostly used as a sedative (three data entries). *Cannabis* root has most entries (three) for treatment of menstrual disorders. In the past, *Cannabis* roots have been consumed for various uses, such as treatment of inflammation, fever, gout, arthritis, joint pain, skin burns, hard tumours, postpartum haemorrhage, difficult child labour, sexually transmitted disease, gastrointestinal disorders and infections (Ryz et al., 2017). A recent study by Lima et al. (2021) showed that roots of this plant have anti-inflammatory effects in mice models. In the dataset analysed here we also found reports for fever and cancer treatment, indigestion problems, stomach pain, liver disorders, and antiacid, among others. However, we found no indication for the anti-inflammatory use of *Cannabis* roots.

3.3. Psychoactive use

Psychoactive use of *Cannabis* is probably one of its most famous ones. It has been employed for millennials in many cultures and in many different forms (e.g., smoking dried inflorescences or purified resinous products like *charas* or *hashish*, drinking preparations of fresh leaves called *bhang*, etc.) (Clarke and Merlin, 2013). Knapp et al. (2019) indicate that today *Cannabis* is mostly used recreationally and consumed in different ways, most frequently by smoking. However, in the CANNUSE database, psychoactive use only represents 8.35% of all entries. The relatively low percentage of psychoactive uses does not match with the relevance of *Cannabis* cultivation, commerce, and consumption as a recreative drug, which has an important incidence at the worldwide level and frequently falls in the field of illegal activities. Irrespective of these societal considerations, ethnobotanical reports of *Cannabis* toxic activity, which could be linked to side effects of psychoactive consumption, are not very numerous (see section 3.5.).

In the CANNUSE database we found different methods of *Cannabis* administration for psychoactive use: smoking the leaves, inflorescences or resin preparations with different potency (*charas* or *attar, hashish, ganja,* plant powder) and drinking preparations from *Cannabis* leaves, inflorescences and shoots (*tandai, bhang*). The majority of references for psychoactive use did not specify the administration mode (73.10%), but considering only the reports including this information, it was administered by smoking in 56.6%, drunk in 37.74% and ingested as food in 5.66% of cases. For psychoactive use, the most used part of the plant is leaf (46.44%) followed by inflorescence (23.85%) (Fig. 1 and Supplementary data Table S2). Even though inflorescences are the biggest source of THC and other cannabinoids, they are also present in leaves (Jin et al., 2020). This could explain the common use of leaf in psychoactive purposes. Higher percentage of leaf use could also be explained by the common consumption of the traditional Indian drink *bhang* (also called *bang, thandai, tandai,* etc.) that is enjoyed in many religious and festivity ceremonies, but also drunk for its medicinal effects. Even though the use of inflorescence for psychoactive purposes is less frequently represented in the database than leaves, our analyses indicate that inflorescence is significantly associated with psychoactive, but not with other uses ($p < 0.0001$) (Fig. 1 and Supplementary data Table S3).

3.4. Alimentary use

Nowadays *Cannabis* products are becoming recognised as functional food. Seeds have been recognised as valuable food source, rich in easily digestible proteins, polyunsaturated fatty acid (PUFA), lipids, carbohydrates, and insoluble fibre (Rupasinghe et al., 2020). They have a favourable ratio of omega-6 to omega-3 of PUFA, well suited for human diet and have beneficial effect on the cardiovascular health, cancer, atopic dermatitis conditions and constipation problems, among other issues (Callaway et al., 2005; Cerino et al., 2021; Cheng et al., 2011; Rupasinghe et al., 2020). Seeds are mostly pressed for oil, but also available in many other preparations – from energy bars, pralines and chocolates, flavoured yogurt, hemp flour, baked goods, hemp milk, protein seed powder and seasoning sauce (Cerino et al., 2021; Rupasinghe et al., 2020). Although *Cannabis* seeds and its products are mostly used in today's food industry, *Cannabis* sprouts, leaves and flowers are also eaten raw in juices or in salads. They contain additional bioactive compounds (e.g., polyphenols and cannabinoids) not found, or less abundantly found in seeds (Cerino et al., 2021; Rupasinghe et al., 2020).

In our dataset, *Cannabis* alimentary use comprised 7.29% of all uses (Fig. 1 Supplementary data Table S2); 58.72% of them corresponded to traditional food and 41.28% to traditional drinks. As expected, the most used plant part for alimentary purposes are seeds (43.60%), which also proved to be significantly associated with alimentary use ($p < 0.0001$) (Supplementary data Table S3). Seeds are still considered as a good food source for elderly people throughout Asia because they contain plenty of

easy digestible protein and dietary roughage (Clarke and Merlin, 2013). Our analysis showed that traditionally seeds are most commonly pressed for oil (17.65%) or pickled (14.71%). We also found references of their use in beverages, as a condiment, they are roasted, or processed in flour or curd. Leaves are the second most used plant part for alimentary purposes (37.18%), mostly consumed in traditional beverages (e.g., *bhang*; 60.34%), but also fried, or otherwise included in the dishes.

3.5. Fibre and other uses

Regarding the fibre uses, as expected, the most likely used *Cannabis* parts are fibre, stem, and bark (grouped inside other plant parts and products; $p < 0.0001$), which represent over 90% of data entries in this category (Supplementary data Table S3). *Cannabis* fibres were most often used for making ropes (27.40%) and fabric (24.66%). Even though *Cannabis* used to be a very important fibre plant (Clarke and Merlin, 2013), in the CANNUSE database, fibre use represents only 3.82% of all data entries (Fig. 1 and Supplementary data Table S3). Many ethnobotanical papers included in the database almost exclusively focused on medicinal plants in the area, so the traditional uses of *Cannabis* fibres are probably underrepresented in our results. In the last decades, this use has almost disappeared because of the discovery of synthetic materials, but it remained strong in some areas, like China (Clarke and Merlin, 2013). In recent years it is again being rediscovered due to durability of fibres and sustainable production (Gedik and Avinc, 2020).

Besides the most known and common uses mentioned above, we also recorded *Cannabis* magicoreligious and cosmetic uses, use for firewood and other miscellaneous ones, which together represented 5.13% of all data entries (Fig. 1). Most frequently used parts in these cases are leaves (39.22%) and seeds (27.45%) (Supplementary data Table S2). Magicoreligious use represents 23.14% of reports in the category other uses. Due to the mind-altering purposes, *Cannabis* has been a vital element of many religious ceremonies. In India, *Cannabis* is considered a holy plant, and it is a vital element in many religious rituals, mainly regarding the worship of Lord Shiva. The traditional drink *bhang* is often consumed during Indian festivals like Shivratri and Holi (Chopra and Chopra, 1957). Due to the high content of oil (especially polyunsaturated fatty acids) in seeds, *Cannabis* was also used in traditional cosmetic preparations (16.53%), especially in hair care. A study in 2005 (Callaway et al., 2005) found that the addition of modest amounts of hemp seed oil in everyday diet significantly improved the strength of fingernails and hair thickness. Although *Cannabis* is an herbaceous plant, its stems are also used for firewood or torch wood (13.22%), especially in Pakistan, where 62.5% of records comes from. The other 47.11% of data entries in this category are comprised of miscellaneous uses. Leaves and above-ground parts of *Cannabis* were used in apiculture, pest control and for fish poisoning, while oil made from seeds was used for production of soaps, paints, varnishes and for lightning.

3.6. Potential toxic effects

Even though in many regions of the world *Cannabis* is considered a valuable medicinal plant, it is considered toxic (or toxic if used in excess) in others. There are still opposing opinions about the extent of negative effects of *Cannabis* consumptions between scientists. Results of some studies indicate that long-term consumption of *Cannabis* has harmful effects on developing brain (e.g., neuroanatomic changes, metabolic and neurotransmitter activity, and neuronal activation), especially in people with specific genetic polymorphisms, which indicates that *Cannabis* use can interact with genotype to increase the risk of mental health issues (Hurd et al., 2019). A recent review by Thomas et al. (2014) indicated adverse effects of *Cannabis* on cardiovascular activity (e.g., myocardial infarction, sudden cardiac death, cardiomyopathy, stroke, transient ischemic attack, and *Cannabis* arteritis). Additional adverse effects in other body systems, such as ophthalmological, gastrointestinal, respiratory, immune, and hormonal system were also connected with exposure to high THC concentrations, mainly related with recreational use. However, significant toxicity is infrequent in adults, intoxication symptoms are normally short-lived and do not pose a significant risk of death (Breijyeh et al., 2021; Cabral and Staab, 2005). Regarding our data, only 3.24% of data entries reported toxic effects. They were mostly caused using the inflorescence (42.86%) and leaf (40.82%). We found 45 side effects, the most frequent were hallucination, poisoning, drowsiness, nausea, and vomiting. Only one reference mentioned death.

Many references also stated that *Cannabis* was only considered toxic if used in excess. The importance of the correct dosage and negative consequences of extensive and prolonged abuse of *Cannabis* were well known in traditional medicine (Chopra and Chopra, 1957). We can find records that differentiate between early effects (reviving heat, exhilaration, improvement of complexion, excitement of imagination, appetite increase, aphrodisiac) and late effects of *Cannabis* consumption (refrigerant and sedative effect) in Pharmacographia Indica (Dymock et al., 1893). The authors also warned that prolonged use can cause unwanted negative effects like indigestion, wasting of the body, melancholy, impotence, and dropsy (swelling, accumulation of water). Today we know that cannabinoids display bell-shape dose-response curves (Jamontt et al., 2010; Zuardi et al., 2017), and so the correct dosing is crucial in therapeutic and recreational use to avoid undesired effects. The conflicting evidence of *Cannabis* effects are probably the reason why at the end of 2020 UNDOC Commission followed the WHO recommendation and removed *Cannabis* from the Schedule IV drug list, but it remained listed as a Schedule I drug (UNDOC, 2020).

4. Conclusion

Today, *Cannabis* is mostly associated to recreational use due to its mind-altering effects. However, this is not reflected in the dataset of the CANNUSE database analysed here, where 92% of data entries correspond to non-psychoactive uses. Over two thirds of this data are comprised of *Cannabis* medicinal uses – most of them human medicinal uses – representing treatments for 210 human ailments. Together, our study confirms that *Cannabis* shows a large number and diversity of traditional medicinal uses. The majority of data analysed here come from a determined geographic region (i.e., India and Pakistan), so the results obtained here could be biased towards the uses from those areas. The chemical composition of the plants used in certain regions is expected to vary, therefore the associations between plant parts and medicinal use could also be different in other areas of the world where *Cannabis* is traditionally used. Unfortunately, ethnobotanical papers rarely contain information about the chemical composition, or the cultivars of the plants studied, hence this information was not available. Regardless, describing the specific chemical components and the exact phytochemical pathways responsible for medicinal effects was beyond the scope of this paper. Our aim was to shed light to the less known traditional uses of *Cannabis* and connect them with the use of different plant parts on a global scale. We believe that this study revealed some new potential uses that could be further chemically and pharmacologically explored for potential drug development.

Many pharmaceutical companies are intensely working on developing new drugs with isolated natural products or crude extracts of *Cannabis*, almost exclusively based on inflorescences from commercial varieties. In contrast, references included in the CANNUSE database show that 89% of all traditional medicinal uses are related to other plant parts. *Cannabis* inflorescences are of great importance for drug development because of their high content of cannabinoids. However, other plant parts also contain a diverse composition of valuable secondary metabolites that could make them effective for treatment of a variety of illnesses. In this study, we prove that *Cannabis* parts are not randomly used in the traditional treatment of different body systems and ailments. Instead, our results clearly show that certain plant parts are significantly associated with particular body systems and ailments. Some of these relationships (e.g., inflorescences and treatment of nervous system) have

already been confirmed in previous clinical studies, but others (e.g., leaves for treatment of haemorrhoids; or seeds for treatment of musculoskeletal system disorders and traumas) still need to be further explored. As more information becomes available on *Cannabis* diversity (e.g., genetic, biochemical, and clinical studies) and more comprehensive ethnobotanical dataset is gathered (in terms of geographic regions and local landraces surveyed), the usefulness of the CANNUSE database is poised to be of much greater significance.

Acknowledgements:

This research was financially supported by project WECANN (CGL2017-84297-R, AEI/FEDER, UE) from the Spanish government, Generalitat de Catalunya (2017SGR1116), Institut d'Estudis Catalans (PRO2021-S02-VALLES) and FPI predoctoral contract of the Ministerio de Ciencia, Innovación y Universidades (PRE2018-083226). We thank CIJA Preservation S.L., and Dr. Joan Uriach, from Uriach laboratories, for their support to the project. We further thank Inés Fuentes Garcia for her help with the design and Vinayak Ramasahayam for help with improving the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jep.2021.114362.

Author contributions

T.G., J.V. and A.G. designed the research, M.B. and M.R. analysed the dataset, and M.B. and D.V. wrote the manuscript; all authors discussed the results and commented on the manuscript.

References

Abdollahi, M., Sefidkon, F., Calagari, M., Mousavi, A., Mahomoodally, M.F., 2020. Impact of four hemp (*Cannabis sativa* L.) varieties and stage of plant growth on yield and composition of essential oils. Ind. Crop. Prod. 155, 112793. https://doi.org/10.1016/j.indcrop.2020.112793.

Andre, C.M., Hausman, J.F., Guerriero, G., 2016. *Cannabis sativa*: the plant of the thousand and one molecules. Front. Plant Sci. 7, 19. https://doi.org/10.3389/fpls.2016.00019.

Arie, E., Rosen, B., Namdar, D., 2020. *Cannabis* and frankincense at the judahite shrine of Arad. Tel Aviv 47, 5–28. https://doi.org/10.1080/03344355.2020.1732046.

Aviram, J., Vysotski, Y., Berman, P., Lewitus, G.M., Eisenberg, E., Meiri, D., 2020. Migraine frequency decrease following frolonged medical cannabis treatment: A cross-sectional study. Brain Sci. 10, 360. https://doi.org/10.3390/brainsci10060360.

Balant, M., Gras, A., Gálvez, F., Garnatje, T., Vallès, J., Vitales, D., 2021a. CANNUSE, the Database of Traditional Uses. DIGITAL.CSIC. https://digital.csic.es/handle/10261/226973?locale=en.

Balant, M., Gras, A., Francisco, G., Garnatje, T., Vallès, J., Vitales, D., 2021b. CANNUSE, a database of traditional *Cannabis* uses—an opportunity for new research. Database 2021. https://doi.org/10.1093/database/baab024 baab024.

Bernstein, N., Gorelick, J., Koch, S., 2019. Interplay between chemistry and morphology in medical cannabis (*Cannabis sativa* L.). Ind. Crop. Prod. 129, 185–194. https://doi.org/10.1016/j.indcrop.2018.11.039.

Blake, A., Wan, B.A., Malek, L., DeAngelis, C., Diaz, P., Lao, N., Chow, E., O'Hearn, S., 2017. A selective review of medical cannabis in cancer pain management. Ann. Palliat. Med. 6 (Suppl. 2), S215–S222. https://doi.org/10.21037/apm.2017.08.05.

Blake, D.R., Robson, P., Ho, M., Jubb, R.W., McCabe, C.S., 2006. Preliminary assessment of the efficacy, tolerability and safety of a cannabis-based medicine (Sativex) in the treatment of pain caused by rheumatoid arthritis. Rheumatology 45, 50–52. https://doi.org/10.1093/rheumatology/kei183.

Bonini, S.A., Premoli, M., Tambaro, S., Kumar, A., Maccarinelli, G., Memo, M., Mastinu, A., 2018. *Cannabis sativa*: a comprehensive ethnopharmacological review of a medicinal plant with a long history. J. Ethnopharmacol. 227, 300–315. https://doi.org/10.1016/j.jep.2018.09.004.

Breijyeh, Z., Jubeh, B., Bufo, S.A., Karaman, R., Scrano, L., 2021. *Cannabis*: a toxin-producing plant with potential therapeutic uses. Toxins 13, 117. https://doi.org/10.3390/toxins13020117.

Bueno, J., Greenbaum, E.A., 2021. (−)- trans - Δ 9 - tetrahydrocannabiphorol content of *Cannabis sativa* inflorescence from various chemotypes. J. Nat. Prod. 84, 531–536. https://doi.org/10.1021/acs.jnatprod.0c01034.

Buggy, D.J., Toogood, L., Maric, S., Sharpe, P., Lambert, D.G., Rowbotham, D.J., 2003. Lack of analgesic efficacy of oral δ-9-tetrahydrocannabinol in postoperative pain. Pain 106, 169–172. https://doi.org/10.1016/S0304-3959(03)00331-2.

Burgel, L., Hartung, J., Graeff-Hönninger, S., 2020. Impact of different growing substrates on growth, yield and cannabinoid content of two *Cannabis sativa* L. genotypes in a pot culture. Horticulturae 6, 62. https://doi.org/10.3390/horticulturae6040062.

Cabral, G.A., Staab, A., 2005. Effects on the immune system. In: Pertwee, R.G. (Ed.), Cannabinoids, vol. 168. Springer-Verlag Berlin Heidelberg, Germany, pp. 385–423. https://doi.org/10.1007/b137831.

Callaway, J., Schwab, U., Harvima, I., Halonen, P., Mykkänen, O., Hyvönen, P., Järvinen, T., 2005. Efficacy of dietary hempseed oil in patients with atopic dermatitis. J. Dermatol. Treat. 16, 87–94. https://doi.org/10.1080/09546630510035832.

Callaway, J.C., 2004. Hempseed as a nutritional resource: an overview. Euphytica 140, 65–72. https://doi.org/10.1007/s10681-004-4811-6.

Cameron, E.C., Hemingway, S.L., 2020. Cannabinoids for fibromyalgia pain: a critical review of recent studies (2015 – 2019). J. Cannabis Res. 2, 19. https://doi.org/10.1186/s42238-020-00024-2.

Cerino, P., Buonerba, C., Cannazza, G., D'Auria, J., Ottoni, E., Fulgione, A., Di Stasio, A., Pierri, B., Gallo, A., 2021. A review of hemp as food and nutritional supplement. Cannabis Cannabinoid Res 6, 19–27. https://doi.org/10.1089/can.2020.0001.

Cheng, C.W., Bian, Z.X., Zhu, L.X., Wu, J.C.Y., Sung, J.J.Y., 2011. Efficacy of a Chinese herbal proprietary medicine (Hemp Seed Pill) for functional constipation. Am. J. Gastroenterol. 106, 120–129. https://doi.org/10.1038/ajg.2010.305.

Choi, S., Huang, B.C., Gamaldo, C.E., 2020. Therapeutic uses of Cannabis on sleep disorders and related conditions. J. Clin. Neurophysiol. 37, 39–49. https://doi.org/10.1097/WNP.0000000000000617.

Chopra, I.C., Chopra, R.N., 1957. The use of *Cannabis* drugs in India. Bull. Narc. 9, 4–29.

Clarke, R., Merlin, M., 2013. *Cannabis*: Evolution and Ethnobotany, first ed. ed. University of California Press, Berkley, Los Angeles, London.

Clarke, R.C., 2010a. Traditional fiber hemp (*Cannabis*) production, processing, yarn making, and weaving strategies-functional constraints and regional responses. Part 2. J. Nat. Fibers 7, 229–250. https://doi.org/10.1080/15440478.2010.504043.

Clarke, R.C., 2010b. Traditional fiber hemp (*Cannabis*) production, processing, yarn making, and weaving strategies-functional constraints and regional responses. Part 1. J. Nat. Fibers 7, 118–153. https://doi.org/10.1080/15440478.2010.482324.

Clarke, R.C., Merlin, M.D., 2016. *Cannabis* domestication, breeding history, present-day genetic diversity, and future prospects. Crit. Rev. Plant Sci. 35, 293–327. https://doi.org/10.1080/07352689.2016.1267498.

Comitas, L., 2011. The social nexus of Ganja in Jamaica. In: Vera, R. (Ed.), Cannabis and Culture. De Gruyter Mouton, Berlin, New York, pp. 119–132. https://doi.org/10.1515/9783110812060.119.

Convention on Biological Diversity [WWW Document], 2011. URL https://www.cbd.int/traditional/intro.shtml [accessed 28.5.2021].

Cook, F.E.M., 1995. Economic Botany Data Collection Standard. Royal Botanic Gardens, Kew. https://doi.org/10.1111/j.1756-1051.1997.tb00317.x.

Dymock, W., Warden, C.J.H., Hooper, D., 1893. Pharmacographia Indica [electronic Resource]: A History of the Principal Drugs of Vegetable Origin, Met with in British India. K. Paul, Trench, Trübner. London. https://archive.org/details/b212973 7x_0003/page/324/mode/2up?q=Cannabis.

Eržen, M., Košir, I.J., Ocvirk, M., Kreft, S., Čerenak, A., 2021. Metabolomic analysis of cannabinoid and essential oil profiles in different hemp (*Cannabis sativa* L.) phenotypes. Plants 10, 966.

Gallily, R., Yekhtin, Z., Hanuš, L.O., 2018. The anti-inflammatory properties of terpenoids from *Cannabis*. Cannabis Cannabinoid Res 3, 282–290. https://doi.org/10.1089/can.2018.0014.

Gedik, G., Avinc, O., 2020. Hemp fiber as a sustainable raw material source for textile industry: can we use its potential for more eco-friendly production? In: Muthu, S., Gardetti, M. (Eds.), Sustainability in the Textile and Apparel Industries. Sustainable Textiles: Production, Processing, Manufacturing & Chemistry. Springer, Cham, pp. 87–109. https://doi.org/10.1007/978-3-030-38541-5_4.

Gertsch, J., 2018. The intricate influence of the placebo effect on medical *cannabis* and cannabinoids. Med. Cannabis Cannabinoids 1, 60–64. https://doi.org/10.1159/000489291.

Gorelick, J., Bernstein, N., 2014. Elicitation: an underutilized tool in the development of medicinal plants as a source of therapeutic secondary metabolites. In: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 201–230. https://doi.org/10.1016/B978-0-12-800138-7.00005-X.

Goyal, H., Singla, U., Gupta, U., May, E., 2017. Role of cannabis in digestive disorders. Eur. J. Gastroenterol. Hepatol. 29, 135–143. https://doi.org/10.1097/MEG.0000000000000779.

Hanuš, L.O., Meyer, S.M., Muñoz, E., Taglialatela-Scafati, O., Appendino, G., 2016. Phytocannabinoids: a unified critical inventory. Nat. Prod. Rep. 33, 1357–1392. https://doi.org/10.1039/c6np00074f.

Hazekamp, A., Fischedick, J.T., Llano, M.D., Lubbe, A., Ruhaak, R.L., 2010. Chemistry of cannabis. In: Liu, H.-W., Ben), Mander, L. (Eds.), Comprehensive Natural Products Chemistry II. Elsevier Ldt., Oxford, pp. 1033–1084. https://doi.org/10.1016/b978-008045382-8.00091-5.

Hillig, K.W., 2005. Genetic evidence for speciation in *cannabis* (cannabaceae). Genet. Resour. Crop Evol. 52, 161–180. https://doi.org/10.1007/s10722-003-4452-y.

Hurd, Y.L., Manzoni, O.J., Pletnikov, M.V., Lee, F.S., Bhattacharyya, S., Melis, M., 2019. *Cannabis* and the developing brain: insights into its long-lasting effects. J. Neurosci. 39, 8250–8258. https://doi.org/10.1523/JNEUROSCI.1165-19.2019.

Hurgobin, B., Tamiru-Oli, M., Welling, M.T., Doblin, M.S., Bacic, A., Whelan, J., Lewsey, M.G., 2021. Recent advances in *Cannabis sativa* genomics research. New Phytol. 230, 73–89. https://doi.org/10.1111/nph.17140.

Jamontt, J.M., Molleman, A., Pertwee, R.G., Parsons, M.E., 2010. The effects of Δ 9-tetrahydrocannabinol and cannabidiol alone and in combination on damage,

inflammation and *in vitro* motility disturbances in rat colitis. Br. J. Pharmacol. 160, 712–723. https://doi.org/10.1111/j.1476-5381.2010.00791.x.
Jiang, H.-E., Li, X., Zhao, Y.-X., Ferguson, D.K., Hueber, F., Bera, S., Wang, Y.-F., Zhao, L.-C., Liu, C.-J., Li, C.-S., 2006. A new insight into *Cannabis sativa* (cannabaceae) utilization from 2500-year-old yanghai tombs, xinjiang, China. J. Ethnopharmacol. 108, 414–422. https://doi.org/10.1016/j.jep.2006.05.034.
Jiang, H., Wang, L., Merlin, M.D., Clarke, R.C., Zhang, Y., Pan, Y., Xiao, G., Ding, Ding, X., 2016. Ancient *cannabis* burial shroud in a central eurasian cemetery. Econ. Bot. 70, 213–221. https://doi.org/10.1007/s12231-016-9351-1.
Jin, D., Dai, K., Xie, Z., Jie, C., 2020. Secondary metabolites profiled in Cannabis inflorescences, leaves, stem barks, and roots for medicinal purposes. Sci. Rep. 10, 3309. https://doi.org/10.1038/s41598-020-60172-6.
Kienzl, M., Storr, M., Schicho, R., 2020. Cannabinoids and opioids in the treatment of inflammatory bowel diseases. Clin. Transl. Gastroenterol. 11, e00120 https://doi.org/10.14309/ctg.0000000000000120.
Knapp, A.A., Lee, D.C., Borodovsky, J.T., Auty, S.G., Gabrielli, J., Budney, A.J., 2019. Emerging trends in cannabis administration among adolescent cannabis users. J. Adolesc. Health 64, 487–493. https://doi.org/10.1016/j.jadohealth.2018.07.012.
Koltai, H., Namdar, D., 2020. Cannabis phytomolecule "entourage": from domestication to medical use. Trends Plant Sci. 25, 976–984. https://doi.org/10.1016/j.tplants.2020.04.007.
Kornpointner, C., Sainz Martinez, A., Marinovic, S., Haselmair-Gosch, C., Jamnik, P., Schröder, K., Löfke, C., Halbwirth, H., 2021. Chemical composition and antioxidant potential of *Cannabis sativa* L. roots. Ind. Crop. Prod. 165, 113422. https://doi.org/10.1016/j.indcrop.2021.113422.
Kovalchuk, I., Pellino, M., Rigault, P., Velzen, R. van, Ebersbach, J., Ashnest, J.R., Mau, M., Schranz, M.E., Alcorn, J., Laprairie, R.B., McKay, J.K., Burbridge, C., Schneider, D., Vergara, D., Kane, N.C., Sharbel, T.F., 2020. The genomics of *Cannabis* and its close relatives. Annu. Rev. Plant Biol. 71, 20. https://doi.org/10.20506/rst.35.1.2429, 1–20.27.
Kudo, Y., Kobayashi, M., Momohara, A., Noshiro, S., Nakamura, T., Okitsu, S., Yanagisawa, S., Okamoto, T., 2009. Radiocarbon dating of fossil hemp fruits in the earliest Jomon period excavated from the Okinoshima site. Chiba, Japan. Japanese J. Hist. Bot. 17, 27–31 (in Japanese with English abstract).
Kunwar, R.M., Shrestha, K., Malla, S., Acharya, T., Sementelli, A.J., Kutal, D., Bussmann, R.W., 2019. Relation of medicinal plants, their use patterns and availability in the lower Kailash Sacred Landscape, Nepal. Ethnobot. Res. Appl. 18, 6. https://doi.org/10.32859/era.18.7.1-14.
Kupczyk, P., Reich, A., Szepietowski, J.C., 2009. Cannabinoid system in the skin - a possible target for future therapies in dermatology. Exp. Dermatol. 18, 669–679. https://doi.org/10.1111/j.1600-0625.2009.00923.x.
Li, J., Wang, G., Qin, Y., Zhang, X., Wang, H., Liu, H., Zhu, L., Yao, X., 2020. Neuroprotective constituents from the aerial parts of *Cannabis sativa* L. subsp. *sativa*. RSC Adv. 10, 32043–32049. https://doi.org/10.1039/d0ra04565a.
Lima, K.S.B., Silva, M.E.G. da C., Araújo, T.C. de L., Silva, C.P. da F., Santos, B.L., Ribeiro, L.A. de A., Menezes, P.M.N., Silva, M.G., Lavor, É.M., Silva, F.S., Nunes, X. P., Rolim, L.A., 2021. *Cannabis* roots: pharmacological and toxicological studies in mice. J. Ethnopharmacol. 271, 113868. https://doi.org/10.1016/j.jep.2021.113868.
Linné, C. von, 1737. Hortus Cliffortianus :plantas exhibens quas in hortis tam vivis quam siccis, Hartecampi in Hollandia, Hortus Cliffortianus. Amsterdam. https://doi.org/10.5962/bhl.title.690.
Liu, F.-H., Hu, H.-R., Du, G.-H., Deng, G., Yang, Y., 2017. Ethnobotanical research on origin, cultivation, distribution and utilization of hemp (Cannabis sativa L.) in China. Indian J. Tradit. Knowl. 16, 235–242.
Livingston, S.J., Quilichini, T.D., Booth, J.K., Wong, D.C.J., Rensing, K.H., Laflamme-Yonkman, J., Castellarin, S.D., Bohlmann, J., Page, J.E., Samuels, A.L., 2020. Cannabis glandular trichomes alter morphology and metabolite content during flower maturation. Plant J. 101, 37–56. https://doi.org/10.1111/tpj.14516.
Long, T., Wagner, M., Demske, D., Leipe, C., Tarasov, P.E., 2017. *Cannabis* in Eurasia: origin of human use and Bronze Age trans-continental connections. Veg. Hist. Archaeobotany. https://doi.org/10.1007/s00334-016-0579-6.
Lozano, I., 2003. The therapeutic use of Cannabis sativa (L.) in Arabic medicine. J. Cannabis Ther. 1, 63–70. https://doi.org/10.1300/j175v01n01_05.
Lynch, M.E., Ware, M.A., 2015. Cannabinoids for the treatment of chronic non-cancer pain: an updated systematic review of randomized controlled trials. J. Neuroimmune Pharmacol. 10, 293–301. https://doi.org/10.1007/s11481-015-9600-6.
Maharajan, M.K., Yong, Y.J., Yip, H.Y., Woon, S.S., Yeap, K.M., Yap, K.Y., Yip, S.C., Yap, K.X., 2020. Medical cannabis for chronic pain: can it make a difference in pain management? J. Anesth. 34, 95–103. https://doi.org/10.1007/s00540-019-02680-y.
Malfait, A.M., Gallily, R., Sumariwalla, P.F., Malik, A.S., Andreakos, E., Mechoulam, R., Feldmann, M., 2000. The nonpsychoactive cannabis constituent cannabidiol is an oral anti-arthritic therapeutic in murine collagen-induced arthritis. Proc. Natl. Acad. Sci. U.S.A. 97, 9561–9566. https://doi.org/10.1073/pnas.160105897.
McPartland, J.M., 2018. *Cannabis* systematics at the levels of family, genus, and species. Cannabis Cannabinoid Res 3, 203–212. https://doi.org/10.1089/can.2018.0039.
McPartland, J.M., Hegman, W., Long, T., 2019. *Cannabis* in Asia: its center of origin and early cultivation, based on a synthesis of subfossil pollen and archaeobotanical studies. Veg. Hist. Archaeobotany 28, 691–702. https://doi.org/10.1007/s00334-019-00731-8.
McPartland, J.M., Small, E., 2020. A classification of endangered high-THC cannabis (*Cannabis sativa* subsp. *indica*) domesticates and their wild relatives. PhytoKeyes 144, 81–112. https://doi.org/10.3897/phytokeys.144.46700.
Mecha, M., Carrillo-Salinas, F.J., Feliú, A., Mestre, L., Guaza, C., 2020. Perspectives on cannabis-based therapy of multiple sclerosis: a mini-review. Front. Cell. Neurosci. 14, 35. https://doi.org/10.3389/fncel.2020.00034.
Mechoulam, R., Hanuš, L., 2001. The cannabinoids: an overview. Therapeutic implications in vomiting and nausea after cancer chemotherapy, in appetite promotion, in multiple sclerosis and in neuroprotection. Pain Res. Manag. 6, 67–73. https://doi.org/10.1155/2001/183057.
Minghetti, P., Marini, V., Zaccara, V., Casiraghi, A., 2019. Regulation for prescribing and dispensing system of Cannabis: the Italian case. Curr. Bioact. Compd. 15, 196–200. https://doi.org/10.2174/1573407214666181011120031.
Mondino, A., Cavelli, M., González, J., Santana, N., Castro-Zaballa, S., Mechoso, B., Bracesco, N., Fernandez, S., Garcia-Carnelli, C., Castro, M.J., Umpierrez, E., Murillo-Rodriguez, E., Torterolo, P., Falconi, A., 2019. Acute effect of vaporized *Cannabis* on sleep and electrocortical activity. Pharmacol. Biochem. Behav. 179, 113–123. https://doi.org/10.1016/j.pbb.2019.02.012.
Mudge, E.M., Brown, P.N., Murch, S.J., 2019. The terroir of *Cannabis*: terpene metabolomics as a tool to understand *Cannabis sativa* selections. Planta Med. 85, 781–796. https://doi.org/10.1055/a-0915-2550.
Nagy, D.U., Cianfaglione, K., Maggi, F., Sut, S., Dall'Acqua, S., 2019. Chemical characterization of leaves, male and female flowers from spontaneous Cannabis (*Cannabis sativa* L.) growing in Hungary. Chem. Biodivers. 16, e1800562 https://doi.org/10.1002/cbdv.201800562.
Namdar, D., Mazuz, M., Ion, A., Koltai, H., 2018. Variation in the compositions of cannabinoid and terpenoids in Cannabis sativa derived from inflorescence position along the stem and extraction methods. Ind. Crops Prod. 113, 376–382. https://doi.org/10.1016/j.indcrop.2018.01.060.
Namdar, D., Voet, H., Ajjampura, V., Nadarajan, S., Mayzlish-Gati, E., Mazuz, M., Shalev, N., Koltai, H., 2019. Terpenoids and phytocannabinoids co-produced in *Cannabis sativa* strains show specific interaction for cell cytotoxic activity. Molecules 24, 3031. https://doi.org/10.3390/molecules24173031.
Naraine, S.G.U., Small, E., Laursen, A.E., Campbell, L.G., 2020. A multivariate analysis of morphological divergence of "seeds" (achenes) among ruderal, fibre, oilseed, dioecious/monoecious and marijuana variants of *Cannabis sativa* L. Genet. Resour. Crop Evol. 67, 703–714. https://doi.org/10.1007/s10722-019-00848-9.
Nissen, L., Zatta, A., Stefanini, I., Grandi, S., Sgorbati, B., Biavati, B., Monti, A., 2010. Characterization and antimicrobial activity of essential oils of industrial hemp varieties (*Cannabis sativa* L.). Fitoterapia 81, 413–419. https://doi.org/10.1016/j.fitote.2009.11.010.
Novak, J., Zitterl-Eglseer, K., Deans, S.G., Franz, C.M., 2001. Essential oils of different cultivars of *Cannabis sativa* L. and their antimicrobial activity. Flavour Fragrance J. 16, 259–262. https://doi.org/10.1002/ffj.993.
Nuutinen, T., 2018. Medicinal properties of terpenes found in Cannabis sativa and Humulus lupulus. Eur. J. Med. Chem. 157, 198–228. https://doi.org/10.1016/j.ejmech.2018.07.076.
Olson, D., 1997. Hemp culture in Japan. J. Int. Hemp Assoc. 4, 40–50.
Pellesi, L., Verga, M.C., De Maria, N., Villa, E., Pini, L.A., Guerzoni, S., 2019. Nabilone administration in refractory chronic diarrhea: a case series. BMC Gastroenterol. 19, 105. https://doi.org/10.1186/s12876-019-1024-y.
Perisetti, A., Rimu, A.H., Khan, S.A., Bansal, P., Goyal, H., 2020. Role of cannabis in inflammatory bowel diseases. Ann. Gastroenterol. 33, 134–144. https://doi.org/10.20524/aog.2020.0452.
Picardo, S., Kaplan, G.G., Sharkey, K.A., Seow, C.H., 2019. Insights into the role of cannabis in the management of inflammatory bowel disease. Ther. Adv. Gastroenterol. Rev. 12, 1–13. https://doi.org/10.1177/1756284819870977.
Pisanti, S., Bifulco, M., 2019. Medical *Cannabis*: a plurimillennial history of an evergreen. J. Cell. Physiol. 234, 8342–8351. https://doi.org/10.1002/jcp.27725.
R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rabgay, K., Waranuch, N., Chaiyakunapruk, N., Sawangjit, R., Ingkaninan, K., Dilokthornsakul, P., 2020. The effects of cannabis, cannabinoids, and their administration routes on pain control efficacy and safety: a systematic review and network meta-analysis. J. Am. Pharmaceut. Assoc. 60, 225–234. https://doi.org/10.1016/j.japh.2019.07.015.
Ren, M., Tang, Z., Wu, X., Spengler, R., Jiang, H., Yang, Y., Boivin, N., 2019. The origins of cannabis smoking: chemical residue evidence from the first millennium BCE in the Pamirs. Sci. Adv. 5, eaaw1391 https://doi.org/10.1126/sciadv.aaw1391.
Rupasinghe, H.P.V., Davis, A., Kumar, S.K., Murray, B., Zheljazkov, V.D., 2020. Industrial hemp (*Cannabis sativa* subsp. *sativa*) as an emerging source for value-added functional food ingredients and nutraceuticals. Molecules 25, 4078.
Russo, E., 2005. Cannabis in India: ancient lore and modern medicine. In: Mechoulam, R. (Ed.), Cannabinoids as Therapeutics. Birkhäuser Basel, Basel, pp. 1–22. https://doi.org/10.1007/3-7643-7358-X_1.
Russo, E.B., 2016. Current therapeutic cannabis controversies and clinical trial design issues. Front. Pharmacol. 7, 309. https://doi.org/10.3389/fphar.2016.00309.
Ryz, N.R., Remillard, D.J., Russo, E.B., 2017. *Cannabis* roots: a Traditional therapy with future potential for treating inflammation and pain. Cannabis Cannabinoid Res 2, 210–216. https://doi.org/10.1089/can.2017.0028.
Saloner, A., Bernstein, N., 2021. Nitrogen supply affects cannabinoid and terpenoid profile in medical cannabis (*Cannabis sativa* L.). Ind. Crop. Prod. 167, 113516. https://doi.org/10.1016/j.indcrop.2021.113516.
Sawler, J., Stout, J.M., Gardner, K.M., Hudson, D., Vidmar, J., Butler, L., Page, J.E., Myles, S., 2015. The genetic structure of marijuana and hemp. PloS One 10, 9. https://doi.org/10.1371/journal.pone.0133292.
Shou-zhong, Y., 1998. The Divine Farmer's Materia Medica. A Translation of the ShenNong Ben Cao Ling. Blue Poppy Press, Inc., Boulder, Colorado.
Small, E., 2015. Evolution and classification of *Cannabis sativa* (marijuana, hemp) in relation to human utilization. Bot. Rev. 81, 189–294. https://doi.org/10.1007/s12229-015-9157-3.

Small, E., Cronquist, A., 1976. A practical and natural taxonomy for *cannabis*. Taxon 25, 405–435. https://doi.org/10.2307/1220524.

Stack, G.M., Toth, J.A., Carlson, C.H., Cala, A.R., Rebecca, M.I.M., Deanna, L.W., Crawford, J.L., Philippe, G., Rose, J.K.C., Viands, D.R., Smart, C.D., Smart, L.B., 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. https://doi.org/10.1111/gcbb.12793.

Stuart, G.A., Smith, F.P., 1911. Chinese materia medica. Part 1. Vegetable kingdom. Shanghai? :Presbyterian mission Press?,1911? Online source. https://doi.org/10.5962/bhl.title.25114.

Thomas, G., Kloner, R.A., Rezkalla, S., 2014. Adverse cardiovascular, cerebrovascular, and peripheral vascular effects of marijuana inhalation: what cardiologists need to know. Am. J. Cardiol. 113, 187–190. https://doi.org/10.1016/j.amjcard.2013.09.042.

Touw, M., 1981. The religious and medicinal uses of *Cannabis* in China, India and Tibet. J. Psychoact. Drugs 13, 23–34. https://doi.org/10.1080/02791072.1981.10471447.

UNDOC, 2020. UNDOC [WWW Document E/CN.7/2020/CRP.1]. Consideration of the Notification from the World Health Organization Concerning Scheduling under the Single Convention on Narcotic Drugs of 1961 as Amended by the 1972 Protocol, and the Convention on Psychotropic Substances of 1971**; Commission on Narcotic Drugs Reconvened Sixty-Third Session Vienna, pp. 2–4. December 2020. https://www.unodc.org/documents/commissions/CND/CND_Sessions/CND_62Reconvened/ECN72020_CRP19_V2006823.pdf.

UNODC, 2016. World Drug Report 2016, United Nations Office on Drugs and Crime. https://doi.org/10.1142/9789814508957_0005.

Vavilov, N.I., 1992. Origin and Geography of Cultivated Plants (translated by Doris Löve). Cambridge University Press, Cambridge.

Vergara, D., Huscher, E.L., Keepers, K.G., Givens, R.M., Cizek, C.G., Torres, A., Gaudino, R., Kane, N.C., 2019. Gene copy number is associated with phytochemistry in *Cannabis sativa*. AoB Plants 11, 1–12. https://doi.org/10.1093/aobpla/plz074.

Wei, T., Simko, V., 2017. R package "corrplot": visualization of a correlation matrix. Version 0.84.

Wei, X., Zhao, X., Long, S., Xiao, Q., Guo, Y., Qiu, C., Qiu, H., Wang, Y., 2021. Wavelengths of LED light affect the growth and cannabidiol content in *Cannabis sativa* L. Ind. Crop. Prod. 165, 113433. https://doi.org/10.1016/j.indcrop.2021.113433.

Wilkinson, J.D., Williamson, E.M., 2007. Cannabinoids inhibit human keratinocyte proliferation through a non-CB1/CB2 mechanism and have a potential therapeutic value in the treatment of psoriasis. J. Dermatol. Sci. 45, 87–92. https://doi.org/10.1016/j.jdermsci.2006.10.009.

Wright, K., Rooney, N., Feeney, M., Tate, J., Robertson, D., Welham, M., Ward, S., 2005. Differential expression of cannabinoid receptors in the human colon: cannabinoids promote epithelial wound healing. Gastroenterology 129, 437–453. https://doi.org/10.1053/j.gastro.2005.05.026.

Xue, R., Du, M., Zhou, T., Ai, W., Zhang, Z., Xiang, X., Zhou, Y., Wen, Z., 2020. Polysaccharides from hemp seed protect against cyclophosphamide-induced intestinal oxidative damage via Nrf2-Keap1 signaling pathway in mice. Oxid. Med. Cell. Longev. 2020, 1813798. https://doi.org/10.1155/2020/1813798.

Zhang, Q., Chen, X., Guo, H., Trindade, L.M., Salentijn, E.M.J., Guo, R., Guo, M., Xu, Y., Yang, M., 2018. Latitudinal adaptation and genetic insights into the origins of *Cannabis sativa* L. Front. Plant Sci. 9 https://doi.org/10.3389/fpls.2018.01876.

Zuardi, A.W., Rodrigues, N.P., Silva, A.L., Bernardo, S.A., Hallak, J.E.C., Guimarães, F.S., Crippa, J.A.S., 2017. Inverted U-shaped dose-response curve of the anxiolytic effect of cannabidiol during public speaking in real life. Front. Pharmacol. 8, 259. https://doi.org/10.3389/fphar.2017.00259.