eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Department of Analytical Chemistry - Kimika Analitikoa saila

Analytical approaches in technology development for quality medical cannabis production

Kalitatezko kannabis medizinalaren produkziorako teknologiaren garapenerako erraminta analitikoak

Markel San Nicolas Oruetxebarria

2024

Doktorego tesi honek iraun dituen lau urteak gorabehera profesional eta emozionalez beteta egon dira, denbora luzez hedatu den errusiar mendi baten antzera. Begirada denboran atzera botaz, aldi oso pozgarria izan dela ikus dezaket, zeinetan, azkenean, aldi hau osatzeko gai sentitzeak eta merezi izan dezakedala uste izateak, poztasunez bete eta harro sentiarazten nauen. Izan ere, tesiak iraun dituen urte hauek ez dira guztiz errazak izan, erori eta altxatzearen legeak gidatu baititu joan-etorrien itsasaldiak, baina tesia paper gainean ikustea eta azken helburu hau lortzea ezinezkoa izango zitzaidakeen ondoan izan zaituztedanen laguntzarik gabe. Horregatik, ezer baino lehen, zuei bihotzez eskertzea dagokit.

Eskerrik asko, lehenik eta behin, Pot Sistemak S.L.-ko jendeari, zuengatik izango ez balitz ez nintzateke honaino heldu izango eta. Zuek eman zenidaten lehen aukera eta zuek helarazi zenidaten tesi proiektu hau abian jartzeko lehen proposamena. Ikasketak amaitu berri, une hartan zuen talde atsegingarri hartan sartzeko aukera eman izanagatik ikaragarri esker oneko nauzue gaur eta, esan bezala, lehen aukera horregatik izan ez balitz, gaur ez nintzateke doktorego tesi honek azken hitzak idazten egongo. Ezin ahaztu ere gerora ezagutu nuen Sovereign Fields-eko taldea, haien gabe ikerketa gauzatzeko landare kultiborik ere ez litzateke egongo eta.

Ibilbide hau ezin izango nuke burutu IBeA ikerketa taldeko lagun eta lankideengatik izango ez balitz ere. Tesi hau aurrera ateratzeko ikasi behar izan dudan dena zuek erakutsi didazue eta zuen denen elkartasunagatik izango ez balitz, gaur ez nintzateke orrialde honen aurrean hau idazten egongo. Eskerrik asko bihotzez, beraz, ikasketa prozesu honetan tesi hau aurrera ateratzeko laguntza eta indar gogoa jarri duzuenei, nahiz eta zuek txikia izan dela uste, niretzako esfortzu horrek ikaragarri suposatu du eta. Denoi mila esker, guztion arteko esfortzua izan da hau eta. Esker oneko nauzue denok. Horregatik ere, nola ahaztu mintegiko lankideez, egunerokotasunari aurrera jarraitu nahi izateko txinparta eta arima eman diozuenak. Hau zuengatik doa ere, elkarlana aburu, bakarkako helburuen lorpenak haiek partekatzeko talderik gabe zentzurik ez duela erakusteagatik. Eskerrik asko laneko basamortuaren erdian mintegiaz ongizate oasia egitegatik.

Baina barkatuko nauzue, batez ere, nire esker onak Aresatz eta Oierri zuzentzen badizkiet. Eskerrik asko proiektu hau burutzeko jarri dituzuen gogo, indar, ahalegin eta buruhauste denengatik. Izan ere, ibilbide hau errusiar mendia izan da denontzako, zuek barne, eta bihotzez eskertzen dizuet uneoro buru-belarri hor egoteko erabakia hartu eta mantendu izanagatik. Zuen eskutik ikasketa eta formakuntza prozesua izan den tesi honetan erratzeko guztizko askatasuna topatu dut, norberak bidean agertzen diren galde rei erantzuna bilatzeko libertateak aurkezten dituen erronkekin, baina momentu oro zuen sostengua ikusiz. Denbora honetan, asko ikasi dut horrela, zientzia esparrutik haratago, eta horren hein handi bat zuei zor dizuedala uste dut. Mila mila esker bihotzez, beraz.

Azkenik, baina ez garrantzi gutxiagoz noski, familia, bizitzak aurrean jartzen dizkigun oztopoen artean, aurrera jarraitzeko arrazoirik garrantzitsuena. Eskerrik asko aita, ama, Arrate eta etxeko guztiei, zuen betiko baldintzagabeko euskarriak gaur naizen pertsona bihurtu naute eta. Eskerrik asko nigan momentu oro sinistu izanagatik eta aurrera jarraitzeko indarrak emateagatik. Eta ez aparte, ezin lerro hauek itxi, nola ez, Bego, zuk merezi dituzun meritu mugagabeengatik bihotzez eskerrak eman gabe, tesi oso honek iraun dituen egun guzti eta bakoitzean alboan uneoro egon zarena, nire errusiar mendi honen gorabeherak zureganatuz, eta zure indar, alaitasun eta bizi grina elkarbanatuz. Ibilbide oso honen helmugara heltzea ezinezkoa izango litzateke alboan eduki izan ez bazintut, horregatik, bihotzez, eskerrik asko. Espero dut ni zutaz nagoen bezain

harro egotea. Hau zuengatik eta zuentzako doa batez ere, zuek gabe ez bainintzen gaur naizena izatera heldu izango eta. Bihotz bihotzez, mila esker.

Overview

This research was carried out under the premise of exploring new analytical technologies regarding the cultivation of cannabis for medical purposes, due to the remarkable growth in the popularity of this plant over the last decade. As the forecasts of its expansion and integration into worldwide health systems are positive, accurate and precise analytical methodologies need to be developed to ensure proper control of products for medical and pharmaceutical purposes of this provenance. This research project therefore studied the feasibility of two analytical approaches for this purpose.

On the one hand, an untargeted metabolomic research of the cannabis plant was conducted with the aim for searching constitutive metabolic markers to predict the adaptability of cannabis varieties to tropical climates, since some of the countries that are integrating cannabis into their regulatory legislation exhibit climatic features such as high temperature and humidity, which may influence the bioproduction yield of bioactive compounds in the plant. To perform this research accurately, an optimisation of the methodology to be applied was previously performed, so that the determined procedure would offer the broadest possible metabolic coverage and thus capture metabolites from a wide polarity range through a single analytical method, based on liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS) analysis.

On the other hand, the viability of hyperspectral imaging as a non-invasive analytical technique was studied for its integration into the quality control and traceability assurance workflows of cannabis chemotypes in the cultivation facilities. Since the expertise around cannabis cultivation methodologies originates from a clandestine background, the practices to accomplish it might questionably lack strict control, so this methodology is presented as an opportunity for its integration in the automated processes of the medical cannabis production industry in compliance with the product control standards.

Laburpena

Ikerketa lan hau erabilpen medizinalerako kannabisaren kultibazioan teknologia analitiko berritzaileak aztertzeko beharrizanaren pean egin zen, azken hamarkadan zehar landare honek izan duen ospearen hazkunde nabarmenaren ondorioz. Kannabisa munduko osasun-sistemetan hedatzeko eta integratzeko aurreikuspenak positiboak direnez, metodologia analitiko doi eta zehatzak garatzea beharrezkoa da, jatorri honetako aplikazio medizinal eta farmakologikorako produktuen kontrol egokia bermatzeko. Ikerketa proiektu honetan, beraz, jomuga honetarako bi erraminta analitikoren bideragarritasuna aztertu zen.

Alde batetik, kannabis landarearen ikerketa metabolomiko ez-zuzendua gauzatu zen, kannabis barietate ezberdinek klima tropikaletara egokitzeko duten gaitasuna aurreikusteko markatzaile metaboliko eratzaileak bilatzeko. Izan ere, kannabisa beren legeria erregulatzaileetan integratzen ari diren herrialde batzuek ezaugarri klimatiko hauek aurkezten dituzte, hala nola, tenperatura eta hezetasun altuak, eta giro-faktore hauek eragina izan dezakete landarearen konposatu bioaktiboen produkzioaren errendimenduan. Ikerketa hau zehatz-mehatz gauzatzeko, aplikatu beharreko metodologia aldez aurretik optimizatu zen, jarraitu beharreko prozedurak ahalik eta estaldura metaboliko zabalena eskaintzeko eta, horrela, polaritate izaera desberdineko metabolito sortak metodo analitiko bakar baten bidez eskuratzeko, bereizmen altuko masa-espektrometria eta kromatografia likidoan oinarrituta.

Bestalde, irudi hiperespektralen bideragarritasuna aztertu zen teknika analitiko ez-inbaditzaile gisa, kultibazio-instalazioetan kannabis kemotipoen trazabilitatea bermatzeko kalitate-kontroleko lan-fluxuetan integratzeko. Kannabisa kultibatzeko metodologien inguruko ezagutza legez-kanpoko aurrekarietatik eratorria denez, hura gauzatzeko praktikek kontrol-zorrotzerako neurriak faltan izan ditzakete eta, beraz, metodologia hau aukera gisa aurkezten da erabilpen medizinalerako kannabisaren produkzio-industriako prozesu automatizatuetan integratzeko, horrela produktuen kontrol-arauak betez.

Table of contents

SARRERA - INTRODUCTION

1. Kannabisaren eta haren osagai bereizgarrien sarrera

Kannabisa, zientifikoki *Cannabis sativa L*. bezala ezagutzen dena, urteko landare dioiko bat da, hasiera batean Asiako mendebaldeko lehen nekazal gizarteetan landu zena [1]. Bere leinu taxonomikoa honela egituratuta dago [2]:

Erreinua: Viridiplantae

Dibisioa: Streptophyta

Klasea: Eudicotyledons

Azpiklasea: Rosids

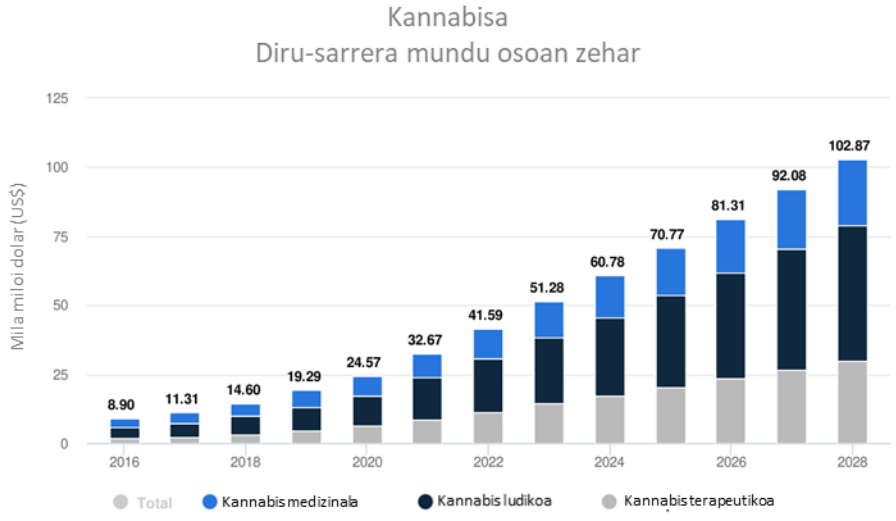
Ordena: Rosales

Familia: Cannabinaceae

Generoa: *Cannabis*

Espeziea: *Cannabis sativa* L.

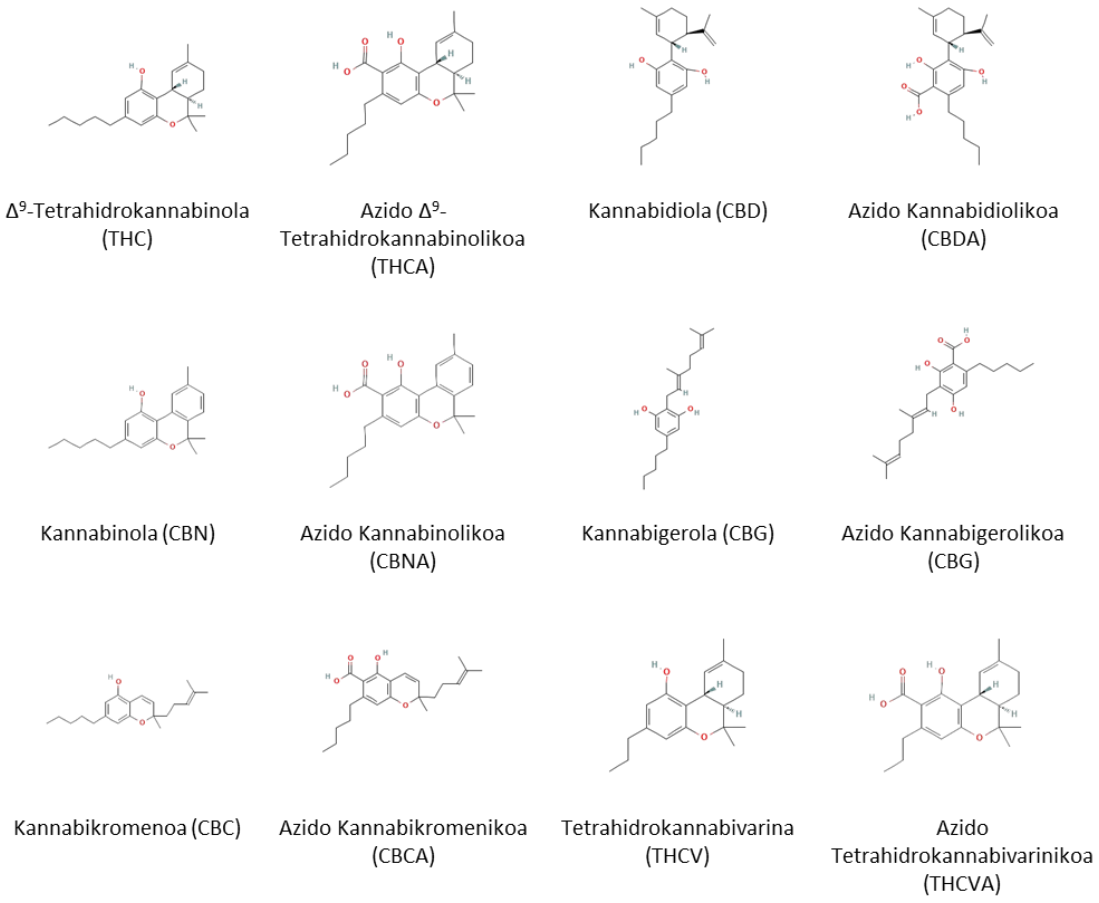
Landare honen lehen erabilera medikoen erregistroak K.a. 2800 ingurukoak dira. Osasun-arazo ugaritarako tratamendu gisa erabiltzen zen garaia izan zen hura, nahiz eta uste den lehenago ehungintzarako material gisa ere erabiltzen zela. Kannabisa Txinako medikuntzaren aitatzat hartzen den Shen Nung enperadorearen farmakopean zegoen [3]. Geroztik, kannabisaren erabilera eta landaketa munduko beste eskualde batzuetara hedatu zen pixkanaka, nazioarteko harreman komertzialak medio. Horrela, denboraren poderioz, kannabisa munduko drogarik zabalduenetako bat izatera iritsi da. Gaur egun, azken estimazioen arabera, kannabisaren munduko merkatuak 2024an 60.79 mila milioi dolarreko fakturazioa izatea espero da, eta urteko hazkunde-tasa estimatua % 14.06koa izanik (CAGR 2024-2028), 2028rako 102.90 mila milioi dolarreko guztizko merkatu-bolumenera iritsi daitekeela uste da (*1. Irudia*) [4]. Beraz, erabileraren eta laborantzaren inguruko iragarpenak oso positiboak dira, eta kultibo honi loturiko enplegua ere are gehiago hedatuko dela aurreikusten da.



1. Irudia: Aurreikusitako mundu mailako kannabis merkatuaren irabaziak

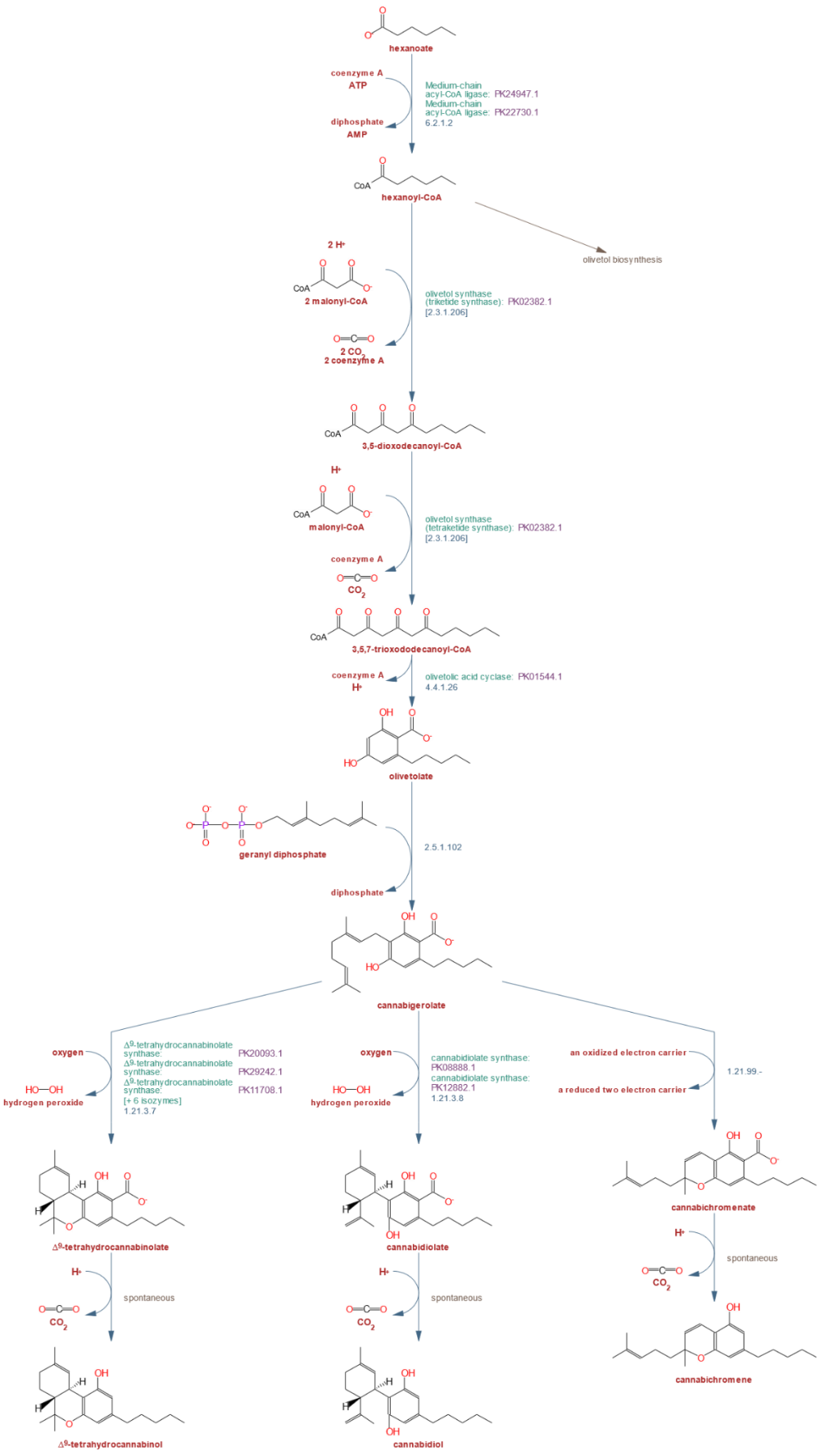
Baina, zergatik izan da hain ezaguna kannabisa historian zehar, eta zergatik jarraitzen du oraindik hala izaten? Beno, gaur egun, ezaguna da kannabisak efektu psikoaktiboak eragiten dituela kontsumitzen denean, Δ^9-tetrahidrokannabinolaren (THC) presentziaren ondorioz, baina 1964. urtera arte hau ez zen gauza jakina izan, Dr. R. Mechoulamek eta Dr. Y. Gaonik konposatua isolatu eta identifikatu zuten arte hain zuzen ere [5]. Gainera, Mechoulam izan zen, Y. Shvo doktorearekin batera, urte bete lehenago kannabis landarean kannabidiola (CBD) isolatu eta identifikatu zuena [6]. Dena den, kannabinola (CBN) izan zen kannabisetik bere forma puruan isolatu zen lehen konposatua, 1899an hain zuzen ere, eta garai hartan modu okerrean kannabisaren efektu psikoaktiboen erantzule nagusi gisa aurkeztu zen [7], [8]. Aurkikuntza horien ondorengo gertaerek kannabisean zeuden antzeko konposatu gehiago azaleratu zituzten, modu honetan konposatu talde bat identifikatuz, fitokannabinoide edo kannabinoide gisa deiturikoak.

Kannabinoideak isoprenilatutako erresorzinil poliketido gisako konposatuak dira. Kannabis landarean gain beste landare batzuetan ere aurkitu diren arren [9], oso modu bereizgarrian aurkitzen dira landare-espezie honetan. Kannabinoide garrantzitsuenak *2. Irudian* agertzen diren 12ak dira, baina kopuru osoa nabarmenki handiagoa da, gutxienez 113 kannabinoide identifikatu baitira kannabis landarean [10].



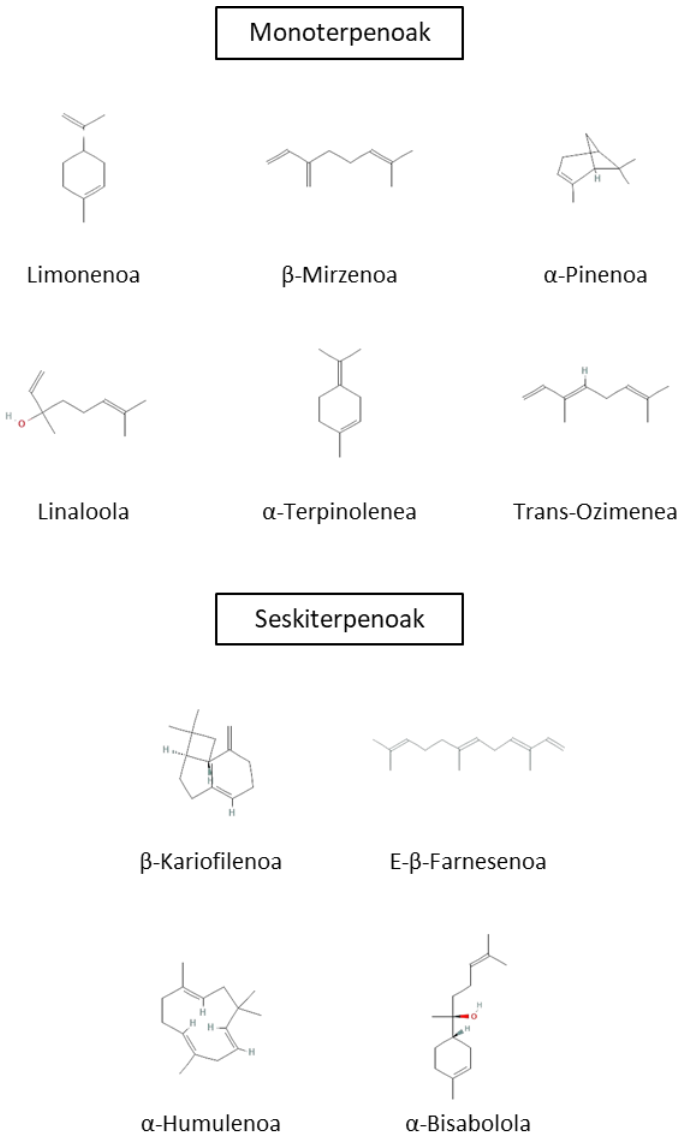
<u>2. Irudia:</u> Kannabisean aurkitzen diren kannabinoide garrantzitsuenak

2. Irudian ikus daitekeen bezala, erakutsitako konposatuak 6 kannabinoide pare dira, bakoitzari dagokion bere forma neutroaz eta azido organikoaz osatuak; izan ere, kannabinoideak nagusiki azken modu horretan sintetizatzen dira kannabis landare emeen infloreszentzietako trikoma glandularretan, bide metaboliko arrunt baten bidez [11]. Lore arrek kannabinoideen oso kontzentrazio txikia izan ohi dute, hauetan trikomen edukia oso mugatua izaten baita [12]. Bere sintesia hexanoil-CoA-tik abiatzen da substratu kimiko bezala, non tetraketido sintasa entzimak (TKS) bere kondentsazioa katalizatzen duen 3 malonil-CoA molekulekin [13], modu honetan 3,5,7-trioxododekanoil-CoA osatuz [14]. Molekula hau jarraian ziklatu egiten da azido olibetolikoan (OLA) bihurtuz, azido olibetoliko ziklasa (OAC) bidezko katalizazioaren ondorioz [15], eta, ondoren, preniltransferasa aromatiko batek geranilo pirofosfatoaren (GPP) prenil taldea 2-erresorzinol posizioan txertatzea katalizatzen du, modu honetan kannabinoide aitzindaria sortuz, CBGA [16]. CBGA beste kannabinoide batzuen sintesirako oinarrizko konposatua da; eta bere sintesi bidea hiru norabidetan banatzen da THCA, CBDA eta CBCA kannabinoideen biosintesiak ahalbidetuz [17]. Kannabinoide azido horiek dekarboxilatu egiten dira gero, erreakzio espontaneo ez-entzimatikoen bidez, gehiegizko argiztapenaren, beroaren edo denboraren joanaren ondorioz, eta, horrela, THC, CBD eta CBC eratzen dira, hurrenez hurren [18]. Gainera, CBNA eta CBN-a THCA-ren eta THC-aren degradazioaren ondorioz sortzen dira, hurrenez hurren [19]. Bestalde, aurrez aipatutako gainerako kannabinoideak sortzeko, CBCA-tik abiatutako biosintesi bide gehigarriak gertatzen dira. Kannabinoideen biosintesi-bidea *3. Irudian* laburbiltzen da.



3. Irudia: *Kannabinoideen biosintesi-bidea* [20]

Baina kannabinoideez gain beste konposatu asko ere aurkitu dira kannabis landarean. Izaera desberdineko 550 konposatu baino gehiago identifikatu dira guztira [21], 200 konposatu hegazkor barne [22]. Identifikatutako konposatu horietatik guztietatik, terpenoak eta terpenoideak bereziki nabarmentzen dira. Terpenoak isoprenozko unitate txikiz osatutako hidrokarburoak dira; terpenoideak, berriz, oxigenoa duten terpenoak dira. Olio esentzialen osagai nagusiak dira, eta kannabisaren usain bereizgarriaren eta landareak duen famaren arduradun nagusiak dira [23]. Badakigu kannabisean gutxienez 120 terpeno daudela, eta normalean honela sailkatzen dira: monoterpenoak, bi isopreno unitatez osatuak (10 C); seskiterpenoak, hiru isoprenoz (15 C); diterpenoak, lau isoprenoz (20 C); eta triterpenoak, sei isoprenoz (30 C) osatuak [24]. Horietatik, 58 monoterpeno eta 38 seskiterpeno karakterizatu dira orain arte [25]. Monoterpenorik aipagarrienak limonenoa, β-mirzenoa, α-pinenoa eta linalola dira, eta, agian, baita α-terpinolenoa eta trans-ozimenoa ere [26]. Bestalde, seskiterpeno nabarmenenak honako hauek dira: β-kariofilenoa, E-β-farnesenoa, α-humulenoa eta α-bisabolola [27], [28]. Terpeno hauen egiturak *4. Irudian* ikus daitezke.



<u>4. Irudia:</u> Kannabisean aurkitzen diren terpeno garrantzitsuenak

Geroz eta eztabaida zientifiko handiagoa dago terpenoek, kannabinoideekin elkarreraginean, efektu sinergikoa eragin dezaketen ala ez. Hala balitz, profil terpeniko ezberdinek kannabinoideen profil jakin batek eskainiko lituzkeen efektu medizinalak aldatuko lituzkete [25]. Horregatik, duela gutxi, terpenoak kannabisaren erabilpen medikorako ikerketa eta garapeneko (I+G) funtsezko helburu bihurtu dira.

Horrela, kannabinoideak eta terpenoak dira kannabisaren marka bereizgarriaren konposaturik ezagunenak, landarea ezaugarritzen duten potentzial medikoaren eta terapeutikoaren erantzule baitira. Hala ere, beste konposatu-mota asko ere badaude kannabis landarean, eta horiek aktiboki inplikatuta daude landarearen metabolismoan. Konposatu horien adibide nagusiak *1. Taulan* agertzen dira [10], [21], [23], [29].

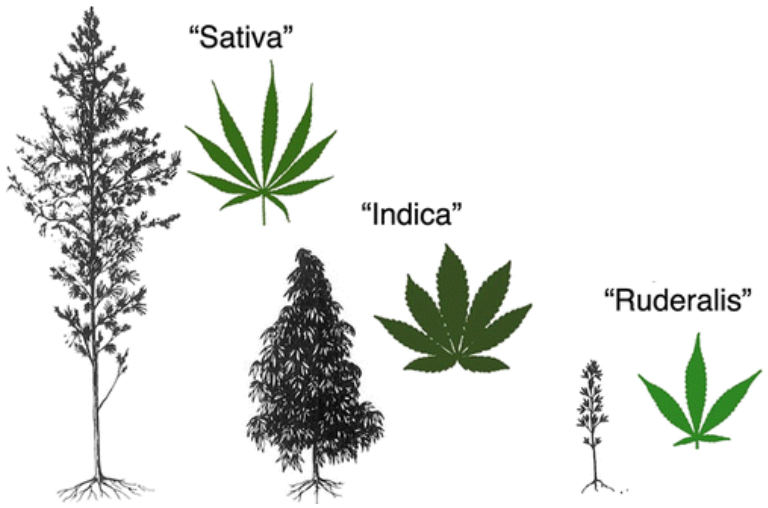
1. Taula: Kannabis landarean identifikatutako konposatu-mota ezberdinak

Konposatu taldea	**Aurkitutako konposatu kopurua**
Terpenoak	120
Kannabinoideak	113
Hidrocarburoak	50
Polifenolak	48
Azukreak	34
Konposatu nitrogenatuak	27
Gantz azidoak	22
Azido sinpleak	21
Amina sinpleak	18
Zetona sinpleak	13
Ester eta laktona sinpleak	13
Aldehido sinpleak	12
Esteroideak	11
Alkohol sinpleak	7

Beraz, ziurtasunez esan daiteke kannabisak konplexutasun kimiko zabala duela, orokorrean ezagutzen diren THC-tik eta CBD-tik haratago. Hori dela eta, kannabisa interes publikoaren arreta-zentroan egon da azkenaldian, gizartean finkatuta dauden sendagai sintetikoak ordezkatzeko jatorri naturaleko produktu medizinal gisa duen potentzialagatik.

2. Kannabisaren kemotipo aniztasuna

Kannabisaren konplexutasun kimikoa alboratuta, kannabisaren espeziearen barruan bi talde nagusiren arteko bereizketa orokorra egin zen bere taxonomian oinarrituta: *cannabis sativa* eta *cannabis indica*. Sailkapen hori Meijer eta van Soest-ek sartu zuten literatura zientifikoan [30]. *Sativa* gisa sailkatutako kannabis-landareen ezaugarri nagusiak hauek dira: foliolo estuak, altuera handiko hazkuntza eta heltze berantiarra. *Indica* landareak aldiz, foliolo zabaleko eta zuhaixka gisako hazkuntzadun heltze goiztiarreko landare gisa definitu ziren [31]. Hain ezaguna ez den beste azpiespeziea *rudelaris* gisa izendatu zen. Europa ekialdean eta Errusian dokumentatua, morfologiaz tamaina txikikoa eta hiru folioloz osatutako hostoak dituena, eta THC-ren eduki nahiko baxuak aurkezten dituena. Azpiespezie honen ezaugarri nagusia autolorakuntzaren atributua da, ez baitu fotoperiodoaren indukzioarekiko menpekotasunik (*5. Irudia*).



5. Irudia: Cannabis sativa, indica eta ruderalis azpiespezieen arteko ezberdintasun morfologikoak. Irudiaren iturria: John M. McPartland. Cannabis and Cannabinoid Research. Dec 2018.203-212. http://doi.org/10.1089/can.2018.0039 [31]

Hala ere, kannabisaren legez-kanpoko kultibazio ezaugarri nagusia genetika-Mendeliarrean oinarritutako garapen-hazkuntza izan da beti [32]. Hala, landarearen ezaugarri psikoaktiboak hobetzeko helburuarekin egindako garapenaren ondorioz, gaur egungo *sativa* eta *indica* azpiespezieen arteko desberdintasuna ezerezera diluitu da. Modu honetan, barietate ezberdinen arteko bereizketarako gaur egungo irizpidea kemotipoan oinarritzen da, kannabis landareko kannabinoide nagusien proportzioen arabera definitzen dena, batez ere THC-an eta CBD-an oinarrituz. Kannabisaren barruan bost kemotipo bereiz daitezke, *2. Taulan* laburbiltzen direnak. I, II eta III kemotipoak dira klaserik ezagunenak, THC eta CBD eduki esanguratsuak dituztenak baitira.

2. Taula: Kannabinoide-edukiaren araberako sailkapen irizpideak

Kemotipoa	Kannabinoide edukiaren erlazioen araberako sailkapena
I	≈ [THC]/[CBD] >10
II	≈ 3 > [THC]/[CBD] > 0.3
III	≈ [THC]/[CBD] < 0.1
IV	[CBG] >> [THC] & [CBD]
V	Ia kannabinoiderik ez

Horrela, gaur egun, sailkapen hori zehatzagoa da, *sativa/indica* bereizketa zaharkituta geratu baita. Gainera, azken bereizketa horrek ez du zorroztasun zientifikorik, sailkapen bernakularra baita; beraz, ikuspegi zientifiko batetik, egokiagoa da kannabisa landareak edo barietateak kannabinoideen eta terpenoen profilaren arabera kategorizatzea, profil hori faktore genetikoen menpe baitago.

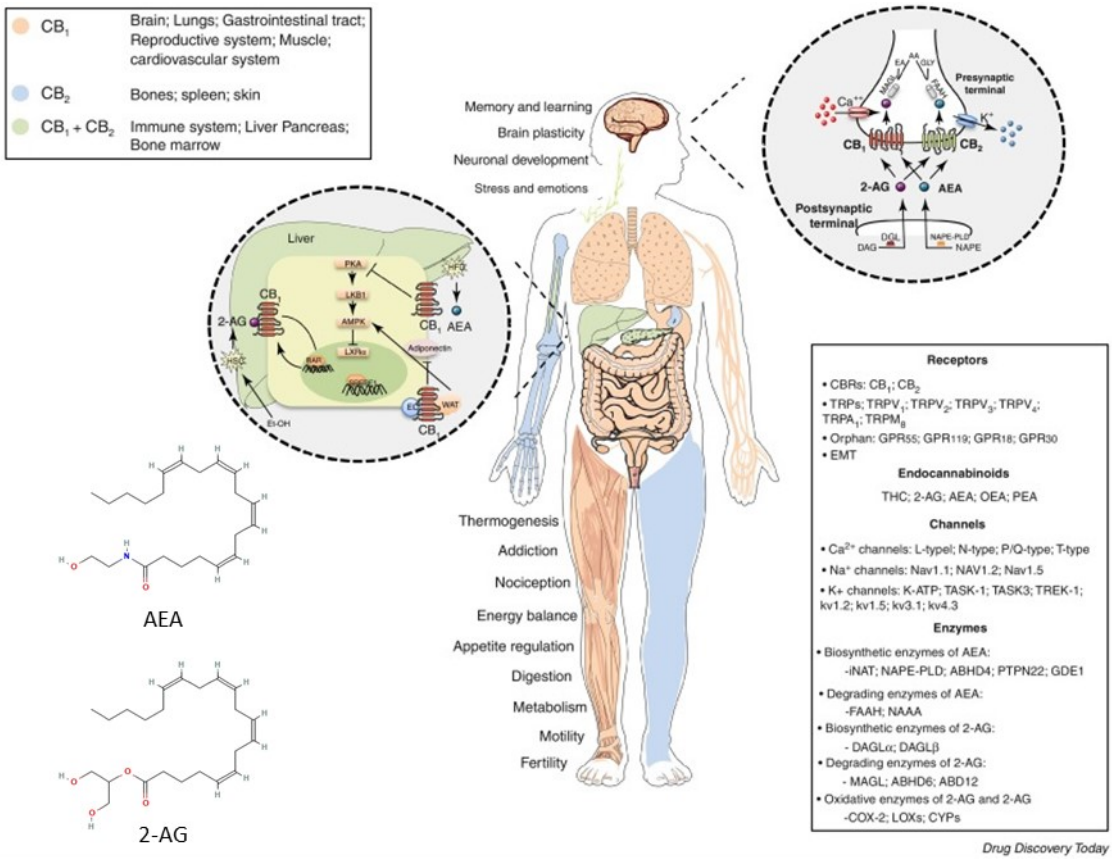
Badakigu kannabis-landare baten kemotipoa B locus-eko bi alelo kodominanteren mende dagoela, B_D and B_T, eta horiek dira, hurrenez hurren, landareko CBDA-ren eta THCA-ren presentziaren erantzuleak [33]. Ondorioz, B_T/B_T alelodun kannabis barietateak I kemotipokoak dira, B_D/B_D dutenak III kemotipokoak dira eta B_D/B_T alelo ezberdinak dituzten barietateek kannabinoide nagusien eduki orekatua sortzen dute, II kemotipokoak, hain zuzen ere [34], [35]. Bestalde, IV kemotipoko kannabis-landareak B locus-ean funtzionalak ez diren aleloen (B_0) emaitza dira. CBGA-tik CBDA-rako bihurketa biosintetikoa inhibitzen dute alelo hauek, emaitza gisa nagusiki CBGA duten landareak emanez [36].

Beraz, kemotipo ezberdinen existentzia oso gai erakargarria bihurtu da erabilera medizinaleko eta farmakologikoko kannabisaren esparruan, espezie honen barietate ezberdinek giza gorputzean eragin dezaketen eragin ezberdinak direla eta.

3. Zerk bihurtzen du kannabisa hain etorkizun handiko alternatiba medikoa?

Kannabisak alternatiba mediko natural gisa hainbesteko fama irabazi izanaren arrazoia kannabinoideek giza nerbio-sistemaren hartzaileekin duten interakzioan datza, giza-endokannabinoide sistema (ECs) bezala ezagutzen den interakzioan [37]. Sistema endokannabinoideak hartzaile neurologikoz (CB), ligando endogenoz (endokannabinoideez) eta hauen biosintesia eta degradazioa eragiten duten entzimez osatzen da. Neurosistema honek funtsezko zeregina betetzen du hainbat prozesu fisiologikotan, hala nola, gosearen erregulazioan, minaren bitartekotzan, umorean edo memorian, eta garuneko plastikotasunaren, garapen neuronalaren eta oreka energetikoaren erregulazioaren arduraduna ere bada [33].

Bi neurohartzaile mota nagusik aktiboki parte hartzen dute ECs-an, hain zuzen ere G proteinetara (GPCR) akoplatutako hartzaileek. GPCR horietatik, CB_1R garuneko eskualdeetan adierazten da nagusiki, eta CB_2R, berriz, sistema immunologikoko zeluletan. Modu honetan, ECs-aren bidez efektu neurologiko bat eragiteko, hartzaile horiek aktibatu egin behar dira, eta hori ligando endogenoekin duten interakzioari esker gertatzen da; horien artean, anandamida (AEA) eta 2-arakidonoil-glizerola (2-AG) dira ordezkari nagusiak (*6. Irudia*) [38]. Gainera, jakina da 1. motako banilloide hartzailea (TRPV1) eta CPR55 ere jomuga posibleak direla ECs-an [39], [40].

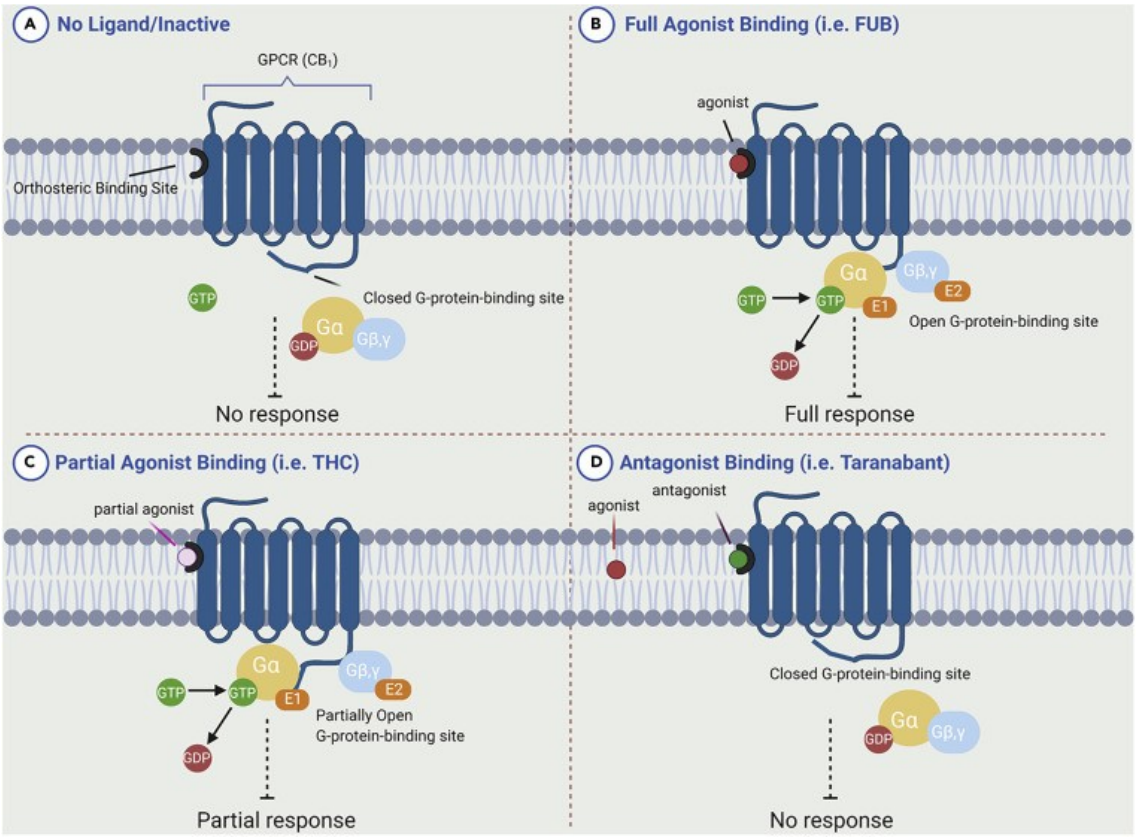


<u>6. Irudia:</u> Endokannabinoide sistemaren irudia, bere funtzio, banaketa eta osagaiekin. Hurrengo iturritik berreskuratutako eta aldatutako irudia: O. Aizpurua-Olaizola et al. "Targeting the endocannabinoid system: future perspectives and strategies", Drug Discovery Today, vol. 22, no. 1, Jan. 2017. http://dx.doi.org/10.1016/j.drudis.2016.08.005

Endokannabinoide sistemak gertaera fisiologiko eta patologiko ugaritan rol homeostatikoa betetzen duela erakutsi du, sistema honek azkar aldatzen diren egoeretara egokitzeko gaitasuna baitu, eskaeraren araberako askapenen ondorioz. Hala ere, hainbat gaixotasun neurologikoren kasuan, hala nola Alzheimer gaixotasuna, Parkinson gaixotasuna, Huntington gaixotasuna edo esklerosi anizkoitza, endokannabinoide sistema aldatu egiten da, sistema honetan homeostasia zailduz. Horregatik, ECs-an eragitea patologia neurologiko horiei aurre egiteko helburu nagusi bihurtu da [38]. Beraz, gaixotasun horiei zuzenean ECs-a jomugan jarriz aurre egiteko, ligando endokannabinoideak modu exogenoan xurgatu beharko lirateke, eta hori zaila da konposatu horien egonkortasun kimiko txikia dela eta. Hala ere, ikusi da fitokannabinoideak agonista selektiboak direla ECs-ko CB hartzaileekiko, eta, beraz, nerbio-sistema iruzurtzeko gaitasuna dutela, giza-prozesu fisiologikoei eraginez. Modu honetan etorkizun oparoa izan dezaketen farmazeutikoki aktibo diren osagaiak (API) bihurtu dira fitokannabinoideak [41].

Ildo horretan, funtsezkoa da kannabinoideek eragindako seinaleen transdukzioak sortutako jardueraren eragina kannabinoideen hartzaileenganako selektibitatearen menpe dagoela jakitea, endokannabinoideekin gertatzen den bezala. Afinitatea aldatu egiten da konposatu bakoitzaren estereokimikaren arabera. Adibidez, THC-a agonista partziala da, bai CB_1R hartzaileentzat, bai CB_2R hartzaileentzat ere, lehenengoentzat afinitate handiagoa duelarik (*7. Irudia*) [42], [43], [44], [45]. CBD-aren kasuan aldiz, bi hartzaileekiko afinitatea oso mugatua dela ikusi da [46], [47], baina kontzentrazio nanomolarretan CB_1R-aren eta CB_2R-aren antagonista gisa jarduten du [48], CBD-ak maila horietan THC-ak eragindako eragin psikotropikoa modulatu dezakeela iradokiz. THCV-a ere CB_1R-aren agonista dela ikusi da, baina kontzentrazio txikietan bi hartzaileekiko antagonista gisa jokatzen du [49]. Bestalde, CBN-a agonista da bi hartzaileentzat [50]. Era berean, CBG-a CB_2R-aren antagonista ahula da [51]. Azkenik, β-kariofilenoaren kasua ere aipagarria da, kannabinoide gisa jokatzen duen terpeno bakarra baita, kasu honetan CB_2R-a aktibatuz [52].

Hori dela eta, kannabisak ECs-an eragindako efektuak erabat ulertzea oso lan zaila da, erantzun fisiologikoak kontsumitutako kannabinoideen profilaren araberakoak baitira erabat. Horregatik, endokannabinoideen inguruko ikerketa sakonak egiten ari dira gaur egun, erantzun fisiologiko posible bakoitzaren aurrean izan dezakeen portaera zehaztasunez deskribatzeko. Argi dago, behintzat, kannabis-kemotipo batzuek erantzun-efektu ezberdinak eragin ditzaketela, eta, beraz, patologia bakoitzean duten eragina ikertu behar da, erremediorik egokiena aurkitzeko. Ondorioz, doktorego-tesi honetan, hainbat ikuspegi analitiko aztertu dira kannabis medizinalerako hazkuntzaren kalitatea bermatzeko; izan ere, kannabis-loreek ere industria farmazeutikoan API-ak ekoizteko gaur egun aplikatzen diren kalitate-estandar guztiak bete beharko baitituzte.

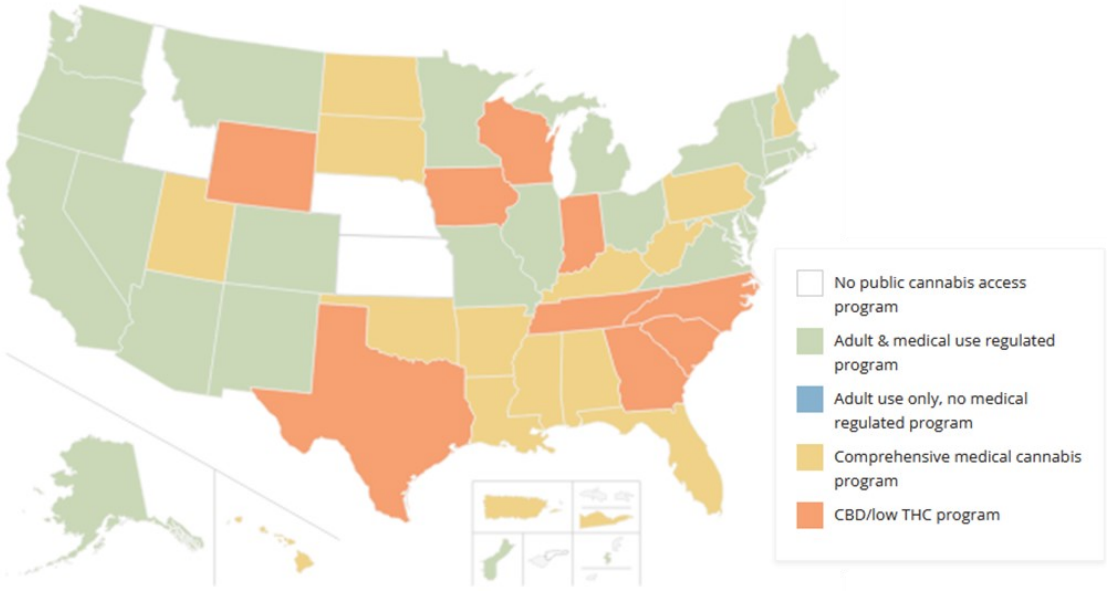


7. Irudia: Ligandoek eta THC-ak CB_1R-a hartzailearekin duten elkarreraginaren bidez sortutako *seinaleen transdukzioaren irudikapen eskematikoa. Irudiaren iturria: F. Shahbazi et al "Cannabinoids and Cannabinoid receptors: The Story so Far", iScience, vol. 23, no. 7, Jul. 2020. https://doi.org/10.1016/j.isci.2020.101301* [45]

4. Erabilera medizinalerako kannabisari buruzko gaur egungo araudia

Azken urteotan, nazioarteko araudi aldaketa esanguratsuek eragin handia izan dute mundu osoko hainbat herrialdetan kannabisa helburu medikoekin ikertzeko eta garatzeko (I+G) lege-jarraibideak onartzerako orduan, halako elur-bola efektu bat sortuz. Osasunaren Mundu Erakundeak kannabisa droga narkotikoen zerrendatik kentzeko gomendioa egin zuen [53], eta, horren ondorioz, kannabisa eta kannabis-erretxina estupefazienteei buruzko Konbentzio Bakarraren (1961) IV. zerrendatik kendu ziren 2021ean, modu honetan, duela gutxi arte hartutako eredu murriztailearekin amaituz [54]. Gertaera hori abiapuntu bat izan zen kannabisak erabilera medikorako zuen potentziala aztertzeko, hainbat hamarkadatan indarrean egon zen debekua desblokeatu baitzuen. Horrela, aurrerapen burokratiko horri esker, zenbait herrialdek kannabisaren balizko efektuen ikerketari ekin diote berriki, azterketa klinikoen eta aurreklinikoen bidez.

Ameriketako Estatu Batuen kasuan (AEB), 2018an Nekazaritza Legea onartu zen [55], eta horri esker onartu ziren Marihuana Medizinalaren Legeak (MML) [56]. Lege horiek kannabisean oinarritutako produktu medikoak estatu mailan erabiltzea eta merkaturatzea ahalbidetu zuten. Gaur egun, AEB-ak osatzen dituzten 50 estatuetatik 38k MML-ak onartu dute, eta kannabisa baimendu dute erabilera sendagarrietarako. Hala ere, estatu bakoitzak berariazko zehaztapen arautzaileak ditu indarrean (*8. Irudia*). Garrantzitsua da ere maila federalean kannabisaren erabilera terapeutikoa oraindik onartu ez dela adieraztea, eta horrek oraindik muga batzuk ezartzen dituela. Hala ere, joera guztiz positiboa da etorkizunerako.



8. Irudia: Kannabisaren araudi egoera ezberdinen AEB-etako mapa [56]

Orokorki ezaguna da Estatu Batuak merkataritza askearen eta merkataritza erregulazioaren epitomea direla, eta, nahiz eta hau ikuspegi moralena edo etikoki zuzenena beti ez izan, ukaezina da munduko potentzia ekonomiko nagusietako bat den aldetik mendebaldeko herrialde askorentzat eredu ere badela. Bestalde, Kanada ere erreferentzia izan da munduan, erabilera

medikorako kannabisa arautzeari dagokionez; izan ere, Kanada izan zen 2018ko Kannabisaren Lege Federalaren bidez bere herritarrentzat kannabisa integratzeko baimena eman zuen lehen herrialdeetako bat [57]. Beraz, Ipar Amerikan erakundeen onarpenerako joera hori hedatu egin da, Europako hainbat herrialdetara ere iritsiz. Osasun-aplikazioetarako kannabisarekin lotutako kanpainak egitea erabaki duten herrialde adierazgarri batzuen artean daude Alemania, Suitza, Italia edo Portugal, besteak beste. 2017an, Alemaniako Estupefazienteen Legea ezarri zen, eta lege horren ondoren, kannabisaren erabilera medikoa onartu zen [58]. Lege hori ezarrita, Alemaniako pazienteek eskura izan dezakete kannabisa, lore lehor edo erauzi moduan, errezetarekin soilik eta gaixotasun larrietarako bakarrik. Izan ere, pazienteari dagozkion baldintzak betetzen badira, osasun-aseguruak kannabisarekin egindako terapiaren kostua itzuli behar du [59]. Italiaren kasuan, medikuen errezeta bidezko kannabisarekin egindako sendagaien baimena errealitate bihurtu zen Osasun Ministerioak onartutako formulario bereziaren bidez [60]. Aurrerago, Portugalen 83/2021 Ministro Aginduak zenbait alderdi praktiko definitu zituen kannabis sendagarriarekin lotutako jarduerak baimentzeko, hala nola kannabisarekin egindako sendagaien laborantza, fabrikazioa, merkataritza, garraioa edo inportazioa/esportazioa [61]. Lege-egoera antzekoa da Suitzan, 2022an Estupefazienteei buruzko Suitzako Lege Federala onartuta [62].

Munduko beste lurralde batzuetan ere, kannabisari buruzko legedia aldatu egin da, eta helburu medikoekin landatzea, errezetatzea edo kontsumitzea ahalbidetu da. Adibide ugari daude: Amerikan, Argentina [63], Uruguai [64], Kolonbia [65], *Saint Vincent and the Grenadines* [66] eta Jamaika [67]; Ozeanian, Australia [68]; Asian, Thailandia [69]; eta, Afrikan, Malawi, Zambia, Hegoafrika, Zimbabwe, Lesotho, Ghana eta Ruanda [70], [71], [72]. Afrikan, herrialde bakoitzak bere berezitasun erregulatzaileak ditu.

Baina kannabis sendagarria hazteko arauditik harago, badira farmakopea orokorrean kannabisaz egindako legezko sendagai batzuk. Esate baterako, Sativex ®, Epidyolex ®, Marinol ®, Syndros ® eta Cesamet™ (*3. Taula*). Sendagai horiek merkaturatzeko baimena mundu osoko hainbat herrialdetan eman da dagoeneko, eta orain arte egindako saiakuntza kliniko eta aurrekliniko gehienak sendagai horietako batzuekin egindakoak dira. Hala ere, metatutako ebidentzia orokorrak adierazten du sendagai horien ondorioek oraindik ez dutela erabat betetzen kannabisari egozten zaion potentzial medikoa, eta, era berean, nahi ez diren albo-ondorio batzuk eragin ditzaketela. Beraz, ikerketa medikoa eta farmazeutikoa oraindik abian dago kannabisetik eratorritako medikamentu eraginkorragoak garatzeko, hala nola min kronikoa, nerbio-sistemako gaixotasunak eta buruko eta portaera-nahasmenduak tratatzeko [73].

3. Taula: Kannabisean oinarritutako medikamentu komertzialak

Botika	Kannabinoide konposizioa	Aplikazio terapeutikoa	Administratzeko modua	Sortzailea	Erref.
Sativex®	THC eta CBD (1:1 erlazioa)	Espastizitatea (Esklerosi anizkoitza)	Ahozko spray bidez	GW Pharmaceuticals Limited (UK)	[74]
Epidyolex®	CBD	Lennox-Gastaut Sindromeari (LGS), Dravet Sindromeari (DS) edo Esklerosi tuberoso konplexuari (TSC) loturiko krisiak	Ahoz likido eran	GW Pharmaceuticals Limited (UK)	[75]
Marinol®	Dronabinola (THC sintetikoa)	Kimioterapiak sorturiko goragaleak eta HIESAk eragindako gose-galera	Ahoz likido edo kapsula eran	AbbVie Inc. (USA)	[76]
Syndros®	Dronabinola (THC sintetikoa)	Kimioterapiak sorturiko goragaleak eta HIESAk eragindako gose-galera	Ahoz likido edo kapsula eran	Benuvia Manufacturing (USA)	[76]
Cesamet™	Nabilona (THC sintetikoaren analogoa)	Kimioterapiak sorturiko goragaleak eta HIESAk eragindako gose-galera	Ahoz kapsula eran	Bausch Health Companies Inc. (Canada)	[77]

Espainiari dagokionez, a priori debekatuta dago kannabisa helburu medikoekin landatzea, Sendagaien eta Osasun Produktuen Espainiako Agentziaren (AEMPS) baimena eduki ezean. Hala ere, 2022ko ekainean, Espainiako Kongresuak kannabisa helburu medikoekin erregulatzeko esperientzien ebaluazio-txosten bat garatzea onartu zuen, kannabisaren erregularizaziorantz urrats txiki bat emanez. Dokumentuak zortzi baieztapen zerrendatzen ditu, "erabilera terapeutikoak izan ditzaketen kannabis-prestakinak" daudela adieraziz. AEMPS gaur egun gomendio-txostena garatzen ari da kannabis medizinala erregulatzeko araudiak betez eta kannabis erauzien eta prestakin estandarizatuen kalitatea bermatuz. Bide honetatik, Kannabis Medizinalaren Espainiako Elkartea (AECAME) eratu da [78]. Nolanahi ere, Espainian badaude AEMPS-ek baimendutako ikerketa-baimenak indarrean. 2024ko urtarrilean, 16 lizentzia daude, hain zuzen ere, ikerketa-jardueretarako enpresa horiei kannabisa landatzeko aukera ematen dietenak: Kannabisaren Ikerketarako Nafarroako Elkartea (ANIC), Bhalutek Sens S.L., Biobizz Worldwide S.L., Helicon Medical S.L., Bartzelonako Institutu Botanikoko Ikerketa Zientifikoen Goi Kontseilua, M.P. (IBB-CSIC), Landareen Biologia Molekular eta Zelularraren Institutua (IBMCP), Nekazaritzako Elikagaien Ikerketa eta Teknologiaren Institutua (IRTA), Inveseed Technologies S.L., Mifco Biobrands S.L., Phytoplant Research S.L., Ploidy and Genomics S.L., Quorum Biomedical S.L., Trichome Pharma S.L., Ucie Ars Innovatio eta Sovereign Fields S.L.. Aurkezten den ikerketa-proiektu hau azken enpresa honekin gauzatu da.

Aurrez aipatutako herrialde horiek neurri handiagoan edo txikiagoan legedi bat garatu dute kannabisa helburu medikoekin landatzea eta erabiltzea baimentzeko, eta eredu positiboak dira mundu osoko beste herrialde askorentzat. Ondorioz, gizartean eta erakundeetan gero eta onarpen handiagoa dago, momenturen batean kannabis-loreak edo erauzkinak jatorri naturaleko sendagai gisa normalizatuko dituena.

5. Kannabis medizinalaren laborantzan kalitatea bermatzeko teknologia analitikoak

Kannabisari egozten zaion potentzial medikoaz gain, kontuan hartu behar da landare hori munduan gehien kontsumitzen den aisialdiko drogetako bat izan dela eta izaten jarraitzen duela. Ondorioz, kultiborako metodoak hobetzeko teknikak eta trebetasunak legez-kanpoko esparruan sortu eta aplikatu izan dira, non normalean ez diren kalitate-arauak betetzen. Hala ere, medikamentu batek, sintetikoa zein naturala izan, zenbait zehaztapen farmakologiko bete behar ditu fabrikazio-prozesuan kalitate-kontrola bermatzeko. Ziurtatu behar da lortutako produktuak, edozein bidetatik administratua delarik, nahi den efektua emango duela eta ez duela ustekabeko ondorio kaltegarririk eragingo. Lehen eztabaidatu den bezala, kannabisa legegintza-trantsizio garai batean dago gaur egun, oro har erregularizatzeko bidean, beraz, oraindik goizegi da kalitatea ziurtatzeko estandarrak edo APIrako betebeharrak zehazteko. Hala ere, metodologia jakin batzuk azter edo gara daitezke azken produktuaren kalitate kimikoa bermatzeko. Kasu honetan kannabisaren infloreszentziez ari gara.

Ildo horretan, ulertu behar da kannabisa, funtsean, labore bat dela, nekazaritza-sektoreko beste edozein bezala. Elikagaien industriako laboreak ekonomia globalaren bultzatzaileak dira, eta, horregatik, kalitatea betetzen dela bermatzeko, teknologiak garatu dira, eta, aldi berean, frutak, barazkiak, zerealak eta abar merkaturatzeko zehaztapenak doitu dira. Emaitza nekazaritzako elikagaien industria globalean oso garatua dagoen I+G arlo bat da, kapital-inbertsio handiek babestua, eta horiek gabe labore horien ekoizpena ezingo litzateke egungo merkaturatze-irizpideetara egokitu. Beraz, helburua labore batean oinarritutako sendagai eraginkorrak eta fidagarriak garatzea bada, horiek ekoizteko zehaztapenak eta jarraibide egokiak inplementatu behar dira, eta horrek ahalegin handia eskatzen du kannabisari egokitutako teknologiak ikertzeko eta garatzeko.

Gainera, kannabisa konplexutasun handiko landarea da, eta konposatu bioaktiboen espektro zabala du, lehen aipatutako kannabinoideak eta terpenoak kasu. Berezko konplexutasun horrek erronka gehigarriak gehitzen dizkio substantzia mediko egonkorrak eta errepikakorrak egiteari, eta hori ezinbesteko baldintza da merkaturatutako produktu farmazeutiko bat onartzeko.

Faktore horiek kontuan hartuta, arrazoizkoa da teknologia analitiko berritzaileak garatzea, edo daudenak egokitzea, kannabisaren laborantzako prozedurak edo metodologiak hobetzeko, farmaziako eta medikuntzako industrietan eskatzen diren ekoizpen-estandarretara pixkanaka hurbiltzeko helburuarekin.

Ildo horretan, funtsean ezberdinak diren bi metodologia analitiko, dagoeneko sektore horretan integratuak daudenak eta onarpen handia dutenak, kannabisaren bidean ere etorkizun handiko hautagai gisa agertzen dira. Alde batetik, irudi hiperespektrala, nekazaritzako elikagaien sektoreko ekoizpen-kateetan aspalditik erabiltzen den teknika espektroskopikoa, aipatutako gaiari heltzeko ikuspegi egokia da, teknika analitiko zuzena baita produktuen trazabilitatea eta ekoizpen-tokiko kalitate-kontrola errazteko eta bermatzeko. Bestalde, kannabis barietateek erakutsitako konplexutasun kimikoa kontuan hartuta, metabolomikaren bidezko ikuspegia landare barietateak modu holistiko edo ez-zuzendu batean aztertzeko aukera ezin hobea da. Ikuspegi horri esker, bere metabolismoa osatzen duten eta bertan aktiboki parte hartzen duten konposatu motak identifika daitezke, landarearen barne-funtzionamendua ezagutzeko, hazkuntza-garapen egoki baterako haren egokitzapen-mekanismoak ulertzen laguntzeko, eta ezagutza orokorra zabaltzeko, etorkizunean erabaki informatuagoak hartu ahal izateko.

Horrenbestez, bi ikuspegi analitiko horien bideragarritasuna aztertu da ikerketa-proiektu honetan, kannabis landareen kalitatea ziurtatuz, hau industria farmazeutiko eta medikoan integratu ahal izateko.

5.1. Irudi hiperespektrala

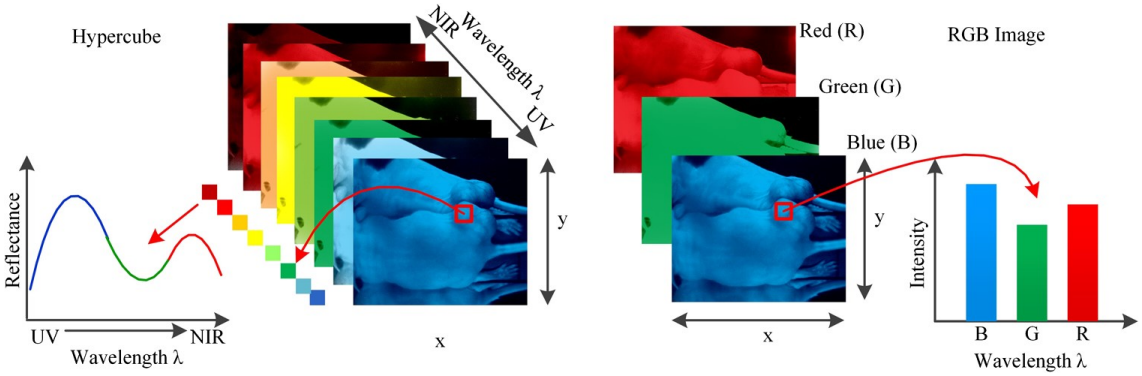
Patroi espazialak eta espektralak aldi berean aztertzearen garrantzia nabarmena da gaur egun hatz-marka kimikoak aztertzeko [79]; horregatik, irudi hiperespektralak (HSI), irudi-espektroskopia ere esaten zaiona, abangoardiako plataforma teknologiko bat da, irudi konbentzionala espektroskopiarekin integratzen baitu objektu baten informazio espazial eta espektral konbinatuak lortzeko. Hasiera batean, HSI teledetekziorako garatu zen, baina azken hamarkadan nekazaritzako elikagaien industrian analisi ez-suntsitzaileak egiteko tresna analitiko boteretsu gisa erabiltzen da [80].

HSI espektroskopia infragorriaren (IRS) aldaera bat dela argudia daiteke, HSI espektroen tartea, oro har, eremu elektromagnetiko horretan baitago. IRS-aren funtzionamendu-oinarriak erradiazio elektromagnetiko infragorriko konposatu baten loturek erakusten duten xurgapenaren detekzioan oinarritzen dira. Absortzio hori gertatzeko, molekula batek lotura kimiko bat izan behar du momentu dipolar handi batekin, hau da, lotura kimiko bat izan behar du elektronegatibitate ezberdineko bi elementuren artean. Espektroskopia infragorriak bibrazio molekularrei buruzko informazioa ematen du. Zehatzago esanda, energia bibrazionalen eta errotazionalen energia mailen arteko trantsizioei buruzko informazioa ematen du [81], [82].

Espektroskopia infragorriaren barruan, gehien erabiltzen den eremua infragorri hurbila da (NIRS). NIRS-ak 750 nm eta 2500 nm arteko eremu elektromagnetikoa estaltzen du, espektroskopiarako eremu erabiliena delarik, C-H, N-H eta O-H loturak dituzten konposatu organiko ugari detekta baititzake, horrela, talde funtzional kimiko desberdinak aztertuz [83], [84]. Faktore horri esker, NIRS oso tresna praktikoa da nekazaritzan errutina analisiak egiteko, oso teknika erraza baita maneiatzeko. Kannabisari dagokionez, NIRS kannabinoideen analisirako ohikoa bihurtu da, metodo azkarra eta merkea baita, zehaztasuna arriskuan jarri gabe [85], [86]. Hori kontuan hartuta, NIRS-aren eta irudien analisiaren arteko konbinazioak NIR-irudi hiperespektralen (NIR-HSI) teknika oso tresna praktikoa bihurtu du nekazaritza-industrian, analisi ez-suntsitzaileen bidezko azterketa kimikoaren balio gehigarria ekarpen garrantzitsua baita. Gainera, NIR-HSIk abantaila handia du NIRS-en ohiko analisiaren aldean. NIRS konbentzionala oso sentikorra da ur-molekulekiko bere O-H loturengatik, eta horrek interferentziak eragin ditzake analisi mota batzuetan. Hori dela eta, Europar Batasunak (EB) komisio erregulazio bat plazaratu zuen (Zb. 2017/1155), eta bertan kannabinoideen analisirako kannabis-laginen aurretratamendu gisa lehortzea eskatzen zuen prozedura esperimental bat deskribatzen da [87]. Hau da normalean jarraitu ohi den prozedura [85], [88], [89], [90], [91], [92]. Horrek esan nahi du NIRS bidezko analisiak laginak aldez aurretik lehortzea eskatzen duela, loreetako hezetasuna ezabatzeko, eta, horrela, kannabis landarean prozedura analitiko suntsitzailea eskatzen du. Bestalde, NIR-HSIk NIRS-en analisian interferentziak eragiten dituzten hezetasun-bandekin lotutako aldagai selektiboak trata ditzake, landare batean in vivo analisi ez-suntsitzailea ahalbidetuz. Hori NIR-HSIren osagai anitzeko nahasketarekiko mendekotasunaren ondorio da; izan ere, lotura espezifikoekin lotutako aldagai espektral selektiboak oso gutxitan izaten ditu, eta horrek, aldi berean, aldagai anitzeko datu tratamenduak eskatzen ditu [93].

Era ulergarriagoan azalduz, irudi hiperespektral bat hiru dimentsioko datu-matrize bat da, non bi dimentsio (X eta Y) espazialak diren, pixeletan banatuta, eta hirugarren dimentsioa (Z) dimentsio espektralari dagokion, non pixel bakoitzari dagozkion NIR espektroak kokatuko liratekeen (*9. irudia*). Horrek esan nahi du irudi hiperespektral baten espektroak ez direla elkarrengandik independenteak; beraz, irudi hiperespektral batek osagai anitzeko nahasketarekiko menpekotasuna erakusten du. Hori abantaila gisa erabil daiteke, kannabis

landareen analisi ez-suntsitzailea ahalbidetuz, eta, horrela, NIR-HSI teknika kannabis medizinalaren kalitate-kontrolerako oso tresna egokia bihurtuz [94].



<u>9. Irudia:</u> Irudi hiperespektral baten eta RGB irudi konbentzional baten arteko alderaketa. Irudiaren iturria: G. Lu and B. Fei "Medical hyperspectral imaging: A review", Journal of Biomedical Optics, vol. 19, Issue 1, 010901, Jan. 2014. https://doi.org/10.1117/1.JBO.19.1.010901 [95]

5.2. Landare metabolomika

Villate A., San Nicolas M.*, Aulas P.A., Gallastegi M., Olivares M., Usobiaga A., Etxebarria N., Aizpurua-Olaizola O., 'Review: Metabolomics as a prediction tool for plants performance under abiotic stress', Plant Science, vol. 303, no. 110789, Feb. 2021.*

DOI: 10.1016/j.plantsci.2020.110789

** A. Villate eta M. San Nicolas autoreek maila berean parte hartu dute lanean*

5.2.1. Sarrera

Klasean, irakasle batek ikasleetako bati galdetu zion: "Ba al dago etorkizuna kartetan aurreikusterik?" Ikasleak erantzun zion: "Nire arrebak posible du" Irakasleak bueltan "Benetan?" Ikasleak berehala erantzun zion "Bai, nire nota-txostena begiratu eta gurasoak etxeratzen direnean zer gertatuko den esaten dit." Txantxa izanik ere, iragarpen eredu arrakastatsu baten adibide garbia da, non notak, iragarpen markatzaileak izateaz gain, puntuagarriak diren.

Metabolomikak, gero eta interes handiagoa sortarazten du landareen hazkuntza-errendimendua diagnostikatzeko eta hazkuntza hobetzeko tresna gisa. Izan ere, markatzaile metabolikoak beste markatzaile molekular tradizional batzuk osatzen eta ordezkatzen hasi dira [96]. Landa-baldintzetan, laboreak estres abiotikoko hainbat faktoreren eraginpean egoten dira aldi berean, eta faktore horiek zuzenean eragiten diote landarearen errendimenduari (biomasaren eta konposatu espezifikoen errendimenduak, landareen morfologia, eta abar). Oraintsu frogatu da landareek estres abiotiko desberdinen aldibereko eraginari ematen dioten erantzuna bakarra dela eta ezin daitekeela besterik gabe estres ezberdin bakoitza bere aldetik aztertze soiletara estrapolatu [97]. Klima-aldaketa globalak (urteko prezipitazioen aldaketak, batez besteko tenperatura, bero-boladak eta abar) modu negatiboan eragiten du lurraren eta nekazaritzaren gainean, eta ondorioz, beharrezkoa da hazkuntza modernoko proiektuak eta laboreen egokitzapenean zentratutako planteamendu bioteknologikoak garatzea [98]. Alde horretatik, metabolomika, behar bezala diseinatutako hazkunde-agertokiekin, laginketa-estrategiekin eta datuen tratamendurako planteamenduekin, ezinbesteko tresna bihur daiteke hazlearen tresna-kutxarako [99].

Hala ere, ikuspegi metabolomikoa oso zorrotza eta konplexua da, baina bariazio genomikoaren, transkriptomikoaren eta proteomikoaren irudikapen holistikoa eskaintzen du. Metaboloma, genomaren azken produktua ezezik, barneko eta kanpoko estresatzaile askoren aztarna ere izan ahal da, eta, beraz, beste "omic" batzuek baino hobeto azal dezake sistema biologikoen eta beren ingurunearen arteko elkarrekintzak eta harekin lotutako perturbazio guztiak [100]. Ezaugarri hau bereziki interesgarria da estres abiotikoko hainbat faktore elkarrekin gertatzen ari diren agertokietan. Testuinguru horretan, metabolomikak hazkuntza-material hobetuen hautaketa erraztuko duten markatzaile espezifikoak aurkitzeko ahalmen o paroa du [101].

Interesgarria da markatzaile metaboliko "eratzaileak" eta "induzigarriak" bereiztea. Lehenengoak baldintza estandarretan neur daitezke, eta bigarrenak, berriz, estres-baldintzak ezartzen direnean bakarrik. Denbora eta baliabideak aurrezteko, aurreikuspen-eredu bat hasierako hazkuntza-fasean neurtutako markagailu metaboliko eratzaileei mugatutako ereduaren bidez eraiki beharko litzateke. Helburu hori lortzeko, beharrezkoa izango da

informazio ezkutua identifikatzeko tamaina handiko dataset horietan bereizgarriak diren metabolitoak hautatzeko metodo matematiko egokiak aplikatzea [102].

Nolanahi ere, iragarpen ereduak eraikitzeko eta baliozkotzeko, estres eszenatokiak aplikatu beharko dira aldibereko kontrol eszenatoki estandarrekin. Ondoren, eta emaitzetan oinarrituta, aurreikuspenaren eraginkortasunaren eta kostuen murrizketaren arteko oreka egokia gakoa izango da. Aldi berean, eredu prediktiboan erabilitako metabolitoen kopurua optimizatu behar da. Markatzaile kopuru handiagoak ez du zertan hertsiki lotu behar korrelazio prediktibo handiagoarekin [103]. Beraz, eredua sortzeko erabili behar den metodo ez-bideratua erabili ordez, puntuatzeko, jarraitzeko, erraza den metodo bideratua erabiltzea da helburua.

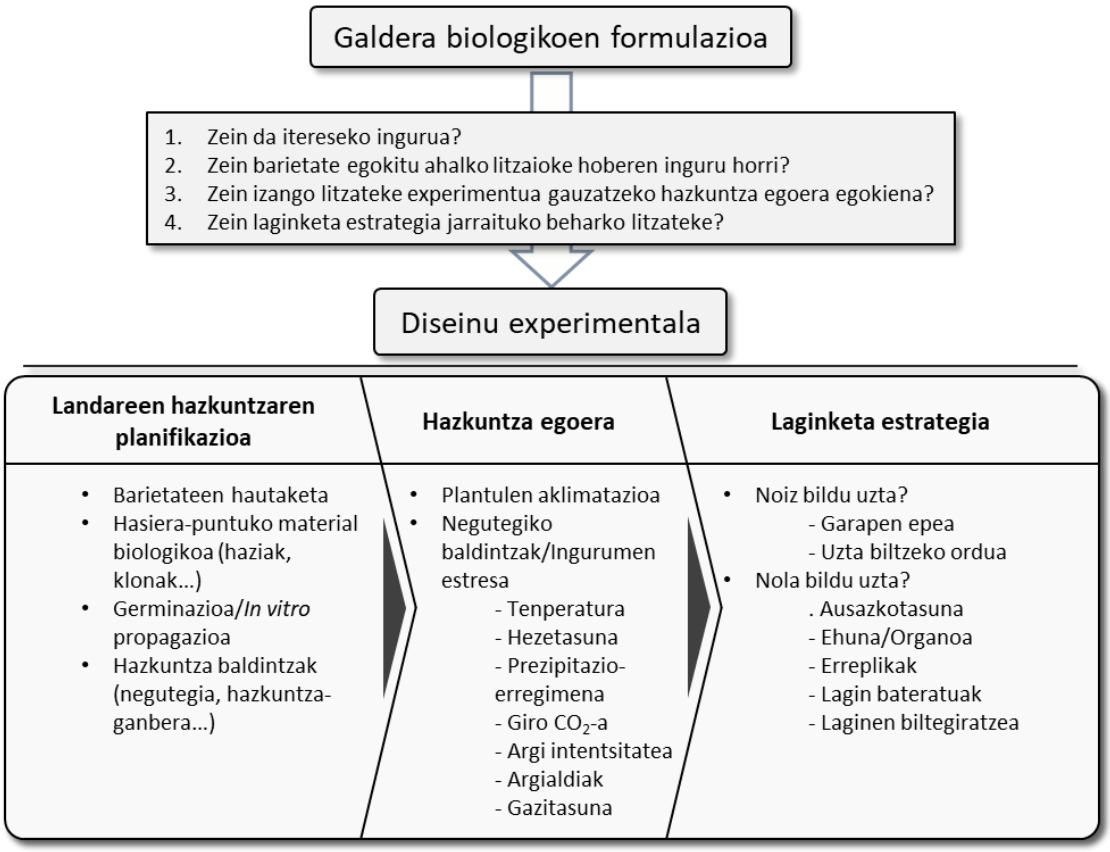
Metabolomika, iragarpen-tresna gisa, nahiko berria da landare-zientzien arloan. Hori dela eta, zenbait ikerketa egin dira ingurumen-estresaren pean egondako laboreen errendimendua hobetzeko [96], [100], [104]. Gaur egun ez dago metodo unibertsalik landareen errendimenduaren markatzaile metaboliko prediktiboak ingurumen-estresaren pean identifikatzeko. Ondorioz, ikerketa-helburuek metodo onena edo metodo egokien konbinazioa orekatu behar dute. Horretarako, hazkunde-agertokiei eta laginketa-estrategiei buruz hitz egingo dugu hurrengo azpiataletan, eta ikuspegi analitiko handiak eta ikuspegi bakoitzak eskatzen dituen tresna kemometrikoak aurkeztuko ditugu.

5.2.2. Hazkunde-egoerak eta laginketa-estrategiak

Estresaren tolerantzia ardatz duten landare-hazkuntzako proiektuetan, estres abiotikoko faktoreak aplikatuz ingurumen-baldintzak simulatzen dira. Horien artean, argi-intentsitatearekin eta fotoperiodoarekin batera (adibidez, eguneko eta gaueko argiztapen-aldien luzera), bereizgarrienak tenperatura, hezetasuna eta prezipitazio-erregimenak dira. Hala ere, beste aldagai batzuek ere, hala nola gazitasunak eta ingurumeneko CO_2 kontzentrazioak, aldaketak eragiten dituzte landare-metaboloman, eta, beraz, horien eragina ere aztertu behar da [105], [106]. Funtsezkoa da estres-faktoreek, konbinatuta edo modu independentean aplikatzen direnean, ez dietela berdin eragiten landareen estres-tolerantzia mekanismoak aktibatzen dituzten geneei kontuan hartzea. Landarearen estres-tolerantzia bere material genetikoaren arabera arautzen denez, aplikatutako estimuluen eragin zehatza ebaluatzeko, aldagarritasun biologikoa minimizatu behar da biomarkatzaileen eta ingurumen-estresaren arteko korrelazioa zehaztu ahal izateko [107]. Horrekin bat etorriz, landareen errendimendu prediktiboko eredu sendo bat eraikitzeko diseinu esperimental zorrotza behar da. *10. Irudian* laburbiltzen denez, diseinu esperimentalak ikerketaren alderdi guztiak definitu behar ditu, eta, estres abiotikoko faktoreak alde batera utzita, erantzun beharreko galdera biologikoaren arabera diseinatu behar da [108]. Planifikazio optimizatu batek landare-barietatea edo -laborea, ernetzea, haztea, biltzea eta lan-fluxua hautatzea dakarten urrats guztiak azaldu behar ditu, akats esperimentala minimizatuz eta proiektuaren bideragarritasun tenporala eta ekonomikoa ahaztu gabe [109].

Estres abiotikoen simulazio egokia funtsezkoa da markatzaile fidagarriak lortzeko. Jarduteko modu egokia aztertutako faktoreak baldintza oso kontrolatuetan mantentzea da gainerako parametroak maila optimoan mantenduz, landareen erantzunetan beraien eragina nabarmena ez izateko asmoarekin [110]. Hala ere, ingurune jakin bateko egoera lortu nahi denean, ingurune horren arabera ezarri behar dira estres abiotikoko faktoreak (arestian zerrendatutako horiek). Gainera, estresa sortzeko garapen-momentu egokia aukeratzea ere funtsezkoa da, landareen garapenean metaboloma aldatzen delako. Landareek aklimatazio aldi bat jasan beharko dute tratatu aurretik, espeziearen eta tratamenduaren beraren arabera alda daitekeena. Horrela,

landareen errendimendu-eredu prediktiboa garatzeko nahi diren biomarkatzaileak aurkitzeko eta denboran zeharreko bi baldintzetan hazitako landareen metabolito-profila alderatzeko helburuekin, landareak bai ingurumen-estresaren baldintzetan eta bai baldintza estandarretan hazten dira.



10. Irudia: Estres abiotikopeko ikerketa metabolomiko baterako diseinu esperimentalaren eskema

Diseinuaren funtsezko alderdi bat hautetsitako genotipoa da, landare espezie beraren barruan metaboloma oso ezberdinak ager baitaitezke. Esparru horretan, erabaki bat hartu behar da aztertu beharreko laboreari buruz, egindako galderaren arabera: "Zein da aztertuko den ingurunea? Ba al dago ingurune horretara egokitzeko joera duen genotiporik?" Gaur egun, landare-espezie asko landatzen dira mundu osoan, adibidez, patata edo tomatea; hala ere, landare-espezie beraren berariazko zenbait aldaketek portaera hobeak erakusten dituzte ingurune espezifikoetan, tokian toki ingurumen-faktoreak asko aldatzen baitira. Beraz, laboreen hautaketa, bistan denez, partziala da, ikertzaileek beren aurretiko ezagutzetan oinarritzen baitira zein barietate aztertu eta. Denbora eta baldintza ekonomikoak kontuan hartuta, normalean bi lanketa aztertzen dira horrelako proiektuetan. Material abiarazlearen izaera ere funtsezkoa da. Hazietatik abiatuz gero, espezie bakoitzerako ernetzea baldintza egokietan egin behar da, baina antzeko banakoak lortzeko probabilitateak murriztu egiten dira. Aldizka, aldaxken bidez edo hedapen nodalaren bidez lortutako klonek aukera ematen dute banakoen arteko aldagarritasuna minimizatzeko profil genetiko berdinak dituzten heinean.

Bilketa faseko oinarriak ezartzea kritikoa da, kontrolatu gabeko lan-fluxu batek nahi ez den estres gehigarria eragin baitezake landareetan, eta estres gehigarri honek aldagarritasuna eragin baitezake metabolitoaren profilean. Fase honetan, kontuan hartu beharreko faktoreak hauek dira: errepliken kopurua, hartutako laginen ausazkotasuna, eguneko bilketa, bildu beharreko landareen atalen hautaketa, biltzeko eta laginak biltegiratzeko metodoa garatzeko faseak [111]. Landareen metabolomikan, erreplika biologikoen erabilera iradokitzen da, erreplika teknikoen baino hobeto azaldu dezaketelako esperimentu osoaren aldakortasuna. Emaitzen analisi estatistiko egokia egiteko, erreplikazio-kopuru onargarria sei izan beharko litzateke [112], nahiz eta adierazpen estatistikoa areagotu daitekeen erreplika-kopurua handituz. Sarbide irekiko software-tresna batzuek, adibidez, R [113] edo G* Power [114], potentzia-analisia egiteko paketeak eskaintzen dituzte, efektuaren tamaina zehazteko eta proposatutako esperimenturako lagin-zenbaki optimoa kalkulatzeko, nahiz eta gerta litekeen esperimentuaren denbora- eta ekonomia-eskuragarritasunarekin lagin-tamaina optimoak ezin izatea bat etorri. Landare bakoitzak erreplika gisa balio badu, laginak bateratzeak aldagarritasun biologikoa murrizten ere lagun dezake, baita kostu ekonomikoan ere [109]. Izan ere, helburu nagusia ingurumen-estresaren pean dagoen landare-errendimenduan oinarritzen denez, laginen bateratzea biziki iradokitzen da, eta beraz, aurreko analisiak landare-banako osoen adierazgarri bilakatzen dira. Gainera, laginak landare aleatorioetatik hartu behar dira, eta haien kokapena aldizka aldatu behar da, itzal efektuak saihesteko [112].

Uzta biltzeko egunaren momentua berdina izan behar du definitutako garapen etapetan, landareen aktibitate metabolikoa egunaren momentuaren arabera aldatzen delako. Landareen metabolismoa erloju zirkadianoak erregulatzen du eta, erregulazio horrek, egunean zehar karbohidratoak eta hormona erregulatzaileak metatzea dakar [113], [115]. Horregatik, laginketa-prozesua oraindik jarduera metabolikoa hasi ez denean gauzatzea iradokitzen da, goizean goiz, laginak egoera egonkorrean eta konparagarrian biltzeko.

Bildu beharreko landarearen organoei dagokienez, ez dago protokolo finkaturik bildu beharreko ehunei buruz, baina laginek datu erreproduktiboak eman behar dituzte ezaugarri desiragarrirako [116]. Horrela, landare osoaren aldakortasuna aztertzeko, modu independentean atal bat baino gehiago biltzea iradokitzen da. Gainera, ikertzen ari diren landare espezieei buruzko aldez aurreko daturik ez badago, hiru ehun mota ezberdin lehenesten dira, hala nola hostoa, lorea eta zurtoina. Gauza bera gertatzen da aztertutako garapen-etapekin: egoki definitu behar dira landareen hazkuntza-aldiaren datu adierazgarriak eskuratzeko [111]. Azpimarratu behar da bi faktoreetako edozein gehituz gero analisi estatistiko osoa areagotuko litzatekeela, eta horrek espeziearen aldakortasun biologikoaren definizio hobetua ekarriko lukeela.

Laginen biltegiratzeak ez du metaboloma aldatu behar bilketa unean, beraz, laginak berehala izoztu behar dira. Alde horretatik, lagina nitrogeno likidoan kontserbatzea da aukerarik onena, azkarra delako eta jarduera metabolikoa gelditzen duelako. Gero, laginak -80 °C-tan gordetzen dira, eta bertan mantendu analisirako erabili arte inolako degradazio metabolitikorik jasan gabe.

Landare bakoitzaren azterketa metabolomikoa nahiko berezia denez, ez dago diseinu esperimental osoa oinarrituko den protokolo estandarizaturik landareak nola hazi behar diren azaltzen duenik. Hala eta guztiz ere, literaturaren aurkitu daitezkeen aurreko ikerketak oso lagungarriak izan daitezke landare errendimendu berriko ikerketa metabolomikoetan ikertzaileak bideratzeko. Adibidez, *Astrid Junker et al.*, 2015 [117], egileek informazio osotua eskaintzen dute uraren erabileraren eraginkortasunari buruz Arabidopsis eta artoaren

hazkuntzan. Artikulu honetan, egileek adierazten dute 1 cm inguruko lodiera duen harri-hartxintxarrezko geruza batek eta geruza urdineko hiru geruza zulatuk lurzoruaren gaineko lurrunketa nabarmen murrizten dutela. Gainera, esaten da banakako ontziek plastikozko poltsak eduki beharko lituzketela barruan, eta, horrela, ez direla drainatzen eltzeen hondotik. Gainera, artikulu horretan bertan [117], landareak hazteko abiapuntua haziak ziren, bai hazkuntza-baldintza estandarretan, bai baldintza optimizatuetan, eta ilunpetan eta 5 °C-tan ernatzen ziren, paper iragazle hezez estalietan 48 orduz mantenduz loreontzi txikietan (9 cm) lurrera eraman aurretik. Lurrean landatuta, plastikozko tapakiz estali zituzten hezetasuna gordetzeko, eta ernamuindu ondoren kendu egin zituzten tapakiak. Horretaz gain, argialdiak 16 ordura ezarri ziren bi hazkuntza-baldintzetan, eta argiaren intentsitate artifiziala 205-245 $\mu mol \cdot m^{-2} \cdot s^{-1}$ irismenean ezarri zen; beraz, argiaren guztizko intentsitateak (eguzki-argi naturala + argiztapen osagarria) nekez gainditzen zuen 380 $\mu mol \cdot m^{-2} \cdot s^{-1}$ balioa. Azkenik, tenperatura-erregimena intereseko eskualdeko udaberriko tenperaturaren monitorizazioaren arabera egokitu zen. Hau bezalako artikuluek [117] ekarpen handia egin dezakete landare berrien errendimenduaren ikerketetarako, kontuan hartu beharreko aldagaia enpirikoki optimizatzen baita landareen hazkunde-aldian. Hala ere, prozedura esperimentalen optimizazioari buruzko ikerketa gehiago egin behar da, etorkizunean landareen hazkuntzako lan-fluxuen protokolo estandarizatuak gara daitezke.

5.2.3. Landare-metabolomikarako teknika analitikoen bibliografiaren egoera

Lagin-tamaina bat gainontzeko denekin bat ez etortzeak prozedura analitikoen edo lan-fluxuen funtzioaren garrantzia islatzen du metabolomikan. Laginen aurre-tratamendua urrats erabakigarria da landareen metabolomikan, analisi eta behaketa orokorren kalitatea arriskuan jar baitezake [112], [118], [119], [120]. Izan ere, lan-fluxu analitikoaren edozein urratsek eragin partziala du metabolito neurgarrien mailan, eta, beraz, urrats horien berri izan beharko genuke azterketa metabolomikoaren irismena ulertu ahal izateko. Adibidez, erabilitako erauzketa-metodoa kontuan hartu gabe, eta -80 °C-tan berehalako izozketa egiteaz gain, herdoilaren aurkako erreaktiboak erabiltzea iradokitzen da, hala nola butilatutako hidroxitoluenoa, oxidazioarekiko sentikorrak diren metabolitoak kontserbatzeko [120]. Prozedura osoan hotz-katea ziurtatzeko, laginak tenperatura baxuetan homogeneizatu ohi dira (adibidez, nitrogeno likidoa edo ganbera hoztuak erabiliz) aurrez hoztutako mortero eta pestle bat erabiliz [112], [118], [119], [121]. Hala ere, homogeneizazio-teknika berriak, hala nola errotagailu elektrikoa, ehunen lyserra eta osziladore ultrasonikoa, arestian aipatutako metodo horiek ordezten hasi dira azken urteetan [98]. Horrez gain, laginak lehortzea gomendatzen da, entzimek eragindako eta metabolitoen deskonposizioa eragiten duten erreakzioak saihesteko. Izoztea da hozte-kateari eustea ahalbidetzen duen lehortze-metodo gogokoena, ura lurruntzeko beroa erabiltzen den beste lehortze-metodo batzuen aldean [111].

Mikrouhinen bidezko erauzketa eta likido superkritikoen erauzketa metabolitoa isolatzeko erabil daitezke [122] baina, jakina, erauzketa energetiko hori aplikatzeak metabolitoen egonkortasuna arriskuan jarriko luke, prozedurarik ezagunena disolbatzaileetan oinarritutako erauzketa delarik. Erauzketaren eraginkortasuna hobetu daiteke ekografien energia erabiliz [119], [120] edo tenperatura baxuetan erauziz [123], hotz-katea ziurtatuz eta metabolitoen eraldaketa edota degradazioa saihestuz. Kloroformoa ($CHCl_3$) edo metil teril butilozko eterra (MTBE) [118], [120] eta metanolaren eta uraren nahasketak (adibidez, maiz erabiltzen diren MeOH: H_2O (1:2 v/v) [120] edo MeOH: H_2O: azido formikoa (70:30: 0,1 v/v/v) erabiltzen dira, hurrenez hurren, polar eta polar ez diren metabolitoak isolatzeko). Hala ere, disolbatzaile

ternariotan, MeOH: $CHCl_3$: H_2On, oinarritutako nahasketak polaritate sorta zabaleko metabolitoak urrats bakar batean erauzteko erabiltzen dira [112], [118], [119]. Nahasketako disolbatzaileen proportzioak metabolito guztiak dituen fase bakar batera garamatza [120], edo ongi ezarritako $CHCl_3$: MeOH: H_2O (2:1:1 v/v) nahasketa erabiliz lor daitezkeen eta metabolito polarrak eta ez-polarrak dituzten bi fasetan ere lan egin daiteke [112], [118]. Azken honek estraktu garbiagoak lor badaitezke ere, batzuetan, zenbait metabolito galtzearen kontura egiten diren beste garbiketa urrats batzuk aplikatzea beharrezkoa da [121]. Disolbatzaileetan oinarritutako erauzketaz gain, absortzioan oinarritutako erauzketa-teknikak, hala nola fase solidoko mikroerauzketa (SPME) eta stir-barrako sortzio-erauzketa (SBSE), teknika osagarri interesgarriak izan daitezke profil metaboliko zuzenduetarako zein ez-zuzenduetarako, batez ere metabolito ez-polar lurrunkorren eta erdilurrunkorren detekziorako [124].

Nahiz eta instrumentazio analitikoaren etengabeko aurrerapenek egungo paradigmak neurtzen ari dena behar bezala ulertzeko eguneratzera eta tresna osagarriak garatzera bultzatu gaituzten, landare-metabolito ezagunak eta ezezagunak identifikatzea erronka bat da oraindik. Teknika analitikoen konbinazioa beharrezkoa da metabolomaren analisi integrala lortzeko [107]. Landareen metaboloman espero diren susmagarrien propietate fisiko-kimikoen ondorioz (adibidez, fitohormonak, aminoazidoak, oligosakaridoak, etab.), kromatografia likidoa - masa espektrometria (LC-MS) da, alde handiz, landareen metabolomikan gehien erabiltzen den teknika [107], [112], [118], [125]. Merkatuan dauden bereizte-zutabeen barietate handiak konposatuak bereizteko aukera ematen du. Alderantzizko faseko zutabeak dira gehien erabiltzen diren zutabeak landareen azterketa metabolomikoetan [118], [123], [125], baina baita fitohormonak bezalako metabolito garrantzitsuen analisi zuzenduan ere [112], [121]. Emaitza horiek elkarrekintza hidrofilikoko kromatografia-zutabeak (HILIC) erabiliz lortutakoekin osatzen dira. Azken zutabe horiek, metabolito polar gehiago bereizteko aukera ematen dute, hala nola aminoazidoak, azukreak, oligosakarido raffinosa eta metabolitoak dituzten sufreak; eta grafitozko karbono fase porotsuetan (PGC) oinarritutako zutabeak, nukleotidoak eta horien deribatu azukretsuak eta fosfatoak bereizteko aukera ematen duten [118], [125]. Masa-espektrometriari lotutako elektroforesi kapilarraren (CE-MS-aren) erabilera landareen pisu polar handiko eta karga molekular txikiko metabolitoak bereizteko alternatiba bat da, hala nola koentzimak, aminak, minda-baseak, azido organiko txikiak eta nukleotidoak [112], [118], baina CE-MS ez da ia erabiltzen landareen metabolomikan. Neurri txikiagoan bada ere, masa-espektrometriari (GC-MS) lotutako gas-kromatografia teknika analitiko indartsua da landareen metabolomikan. LC-MS [107], [112] baino erreproduzigarritasun hobea eta informazio estrukturalagoa ematen du, baina deribazio-urrats bat behar du metabolito ez-lurrunkorren analisiaren aurretik [118]. Horregatik, GC-MS erabiltzen da fitohormona erdilurrunkorrak edo deribazio ondorengo metabolito primario ezberdinak (normalean sililazio baten ondorengo metoximazioa) zehazteko, behar direnean [107], [112], [126].

Analisi bideratua da zalantzarik gabeko identifikazio-hurbilketa metabolito ezagunak detektatzeko, baina ez da hain erabilgarria metabolitoen detekzio-anitzeko eta landareen metabolomiako azterketetan, askotan argi gabeko hurbilketak eskatzen baitituzte. Ondorioz, erabilitako bereizte-teknika alde batera utzita, bereizmen handiko masa-espektrometrian (HRMS) oinarritutako teknikak dira landareen metabolito-profilean hobesten direnak [112], [118], [127]. Bereizmen handiko/masa-doitasun handiko, masa zehatzeko eta profil isotopikoko neurketek, HRMS detektagailuek egindakoak (adibidez, hegaldi-denbora (TOF), Fourier Transformazioa - Ion Cyclotron Erresonantzia (FT-ICR) eta Orbitrap masa-espektrometroak), konposatu ezezagunen formula molekularra ondorioztatzea ahalbidetzen dute [126], [127],

[128]. MS datu osoak egiturazko informazioa ematen dituzten masa-espektrometro hibridoetan lortutako MS/MS espektroekin osatzen direnean (hau da, qTOF, qOrbitrap eta antzeko laukoitz bakarrekin elkartuta), metabolito ezezagunetarako behin-behineko hautagaiak proposa daitezke. *4. Taulan* metabolomikarako erabiltzen diren analizatzaile nagusiak laburbiltzen dira, eta horien selektibitatea, sentsibilitatea, neurketa kuantitatiboak egiteko gaitasuna, mugak eta abantailak eztabaidatzen dira.

HRMSen artean, FT-ICR tresnek dute ebazteko ahalmen handiena, baina eskuratze-ahalmen nahiko baxua dutenez [118], landareen metabolomikan dituzten aplikazio nagusiak infusio zuzeneko MS (DIMS) esperimentuetan daude [118], [128]. DIMSen helburu nagusia ez da metabolitoen profilak egitea, eta hatz-markaketa metabolikorako erabiltzen dira. Izan ere, landareek estres biotiko eta abiotiko ezberdinei emandako erantzunak eta landare espezieen karakterizazioa ikertzeko erabili da DIMS. Orbitrap MS detektagailuek, normalean LC edo GC bereizmen tekniketara lotuta datozenak, aplikazio-eremu zabalagoa dutela ikusi da, suntsipen ahalmen handia (batez ere m/z-ko baxuan), doitasun handia eta eskuratze-tasa handiagoa baitute FT-ICR teknikekin alderatuta [112], [118], [126], [128]. LC-qOrbitrap sistemak metabolito sekundario txikiak identifikatzeko erabili dira, hala nola, metabolito fenolikoak eta alkaloideak, baita metabolitoen profilak eta erauzkin natural konplexuetan metaboloma edo dereplikazio ikerketak egiteko ere [126]. Hala ere, HRMSko beste tresna batzuekin alderatuta erresoluzio baxukokoak izan arren m/z baxuetan [130], landareen azterketa metabolomikoetan erabiltzen diren detektagailu ezagunenak TOF analizatzaileak dira, batez ere, kuadrupolo bakarrei eta LC, GC edo Matrizeak lagundutako laser desortzio/ionizazio (MALDI) tresnei lotuta daudenean [126]. LC-qTOFen oinarritutako metodoak oso erabiliak dira nahaste konplexuen azterketa metabolomiko zuzenduetarako eta ez-zuzenduetarako [123], [126]. Aplikazio-eremua, aminoazidoak eta azido organikoak bezalako metabolito primarioak identifikatzetik [123] zenbait fitohormonen [121] eta beste metabolito sekundario askoren identifikaziora hedatzen da [126]. GC-qTOF tresnak poliaminen [118] eta beste metabolito lurrunkorren eta erdilurrunkorren analisian erabili ohi dira [126].

Landareen metabolomikaren etorkizuneko azterketetan erabilgarria den MSren beste aukera bat, masa-espektrometroei akoplatutako ioi-mugikortasuneko espektrometroetan (IM-MSetan) oinarritzen da. Instrumentazio horren bidez, bereizte-maila gehigarria lortzen da (ortogonaletik MSra), eta egitura molekularrari buruzko informazio zehatza ematen du, bereizketa isomerikoa ahalbidetzen du, zarata kimikoa murrizten du, konfiantza handiko metabolitoa identifikatzea ahalbidetzen du eta estaldura metabolomikoa zabaltzen du [131]. Abantaila horiek guztiek IM-MSa tresna bikaina bihurtzen dute metabolomikako hainbat aplikaziotan erabiltzeko, batez ere LCari lotuta dagoenean. Azken lanetan, jariakin biologikoetan eta ehunen metabolomikan IM-MS eta LC-IM-MS erabili dira [107], [125], [131], baina oraindik ez daude oso hedatuta landare-metabolomikan. Izan ere, teknika horiek metabolomikako lan-fluxu guztietan aplikatzeak aurrerapen handiak ekartzen ditu LC-IM-MS datu multidimentsional konplexuak eta datu-tratamenduko lan-fluxuak prozesatzeko gai diren software-tresna altuen garapenean [131].

MALDI-masen irudi espektrometria (MALDI-MSI) ere landareen metabolomikan etorkizun handiko teknika gisa aztertu da, metabolitoen eta ehunen azaleko analisien banaketa espaziala eman baitezake [107], [132]. Pisu molekular handiko metabolitoak (>500 Da), lipidoak eta bigarren mailako metabolito batzuk (hortatinak, flavanolak, etab.) aztertzeko erabiltzen den arren, oraintsuko aurrerapenek aukera ematen dute metabolito txikiak (adibidez, fitohormonak eta metabolito primarioak) aztertzeko [132].

4. Taula: Metabolomikarako erabiltzen diren analisi-metodoen laburpena, bakoitzaren selektibitatea, sentsibilitatea, neurketarako-ahalmena, mugak eta abantailak ere jasotzen dira

Analisi-metodoa [ref]	Selektibitatea/ Metabolitoen identifikazioa	Sentikortasuna	Analisi kuantitatiboa	Mugak	Abantailak eta bestelako informazioa
NMR [127], [129]	Ona	Baxua	Altua	Seinaleen gainezarmena, MS-arekiko konparatuz data baseetan jasotako/identifikatutako metabolito kopuru baxua	Laginen prestaketa sinplea, neurketa azkarra, metabolitoen identifikazioa LC edo 2D/multidimentsionalera modu errazean lotu daiteke
FT-ICR [118], [130]	Altua	Altua	Nahikoa	Neurketa ratio baxua, ekipamenduaren kostu altua (erosketa eta mantenua)	Erresoluzio altuak eta masa ziurgabetasun baxuak formula enpirikoaren determinazioa ahalbideratzen du
TOF [118], [126], [130]	Ona	Ona	Nahikoa	Masa-ardatzak maiz kalibratu behar dira, m/z baxuko potentzia galerak konponduz.	Kuadropolo bakunera lotuz selektibitatea, sentsibilitatea eta kuantitibitatea hobetu daiteke. Analisi azkarrak banaketa tekniketara erraz akoplatzea ahalbideratzen du.
Orbitrap [118], [126], [130]	Altua	Altua	Nahikoa	Hornitzaile bakarraren teknologia, beste MS batzuekin alderatuta m/z altuetan erresoluzio-maila baxuagoa.	Selektiboa, sentikorra eta analisi kuantitatiboa indartu egin daitezke kuadrupolo bakarrari (q-Orbitrap) lotuz gero, epe luzerako masa-ardatzaren egonkortasun bikaina, erabiltzeko erraza.
IM-MS [131]	Altua	Altua	Ona	Sofware espezifikoa behar du datu multidimentsionalen tratamendurako	Beste MS batzuekin alderatuz, banaketa isomerikoa ahalbideratzen du

Azkenik, baina ez gutxi, metodo sinpleek, errepikakortasun handiek eta neurketa azkarrek erresonantzia magnetiko nuklearra (NMR) landareen metabolomikan erabilera eraginkorreko teknikan bihurtzen dute, hala nola, kimioaxonomiarekin lotutako ikerketetan, elikagaien kalitatearen kontroletan eta estres biotiko/abiotiko ezberdinen aurrean landareen erantzunen ikerketetan. NMRaren erabilera hatz-marka metabolikoetara mugatzen da, sentikortasun txikia eta seinalearen koeluzioa direla eta. Bi NMR dimentsiodun eta multidimentsional erabiltzeak [133], erradiofrekuentzia miniaturizatuko bobinak, imanak eta zunda kriogenikoak gaingidatzeaz gain, NMRren muga handiak gainditzea ahalbidetzen dute [134] eta NMR profil metabolikorako tresna egokia eta etorkizuneko azterketa metabolomikoetarako teknika oparoa izatea ahalbidetzen dute.

Berrikusi denez, zenbait metodo eta plataforma analitiko erabil daitezke landareen metabolomikan, baina ez dago protokolo estandarizatu eta bakar bat metaboloma osoaren estaldura ahalbidetzen duenik. Matrizearen konplexutasunak, metabolitoen izaera kimikoak (hau da, konposatu kopuru handiak, eta, propietate eta kontzentrazio kimiko ugariak) eta egungo teknologia analitikoen mugek zaildu egiten dute helburu hori lortzea, teknika desberdinak konbinatuta ere. Instrumentazio analitikoaren azken aurrerapenek, hala nola HR-MS, IM-MS eta NMRren azken garapenak, metabolomaren profilaren azterketa hobetzen dutela dirudien bitartean, oraindik lan handia egin behar da laginen bilketan, laginen tratamenduan/erauzketan eta analisien optimizazioan.

5.2.4. Landare-metabolomikako datuen tratamendua

Landare zientzia eta bioteknologiako datu metabolomikoen analisia oztopo ugariz betetako bidea da. Esperimentuen diseinuaren konplexutasuna alde batera uzten badugu, landare-metabolomikan bereziki garrantzitsua dena [108], oztopo horietako asko LC-HRMS datuen tratamenduan aurkituko ditugu, batez ere ezaugarri molekularrak ateratzeko lan-fluxu multzo egoki baten diseinuan eta aplikazioan, konposatuen detekzioan eta identifikazioan, eta, konposatu horiek bide biokimikoetan integratzean [135]. Aplikazio-eremua alde batera utzita, gaur egungo arazo eta erronka asko sakon deskribatzen dira azken berrikuspenetan [136], [137], [138], [139]. *5. Taulan*, kode irekiko tresna erabilgarri batzuk nabarmendu ditugu, bai detekzioan eta identifikazioan zentratuak, bai sare molekularrean zentratutakoak ere.

HRMSren pakete saltzaileen aldean, tratamendu-katearen ia urrats operatibo guztiak barne hartzen dituztenak, kode irekiko aplikazio gutxi daude, hala nola Open MS (www.openms.de) [140], envibee (www.envibee.ch), MZmine2 (mzmine.github.io), patroon (rickhelmus.githubio/patRoon /) edo MS-DIAL (prime.psc.riken.jp). Datuen tratamenduan zehar, lehen oztopoa ezaugarrien finkatzea da, formula molekularraren eta egituraren identifikazioa barne [141]. Hurbilketa gehienek, hala nola Sirius eta CSI:FingerID arteko konbinazioak, edo, XCMS familiak edo MetFragek, analisi-prozedura automatikoak ahalbidetzen dituzte.

Sare-tresna molekularrak ikerketa metabolomikoetan interesa pizten ari dira, ezkutuko metabolito askoren detekzio-tasak handitzeko. Printzipio gidaria da familia bioaktibo askoren arteko harreman estruktural estuek zatiketa-ereduak partekatzen dituzten molekula multzoak sortzen dituztela, eta ideia hau errazago detektatu daitezen eta, azkenean, identifika daitezen erabili daiteke [142], [143]. Zentzu honetan, Global Natural Product Social Molecular Networking-a (GNPS, gnps.ucsd.edu) web plataforma bat da, Network Annotation Propagation (NAP) honi software mota askorekin hurbiltzeko aukera ematen diona, detekzio-eta bistaratze-tresnak barne, eta, azkenik, Cytoscape [144] bezalako kanpoko softwarean lortutako sareen

esplorazioa ahalbidetzen duena. *5. Taulako* Sare Molekularren barruan dauden lau aplikazioetatik, GNPSren kanalizaziotik kanpo MetGenek bakarrik funtzionatzen du. Beste hirurak sare molekularren lan-fluxuan erabili ohi dira. Dereplicator MS detekzio-tresna da, bai aitzindaria, bai zatikatze-informazioa erabiltzen duena, bai liburutegi estandarretan bilaketak egiten dituena (MassBank [massbank.eu], NIST [chemdata.nist.gov], METLIN [metlin.scripps.edu] edo HMDB [hmdb.ca/hml]), edo MSn espektro liburutegi bat duena, *In-silico* zatiketa-tresnetan, hala nola Sirius+CSI: FINGER, oinarritutako hautagai potentzial multzo batetik lortutakoa. Horrez gain, MS2LDAk aukera ematen du zatien eta galera neutroen gainbegiratu gabeko analisiek egiturazko eredu komunak aurkitzeko. ClassyFirek konposatu kimikoak egitura-ezaugarrien arabera sailkatzeko aukera ematen du, aldez aurretik esleitutako taxonomia baten arabera, konposatu organikoen eta inorganikoen 4800 konposatu kimiko baino gehiago barneratzen dituena. Azkenik, aipatutako gehienak barne hartzen dituen (Dereplicator, MS2LDA eta sare molekularrak) datu metabolomikoen ikuspegi sakonagoa emateko, MolNetEnhancer sare-tresna gune bat da.

Lehen aipatutako erreminta guztiek, metabolito ezezagunen identifikazioa hobetzeko eta ezaugarri molekularra erabat detektatzeko eta identifikatzeko, tresna osagarriak behar dituzte. *5. Taulako* datu-baseetatik, Produktu Natural Unibertsalen Datu-basea [158] egokiena zirudiena zen, baina ez dago aktibo. Landare metabolomaren datu-basean 1000 metabolito inguru daude, baina egiturak eta konposatuen bilatzaile batzuk bakarrik daude aktibo. KNApSAcK deskarga daitekeen Java aplikazio bat da, espezieak eta metabolitoak aurkitzea ahalbidetzen duen web aplikazio bat. Azkenik, ReSpect for Phytochemicals MS espektro datu-basea da, 9000 espektro baino gehiago eta prozedura bilatzaile asko dituen erreminta da.

Oztoporik korapilatsuenetako bat metabolitoak dagozkien bide metabolikoetan lotzea da, hau, atzean dagoen transkripzioa eta epigenomika ulertzeko funtsezko urratsa delarik [159]. Maila honetan, aipatzekoa da MATLABen oinarritutako MetaboTools delakoan COBRA tresnaren aplikazioa [160]. Horrez gain, PlantMAT Excel makro bat da, metabolito ezezagunetan egiturazko osagaien presentzia aurreikusteko, egiturei buruzko informazio fitokimikoa eta bilaketa konbinatorioko estrategia konbinatzen dituena. Gainera, Metacrop 2.0-k informazio osagarria eskaintzen du laborantzako landare nagusien metabolismo primarioari buruz. Azkenik, landareen sare metabolomikoa (PMN) datu-base bat da, landare mota askotako geneak, entzimak eta metabolitoak integratzen dituena.

Tresna eta planteamendu metodologikoez gain, *6. Taulan* bereziki garrantzitsuak diren aplikazioak sartu ditugu, landareen hazkundeari buruzko azterlanekin zerikusia dutenak. Hautatutako azterlanek ingurumena azpimarratzen duten faktore orokorrak dituzte ardatz, hala nola, ura [161], gatza [162] eta lehorteak eragindako tolerantzia [163], [164], eta klima-aldaketara [165] edo xenobiotikoekiko esposiziora [136] egokitzeko gaitasuna ere badute.

Azpimarratzekoa da lan hauen azpian dagoen ikuspegi omikoa, batez ere ezaugarri kuantitatiboen bidez loci (mQTLs), perturbazioa eta azpiko erregulazio genetikoa azpimarratzen dituen ingurumena aztertzeko eta metabolito-ugaritasuna entzima-kodifikazioko geneekin lotzeko, metabolomika ez-bideratuen erabileran oinarritutakoa, [105], [166]. *6. Taulan* adierazten den bezala, arroz [167], [168], tomate [166], [169] edo belarretan [170] aztertu izan da.

5. Taula: Bai detekziora eta identifikaziora eta bai sare-molekularrera bideratutako erabilera askeko erraminten laburpena, eta, landare-metabolomikan garrantzitsuak diren datu-baseak

Lan-fluxua	Programa / eremua	Web-a / Ikerketa	Bibliografia
Formula molekularra eta egituraren identifikazioa	Sirius + CSI:FingerID	bio.informatik.uni-jena.de/software/sirius/	[145], [146]
	XCMS family	www.bioconductor.org/packages	[147]
	MetFrag	ipb-halle.github.io/MetFrag/	[148]
GNPS Sare molekularra	Dereplicator	ccms-ucsd.github.io/GNPSDocumentation/dereplicator/	[149]
	MolNetEnhancer	ccms-ucsd.github.io/GNPSDocumentation/molnetenhancer/	[150]
	MetGen	metgem.github.io	[151]
	MS2LDA	ms2lda.org/	[152]
	ClassyFire	classyfire.wishartlab.com/	
Datu-base espezifikoak	Plant Metabolome database (PMDB)	scbt.sastra.edu/pmdb/	[153]
	KNApSAcK	www.knapsackfamily.com/KNApSAcK/	[154]
	Dictionary of Natural Products	dnp.chemnetbase.com	
	ReSpect for phytochemicals	spectra.psc.riken.jp	
	Plant Metabolic Network (PMN)	plantcyc.org/	
Detekzio- eta integrazio-herramintak	PlanMAT	sourceforge.net/projects/plantmat/	[155]
	MetaboTools	opencobra.github.io/cobratoolbox/stable/modules/dataIntegration/metabotools/index.html	[156]
	MetaCrop 2.0	metacrop.ipk-gatersleben.de	[157]

6. Taula: Estres abiotikoan oinarritutako landareen hazkuntzan egindako azken lanen laburpena

Aplikazio esparrua	Hazkuntza eta errendimendua	Bibliografia
Aldaketa klimatikoa	Tenperatura eta CO_2 maila altuak	[165]
Xenobiotikoei esposizioa	Konposatu per eta polifluoroalkilak letxugan	[136]
Ingurugiroko estresa: ura	Uraren estresa hirustan eta β-sitosterola hazkuntza erregulatzaile gisa	[161]
Ingurugiroko estresa	Landare medizinalen konposatu hegazkorretan tenperatura, argia eta ura	[164]
Ingurugiroko estresa: lehortea	Soja eta lehortearekiko tolerantzia	[163]
Ingurugiroko estresa: ura eta gatza	Hurbilketa metaboliko bidezko analisia	[162]
Loci ezaugarri kuantitatiboa (mQTLs)	Asiako arroza (*Oryza sativa*)	[167]
	Asiako arroza (*Oryza sativa*) eta uztaren hobekuntza	[168]
	Llollobelarra (*Lolium perenne*) eta mikrobioma	[170]
	Tomatea (*S. lycopersicum*) eta landarean hostoa eta lorea	[169]
	Tomatea (*S. lycopersicum*) eta Gibberellinen papera lehortearekiko tolerantzian	[166]
Egokitzea	Rhamnaceae mekanismo kimioebolutiboak eta sare molekularren aplikazioa	[171]

5.2.5. Etorkizunerako perspektibak

Metabolomikak ahalmen handia du estresari aurre egiteko landareak garatzen dituzten konposatuak aurreikusteko, hots, moldakortasuna aurreikusteko erabil daitezkeen markatzaileak aurkitzeko. Metaboloma aztertuz, landarearen egoera fisiologikoaren irudi globala ondoriozta daiteke, eta egoera hori erabakigarria da ingurumen-estresaren hainbat aldagaik batera parte hartzen duten agertokietan.

Hala ere, metabolomikaren eskaintza potentzial guztiak aprobetxatu ahal izateko, laborariek kontuan izan behar dituzte hazkunde-agertokiei buruzko zenbait baldintza. Gakoa, diseinu esperimentala egoera biologikoak gidatu behar duela ulertzea da. Landarearen genotipoa, garapen-faseak, uzta jasotzeko ehun-mota, eta erreplika biologikoen kantitatea eta aldakortasuna (haziak, klonak, etab.), beste aldagai batzuen artean, kontu handiz aukeratu behar dira formulatutako gakoen arabera. Gainera, estres abiotikoak behar bezala simulatu behar dira, aztertutako faktoreak baldintza oso kontrolatuetan eta gainerako parametroak konfiantza maila ezin hobean mantenduz

Bestalde, mugarri teknologiko batzuk lortu behar dira datuen analisiari eta tratamenduari dagokienez. Uste denez, landareek 100.000 eta milioi bat metabolito artean sortzen dituzte, eta horietako askok funtsezko rolak betetzen dituzte estresari aurre egiteko [172]. Hala ere, gaur egun ikerketa metabolomikoetan horien zati txiki bat baino ez da detektatzen (ehunka edo milaka metabolito gutxi), eta, beraz, aurreikusgarritasunari dagokionez, hobetzeko tartea handia da.

Espero dugu metabolomikak, bakarrik edo beste ikuspegi "omiko" edo irudi-teknika batzuekin batera, funtsezko zeregina izango duela landareen errendimendua ingurumenaren estrespean zentratzen diren hazkuntza-proiektuetan. Gero eta populazio eta klima aldaketa handiagoak eraginda, eta, lur eremuak dituen mugak direla eta, metabolomika proiektu hauen beharra areagotu egingo da [98], [173]. Gainera, badira sortzen ari diren hazkuntza-eremu batzuk, kannabisaren industria adibidez, hazkuntza-prozesuetan teknologia hau ezarriko dutenak. Zentzu horretan, puntako tresna analitikoen demokratizazioak eta datuen tratamendurako softwareak teknologia hau hurrengo urteetan erabiltzeko aukera emango dute.

Bibliografia

[1] S. A. Bonini *et al.*, 'Cannabis sativa: A comprehensive ethnopharmacological review of a medicinal plant with a long history', *Journal of Ethnopharmacology*, vol. 227, pp. 300–315, 2018, doi: 10.1016/j.jep.2018.09.004.

[2] 'Summary of Cannabis sativa, version 27.0'. Accessed: Jan. 11, 2024. [Online]. Available: https://pmn.plantcyc.org/organism-summary?object=CSATIVA_PK&_gl=1*t3ze85*_ga*OTk3NzcyMzE2LjE2OTkwMjUyMzI.*_ga_924HD6E7CQ*MTcwNDk5OTM4OS40LjEuMTcwNTAwMTE5My4wLjAuMA..

[3] 'History of cannabis', The University of Sydney. Accessed: Jan. 08, 2024. [Online]. Available: https://www.sydney.edu.au/lambert/medicinal-cannabis/history-of-cannabis.html

[4] 'Cannabis - Worldwide | Statista Market Forecast', Statista. Accessed: Jan. 12, 2024. [Online]. Available: https://www.statista.com/outlook/hmo/cannabis/worldwide

[5] Y. Gaoni and R. Mechoulam, 'Isolation, Structure, and Partial Synthesis of an Active Constituent of Hashish', *J. Am. Chem. Soc.*, vol. 86, no. 8, pp. 1646–1647, Apr. 1964, doi: 10.1021/ja01062a046.

[6] R. Mechoulam and Y. Shvo, 'Hashish—I: The structure of Cannabidiol', *Tetrahedron*, vol. 19, no. 12, pp. 2073–2078, Jan. 1963, doi: 10.1016/0040-4020(63)85022-X.

[7] T. B. Wood, W. T. Newton Spivey, and T. H. Easterfield, 'Cannabinol, part I', *Journal of Chemical Society*, vol. 75, pp. 20–36, 1899.

[8] Z. Atakan, 'Cannabis, a complex plant: different compounds and different effects on individuals', *Ther Adv Psychopharmacol*, vol. 2, no. 6, pp. 241–254, Dec. 2012, doi: 10.1177/2045125312457586.

[9] T. Gülck and B. L. Møller, 'Phytocannabinoids: Origins and Biosynthesis', *Trends in Plant Science*, vol. 25, no. 10, pp. 985–1004, 2020, doi: 10.1016/j.tplants.2020.05.005.

[10] S. A. Ahmed, S. A. Ross, D. Slade, M. M. Radwan, I. A. Khan, and M. A. ElSohly, 'Minor oxygenated cannabinoids from high potency Cannabis sativa L.', *Phytochemistry*, vol. 117, pp. 194–199, Sep. 2015, doi: 10.1016/j.phytochem.2015.04.007.

[11] M. N. Tahir, F. Shahbazi, S. Rondeau-Gagné, and J. F. Trant, 'The biosynthesis of the cannabinoids', *Journal of Cannabis Research*, vol. 3, no. 1, 2021, doi: 10.1186/s42238-021-00062-4.

[12] S. J. Livingston *et al.*, 'Cannabis glandular trichomes alter morphology and metabolite content during flower maturation', *The Plant Journal*, vol. 101, no. 1, pp. 37–56, 2020, doi: 10.1111/tpj.14516.

[13] L. J. Kearsey *et al.*, 'Structure of the Cannabis sativa olivetol-producing enzyme reveals cyclization plasticity in type III polyketide synthases', *The FEBS Journal*, vol. 287, no. 8, pp. 1511–1524, 2020, doi: 10.1111/febs.15089.

[14] F. Taura, S. Sirikantaramas, Y. Shoyama, K. Yoshikai, Y. Shoyama, and S. Morimoto, 'Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type

Cannabis sativa', *FEBS Letters*, vol. 581, no. 16, pp. 2929–2934, Jun. 2007, doi: 10.1016/j.febslet.2007.05.043.

[15] S. J. Gagne, J. M. Stout, E. Liu, Z. Boubakir, S. M. Clark, and J. E. Page, 'Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides', *Proceedings of the National Academy of Sciences*, vol. 109, no. 31, pp. 12811–12816, Jul. 2012, doi: 10.1073/pnas.1200330109.

[16] M. Fellermeier and M. H. Zenk, 'Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol', *FEBS Letters*, vol. 427, no. 2, pp. 283–285, May 1998, doi: 10.1016/S0014-5793(98)00450-5.

[17] I. J. Flores-Sanchez and R. Verpoorte, 'Plant Polyketide Synthases: A fascinating group of enzymes', *Plant Physiology and Biochemistry*, vol. 47, no. 3, pp. 167–174, Mar. 2009, doi: 10.1016/j.plaphy.2008.11.005.

[18] E. B. Russo, 'History of cannabis and its preparations in saga, science, and sobriquet', *Chemistry and Biodiversity*, vol. 4, no. 8, pp. 1614–1648, 2007, doi: 10.1002/cbdv.200790144.

[19] R. Mechoulam, 'Marihuana chemistry', *Science*, vol. 168, no. 3936, pp. 1159–1166, 1970, doi: 10.1126/science.168.3936.1159.

[20] 'Cannabis sativa cannabinoid biosynthesis'. Accessed: Jan. 13, 2024. [Online]. Available: https://pmn.plantcyc.org/CSATIVA_PK/NEW-IMAGE?type=PATHWAY&object=PWY-5140&detail-level=3

[21] E. M. Rock and L. A. Parker, 'Constituents of Cannabis Sativa', in *Cannabinoids and Neuropsychiatric Disorders*, E. Murillo-Rodriguez, S. R. Pandi-Perumal, and J. M. Monti, Eds., in Advances in Experimental Medicine and Biology. , Cham: Springer International Publishing, 2021, pp. 1–13. doi: 10.1007/978-3-030-57369-0_1.

[22] S. A. Ross and M. A. ElSohly, 'The Volatile Oil Composition of Fresh and Air-Dried Buds of Cannabis sativa', *J. Nat. Prod.*, vol. 59, no. 1, pp. 49–51, Jan. 1996, doi: 10.1021/np960004a.

[23] M. A. ElSohly and D. Slade, 'Chemical constituents of marijuana: The complex mixture of natural cannabinoids', *Life Sciences*, vol. 78, no. 5, pp. 539–548, Dec. 2005, doi: 10.1016/j.lfs.2005.09.011.

[24] L. O. Hanuš and Y. Hod, 'Terpenes/Terpenoids in Cannabis: Are They Important?', *Medical Cannabis and Cannabinoids*, vol. 3, no. 1, pp. 25–60, Aug. 2020, doi: 10.1159/000509733.

[25] S. R. Sommano, C. Chittasupho, W. Ruksiriwanich, and P. Jantrawut, 'The Cannabis Terpenes', *Molecules*, vol. 25, no. 24, p. 5792, Dec. 2020, doi: 10.3390/molecules25245792.

[26] M. Ternelli *et al.*, 'Innovative methods for the preparation of medical Cannabis oils with a high content of both cannabinoids and terpenes', *Journal of Pharmaceutical and Biomedical Analysis*, vol. 186, p. 113296, Jul. 2020, doi: 10.1016/j.jpba.2020.113296.

[27] M. Abdollahi, F. Sefidkon, M. Calagari, A. Mousavi, and M. F. Mahomoodally, 'Impact of four hemp (Cannabis sativa L.) varieties and stage of plant growth on yield and composition of essential oils', *Industrial Crops and Products*, vol. 155, p. 112793, Nov. 2020, doi: 10.1016/j.indcrop.2020.112793.

[28] J. K. Booth and J. Bohlmann, 'Terpenes in Cannabis sativa – From plant genome to humans', *Plant Science*, vol. 284, pp. 67–72, 2019, doi: 10.1016/j.plantsci.2019.03.022.

[29] M. M. Radwan, S. Chandra, S. Gul, and M. A. ElSohly, 'Cannabinoids, Phenolics, Terpenes and Alkaloids of Cannabis', *Molecules*, vol. 26, no. 9, Art. no. 9, Jan. 2021, doi: 10.3390/molecules26092774.

[30] E. P. M. de Meijer and L. J. M. van Soest, 'The CPRO Cannabis germplasm collection', *Euphytica*, vol. 62, no. 3, pp. 201–211, 1992, doi: 10.1007/BF00041754.

[31] J. M. McPartland, 'Cannabis Systematics at the Levels of Family, Genus, and Species', *Cannabis and Cannabinoid Research*, vol. 3, no. 1, pp. 203–212, 2018, doi: 10.1089/can.2018.0039.

[32] W. Bateson, *Mendel's Principles of Heredity - A Defence, with a Translation of Mendel's Original Papers on Hybridisation*. Cambridge University Press, 2009.

[33] O. Aizpurua-Olaizola, I. Elezgarai, I. Rico-Barrio, I. Zarandona, N. Etxebarria, and A. Usobiaga, 'Targeting the endocannabinoid system: future therapeutic strategies', *Drug Discovery Today*, vol. 22, no. 1, pp. 105–110, 2017, doi: 10.1016/j.drudis.2016.08.005.

[34] G. Mandolino, M. Bagatta, A. Carboni, P. Ranalli, and E. De Meijer, 'Qualitative and quantitative aspects of the inheritance of chemical phenotype in cannabis', *Journal of Industrial Hemp*, vol. 8, no. 2, pp. 51–72, 2003, doi: 10.1300/J237v08n02_04.

[35] E. P. M. de Meijer *et al.*, 'The Inheritance of Chemical Phenotype in Cannabis sativa L.', *Genetics*, vol. 163, no. 1, pp. 335–346, Jan. 2003, doi: 10.1093/genetics/163.1.335.

[36] E. P. M. de Meijer and K. M. Hammond, 'The inheritance of chemical phenotype in Cannabis sativa L. (II): Cannabigerol predominant plants', *Euphytica*, vol. 145, no. 1, pp. 189–198, Sep. 2005, doi: 10.1007/s10681-005-1164-8.

[37] V. Di Marzo, 'New approaches and challenges to targeting the endocannabinoid system', *Nature Reviews Drug Discovery*, vol. 17, no. 9, pp. 623–639, 2018, doi: 10.1038/nrd.2018.115.

[38] F. A. Iannotti, V. Di Marzo, and S. Petrosino, 'Endocannabinoids and endocannabinoid-related mediators: Targets, metabolism and role in neurological disorders', *Progress in Lipid Research*, vol. 62, pp. 107–128, 2016, doi: 10.1016/j.plipres.2016.02.002.

[39] Y. Li, X. Chen, Y. Nie, Y. Tian, X. Xiao, and F. Yang, 'Endocannabinoid activation of the TRPV1 ion channel is distinct from activation by capsaicin', *Journal of Biological Chemistry*, vol. 297, no. 3, Sep. 2021, doi: 10.1016/j.jbc.2021.101022.

[40] J. E. Lauckner, J. B. Jensen, H.-Y. Chen, H.-C. Lu, B. Hille, and K. Mackie, 'GPR55 is a cannabinoid receptor that increases intracellular calcium and inhibits M current', *Proc Natl Acad Sci U S A*, vol. 105, no. 7, pp. 2699–2704, Feb. 2008, doi: 10.1073/pnas.0711278105.

[41] V. Di Marzo and F. Piscitelli, 'The Endocannabinoid System and its Modulation by Phytocannabinoids', *Neurotherapeutics*, vol. 12, no. 4, pp. 692–698, Oct. 2015, doi: 10.1007/s13311-015-0374-6.

[42] C. S. Breivogel, D. E. Selley, and S. R. Childers, 'Cannabinoid receptor agonist efficacy for stimulating [35S]GTPgammaS binding to rat cerebellar membranes correlates with agonist-induced decreases in GDP affinity', *J Biol Chem*, vol. 273, no. 27, pp. 16865–16873, Jul. 1998, doi: 10.1074/jbc.273.27.16865.

[43] G. Griffin, P. J. Atkinson, V. M. Showalter, B. R. Martin, and M. E. Abood, 'Evaluation of cannabinoid receptor agonists and antagonists using the guanosine-5'-O-(3-[35S]thio)-triphosphate binding assay in rat cerebellar membranes', *J Pharmacol Exp Ther*, vol. 285, no. 2, pp. 553–560, May 1998.

[44] R. G. Pertwee, 'Pharmacology of cannabinoid receptor ligands', *Curr Med Chem*, vol. 6, no. 8, pp. 635–664, Aug. 1999.

[45] F. Shahbazi, V. Grandi, A. Banerjee, and J. F. Trant, 'Cannabinoids and Cannabinoid Receptors: The Story so Far', *iScience*, vol. 23, no. 7, p. 101301, Jul. 2020, doi: 10.1016/j.isci.2020.101301.

[46] T. Bisogno *et al.*, 'Molecular targets for cannabidiol and its synthetic analogues: effect on vanilloid VR1 receptors and on the cellular uptake and enzymatic hydrolysis of anandamide', *Br J Pharmacol*, vol. 134, no. 4, pp. 845–852, Oct. 2001, doi: 10.1038/sj.bjp.0704327.

[47] V. M. Showalter, D. R. Compton, B. R. Martin, and M. E. Abood, 'Evaluation of binding in a transfected cell line expressing a peripheral cannabinoid receptor (CB2): identification of cannabinoid receptor subtype selective ligands', *J Pharmacol Exp Ther*, vol. 278, no. 3, pp. 989–999, Sep. 1996.

[48] A. Thomas, G. L. Baillie, A. M. Phillips, R. K. Razdan, R. A. Ross, and R. G. Pertwee, 'Cannabidiol displays unexpectedly high potency as an antagonist of CB1 and CB2 receptor agonists in vitro', *Br J Pharmacol*, vol. 150, no. 5, pp. 613–623, Mar. 2007, doi: 10.1038/sj.bjp.0707133.

[49] A. Thomas *et al.*, 'Evidence that the plant cannabinoid Delta9-tetrahydrocannabivarin is a cannabinoid CB1 and CB2 receptor antagonist', *Br J Pharmacol*, vol. 146, no. 7, pp. 917–926, Dec. 2005, doi: 10.1038/sj.bjp.0706414.

[50] M. H. Rhee *et al.*, 'Cannabinol derivatives: binding to cannabinoid receptors and inhibition of adenylylcyclase', *J Med Chem*, vol. 40, no. 20, pp. 3228–3233, Sep. 1997, doi: 10.1021/jm970126f.

[51] M. Cascio, L. Gauson, L. Stevenson, R. Ross, and R. Pertwee, 'Evidence that the plant cannabinoid cannabigerol is a highly potent α2-adrenoceptor agonist and moderately potent 5HT1A receptor antagonist', *Br J Pharmacol*, vol. 159, no. 1, pp. 129–141, Jan. 2010, doi: 10.1111/j.1476-5381.2009.00515.x.

[52] A. F. Bento *et al.*, 'β-Caryophyllene Inhibits Dextran Sulfate Sodium-Induced Colitis in Mice through CB2 Receptor Activation and PPARγ Pathway', *Am J Pathol*, vol. 178, no. 3, pp. 1153–1166, Mar. 2011, doi: 10.1016/j.ajpath.2010.11.052.

[53] K. Riboulet-Zemouli and M. A. Krawitz, 'WHO's first scientific review of medicinal Cannabis: from global struggle to patient implications', *Drugs, Habits and Social Policy*, vol. 23, no. 1, pp. 5–21, Jan. 2022, doi: 10.1108/DHS-11-2021-0060.

[54] 'Recommendation of removal for cannabis and cannabis resin from Schedule IV of 1961 Single Convention on Narcotic Drugs', presented at the United Nations Comission on Narcotic Drugs, Vienna, Dec. 2020. Accessed: Jun. 02, 2023. [Online]. Available: https://www.who.int/news/item/04-12-2020-un-commission-on-narcotic-drugs-reclassifies-cannabis-to-recognize-its-therapeutic-uses

[55] '2018 Farm Bill Overview', National Sustainable Agriculture Coalition. Accessed: Jan. 16, 2024. [Online]. Available: https://sustainableagriculture.net/our-work/campaigns/fbcampaign/

[56] 'State Medical Cannabis Laws'. Accessed: Jun. 02, 2023. [Online]. Available: https://www.ncsl.org/health/state-medical-cannabis-laws

[57] 'Legislative Review of the Cannabis Act: What We Heard Report'. Accessed: Jan. 22, 2024. [Online]. Available: https://www.canada.ca/en/health-canada/services/drugs-medication/cannabis/laws-regulations/cannabis-act-legislative-review/expert-panel/legislative-review-cannabis-act-report.html

[58] *Kontrollierte Abgabe von Cannabis: Eckpunktepapier der Bundesregierung liegt vor*. 2022. Accessed: Jun. 02, 2023. [Online]. Available: https://www.bundesgesundheitsministerium.de/ministerium/meldungen/kontrollierte-abgabe-von-cannabis-eckpunktepapier-der-bundesregierung-liegt-vor.html

[59] 'Cannabis law and legislation in Germany | CMS Expert Guides'. Accessed: Jan. 17, 2024. [Online]. Available: https://cms.law/en/int/expert-guides/cms-expert-guide-to-a-legal-roadmap-to-cannabis/germany

[60] 'Cannabis law and legislation in Italy | CMS Expert Guides'. Accessed: Jan. 17, 2024. [Online]. Available: https://cms.law/en/int/expert-guides/cms-expert-guide-to-a-legal-roadmap-to-cannabis/italy

[61] 'Cannabis law and legislation in Portugal | CMS Expert Guides'. Accessed: Jan. 17, 2024. [Online]. Available: https://cms.law/en/int/expert-guides/cms-expert-guide-to-a-legal-roadmap-to-cannabis/portugal

[62] *Federal Act on Narcotics and Psychotropic Substances (Narcotics Act, NarcA)*. 2022. Accessed: Jun. 02, 2023. [Online]. Available: https://www.fedlex.admin.ch/eli/cc/1952/241_241_245/en

[63] 'Cannabis medicinal', Argentina.gob.ar. Accessed: Jan. 08, 2024. [Online]. Available: https://www.argentina.gob.ar/justicia/derechofacil/leysimple/cannabis-medicinal

[64] C. J. C. F. Fijnaut and B. de Ruyver, *The Third Way: A Plea for a Balanced Cannabis Policy*. BRILL, 2015.

[65] 'Cannabis law and legislation in Colombia | CMS Expert Guides'. Accessed: Jan. 07, 2024. [Online]. Available: https://cms.law/en/int/expert-guides/cms-expert-guide-to-a-legal-roadmap-to-cannabis/colombia

[66] 'Medicinal Cannabis Industry Given Legal Footing - Ministry of Agriculture, Forestry, Fisheries and Rural Transformation'. Accessed: Jan. 07, 2024. [Online]. Available:

https://agriculture.gov.vc/agriculture/index.php?option=com_content&view=article&id=419:medicinal-cannabis-industry-given-legal-footing-3&catid=104&Itemid=288

[67] C. M. Lindsay *et al.*, 'Potency trends of cannabis in Jamaica during the period of 2014 to 2020', *Drug Test Anal*, Jun. 2023, doi: 10.1002/dta.3527.

[68] T. G. Administration (TGA), 'Guidance for the use of medicinal cannabis in Australia: Overview | Therapeutic Goods Administration (TGA)'. Accessed: Jan. 08, 2024. [Online]. Available: https://www.tga.gov.au/resources/resource/guidance/guidance-use-medicinal-cannabis-australia-overview

[69] S. Assanangkornchai, K. Thaikla, M. Talek, and D. Saingam, 'Medical cannabis use in Thailand after its legalization: a respondent-driven sample survey', *PeerJ*, vol. 10, p. e12809, Jan. 2022, doi: 10.7717/peerj.12809.

[70] C. Rusenga, G. Klantschnig, N. Carrier, and S. Howell, 'Cannabis policy changes in Africa are welcome. But small producers are the losers', The Conversation. Accessed: Jan. 08, 2024. [Online]. Available: http://theconversation.com/cannabis-policy-changes-in-africa-are-welcome-but-small-producers-are-the-losers-179681

[71] AfricaNews, 'Morocco joins growing list of African countries to legalize cannabis', Africanews. Accessed: Jan. 08, 2024. [Online]. Available: https://www.africanews.com/2021/05/27/morocco-joins-growing-list-of-african-countries-to-legalize-cannabis/

[72] Y. A. Adebisi and D. Quazeem Olaoye, 'Medical Use of Cannabis in Africa: The Pharmacists' Perspective', *Innov Pharm*, vol. 13, no. 1, p. 10.24926/iip.v13i1.4430, Apr. 2022, doi: 10.24926/iip.v13i1.4430.

[73] F. Modaresi and K. Talachian, 'The Characteristics of Clinical Trials on Cannabis and Cannabinoids: A Review of Trials for Therapeutic or Drug Development Purposes', *Pharmaceut Med*, vol. 36, no. 6, pp. 387–400, Dec. 2022, doi: 10.1007/s40290-022-00447-7.

[74]. :: 'CIMA ::. PROSPECTO SATIVEX 2,7 MG/2,5 MG SOLUCION PARA PULVERIZACION BUCAL'. Accessed: Jan. 22, 2024. [Online]. Available: https://cima.aemps.es/cima/dochtml/p/72544/Prospecto_72544.html

[75]. :: 'CIMA ::. FICHA TECNICA EPIDYOLEX 100 MG/ML SOLUCION ORAL'. Accessed: Jan. 22, 2024. [Online]. Available: https://cima.aemps.es/cima/dochtml/ft/1191389001/

[76] 'Dronabinol: MedlinePlus Drug Information'. Accessed: Jan. 22, 2024. [Online]. Available: https://medlineplus.gov/druginfo/meds/a607054.html

[77] 'Nabilone: MedlinePlus Drug Information'. Accessed: Jan. 22, 2024. [Online]. Available: https://medlineplus.gov/druginfo/meds/a607048.html

[78] 'Cannabis law and legislation in Spain | CMS Expert Guides'. Accessed: Jan. 17, 2024. [Online]. Available: https://cms.law/en/int/expert-guides/cms-expert-guide-to-a-legal-roadmap-to-cannabis/spain

[79] A. Plaza *et al.*, 'Recent advances in techniques for hyperspectral image processing', *Remote Sensing of Environment*, vol. 113, no. SUPPL. 1, pp. S110–S122, 2009, doi: 10.1016/j.rse.2007.07.028.

[80] A. A. Gowen, C. P. O'Donnell, P. J. Cullen, G. Downey, and J. M. Frias, 'Hyperspectral imaging - an emerging process analytical tool for food quality and safety control', *Trends in Food Science and Technology*, vol. 18, no. 12, pp. 590–598, 2007, doi: 10.1016/j.tifs.2007.06.001.

[81] D. Cozzolino, 'The Ability of Near Infrared (NIR) Spectroscopy to Predict Functional Properties in Foods: Challenges and Opportunities', *Molecules*, vol. 26, no. 22, p. 6981, Nov. 2021, doi: 10.3390/molecules26226981.

[82] K. B. Beć, J. Grabska, and C. W. Huck, 'Near-Infrared Spectroscopy in Bio-Applications', *Molecules*, vol. 25, no. 12, Art. no. 12, Jan. 2020, doi: 10.3390/molecules25122948.

[83] K. B. Beć and C. W. Huck, 'Breakthrough Potential in Near-Infrared Spectroscopy: Spectra Simulation. A Review of Recent Developments', *Frontiers in Chemistry*, vol. 7, 2019, Accessed: Jan. 18, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fchem.2019.00048

[84] T. M. P. Cattaneo and A. Stellari, 'Review: NIR Spectroscopy as a Suitable Tool for the Investigation of the Horticultural Field', *Agronomy*, vol. 9, no. 9, Art. no. 9, Sep. 2019, doi: 10.3390/agronomy9090503.

[85] C. Sánchez-Carnerero Callado, N. Núñez-Sánchez, S. Casano, and C. Ferreiro-Vera, 'The potential of near infrared spectroscopy to estimate the content of cannabinoids in Cannabis sativa L.: A comparative study', *Talanta*, vol. 190, pp. 147–157, 2018, doi: 10.1016/j.talanta.2018.07.085.

[86] R. Deidda *et al.*, 'New perspective for the in-field analysis of cannabis samples using handheld near-infrared spectroscopy: A case study focusing on the determination of Δ9-tetrahydrocannabinol', *Journal of Pharmaceutical and Biomedical Analysis*, vol. 202, 2021, doi: 10.1016/j.jpba.2021.114150.

[87] *Commission Delegated Regulation (EU) 2017/1155 of 15 February 2017 amending Delegated Regulation (EU) No 639/2014 as regards the control measures relating to the cultivation of hemp, certain provisions on the greening payment, the payment for young farmers in control of a legal person, the calculation of the per unit amount in the framework of voluntary coupled support, the fractions of payment entitlements and certain notification requirements relating to the single area payment scheme and the voluntary coupled support, and amending Annex X to Regulation (EU) No 1307/2013 of the European Parliament and of the Council*, vol. 167. 2017. Accessed: Nov. 08, 2023. [Online]. Available: http://data.europa.eu/eli/reg_del/2017/1155/oj/eng

[88] C. Duchateau, J.-M. Kauffmann, M. Canfyn, C. Stévigny, K. De Braekeleer, and E. Deconinck, 'Discrimination of legal and illegal Cannabis spp. according to European legislation using near infrared spectroscopy and chemometrics', *Drug Testing and Analysis*, vol. 12, no. 9, pp. 1309–1319, 2020, doi: 10.1002/dta.2865.

[89] K. Su *et al.*, 'NIR spectroscopy for rapid measurement of moisture and cannabinoid contents of industrial hemp (Cannabis sativa)', *Industrial Crops and Products*, vol. 184, 2022, doi: 10.1016/j.indcrop.2022.115007.

[90] S. Yao, C. Ball, G. Miyagusuku-Cruzado, M. M. Giusti, D. P. Aykas, and L. E. Rodriguez-Saona, 'A novel handheld FT-NIR spectroscopic approach for real-time screening of major cannabinoids content in hemp', *Talanta*, vol. 247, 2022, doi: 10.1016/j.talanta.2022.123559.

[91] C. Jarén, P. C. Zambrana, C. Pérez-Roncal, A. López-Maestresalas, A. Ábrego, and S. Arazuri, 'Potential of NIRS Technology for the Determination of Cannabinoid Content in Industrial Hemp (Cannabis sativa L.)', *Agronomy*, vol. 12, no. 4, 2022, doi: 10.3390/agronomy12040938.

[92] D. Valinger *et al.*, 'Development of ANN models based on combined UV-vis-NIR spectra for rapid quantification of physical and chemical properties of industrial hemp extracts', *Phytochemical Analysis*, vol. 32, no. 3, pp. 326–338, 2021, doi: 10.1002/pca.2979.

[93] J. M. Amigo, H. Babamoradi, and S. Elcoroaristizabal, 'Hyperspectral image analysis. A tutorial', *Analytica Chimica Acta*, vol. 896, pp. 34–51, 2015, doi: 10.1016/j.aca.2015.09.030.

[94] M. San Nicolas *et al.*, 'NIR-hyperspectral imaging and machine learning for non-invasive chemotype classification in Cannabis sativa L', *Computers and Electronics in Agriculture*, vol. 217, 2024, doi: 10.1016/j.compag.2023.108551.

[95] G. Lu and B. Fei, 'Medical hyperspectral imaging: a review', *JBO*, vol. 19, no. 1, p. 010901, Jan. 2014, doi: 10.1117/1.JBO.19.1.010901.

[96] O. Fernandez *et al.*, 'Fortune telling: metabolic markers of plant performance', *Metabolomics*, vol. 12, no. 10, 2016, doi: 10.1007/s11306-016-1099-1.

[97] S. I. Zandalinas, R. Mittler, D. Balfagón, V. Arbona, and A. Gómez-Cadenas, 'Plant adaptations to the combination of drought and high temperatures', *Physiologia Plantarum*, vol. 162, no. 1, pp. 2–12, 2018, doi: 10.1111/ppl.12540.

[98] A. Raza *et al.*, 'Impact of Climate Change on Crops Adaptation and Strategies to Tackle Its Outcome: A Review', *Plants*, vol. 8, no. 2, p. 34, Feb. 2019, doi: 10.3390/plants8020034.

[99] M. Abdelrahman, D. J. Burritt, and L.-S. P. Tran, 'The use of metabolomic quantitative trait locus mapping and osmotic adjustment traits for the improvement of crop yields under environmental stresses', *Seminars in Cell and Developmental Biology*, vol. 83, pp. 86–94, 2018, doi: 10.1016/j.semcdb.2017.06.020.

[100] R. Kumar, A. Bohra, A. K. Pandey, M. K. Pandey, and A. Kumar, 'Metabolomics for Plant Improvement: Status and Prospects', *Front Plant Sci*, vol. 8, Aug. 2017, doi: 10.3389/fpls.2017.01302.

[101] M. Zivy, S. Wienkoop, J. Renaut, C. Pinheiro, E. Goulas, and S. Carpentier, 'The quest for tolerant varieties: the importance of integrating "omics" techniques to phenotyping', *Front Plant Sci*, vol. 6, p. 448, 2015, doi: 10.3389/fpls.2015.00448.

[102] D. Grissa, M. Pétéra, M. Brandolini, A. Napoli, B. Comte, and E. Pujos-Guillot, 'Feature Selection Methods for Early Predictive Biomarker Discovery Using Untargeted Metabolomic Data', *Front. Mol. Biosci.*, vol. 3, 2016, doi: 10.3389/fmolb.2016.00030.

[103]O. Fernandez *et al.*, 'Metabolomic characterization of sunflower leaf allows discriminating genotype groups or stress levels with a minimal set of metabolic markers', *Metabolomics*, vol. 15, no. 4, 2019, doi: 10.1007/s11306-019-1515-4.

[104]S. Alseekh, L. Bermudez, L. A. de Haro, A. R. Fernie, and F. Carrari, 'Crop metabolomics: from diagnostics to assisted breeding', *Metabolomics*, vol. 14, no. 11, p. 148, Nov. 2018, doi: 10.1007/s11306-018-1446-5.

[105]S. Wu *et al.*, 'Mapping the Arabidopsis Metabolic Landscape by Untargeted Metabolomics at Different Environmental Conditions', *Molecular Plant*, vol. 11, no. 1, pp. 118–134, 2018, doi: 10.1016/j.molp.2017.08.012.

[106]K. K. Meena *et al.*, 'Abiotic stress responses and microbe-mediated mitigation in plants: The omics strategies', *Frontiers in Plant Science*, vol. 8, 2017, doi: 10.3389/fpls.2017.00172.

[107]J.-L. Wolfender, S. Rudaz, Y. H. Choi, and H. K. Kim, 'Plant metabolomics: From holistic data to relevant biomarkers', *Current Medicinal Chemistry*, vol. 20, no. 8, pp. 1056–1090, 2013.

[108]A. M. Rodrigues, A. I. Ribeiro-Barros, and C. António, 'Experimental design and sample preparation in forest tree metabolomics', *Metabolites*, vol. 9, no. 12, 2019, doi: 10.3390/metabo9120285.

[109]Y. Gibon and D. Rolin, 'Aspects of experimental design for plant metabolomics experiments and guidelines for growth of plant material', *Methods Mol. Biol.*, vol. 860, pp. 13–30, 2012, doi: 10.1007/978-1-61779-594-7_2.

[110]V. Arbona, M. Manzi, C. de Ollas, and A. Gómez-Cadenas, 'Metabolomics as a tool to investigate abiotic stress tolerance in plants', *International Journal of Molecular Sciences*, vol. 14, no. 3, pp. 4885–4911, 2013, doi: 10.3390/ijms14034885.

[111]H. K. Kim and R. Verpoorte, 'Sample preparation for plant metabolomics', *Phytochem. Anal.*, vol. 21, no. 1, pp. 4–13, Jan. 2010, doi: 10.1002/pca.1188.

[112]T. F. Jorge *et al.*, 'Mass spectrometry-based plant metabolomics: Metabolite responses to abiotic stress', *Mass spectrometry reviews*, vol. 35, no. 5, pp. 620–649, 2016, doi: 10.1002/mas.21449.

[113]S. Champely *et al.*, 'pwr: Basic Functions for Power Analysis'. Mar. 17, 2020. Accessed: Jun. 23, 2020. [Online]. Available: https://CRAN.R-project.org/package=pwr

[114]F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, 'G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences', *Behavior Research Methods*, vol. 39, no. 2, pp. 175–191, 2007, doi: 10.3758/BF03193146.

[115]Y. Fujita, M. Fujita, K. Shinozaki, and K. Yamaguchi-Shinozaki, 'ABA-mediated transcriptional regulation in response to osmotic stress in plants', *Journal of Plant Research*, vol. 124, no. 4, pp. 509–525, 2011, doi: 10.1007/s10265-011-0412-3.

[116]R. C. H. De Vos, S. Moco, A. Lommen, J. J. B. Keurentjes, R. J. Bino, and R. D. Hall, 'Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry', *Nature Protocols*, vol. 2, no. 4, pp. 778–791, 2007, doi: 10.1038/nprot.2007.95.

[117]A. Junker *et al.*, 'Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems', *Frontiers in Plant Science*, vol. 5, 2015, Accessed: Jan. 31, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpls.2014.00770

[118]T. F. Jorge, A. T. Mata, and C. António, 'Mass spectrometry as a quantitative tool in plant metabolomics', *Philos Trans A Math Phys Eng Sci*, vol. 374, no. 2079, Oct. 2016, doi: 10.1098/rsta.2015.0370.

[119]C. Bénard *et al.*, 'Untargeted Analysis of Semipolar Compounds by LC-MS and Targeted Analysis of Fatty Acids by GC-MS/GC-FID: From Plant Cultivation to Extract Preparation', in *Plant Metabolomics: Methods and Protocols*, C. António, Ed., in Methods in Molecular Biology. , New York, NY: Springer, 2018, pp. 101–124. doi: 10.1007/978-1-4939-7819-9_8.

[120]M. Creydt, M. Arndt, D. Hudzik, and M. Fischer, 'Plant Metabolomics: Evaluation of Different Extraction Parameters for Nontargeted UPLC-ESI-QTOF-Mass Spectrometry at the Example of White Asparagus officinalis', *J. Agric. Food Chem.*, vol. 66, no. 48, pp. 12876–12887, Dec. 2018, doi: 10.1021/acs.jafc.8b06037.

[121]J. Chu, S. Fang, P. Xin, Z. Guo, and Y. Chen, '14 - Quantitative analysis of plant hormones based on LC-MS/MS', in *Hormone Metabolism and Signaling in Plants*, J. Li, C. Li, and S. M. Smith, Eds., Academic Press, 2017, pp. 471–537. doi: 10.1016/B978-0-12-811562-6.00014-1.

[122]L. Bernardo *et al.*, 'Metabolomic responses triggered by arbuscular mycorrhiza enhance tolerance to water stress in wheat cultivars', *Plant Physiology and Biochemistry*, vol. 137, pp. 203–212, 2019, doi: 10.1016/j.plaphy.2019.02.007.

[123]H. Ghosson, A. Schwarzenberg, F. Jamois, and J.-C. Yvin, 'Simultaneous untargeted and targeted metabolomics profiling of underivatized primary metabolites in sulfur-deficient barley by ultra-high performance liquid chromatography-quadrupole/time-of-flight mass spectrometry', *Plant Methods*, vol. 14, no. 1, p. 62, Jul. 2018, doi: 10.1186/s13007-018-0329-0.

[124]N. Reyes-Garcés and E. Gionfriddo, 'Recent developments and applications of solid phase microextraction as a sample preparation approach for mass-spectrometry-based metabolomics and lipidomics - ScienceDirect', *Trends Analyt Chem*, vol. 113, pp. 172–181.

[125]H. Gika, C. Virgiliou, G. Theodoridis, R. S. Plumb, and I. D. Wilson, 'Untargeted LC/MS-based metabolic phenotyping (metabonomics/metabolomics): The state of the art', *Journal of Chromatography B*, vol. 1117, pp. 136–147, Jun. 2019, doi: 10.1016/j.jchromb.2019.04.009.

[126]G. Alvarez-Rivera, D. Ballesteros-Vivas, F. Parada-Alfonso, E. Ibañez, and A. Cifuentes, 'Recent applications of high resolution mass spectrometry for the characterization of plant natural products', *TrAC Trends in Analytical Chemistry*, vol. 112, pp. 87–101, Mar. 2019, doi: 10.1016/j.trac.2019.01.002.

[127]A. Piasecka, P. Kachlicki, and M. Stobiecki, 'Analytical Methods for Detection of Plant Metabolomes Changes in Response to Biotic and Abiotic Stresses', *Int J Mol Sci*, vol. 20, no. 2, Jan. 2019, doi: 10.3390/ijms20020379.

[128]M. Ghaste, R. Mistrik, and V. Shulaev, 'Applications of Fourier Transform Ion Cyclotron Resonance (FT-ICR) and Orbitrap Based High Resolution Mass Spectrometry in Metabolomics and Lipidomics', *International Journal of Molecular Sciences*, vol. 17, no. 6, Art. no. 6, Jun. 2016, doi: 10.3390/ijms17060816.

[129]M. Jahangir, T. R. Nuringtyas, K. Ali, E. G. Wilson, Y. H. Choi, and R. Verpoorte, 'CHAPTER 9:NMR-based Metabolomics: Understanding Plant Chemistry and Identification of Biologically Active Compounds', in *NMR-based Metabolomics*, 2018, pp. 246–263. doi: 10.1039/9781782627937-00246.

[130]A. Kaufmann and P. Teale, 'Capabilities and Limitations of High-Resolution Mass Spectrometry (HRMS): Time-of-flight and Orbitrap™', in *Chemical Analysis of Non-antimicrobial Veterinary Drug Residues in Food*, John Wiley & Sons, Ltd, 2016, pp. 93–139. doi: 10.1002/9781118696781.ch3.

[131]A. J. Levy, N. R. Oranzi, A. Ahmadireskety, R. H. J. Kemperman, M. S. Wei, and R. A. Yost, 'Recent progress in metabolomics using ion mobility-mass spectrometry', *TrAC Trends in Analytical Chemistry*, vol. 116, pp. 274–281, Jul. 2019, doi: 10.1016/j.trac.2019.05.001.

[132]D. Sturtevant, Y.-J. Lee, and K. D. Chapman, 'Matrix assisted laser desorption/ionization-mass spectrometry imaging (MALDI-MSI) for direct visualization of plant metabolites in situ', *Current Opinion in Biotechnology*, vol. 37, pp. 53–60, Feb. 2016, doi: 10.1016/j.copbio.2015.10.004.

[133]C. Y. Hamany Djande, C. Pretorius, F. Tugizimana, L. A. Piater, and I. A. Dubery, 'Metabolomics: A Tool for Cultivar Phenotyping and Investigation of Grain Crops', *Agronomy*, vol. 10, no. 6, Art. no. 6, Jun. 2020, doi: 10.3390/agronomy10060831.

[134]A. Razzaq, B. Sadia, A. Raza, M. K. Hameed, and F. Saleem, 'Metabolomics: A Way Forward for Crop Improvement', *Metabolites*, vol. 9, no. 9(12), p. 303, 2019, doi: 10.3390/metabo9120303.

[135]K. Peters *et al.*, 'Current Challenges in Plant Eco-Metabolomics', *Int J Mol Sci*, vol. 19, no. 5, May 2018, doi: 10.3390/ijms19051385.

[136]S. Li, Ed., *Computational Methods and Data Analysis for Metabolomics*. in Methods in Molecular Biology. Springer US, 2020. doi: 10.1007/978-1-0716-0239-3.

[137]C. António, Ed., *Plant Metabolomics: Methods and Protocols*. in Methods in Molecular Biology. Humana Press, 2018. doi: 10.1007/978-1-4939-7819-9.

[138]J. T. Bjerrum, Ed., *Metabonomics: Methods and Protocols*. in Methods in Molecular Biology. Humana Press, 2015. doi: 10.1007/978-1-4939-2377-9.

[139]K. Saito and F. Matsuda, 'Metabolomics for Functional Genomics, Systems Biology, and Biotechnology', *Annual Review of Plant Biology*, vol. 61, no. 1, pp. 463–489, 2010, doi: 10.1146/annurev.arplant.043008.092035.

[140]H. L. Röst *et al.*, 'OpenMS: a flexible open-source software platform for mass spectrometry data analysis', *Nat. Methods*, vol. 13, no. 9, pp. 741–748, 30 2016, doi: 10.1038/nmeth.3959.

[141]I. Blaženović, T. Kind, J. Ji, and O. Fiehn, 'Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics', *Metabolites*, vol. 8, no. 2, May 2018, doi: 10.3390/metabo8020031.

[142]M. Wang *et al.*, 'Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking', *Nature Biotechnology*, vol. 34, no. 8, Art. no. 8, Aug. 2016, doi: 10.1038/nbt.3597.

[143]R. R. Da Silva *et al.*, 'Propagating annotations of molecular networks using in silico fragmentation', *PLOS Computational Biology*, vol. 14, no. 4, p. e1006089, Apr. 2018, doi: 10.1371/journal.pcbi.1006089.

[144]L. F. Nothias *et al.*, 'Feature-based Molecular Networking in the GNPS Analysis Environment', *bioRxiv*, p. 812404, Oct. 2019, doi: 10.1101/812404.

[145]K. Dührkop *et al.*, 'SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information', *Nature Methods*, vol. 16, no. 4, Art. no. 4, Apr. 2019, doi: 10.1038/s41592-019-0344-8.

[146]K. Dührkop, H. Shen, M. Meusel, J. Rousu, and S. Böcker, 'Searching molecular structure databases with tandem mass spectra using CSI:FingerID', *Proc. Natl. Acad. Sci. U.S.A.*, vol. 112, no. 41, pp. 12580–12585, Oct. 2015, doi: 10.1073/pnas.1509788112.

[147]N. G. Mahieu, J. L. Genenbacher, and G. J. Patti, 'A roadmap for the XCMS family of software solutions in metabolomics', *Current Opinion in Chemical Biology*, vol. 30, pp. 87–93, Feb. 2016, doi: 10.1016/j.cbpa.2015.11.009.

[148]C. Ruttkies, E. L. Schymanski, S. Wolf, J. Hollender, and S. Neumann, 'MetFrag relaunched: incorporating strategies beyond in silico fragmentation', *J Cheminform*, vol. 8, p. 3, 2016, doi: 10.1186/s13321-016-0115-9.

[149]H. Mohimani *et al.*, 'Dereplication of microbial metabolites through database search of mass spectra', *Nature Communications*, vol. 9, no. 1, Art. no. 1, Oct. 2018, doi: 10.1038/s41467-018-06082-8.

[150]M. Ernst *et al.*, 'MolNetEnhancer: Enhanced Molecular Networks by Integrating Metabolome Mining and Annotation Tools', *Metabolites*, vol. 9, no. 7, Jul. 2019, doi: 10.3390/metabo9070144.

[151]F. Olivon, N. Elie, G. Grelier, F. Roussi, M. Litaudon, and D. Touboul, 'MetGem Software for the Generation of Molecular Networks Based on the t-SNE Algorithm', *Anal. Chem.*, vol. 90, no. 23, pp. 13900–13908, Dec. 2018, doi: 10.1021/acs.analchem.8b03099.

[152]S. Rogers, C. W. Ong, J. Wandy, M. Ernst, L. Ridder, and J. J. J. van der Hooft, 'Deciphering complex metabolite mixtures by unsupervised and supervised substructure discovery and semi-automated annotation from MS/MS spectra', *Faraday Discuss.*, vol. 218, no. 0, pp. 284–302, Aug. 2019, doi: 10.1039/C8FD00235E.

[153]M. Udayakumar, D. Prem Chandar, N. Arun, J. Mathangi, K. Hemavathi, and R. Seenivasagam, 'PMDB: Plant Metabolome Database—A Metabolomic Approach', *Med Chem Res*, vol. 21, no. 1, pp. 47–52, Jan. 2012, doi: 10.1007/s00044-010-9506-z.

[154]F. M. Afendi *et al.*, 'KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research', *Plant Cell Physiol.*, vol. 53, no. 2, p. e1, Feb. 2012, doi: 10.1093/pcp/pcr165.

[155]F. Qiu, D. D. Fine, D. J. Wherritt, Z. Lei, and L. W. Sumner, 'PlantMAT: A Metabolomics Tool for Predicting the Specialized Metabolic Potential of a System and for Large-Scale Metabolite Identifications', *Anal. Chem.*, vol. 88, no. 23, pp. 11373–11383, Dec. 2016, doi: 10.1021/acs.analchem.6b00906.

[156]M. K. Aurich, R. M. T. Fleming, and I. Thiele, 'MetaboTools: A Comprehensive Toolbox for Analysis of Genome-Scale Metabolic Models', *Front Physiol*, vol. 7, p. 327, 2016, doi: 10.3389/fphys.2016.00327.

[157]F. Schreiber *et al.*, 'MetaCrop 2.0: managing and exploring information about crop plant metabolism', *Nucleic Acids Res.*, vol. 40, no. Database issue, pp. D1173-1177, Jan. 2012, doi: 10.1093/nar/gkr1004.

[158]J. Gu, Y. Gui, L. Chen, G. Yuan, H.-Z. Lu, and X. Xu, 'Use of Natural Products as Chemical Library for Drug Discovery and Network Pharmacology', *PLoS One*, vol. 8, no. 4, Apr. 2013, doi: 10.1371/journal.pone.0062839.

[159]M. M. Rinschen, J. Ivanisevic, M. Giera, and G. Siuzdak, 'Identification of bioactive metabolites using activity metabolomics', *Nat. Rev. Mol. Cell Biol.*, vol. 20, no. 6, pp. 353–367, 2019, doi: 10.1038/s41580-019-0108-4.

[160]L. Heirendt *et al.*, 'Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0', *Nature Protocols*, vol. 14, no. 3, Art. no. 3, Mar. 2019, doi: 10.1038/s41596-018-0098-2.

[161]Z. Li *et al.*, 'Metabolomics and physiological analyses reveal β-sitosterol as an important plant growth regulator inducing tolerance to water stress in white clover', *Planta*, vol. 250, no. 6, pp. 2033–2046, 2019, doi: 10.1007/s00425-019-03277-1.

[162]A. Llanes, A. Andrade, S. Alemano, and V. Luna, *Metabolomic Approach to Understand Plant Adaptations to Water and Salt Stress*. 2018. doi: 10.1016/B978-0-12-812689-9.00006-6.

[163]X. Wang *et al.*, 'Metabolomics reveals the drought-tolerance mechanism in wild soybean (Glycine soja)', *Acta Physiologiae Plantarum*, vol. 41, no. 9, 2019, doi: 10.1007/s11738-019-2939-1.

[164]V. A. Ueno and A. C. H. F. Sawaya, 'Influence of environmental factors on the volatile composition of two Brazilian medicinal plants: Mikania laevigata and Mikania glomerata', *Metabolomics*, vol. 15, no. 6, p. 91, 12 2019, doi: 10.1007/s11306-019-1546-x.

[165]J. M. Wedow, C. R. Yendrek, T. R. Mello, S. Creste, C. A. Martinez, and E. A. Ainsworth, 'Metabolite and transcript profiling of Guinea grass (Panicum maximum Jacq) response to elevated [CO2] and temperature', *Metabolomics*, vol. 15, no. 4, p. 51, Mar. 2019, doi: 10.1007/s11306-019-1511-8.

[166]R. P. Omena-Garcia *et al.*, 'Growth and metabolic adjustments in response to gibberellin deficiency in drought stressed tomato plants', *Environmental and Experimental Botany*, vol. 159, pp. 95–107, 2019, doi: 10.1016/j.envexpbot.2018.12.011.

[167]L. Witjes *et al.*, 'A genetical metabolomics approach for bioprospecting plant biosynthetic gene clusters', *BMC Research Notes*, vol. 12, no. 1, p. 194, Apr. 2019, doi: 10.1186/s13104-019-4222-3.

[168]Y. Okazaki and K. Saito, 'Integrated metabolomics and phytochemical genomics approaches for studies on rice', *GigaScience*, vol. 5, no. 1, p. 11, Mar. 2016, doi: 10.1186/s13742-016-0116-7.

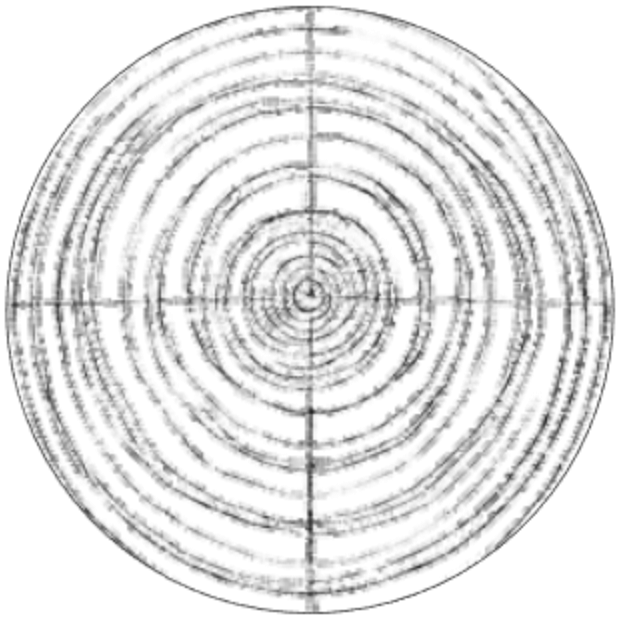
[169]A. Nunes-Nesi *et al.*, 'Identification and characterization of metabolite quantitative trait loci in tomato leaves and comparison with those reported for fruits and seeds', *Metabolomics*, vol. 15, no. 4, p. 46, Mar. 2019, doi: 10.1007/s11306-019-1503-8.

[170]A. Foito *et al.*, 'Quantitative trait loci associated with different polar metabolites in perennial ryegrass - providing scope for breeding towards increasing certain polar metabolites', *BMC Genet*, vol. 18, no. 1, p. 84, Oct. 2017, doi: 10.1186/s12863-017-0552-0.

[171]K. B. Kang *et al.*, 'Comprehensive mass spectrometry-guided phenotyping of plant specialized metabolites reveals metabolic diversity in the cosmopolitan plant family Rhamnaceae', *Plant J.*, vol. 98, no. 6, pp. 1134–1144, 2019, doi: 10.1111/tpj.14292.

[172]C. Fang, A. R. Fernie, and J. Luo, 'Exploring the Diversity of Plant Metabolism', *Trends in Plant Science*, vol. 24, no. 1, pp. 83–98, Jan. 2019, doi: 10.1016/j.tplants.2018.09.006.

[173]C. De Ollas *et al.*, 'Facing climate change: Biotechnology of iconic mediterranean woody crops', *Frontiers in Plant Science*, vol. 10, 2019, doi: 10.3389/fpls.2019.00427.



OBJECTIVES

The current landscape surrounding cannabis for medical purposes is, nowadays, mainly focused on the general discussion around its regulation and the ways for its further implementation. Certain countries, such as the highlighted ones in the introduction, are positioned at the forefront of this policy, as they already have authorised its integration into their public healthcare systems through legislative acts and amendments. Nevertheless, most countries are still reluctant to even begin a debate on its putative feasibility, thereby remaining locked out of its implementation and considering cannabis as a narcotic drug. However, it is undeniable that a soaring tendency does indeed exist, which has led to the opening of public discussion even in those hard-line nations. Since this is the present scenario, it cannot yet be ascertained which precise pharmaceutical manufacturing standards cannabis flower will follow for its authentication as an API, it nonetheless highlights the need for developing analytical methodologies for the quality control compliance that adequate to this natural product. This is why this PhD thesis project was based on the premise of studying analytical methodologies that will facilitate and improve this future quality control procedures implementations. Bearing this in mind, the two introduced analytical approaches were explored for pursuing two general objectives.

The first objective was defined as the need to seek an analytical tool that will guarantee the appropriate growth and development of cannabis plants in those countries with tropical climates that have already authorised the cultivation of cannabis for medicinal purposes, given that unfavourable climatic conditions, such as their excessively high temperature and humidity, can negatively affect the plant's progress and productivity. Since this objective was fundamentally based on the impact of the interaction of the plant with its surroundings, it was decided to pursue this goal by means of metabolomics, as this is an ideal approach to study the effect that the environment triggers on the organism. This first main objective was divided into two sub-objectives:

1. **Exploratory optimisation of a LC-HRMS based analytical method for untargeted metabolomic screening of Cannabis Sativa L. through Data Mining.** This sub-objective was rooted in the need to streamline an optimal analytical method for the appropriate metabolomic screening by untargeted analysis of the cannabis plant, in order to obtain the maximum metabolic coverage and minimum metabolic information loss in the entire process involved from the plant cultivation to the corresponding chromatographic data-analysis.

2. **Determination of predictive constitutive metabolic markers for medical cannabis cultivar adaptability to tropical climates through LC-HRMS based untargeted metabolomics.** A plant possesses certain metabolic mechanisms that enable it to adapt to a stressful environment, given that, unlike animals, it is impossible for them to move from their location. Considering that, this sub-objective was set to search for predictive constitutive metabolic markers that determine the adaptability of cannabis cultivars to tropical climatic conditions, as various countries that have authorised its cultivation exhibit high humidity and temperatures. This sub-objective was directly related to the first one, as the previously developed methodology was adopted to pursue this goal.

The second main objective is based on the search for an analytical tool that facilitates the traceability and quality control process in cannabis cultivation facilities, being this tool non-invasive and easy to implement, so as the corresponding analysis and subsequent conclusion deduction are all carried out in the cultivation sites. Considering these features, NIR hyperspectral imaging was considered to be the most suitable tool to achieve the defined target:

3. **NIR-Hyperspectral imaging and machine learning for non-invasive chemotype classification in Cannabis sativa L.** This objective was set to assess the feasibility of NIR-HSI, for non-invasive chemotype classification in cannabis. Thus, hyperspectral images were coupled with different machine learning approaches, with which spatial patterns of dominant chemotypes would be visually differentiated in a plant individual through a colour map for easy and practical comprehension.

CHAPTER 1:

EXPLORATORY OPTIMISATION OF A LC-HRMS BASED ANALYTICAL METHOD FOR UNTARGETED METABOLOMIC SCREENING OF *CANNABIS SATIVA L*. THROUGH DATA MINING

Analytica Chimica Acta, vol. 1279, no. 341848 (2023).

DOI: 10.1016/j.aca.2023.341848

San Nicolas M., Villate A., Olivares M., Etxeberria N., Zuloaga O., Aizpurua-Olaizola O., Usobiaga A.

Abstract

Background: Recent increase in public acceptance of cannabis as a natural medical alternative for certain neurological pathologies has led to its approval in different regions of the world. However, due to its previous illegal background, little research has been conducted around its biochemical insights. Therefore, in the current framework, metabolomics may be a suitable approach for deepening the knowledge around this plant species. Nevertheless, experimental methods in metabolomics must be carefully handled, as slight modifications can lead to metabolomic coverage loss. Hence, the main objective of this work was to optimise an analytical method for appropriate untargeted metabolomic screening of cannabis.

Results: We present an empirically optimised experimental procedure through which the broadest metabolomic coverage was obtained, in which extraction solvents for metabolite isolation, chromatographic columns for LC-qOrbitrap analysis and plant-representative biological tissues were compared. By exploratory means, it was determined that the solvent combination composed of $CHCl_3$:H_2O:CH_3OH (2:1:1, v/v) provided the highest number of features from diverse chemical classes, as it was a two-phase extractant. In addition, a reverse phase 2.6 µm C18 100 Å (150x3 mm) chromatographic column was determined as the appropriate choice for adequate separation and further detection of the diverse metabolite classes. Apart from that, overall chromatographic peak quality provided by each column was observed and the need for batch correction methods through quality control (QC) samples was confirmed. At last, leaf and flower tissues resulted to provide complementary metabolic information of the plant, to the detriment of stem tissue, which resulted to be negligible.

Significance: It was concluded that the optimised experimental procedure could significantly ease the path for future research works related to cannabis metabolomics by LC-HRMS means, as the work was based on previous plant metabolomics literature. Furthermore, it is crucial to highlight that an optimal analytical method can vary depending on the main objective of the research, as changes in the experimental factors can lead to different outcomes, regardless of whether the results are better or worse.

Keywords: plant metabolomics, experimental factors, *Cannabis sativa L.*, LC-HRMS, metabolomic coverage, data mining

1. Introduction

Up to current times, *Cannabis sativa L.* has been an unauthorised plant species in most countries of the world [1]. As a consequence, little research on its insights has been carried out in the past. However, in the USA, for instance, 28 states have already passed the Medical Marijuana Laws (MML) [2], and even though state and Federal laws currently appear to be in conflict, this outstands an increasing tendency of cannabis application for medical purposes, giving a slight notion of how widespread its use could be in the future. Due to this fact, literature around the inner biochemistry of cannabis is slowly expanding, since this plant species warrants further investigation, on behalf of further research in medicine of natural provenance. Thus, early documented research works in cannabis focused in the characterisation of its bioactive compounds, cannabinoids and terpenes [3]–[6], which seem to have a promising future in palliative medicine [7]–[11], nonetheless, as these compounds are naturally biosynthesised in the cannabis plant, current studies are also trying to document the distinctive features of its inner metabolic pathways. This way, some primary metabolites as amino acids, carbohydrates or organic acids have already been studied for knowledge deepening into the plant's behavioural response mechanisms or chemovars differentiation, whereas some secondary metabolites as alkaloids, plant hormones or phytosteroids have been acknowledged for participating in signalling mechanisms [12]–[21]. Nevertheless, established analytical procedures for the ensuring of quality control with full transferability of results are still lacking.

In this thread, "omics"-related studies occupy one of the most noted positions among current analytical trends, as the deconstruction of molecular mechanisms which make up a biological organism can help to totally comprehend its inner functioning [22]. Moreover, in the metabolome, the last downstream rung of the referred "ome"-s (genome, transcriptome, proteome and metabolome) [23], the fingerprint of both internal and external perturbations can be tracked, making metabolomics a very helpful tool to understand the minutiae of the environment that surrounds a living being [24]. Due to this fact, metabolomics is broadly applied for searching biomarkers in living organisms [25]. Nevertheless, in the plant kingdom, specially, the metabolic diversity is vast, as beyond 200,000 different metabolites can compose the metabolome of a plant individual [26]. Therefore, innovative biotechnological approaches are necessary, for instance, in projects focusing cultivations' adaptability to uneasy surroundings [27], and the application of metabolomics could be a major upgrade [28]. Nevertheless, unless proper analytical methods are applied for biomarkers search, metabolites of interest might be missed or misidentified.

In metabolomics, the dominating analytical platform is mass spectrometry, commonly coupled to a chromatographic system [29]. Among such separation techniques, liquid chromatography (LC) is the most widely used one, as it is a versatile tool for both polar and non-polar metabolite analysis, while gas chromatography (GC) is used for the analysis of volatile metabolites and primary metabolites after derivatisation [29]. In regard to mass spectrometry detection, high- and low-resolution platforms can be distinguished. Broadly speaking, low-resolution platforms (e.g., tandem mass spectrometry, MS/MS) are usually addressed for quantitative target metabolomics, addressing metabolites belonging to certain metabolic pathways, while high-resolution mass-spectrometry (e.g., Orbitrap) is used for unknown metabolite identification in non-targeted metabolomics, providing a broader scope concerning coverage of metabolic pathways [30]. Hence, LC coupled to High Resolution Mass Spectrometry (LC-HRMS) offers versatility and high resolution for an accurate description of the metabolome [31]. Nevertheless,

this approach is never truly unbiased, as different factors or parameters must be defined prior to analysis, such as stationary phases or ionisation modes [30]. Consequently, properly developed analytical methods should be applied when performing metabolomics research.

A properly developed untargeted analytical method should primarily offer a broad metabolomic coverage. As stated, different factors of the whole experimental procedure must be optimised, in order to build a robust analytical method to be applied in future works related to this field. In this regard, works related to methodological optimisation actually exist in the field of metabolomics. For instance, there are examples in the literature focussed on the optimal selection of extraction solvents for the determination of as many metabolites as possible in the exploratory analysis of human fluids [32] or the standardisation of the derivatisation step of volatile metabolites in biological samples [33]. Indeed, this last optimisation approach was later applied in plant metabolic profiling studies [34]. Furthermore, standardised protocols for plant metabolomics have been previously developed based on the objective of covering as many metabolites as possible [35]–[38].

Among the experimental factors affecting the metabolite covering scope in cannabis, solvents for efficient metabolite extraction should be considered. Moreover, if the analytical platform of choice is LC-HRMS, the chromatographic stationary phase should also be taken into account, as efficient separation between compounds would be necessary. In addition, it is stated that plants possess a complex metabolome, thus, exploring the metabolic information in different biological tissues would also be a requirement, as in metabolomic studies organism-representative information should be retrieved.

Bearing all that in mind, multivariate data mining techniques, such as Principal Component Analysis (PCA) and Hierarchical Clustering Analysis (HCA), could be the suitable path to follow for corresponding experimental factor optimisation in the metabolome analysis of cannabis. Both techniques are unsupervised techniques, meaning the variance within the actual data of detected metabolites in the analysed samples is explored, excluding a prediction outcome [39]. Concretely, using those data mining techniques, trends explaining the variance of large datasets, which might not be noticed at first glance due to their dimensionality, can be spotted and relevant information extracted [40]. These are Thus, for factor optimisation, the applied selection criterion was based on the explained variance by the subjects under study. In this work, experimental factors such as extraction solvents, chromatographic columns and biological tissues were studied for appropriate untargeted metabolomic screening of *Cannabis Sativa L.*

2. Materials and methods

2.1. Reagents and solvents

Solvents for chromatographic analysis, acetonitrile (ACN), methanol (MeOH) and water (H_2O) of HPLC grade, were acquired in Panreac/AppliChem (ITW Reagents, S.R.L., Italy). Formic acid (FA) for LC-MS was purchased from Fischer Chemicals (Thermo Fisher Scientific Inc., USA) and ammonium acetate (NH_4Ac, for molecular biology ≥ 98 %) was acquired in Sigma Aldrich (Merck Group, USA). Calibration of the qOrbitrap was performed with Pierce™ LTQ Velos ESI Positive Ion Calibration Solution and Pierce™ ESI Negative Ion Calibration Solution by Thermo Scientific (Thermo Fisher Scientific Inc., USA).

For sample extraction, MeOH (anhydrous), trichloromethane ($CHCl_3$, synthesis grade) and methyl tert-butyl ether (MTBE, HPLC grade) were obtained in Macron Fine (Avantor, Inc., USA), Scharlau (Scharlab S.L., Spain) and Panreac/AppliChem (ITW Reagents S.R.L., Italy), respectively. Water was filtered to a resistivity of 18.2 MΩ.cm @ 25 °C and a total organic content (TOC) less than 3 $ng \cdot mL^{-1}$ using a Q-POD water dispenser and a Millipak express 40 (0.22 µm filter) by Merck (Merck Group, USA).

2.2. Plant samples

The samples were harvested from cannabis plants cultivated in the facilities of Sovereign Fields (Sovereign Fields S.L., Spain). The plants were cultivated indoor, in an isolated 24 m^2 (6 m x 4 m) growth room, built inside the greenhouse. Inside the room, 6 Ceramic Metal Halide (CMH) lamps (315 W/37000 lm) by Lumatek (Lumatek Ltd., UK) were installed in the roof, at 2.10 m height, uniformly distributed through the room, coupled to Adjust-A-Wings large enforcer reflectors.

Two similar cultivations were run for the optimisation procedure. In the first one, just one plant was harvested for the first two experiments described in section 2.4, while in the second one, 10 plants were harvested for the third experiment described in section 2.4. A chemotype III cannabis cultivar was cultivated in both ($C_{total\ THC}$ / $C_{total\ CBD}$ < 0.1) [41].

Plants were grown in 11-L black pots containing a soil/hummus/nutrient mixture. Specifically, the mixture was composed of 80 % of Light mix soil of Biobizz (Biobizz Worldwide S.L., Spain), 20 % of hummus, and 10 g/L of farmer mix nutrient solution by Lurpe (Lurpe Natural Solutions, Spain), which is composed of bat guano, bone meal, kelp meal, Azomite®, organic alfalfa, insect frass, blood meal, dolomite, langbeinite humic and fulvic acids, and a complex blend of rhizobacteria and Trichoderma. The total cultivation time in both cases was 12 weeks, being the first 4 weeks the vegetative stage during which the plants grew, and the next 8 weeks the flowering stage. The vegetative and flowering stages were defined by the photoperiod regime. In the first 4 weeks of growth, the photoperiod was 18 h light/6 h dark, and in the flowering stage, 12 h light/12 h dark. Moreover, the lamps were set to work at 50 % and 80 % of the total lamp intensity during the vegetative and flowering stages, respectively. During the cultivation time, the room temperature was kept between 22-25 °C during the daytime, and it did not decrease below 18 °C at night. The relative humidity of the room was controlled at 60 % (± 5 %).

Finally, when the cultivation time elapsed, the samples were harvested and instantly frozen in liquid nitrogen, following the quenching method. In the first cultivation, a flower sample was taken from a single plant, while in the second cultivation, leaf, flower, and stem samples were collected from each individual (n=10). The collected samples had to be representative of the whole plant individual, as each one of them was a biological replicate. So, the following sampling design was defined:

- Leaf samples: A pool of a total 12 leaves from each plant individual were collected, 6 of them from the upper half part of the plant, and the 6 from the lower half, all of them uniformly distributed across the plant.
- Flower samples: Samples from each biological replicate were composed of 5 different flowers. The first flower sample was taken from the apical end of the plant, in the upper point of it. The second one was collected 30 cm lower than this point, next to the stem. From this point, the two nearest ramifications, which were opposite to each other, were followed until the end-point of them, and the flowers situated there were collected.

Finally, the lowest flower situated next to the stem was collected. The collected 5 flowers were pooled to a representative flower sample of each plant individual.

- Stem sample: The lowest point of the plant stem was set as the reference point. 5 cm above the reference point, the stem was horizontally cut, cutting down the whole plant individual. From this point, a 10 cm long stem piece was taken as a sample. The possible ramifications on the stem sample were cut down, thus, only remaining the principal stem.

Afterwards, the samples were kept at -80 °C temperature prior to analysis.

2.3. General workflow

Figure 1 shows the general workflow followed to achieve optimal parameters of an untargeted analytical method for the determination of as many metabolites as possible in *Cannabis Sativa L.*

The optimisation procedure was divided in three main steps.

The first step was the assessment of the extraction solvent for representative isolation of the metabolites from the plant matrix. In this context, 5 different solvents (i. e., methanol-water, methanol-water acidified with formic acid, methanol-chloroform, chloroform-methanol-water, methyl tert-butyl ether-methanol-water) and literature-based extraction protocols were compared on the same flower sample, using the same analysis conditions such as chromatographic separation and detection setup. Since most of the literature works regarding plant metabolomics use reverse phase liquid chromatography conditions, the extracts were analysed using a reverse phase column (i. e., 2.6 µm C18 100 Å (150x3 mm) column). Once the extraction protocol was set, the second step was the evaluation of three chromatographic columns (i.e., 2.6 µm C18 100 Å (150x3 mm), 2.6 µm PS-C18100 Å (150x3 mm) and BEH 1.7 µm Amide Hydrophilic Interaction Liquid Chromatography (100 x 2.1 mm) columns). In the last step, a small-scale cultivation experiment was designed for the metabolomic-like study of 10 plant-clones, in which leaf, flower and stem tissues of the plant were analysed using the conditions set in the previous experiments.

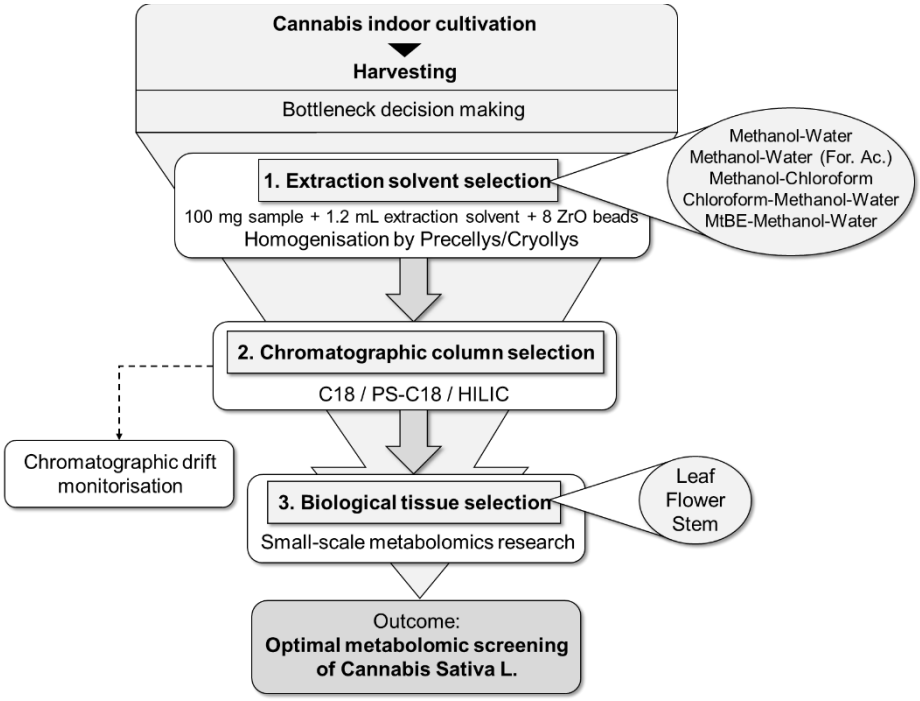


Figure 1: General experimental workflow overview

2.4. Experimental procedure

Prior to any extraction the frozen samples were ground by Spex SamplePrep Freezer Mill 6770 (Thermo Fisher Scientific Inc. USA), using liquid nitrogen, which worked under the following conditions: 1 minute of sample precooling followed by 1 minute grinding cycle at 8 cycles per second (cps). The result was a homogeneous-powder-frozen sample.

As previously mentioned, in the first optimisation step, 5 different literature-based metabolite extraction solvents (see *Table 1*) were compared to choose the extraction solvent or solvent combination that provides the broadest metabolomic coverage of cannabis. The extractions were carried out in a Precellys/Cryollys homogeniser system (Bertin technologies, France). 100 mg ground homogeneous flower sample was weighed, 8 ZrO beads (1.4 mm diameter) were added to each vial and 1.2 mL of the tested solvent were added according to the protocols specified in *Table 1*. All the assays were carried out in triplicate.

Table 1: Studied extraction-solvent combinations for suspect screening of Cannabis

Extraction		Solvents (%, v/v)				Reference
		H_2O	MeOH	$CHCl_3$	MTBE	
1	MW (Single-phase)	20	80	-	-	[42]
2	MWFA (Single-phase)	20 (0.1 % FA)*	80	-	-	[35]
3	MC (Single-phase)	-	80	20	-	[30]
4	CMW (Two-phases)	25	25	50	-	[43]
5	MtMW (Two-phases)	25	-	25	50	[44]

*The extraction procedure Nº 2 is equal to the Nº 1, but acidified at 0.1 % (v/v) with FA. Acronyms: M: Methanol; W: Water; FA: Formic Acid; C: Chloroform; Mt: Methyl tert-Butyl Ether.

The extractions were performed at a 6400 rpm rate, with 3 cycles of 60 seconds run and 15 seconds rest between each run (6400, 3x60, 15). The temperature was kept at 4 °C with liquid nitrogen during the extraction time. In study cases Nº 1, Nº 2 and Nº 3 (see *Table 1*), the respective mixture of solvents resulted in total miscibility at the specified contents. On the contrary, in study cases Nº 4 and Nº 5 (see *Table 1*), the contents of their main non-polar solvents ($CHCl_3$ and MTBE, respectively) caused their immiscibility with the H_2O:MeOH mixture. As a consequence, phase separation occurred in the later cases, where 50% (v/v) of the total extractant consisted of the non-polar phase ($CHCl_3$ in case Nº 4 and MTBE in case Nº 5) and the other 50 % (v/v) consisted of the polar phase (H_2O:MeOH, 1:1 v/v). In the single-phase extractant cases (cases Nº 1, Nº 2 and Nº 3), the respective solvents were added in order of polarity, from the least polar to the most polar. In the case of the two-phase extractions (i.e., Nº 4 and Nº 5, see *Table 1*), the procedure was performed in two steps. First, the sample was treated with 300 µL MeOH and 300 µL H_2O (the polar phase), and after one extraction run, 600 µL of the corresponding main non-polar solvent ($CHCl_3$ or MTBE) were subsequently added for sequential extraction of the sample.

Once the extraction was completed, the vials were centrifuged in the Allegra X-30R centrifuge (Beckman Coulter, USA) for 5 minutes, at 21,000 g force and 4 °C. The supernatants were decanted to another 2 mL vial for a second centrifugation run (15 minutes at 21,000 g and 4 °C). At last, prior to analysis, the polar and non-polar extracts were separated in the case of two-phase extractions, and they were transferred to chromatographic vials and diluted at a 1:20 ratio in MeOH, as well as the single-phase extracts. All the extracts in these assays were analysed by means of LC-qOrbitrap using a reverse phase Kinetex 2.6 µm C18 100 Å (150x3 mm) with AJ0-8782 C-18 pre-column (Phenomenex, USA), as detailed in section 2.5.

Having set the extraction procedure, the appropriate chromatographic column was selected through exploratory analysis of a flower sample. The compared columns were: i) the previously used reverse phase Kinetex 2.6 µm C18 100 Å (150x3 mm) with AJ0-8782 C-18 pre-column (Phenomenex, USA) for enhanced hydrophobic retention, ii) an aqueous stable with positive surface charged reverse phase Kinetex 2.6 µm PS-C18 100 Å (150x3 mm) with AJ0-8950 PS-C18 pre-column (Phenomenex, USA) for improved hydrophilic affinity, and, iii) a normal phase Acquity UPLC BEH 1.7 µm Amide Hydrophilic Interaction Liquid Chromatography (HILIC) 130 Å (100 x 2.1 mm) with Acquity UPLC BEH 1.7 µm Amide Vanguard 130 Å (2.1x5 mm) pre-column (Waters, USA) for enhanced separation and selectivity of polar compounds.

Apart from that, the chromatographic drift and signal fluctuation of the most suitable column sequence were explored across the analysis sequence. For doing so, three replicates of the flower sample were randomly analysed 10 times each, and a pooled sample was periodically analysed between them, serving as a quality control (QC) reference. Variances of the samples were explored throughout the analysis sequence, comparing the results obtained with QC correction and without correction.

In the third optimisation step, as the extractant and the chromatographic column are set up, the small-scale metabolomic study of the 10 plant-clones from the second cultivation (see 2.2. Plant samples section) was conducted for representative biological tissue exploration, comparing leaf, stem and flower samples.

2.5. **Chromatographic methods and general MS settings**

Untargeted analysis in the plant extracts was performed by Thermo Scientific Dionex Ultimate 3000 liquid chromatograph coupled to a Thermo Scientific Q Exactive Focus quadrupole-Orbitrap mass spectrometer (UHPLC-q-Orbitrap), equipped with a heated electrospray ionisation source (HESI, Thermo-Fisher Scientific, CA, USA). Analysis were performed with 5 µL injection from each extract in both ionisation modes, positive and negative, with the automatic injector (at 5 °C). To avoid any carryover, the injection needle was cleaned before and after every injection with 50 µL of H_2O:MeOH (90:10, v/v).

Regardless of the column used, chromatographic analysis consisted of two mobile phases, water (solvent A) as polar solvent and ACN (solvent B) as non-polar solvent, both of them containing either 0.1 % (v/v) of FA or 10 mM NH_4Ac, for respective measurements in positive or negative ionisation.

In the case of the chromatographic separation using the C18 column, the eluent flow rate was kept at 0.3 mL/min. During the analysis time, the solvent gradient started at 95 % A (held for 1 min), then linearly decreased to 5 % A until minute 16 (held for 10 min) and finally returned to the initial conditions in 5 min, where they were maintained for 1 min. Regarding the separation conditions using the PS-C18 column, the eluent flow rate was also 0.3 mL/min and the gradient was stablished as follows. The mobile phase gradient started at 97 % A, which was held for the first 2 min, then decreased to 85 % A until minute 5, followed by a change to 5 % A until minute 11 (held for 14 min), before returning to initial conditions in 4 min (held for 1 min). At last, for the separations using HILIC column, both mobile phase flow and composition were changed. The flow gradient started at 0.2 mL/min (held for 3 min), then it increased to 0.3 mL/min (held for 21 min) and it was again decreased to 0.2 mL/min until minute 35, when the analysis run ended. Regarding the mobile phase composition gradient, the analysis method started at 3 % A for the first 3 min, then increased to 15 % A until minute 5, changing to 25 % A until minute 14 and a final increase to 60 % A until minute 17 (held for 3 min), before returning to initial conditions in 3 min (held for 12 min).

Regardless of the ionisation mode, the HESI ionisation source was operated under the following parameters: spray voltage of 3.20 kV, spray current of 0.50 µA, the capillary temperature was kept at 320 °C and S-lens RF level at 55.0, the sheath gas (nitrogen) flow rate at 48 arbitrary units (au), auxiliary gas flow rate at 11 (au), sweep gas flow rate at 2 (au), and auxiliary gas heater temperature at 310 °C. Every three days Pierce LTQ ESI Calibration Solutions were used for external calibration of the instrument.

Data acquisition in the high-resolution mass spectrometer (qOrbitrap) was done in Full MS-Data dependant MS^2 (Full-MS-dd-MS^2) discovery acquisition mode for every analysis of the optimisation process. Full-scan mass spectrum was collected in a scan range between 70 to 1,000 *m/z* with a resolution of 70,000 FWHM for an *m/z* of 200. Three additional scans were performed in dd-MS^2 mode with a resolution of 17,500 FWHM, an isolation window of 0.8 m/z, and applying an stepped normalized collision energy (nce) of 10 %, 35 % and 75 % in the collision cell, where the first mass fixed was 50.0 *m/z* and intensity of AGC target of $2.00 \cdot 10^5$. The software used was Xcalibur 4.0 (Thermo-Fisher-Scientific).

2.6. Data processing

Data processing was performed using Compound Discoverer 3.3 software by Thermo Fischer. A minimum signal-to-noise ratio (S/N) of 3 was stablished in order to consider a chromatographic peak feasible and subsequently integrate its corresponding under-peak area. In the detected chromatographic features, MS1 was used as precursor for compound detection, with a mass tolerance of 5 ppm after performing 5 scans per feasible peak. Compounds were also grouped across samples with a mass tolerance of 5 ppm and a retention time (RT) tolerance of 0.2 minutes based on the preferred ions $[M+H]^{+1}$ and $[M-H]^{-1}$ for positive ionisation and negative ionisation, respectively. Moreover, a maximum threshold of 1 minute peak width at half peak height was stablished in every detected feature. Additionally, possible gaps were filled with 1.5 S/N ratio and also 5 ppm mass tolerance.

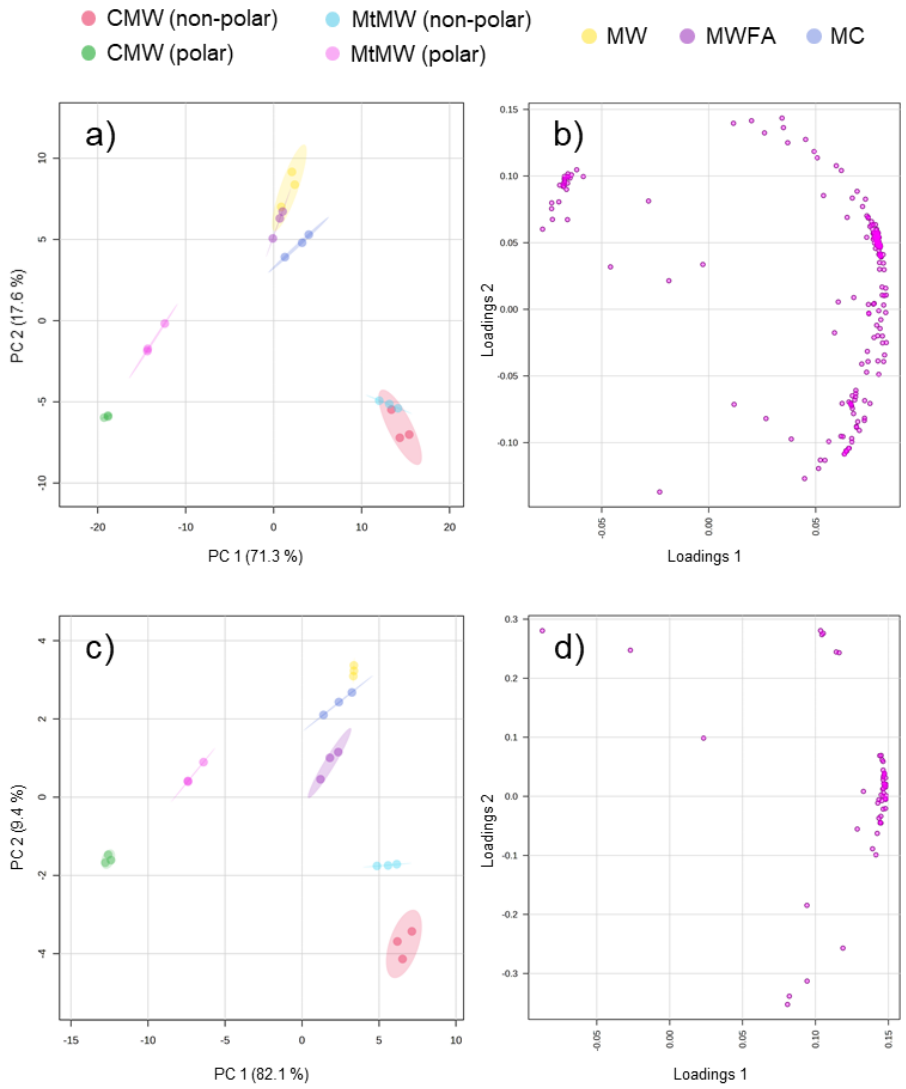
In the study cases of this work, only significant features were considered. To be deemed significant, the detected features had to pass the following filters: greater chromatographic peak area than 1,000,000 intensity counts, less variance than 30 % between corresponding replicates and, at least, partial putative spectral match with a minimum single candidate compound among provided spectral fragmentation libraries. These referred libraries were Cannabis Sativa L. endogen suspect list, retrieved from Plant Metabolic Network database [40], LipidMaps structure database [41] and Endogenous metabolites database of 4,400 compounds provided by Thermo Scientific. Furthermore, detected features were also filtered according to chromatographic peak quality, which was defined by four criteria: the jaggedness, the zig-zag index, the FWHM2base and the modality of the peaks [45]. Each of these metrics was measured at 5/10, being this value, the representation of its contribution compared to the other parameters contributors. Each measured metric contributed equally to the overall peak quality. Therefore, the peak quality filter threshold was set at 5/10, as only peaks surpassing this overall value would be accepted as significant for the study. For metabolite identification in the significant chromatographic features, mzLogic data analysis algorithm and Mass Frontier 7.0 spectral interpretation software were used, both from Thermo Fisher Scientific. In each of the significant chromatographic features, metabolite candidates from spectral libraries were ranked according to the similarity between experimental and estimated MS2 spectrum and the Fragment Ion Search (FISh) score was calculated in the five candidates with highest similarity score for structural elucidation and putative metabolite identification. The FISh score was calculated with a high accuracy mass tolerance of 2.5 mmu, low accuracy mass tolerance of 0.5 Da and a S/N threshold of 3. Feature annotation was performed according to the confidence levels defined in *Schrimpe-Rutledge et al.* [46]. Since no pure reference standard was used, the highest confidence level was 2 (putative identification, MS2 match). The features fulfilling all those previous defined criteria were later analysed through unsupervised multivariate data-analysis to determine the variances and trends between the analysed samples and annotated significant features. These data-analysis approaches were carried out using Metaboanalyst 5.0 [47]–[49].

3. Results and discussion

3.1. Extractant selection

To determine the most suitable extractant for optimal metabolomic screening of cannabis, the range of coverage of each extraction solvent combination was explored. Nonetheless, to avoid the assessment of hypothetical false positives, the processed data was filtered according to the constraints stated in section 2.6., so only chromatographic areas of significant features were evaluated through multivariate data analysis. The results are presented in Table S1 of the Supporting Material (SM), where a total of 171 features were detected in positive ionisation mode and 35 features in negative ionisation mode. The area values were auto scaled and transformed to logarithmic scale (Log_{10}), prior to analysing the data by Principal Component Analysis (PCA), and the trends of greater variance were determined though exploratory means (*see Figure 2*).

According to the PCA scores plot, PC1 explained 71.3 % and 82.1 % of the total variance in the positive and negative ionisation, respectively. Regarding PC1, single-phase extractants (MC, MW and MWFA) seemed to have no relevance at all since most of the variables are related to the antagonist phases (polar and non-polar) of the two-phase extractions (CMW and MtMW) that located at opposed endpoints of PC1. PC2, which explained 17.6 % and 9.4 % of the total variance in positive and negative ionisations, respectively, seemed to be related with the distinction of single-phase and two-phase extractants. According to these facts, antagonist phase extracts of the two-phase extractants presented greater significance over the total variance in the study, which suggested that a two-phase extractant would be more appropriate for representative metabolomic coverage of cannabis. Indeed, phase differentiation could be the reason for the enhancement of the extraction yield, at the expense of single-phase extractions, as both very polar and very non-polar metabolites could be quantitatively isolated from the plant matrix.



<u>Figure 2:</u> PCA scores and loadings of the filtered results in the extractant optimisation step a) Scores plot of positive ionisation results b) Loadings plot of positive ionisation results c) Scores plot of negative ionisation results d) Loadings plot of negative ionisation results

Going deeper into the results obtained using two-phase extractants, it could be noted that scores corresponding to the extracts of CMW appear to be further than the corresponding MtMW extracts from the null relevance point of the variable space according to PC1, which could indicate a greater stronger statistical influence for the CMW extraction. Nevertheless, at first glance, it was difficult to assert that there is actually a significant difference between both two-phase extractions. Thus, the relative extraction yield of the significant chromatographic features got using two-phase extractants was assessed (see *Figure 3*). According to these results, even though the difference among the yield of both extractants was barely significant, in every study case greater average peak area was obtained using CMW as extractant. As the composition of the polar phase was equal in both cases (50 % methanol and 50 % water, v/v), the difference resided in the corresponding non-polar solvent. Chloroform would present greater extraction yield for non-polar compounds than methyl tert-butyl ether, as chloroform possesses greater partition coefficient ($LogP_{CHCl3}$=2.3, $LogP_{C5H12O}$=0.9). Therefore, theoretically,

chloroform presents 2.55 times greater affinity towards non-polar compounds, resulting in an enhanced extraction yield for metabolites of this nature. On the other hand, greater polarity gradient between polar and non-polar phases could also enhance polar compound extraction capability in its corresponding phase, resulting in a greater extraction yield of these metabolites of this nature as well. Hence, based on the exploratory results, the extraction solvent combination composed of chloroform-methanol-water (50 %-25 %-25 %, v/v) was considered the most suitable for representative untargeted metabolomics screening of cannabis. It should also be noted that greater count of significant features was detected in positive ionisation analysis (171) than in negative ionisation analysis (35), thus giving the former case greater statistical weight in the exploratory optimisation process.

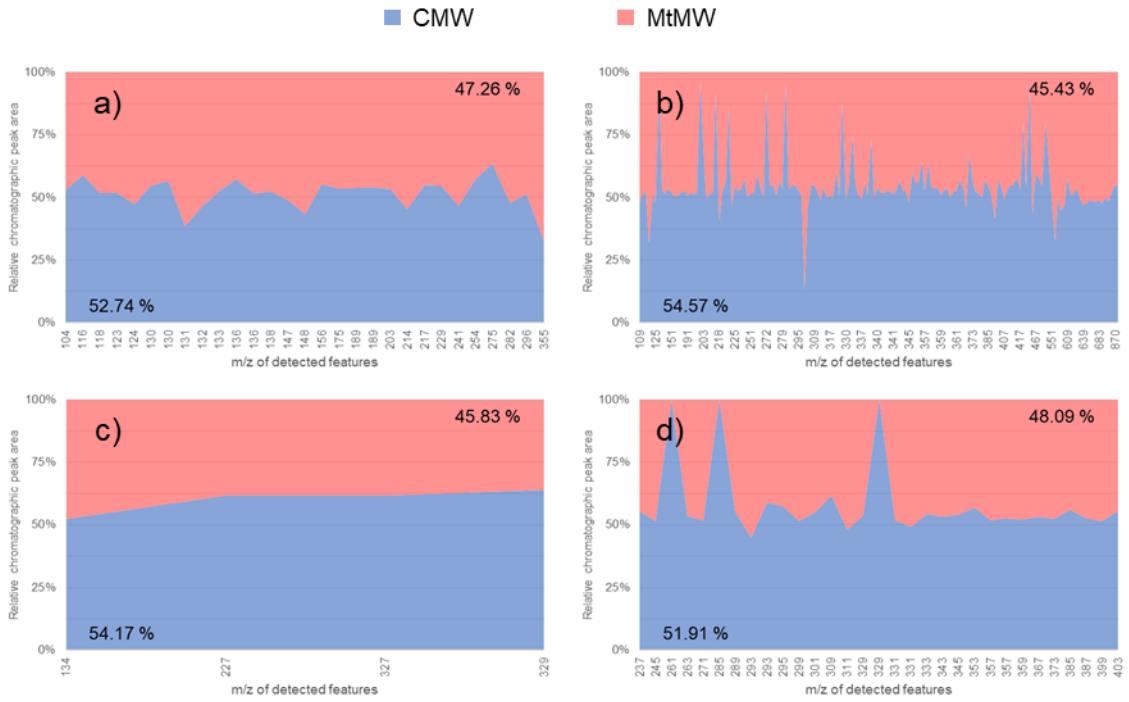
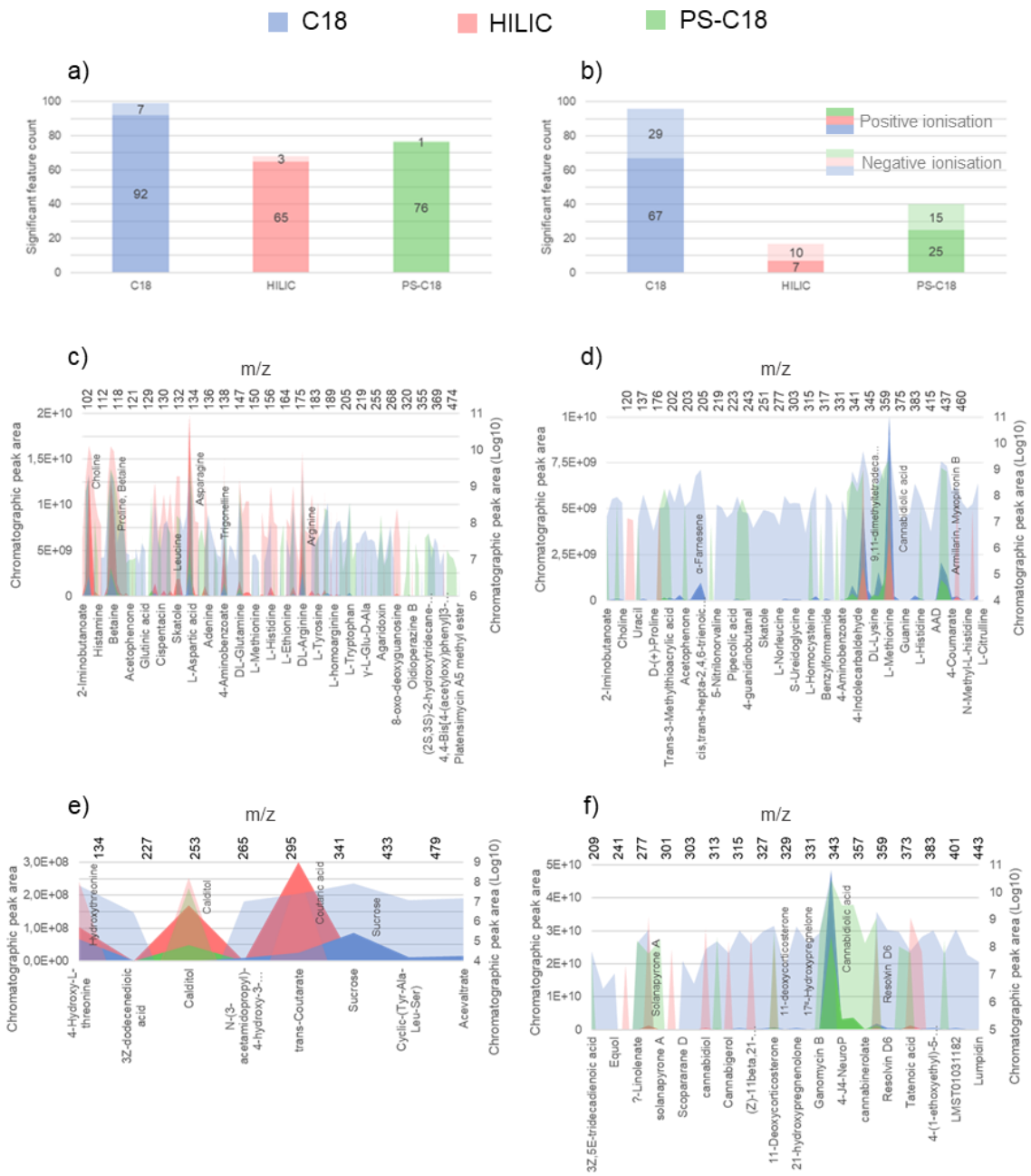


Figure 3: Relative coverage of the detected significant features of the polar and non-polar extracts of the CMW and MtMW extractants: a) Polar extracts through positive ionisation b) Non-polar extracts through positive ionisation c) Polar extracts through negative ionisation d) Non-polar extracts through negative ionisation

3.2. Chromatographic column selection

After performing the data processing and filtering of the results of the studied chromatographic conditions as described in section 2.6, obtained results for polar and non-polar extracts are resumed in Table S2 and Table S3 of SM, respectively. Detected feature counts and corresponding areas obtained with each of the studied chromatographic columns are displayed in *Figure 4*.



<u>Figure 4:</u> Number of significant feature count (SFC) detected by each of the chromatographic columns and their corresponding peak areas (CPA). a) SFC in polar phase, positive and negative ionisation modes b) SFC in non-polar phase, positive and negative ionisation modes c) CPA vs SFC in polar phase, positive ionisation mode d) CPA vs SFC in non-polar phase, positive ionisation mode e) CPA vs SFC in polar phase, negative ionisation mode f) CPA vs SFC in non-polar phase, negative ionisation mode.

The first fact to note was that, either in positive or negative ionisation modes, greater number of significant peaks were annotated when using the C18 chromatographic column. A total of 195 significant features were detected using the C18 column, whereas the HILIC and the PS-C18 columns yielded 85 and 117 significant features, respectively. Secondly, we observed that,

regardless of the chromatographic column used, a higher number of features were detected in the positive ionisation mode, as it happened in the previous optimisation step. At first sight, this fact acknowledges that analysis through positive ionisation mode, regardless of the analytical conditions applied, could provide broader information, as more chromatographic features could be annotated.

In addition to the number of features, the areas of the chromatographic peaks were also a factor to be considered. As expected, the HILIC column provided good separation and large peak areas for characteristic polar compounds, such as amino acids, among others. Concretely, choline (*m/z* 104.1071), proline (*m/z* 116.0707), betaine (*m/z* 118.0863), leucine (*m/z* 132.1018), asparagine (*m/z* 133.0607), trigonelline (*m/z* 138.0549) and glutamine (*m/z* 147.0763) and were detected with the largest peak areas in the positive ionisation mode acquisition, all of them putatively identified (identification confidence level 2), while in the results of negative ionisation mode hydroxy threonine (*m/z* 134.0460), calditol (*m/z* 253.0936) and coutaric acid (*m/z* 295.0457) stood out, identified as tentative structures (identification confidence level 3). On the other hand, the C18 column provided large chromatographic peak areas in the analysis of the non-polar extract, where cannabinoids highlighted, along with lipids and steroids. In this case, the largest peak areas belonged to α-farnesene (*m/z* 205.1948), 6-[(1E,3E,5E,7E,9E,11E)-9,11-dimethyltetradeca-1,3,5,7,9,11-hexaenyl]-5-ethyloxane-2,4-dione (*m/z* 341.2104), cannabidiolic acid (*m/z* 359.2210), armillarin (*m/z* 415.2109) and myxopyronin B (*m/z* 432.2373), putatively identified in positive ionisation (identification confidence level 2), while 11-Deoxycorticosterone (*m/z* 329.2123, level 3), 17α-hydroxypregnelone (*m/z* 331.2280, level 3) cannabidiolic acid (*m/z* 357.2071, level 2) and resolvin D6 (*m/z* 359.2230, level 2) highlighted in negative ionisation.

According to Figure 5, it was also determined that the PS-C18 column, could be deemed as the least relevant among the studied columns to be used in untargeted metabolomics. As its stationary phase relied on a midway polarity nature between the C18 and the HILIC column, it failed to provide more information than the already obtained with the other columns. In particular, the PS-C18 column does not allow the detection of non-polar compounds either in number or type of those already detected by the C18 column, while for the detection of polar compounds it is preferable to use the HILIC column.

Moreover, it is worth mentioning that most of the polar compounds detected in the polar extract using the HILIC were also detected in the measurements using C18 column. Of course, it should not be ignored the fact that the polarity of the reversed-phase column implied a lower retention affinity towards compounds of polar nature, so that both the areas of their corresponding chromatographic peaks and the difference in separation time between them were significantly lower with the C18 column; nevertheless, a higher number of compounds were detected. For instance, some amino acids such as L-methionine (m/z 150.0582), aminoadipic acid (m/z 162.0760) or L-tryptophan (m/z 205.0970), alkaloid derivatives such as nicotinamide (m/z 123.0553), 4-Indolecarbaldehyde (m/z 146.0599) or trans-3-indoleacrylic acid (m/z 188.0704), metabolites that play an important role in the inner regulation of the plants (all of them were identified at confidence level 2) were not detected in the polar extract by the HILIC column, instead they were annotated with the C18 column. That did not happen in the case of the analysis of the non-polar extract with the HILIC column, as only 7 and 10 features were detected in positive and negative ionisation modes, respectively.

Apart from the annotated feature number and their corresponding peak areas, their quality should also be a factor to consider. Further data-analysis of the results of untargeted metabolomic research would depend on the obtained peak areas, hence, in order to avoid undue deviations, detected peak should pass a minimum quality threshold, as poor chromatographic peaks may deviate results from their true values. In *Figure 6,* peak qualities of the detected features are displayed, and the average peak quality was calculated at 95 % confidence for each chromatographic column. The average peak qualities offered by each column were similar in the four study cases shown in the figure. In fact, the internal deviation within each column was greater than the difference between columns, thus, the quality of the detected features was fully comparable in the different study cases. The PS-C18 column offered the smallest deviation in the different study cases, nonetheless, it was previously observed that its contribution, according to the objective of the work, was the least among the three columns. Apart from that, the C18 column had greater statistical weight in the peak quality exploration, as it provided a higher number of significant features.

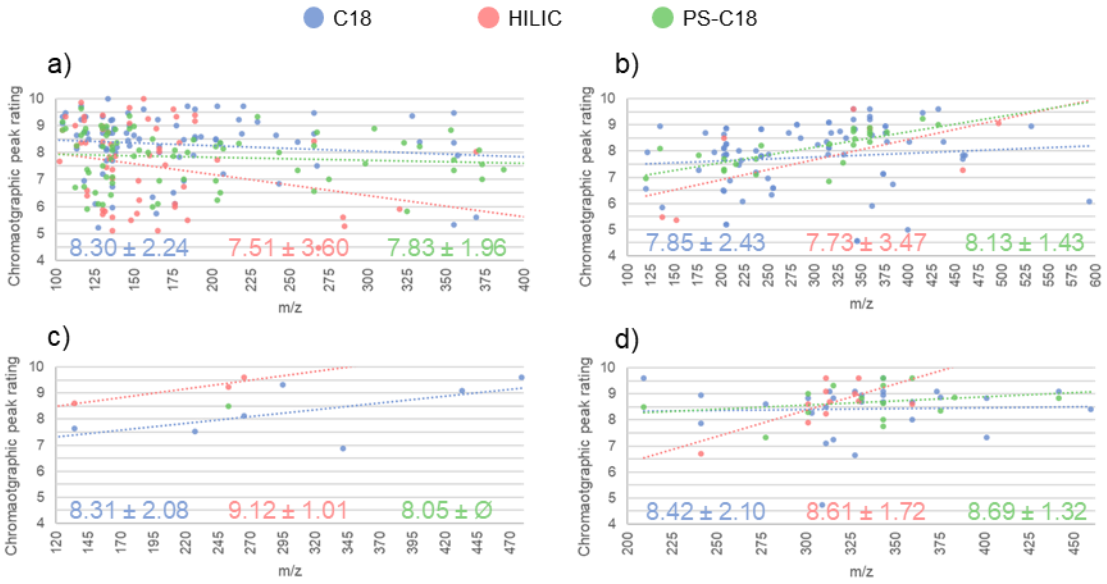


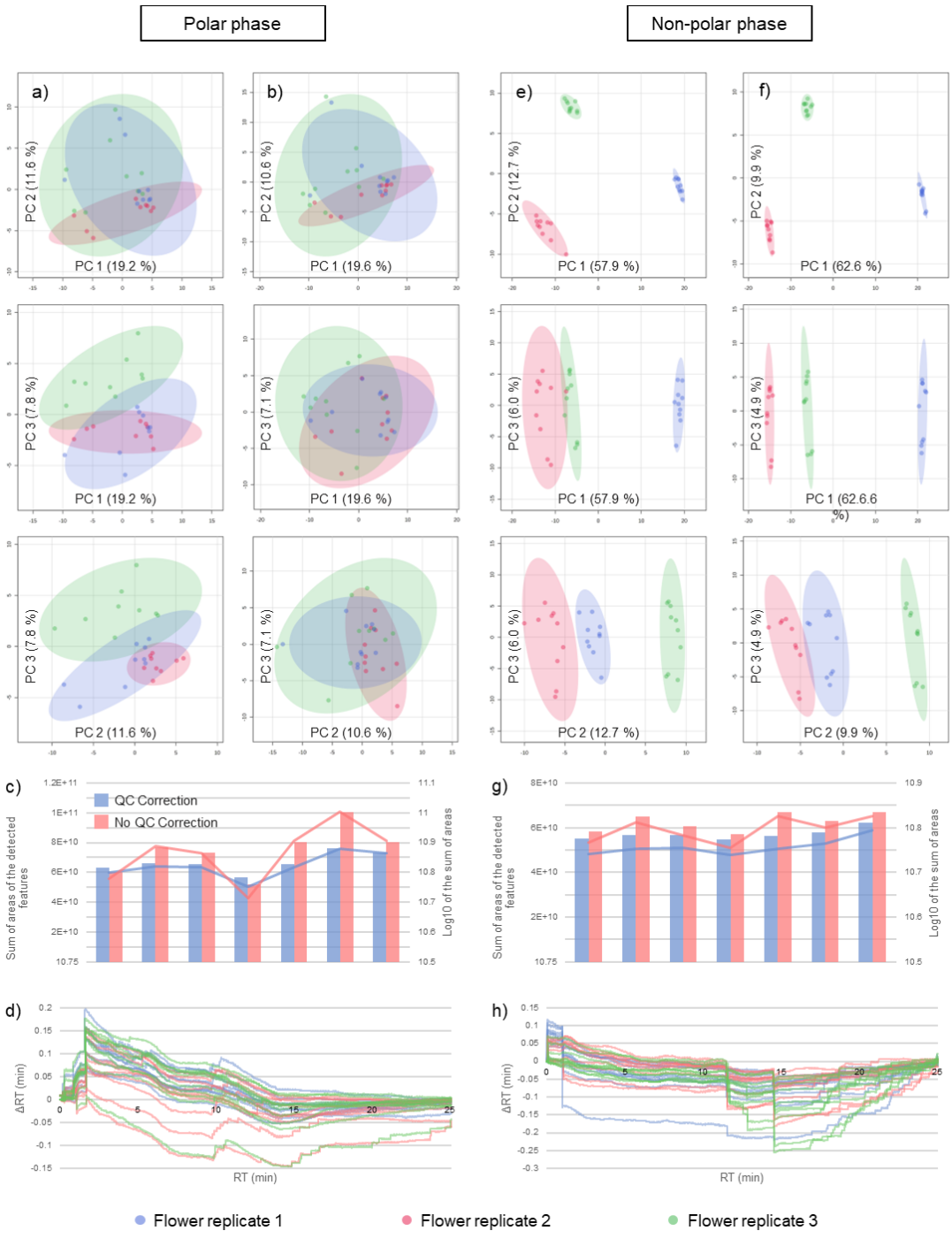
Figure 5: Peak qualities of the detected significant features and their average values in each of the studied cases, in a 95 % confidence interval a) polar phase, positive ionisation mode b) non-polar phase, positive ionisation mode c) polar phase, negative ionisation mode d) non-polar phase, negative ionisation mode

Hence, based on all those results and observations, analysis through the C18 column is suggested for the intended purpose, bearing in mind that its stationary phase will present greater retention affinity towards non-polar compounds. Therefore, it was concluded that C18 column would be the most suitable for a broader metabolomic coverage.

Moreover, it is known that in metabolomic studies, where large sample sets are analysed in a single run, chromatographic signal drift can occur along the sequence. This phenomenon can lead to misleading conclusions since the chromatographic signals may depend on the acquisition-time. Hence, to ensure accurate data acquisition, the chromatographic area drift was

assessed suing the C18 column to define the QC correction effect through the analysis time. The data of unfiltered Log_{10} transformed results can be seen in *Figure 6*. This unfiltered raw data is presented in Table S4 (Polar extract, QC corrected data), Table S5 (Polar extract, non-corrected data), Table S6 (Non-polar extract, QC corrected data) and Table S7 (Non-polar extract, non-corrected data) of SM. Since no filter was applied to the detected chromatographic features, the feature annotation in this section was limited to tentative structure of the metabolites (level 3), molecular formula match (level 4) or to just a unique chromatographic feature (level 5), according to the metabolite identification level specified in *Schrimpe-Rutledge et al.* [46].

At first glance, an apparent difference between QC-corrected and non-corrected results can be distinguished. On the one hand, the Compound Discoverer 3.3 software allowed the correction of the retention time (RT) shifts in the alignment of chromatographic peaks across the samples through ChromAlign algorithm [50], having the QC sample as a reference. As it can be seen in the RT shifts correction plots, in some of the cases RT shifts of up to 0.2 min were corrected, from where it could be deduced that the lack of RT shifts correction could also have a significant impact on the subsequent chromatographic peak area integration, thus, also affecting the area values and, therefore, the subsequent data-analysis. On the other hand, it could also be observed that the cumulative non-corrected signal of the QC sample exhibited significant fluctuations through the time span of the analysis sequence, which appeared to be independent of the data acquisition time, as they did not follow a defined trend. Signal correction through the SERRF (Systematical Error Removal using Random Forest) QC normalization method [51], which was the one implemented in the Compound Discoverer 3.3 data processing software, resulted in more constant area signal results in the chromatographic features through sequence time. Hence, these two correction methods led to lower intra-group variance of the data, as it can be seen in the corresponding PCA score plots. In the case of the polar extract, it was clear that correcting the data using QC samples allowed a clustering of replicates with reduced variance. Since the analysed samples were technical replicates of three biological replicates of the same flower sample, ideal results would be expected to present null variance between either of the replicates, so, based on the empirical results, the smaller the variance between replicates, the more reliable the obtained results would be. In the case of the non-polar extract, this was not so evident, as the variance between replicates did not differ so much between QC-corrected and non-corrected data, so it would be difficult to ascertain from just the PCA plots that there was an apparent difference between the two cases. Nevertheless, as stated before, the QC signal fluctuation plot revealed fluctuations in the summed signals of the QC sample in different analysis times. What it could be determined was that the non-corrected area signals of the QC sample followed a similar overall trend as the QC corrected signals through the analysis time span, but this could be a matter of chance in this exact case, as it did not happen the same in the case of the polar extract. Accordingly, the implementation of signal drift correction and RT shifts correction through QC samples for further research on untargeted cannabis metabolomics is strongly suggested, regardless of the specific research objectives, as the interpretation of non-corrected results could be biased and unreliable.



<u>Figure 6:</u> Chromatographic drift monitorisation results in randomly analysed flower sample's biological and technical replicates: a) PCA scores of the corrected polar extract results (displayed at 95 % confidence regions) b) PCA scores of the non-corrected polar extract results c) Fluctuation of the sum of areas of the features in the polar QC sample across the analysis sequence d) RT shift corrections in chromatographic features of the polar extracts through analysis time e) PCA scores of the corrected non-polar extract results f) PCA scores of the non-corrected non-polar extract results g) Fluctuation of the sum of areas of the features in the non-polar QC sample across the analysis sequence h) RT shift corrections in chromatographic features of the non-polar extracts through analysis time

3.3. Biological tissue selection

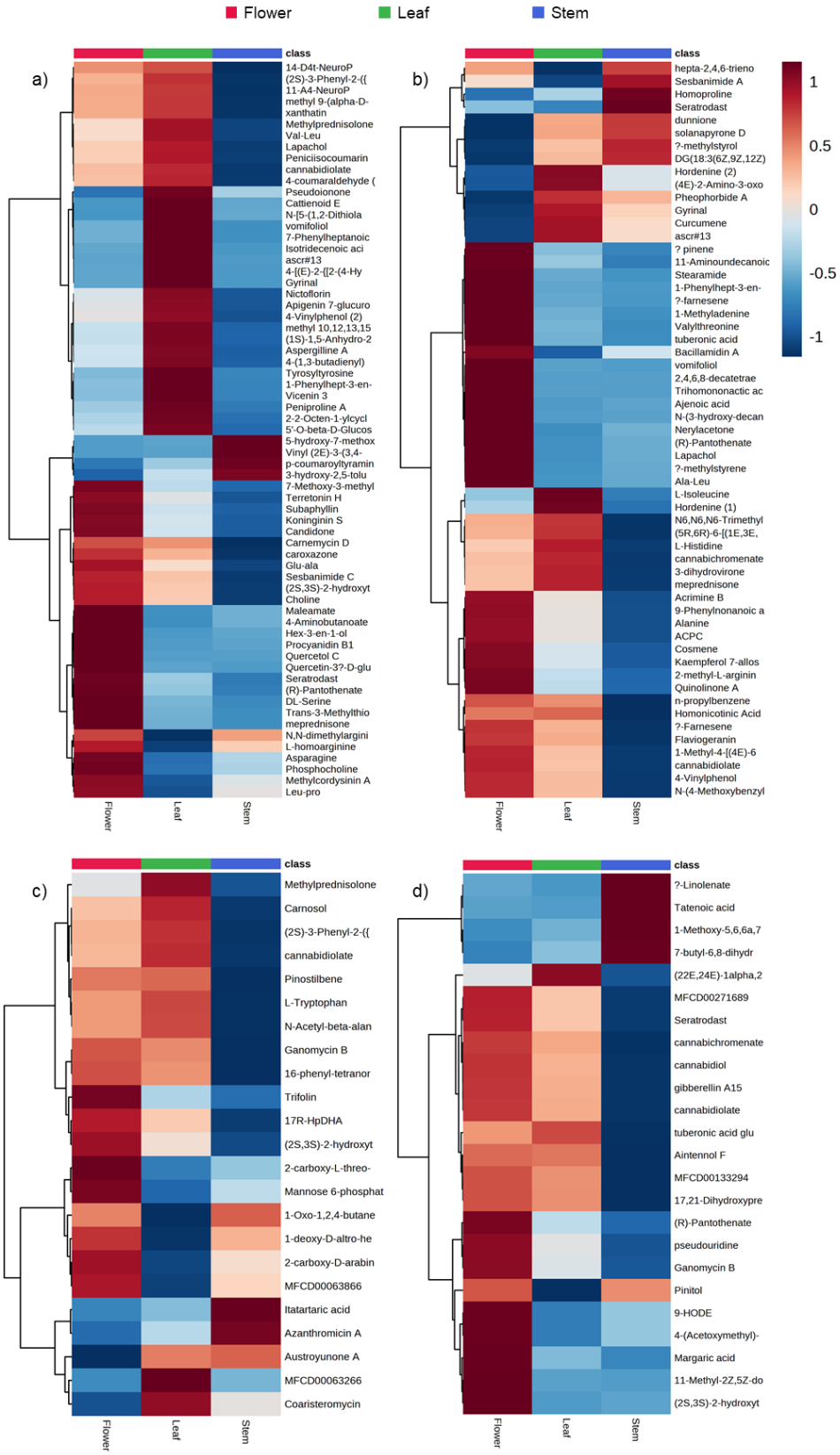
The analysis of leaf, stem and flower tissues from 10 plant clones of the same phenotype was carried out simulating a small-scale experiment for metabolomics study. Samples were extracted using the CMW solvent combination and they were subsequently analysed using the C18 chromatographic column. Periodic analysis of pooled QC samples was performed to enable later signal drift correction. A comparison between different sample classes (leaf, stem and flower) was performed based on unsupervised variance analysis of the filtered results, which can be found in Table S8 (Polar extract results) and Table S9 (Non-polar extract results) of SM. Therefore, a hierarchical cluster analysis (HCA) heatmap was generated with the detected significant features to determine the correlation between the samples and features. The data was auto scaled and transformed to logarithmic scale (Log_{10}), and Ward clustering method was applied using the Euclidean distance measure. The corresponding results, depicting the average peak area values of the detected features for both phases and ionisation modes of the 10 clones, are shown in *Figure 7*, as comparison between the three studied tissues. A total of 166 and 57 significant features were identified in positive ionisation mode analysis of the polar and non-polar extracts, respectively. In the case of negative ionisation mode results, 23 and 24 significant features were identified in the polar and non-polar extracts, respectively. Thereby, to determine significant differences among biological tissues, ANOVA test was performed between tissue-wise sample averages.

Based on the information shown in *Figure 7* it was concluded that most of the identified compounds highly correlate with either flower or leaf tissues, while, in contrast, few of the detected features were directly correlated with the stem tissue, as it is summed in *Table 2*.

Table 2: Significantly correlated features in the studied biological tissues

		Detected features			
Ionisation mode	Extract Phase	Flower	Leaf	Stem	Total
Positive	Polar	77	68	21	166
	Non-polar	34	15	8	57
Negative	Polar	10	9	4	23
	Non-polar	18	2	4	24

Among the detected significant features in the stem tissue, various compounds were identified as metabolites that play diverse roles in plant intern metabolism. For instance, p-coumaroyl tyramine, a derivative of p-coumaric acid, acts as a secondary metabolite involved in defence against stress and pathogens. It also participates in the synthesis of lignin, a characteristic compound that provides structural support to plant cell walls, which explains its abundance in the stem tissue [52]. Moreover, other highlighted metabolites observed in the stem happened to be lipids, such as linolenic acid or diacylglycerol (18:3(6Z,9Z,12Z)/18:3(9Z,12Z,15Z)/0:0). Linolenic acid (Ω-3 fatty acid) plays a key role in plant growth, being part of cell membranes or participating in the synthesis of other growth regulator metabolites such as jasmonic acid. On the other hand, diacyl glycerol contributes to lipid metabolism, playing diverse roles as signal transduction, energy metabolism, and cell-integrity. It is also a component of the hydrophobic cell membranes [53] [54].



<u>Figure 7:</u> HCA on average values of significant features in flower, leaf and stem samples a) polar extracts, positive ionisation (Top 65 features). Complete results of the features detected in the polar phase can be found in Figure S1 of SM b) non-polar extracts, positive ionisation c) polar extracts, negative ionisation d) non-polar extracts, negative ionisation

Nevertheless, in leaf and flower tissues a greater number of key metabolites were identified, which play equal and more significant metabolic roles compared to those identified in the stem. For instance, p-coumaraldehyde, p-coumaric acid and p-coumaroyloctopamine were detected in leaf tissue. p-coumaric acid is created from p-coumaraldehyde, which is an essential part of the phenylpropanoid pathway, the same pathway p-coumaroyl tyramine (detected in the stem) is involved. Coumarins are also a class of phytoalexins, which are defence compounds produced by plants to deal with stress or pathogen attacks [55]. Additionally, metabolites with plant-hormone role such as abscisic acid or some lactones (trihomononactic acid lactone and xanthatin) were also identified. These compounds also significantly impact plant development regulation and response to environmental stimuli [56]. Furthermore, a wide range of secondary metabolites were also annotated, such as alkaloids (hordenine and 2-methylindole), quinones (lapachol) or fatty acids derivatives (ascr#13), metabolites that possess diverse roles as antioxidant, antimicrobial, signalling, plant defence, growth regulation and interaction functions with the environment.

In the case of flower tissue, other essential metabolites were also identified. As in the case of the leaf tissue, some growth regulation plant hormones and defence-related metabolites were observed, such as gibberellins (gibberellin A15) and phytotoxins (vomifoliol), respectively. Regarding secondary metabolites, flavonoids (flaviogeranin, quercetol C, robinetinidol) and their derivatives (kaempferol-7-alloside, quercetin-3-glucoside, nictoflorin) were identified, as well as some phenolic compounds (salicylaldehyde and caffealdehyde), which, similar to their counterparts in leaf tissue, are involved in a wide range of metabolic roles.

Furthermore, several amino acids were identified in both leaf and flower tissues, compounds that play essential roles in protein biosynthesis and signalling. Histidine, isoleucine or N6,N6,N6-trimethyl-lysine could be noted in the leaf tissue, whereas, alanine, choline, proline, aspartic acid or arginine could be observed in the flower, among others. Moreover, some dipeptides were also detected, which can act as signalling actors or storage forms of amino acids, for instance, valine-leucine or alanine-leucine were annotated in the leaf, and leucine-proline or glutamine-alanine in the flower. Essential aromatic amino acids were also detected: tryptophan in leaf and tyrosine (alongside corresponding derivatives tyrosilalanine and tyrosiltyrosine) and N-(Carboxyacetyl)phenylalanine (a phenylalanine derivative) in flower. Tryptophan serves as a precursor to previously stated metabolite classes as alkaloids, phytoalexins and indole glucosinolates [57]; while tyrosine acts as a precursor to isoquinoline alkaloids and quinones [58]. In parallel, flavonoids, tannins, lignins, and many other metabolites originate from phenylalanine [52], as this amino acid is believed to be the precursor of compounds than can constitute up to a third part of the organic compounds present in some plant species [59], [60], therefore, it is crucial to identify them in the tissues under study. These three aromatic amino acids are downstream products of chlorismate, the final compound of the shikimate pathway [61], from which other metabolites, as some vitamins, are also derived. In line with this, it is worth mentioning that pantothenic acid, also known as vitamin B_5, was detected in the flower tissue, which is the acetyl CoA source, an indispensable component in various metabolic pathways [62].

Hence, based on the presence of specific metabolites in each biological tissue, it was determined that stem did not offer a significant contribution in terms of plant-representative metabolic information, whereas flowers and leaves seemed to be essential for the untargeted

metabolomics study of the cannabis plant (all the stated metabolites were putatively identified, at identification confidence level 2).

4. Conclusions

We found that an optimisation of the analytical method was necessary for untargeted studies to achieve meaningful results. For untargeted metabolomic screening of cannabis, we identified chloroform:methanol:water (50:25:25, v/v) as the most suitable extractant, as it offered greater metabolomic coverage by enabling efficient extraction and phase separation. We also determined that LC-qOrbitrap analysis using the Kinetex 2.6 µm C18 100 Å (150x3 mm) column was more appropriate for the defined purpose, as it offered a higher compound number detection capability in both polar and non-polar extracts, with representative metabolomic coverage, and providing correct chromatographic peak qualities for accurate metabolite annotation and further data analysis. Nonetheless, it was found that, even though the C18 column yielded appropriate results, the occurrence of chromatographic drift needed the implementation of batch correction methods during data processing of untargeted metabolomics studies, which are characterised by the analysis of a large number of samples in one analysis-run. Moreover, it was determined that the stem tissue lacks significance in terms of representative metabolic information in comparison to leaf and flower tissues. The conclusions defined in this experimental procedure optimisation path can be contrasted with the methods followed in other research works related to cannabis metabolomics. For instance, in *Li et al.* [18], a methanol-water (75:25, v/v) solvent combination was used for the extraction of the metabolites, method that is very similar to study case Nº 1 of the extraction solvent optimisation step of this work, and as it could be empirically observed, solvent combination of study case Nº 4 ($CHCl_3$:H_2O:CH_3OH, 2:1:1, v/v) could offer broader metabolomic coverage. Furthermore, in *Zheng et al.* [13] it was stated that identification of more polyphenols in cannabis leaves was missed in their targeted assay, as this would turn to be very useful. Fortunately, polyphenols as robitenidol, kaempferol-7-alloside, quercetin-3-glucoside or nictoflorin, among others, were putatively identified in cannabis flower tissue.

Thus, based on the observations in this work, it is suggested that future research on understanding inner functioning of the cannabis plant is conducted according to experimental designs based on empirical evidence, with the purpose of covering the widest possible range of metabolic information. However, this would not mean that the experimental conditions discarded in this optimisation process (i.e., the rest of the tested solvents for metabolite extraction, the studied chromatographic columns and the stem tissue of the plant) are negligible for other works, but they should be optimised according to the specific objective of each research study instead. Therefore, unsupervised multivariate data analysis approaches were used in this work, in order to compare the actual data based on a certain objective, which was to optimise an experimental procedure for an untargeted metabolomics screening analytical method that offered the greatest metabolomic coverage, while maintaining data quality intact.

In this line, it is important to note that each study is unique and may have specific considerations that were not discussed in this work. For instance, factors such as dimensionality of the research plays an important role according to functionality, accessibility or economic matters in experimental design. The sample size to be processed and analysed can be a key factor. In certain studies, a large number of samples may need to be analysed, making it economically and

temporally challenging to perform multiple extractions per sample. Thus, in such cases, a single-phase extraction procedure would be preferable. Similarly, constraints may prevent the analysis of extracts through two ionisation modes, leading to a loss of metabolomic coverage. Therefore, the success of a study resides in designing experiments that maximise resulting information within the available resources.

On the contrary, if hypothetical scenarios where a small sample quantity must be studied and the stated factors are feasible, experiments should be conducted using more than one chromatographic column, without discarding the excluded biological tissue, in order to gather the widest possible metabolic coverage. Ultimately, the design of appropriate experimental approaches should align with the specific research objectives, taking into account the available resources and leveraging empirical results to make informed decisions.

Supplementary Material

The Supplementary Material of Chapter 1 is available in the link or in the QR below:

Table S1: 1st Optimization step. Extraction solvent optimisation

Table S2: 2nd Optimization step. Chromatographic column optimisation (Polar phase)

Table S3: 2nd Optimization step. Chromatographic column optimisation (Non-Polar phase)

Table S4: Chromatographic drift and signal fluctuation study. QC corrected data (Polar phase, positive ionisation)

Table S5: Chromatographic drift and signal fluctuation study. Non-corrected data (Polar phase, positive ionisation)

Table S6: Chromatographic drift and signal fluctuation study. QC corrected data (Non-polar phase, positive ionisation)

Table S7: Chromatographic drift and signal fluctuation study. Non-corrected data (Non-polar phase, positive ionisation)

Table S8: 3rd Optimization step. Biological tissue optimisation (Polar phase)

Table S9: 3rd Optimization step. Biological tissue optimisation (Non-Polar phase)

Figure S1: 3rd Optimization step. Biological tissue optimization (Polar phase, positive ionization)_HCA with every detected compound

Bibliography

[1] D. C. Simiyu, J. H. Jang, and O. R. Lee, 'Understanding Cannabis sativa L.: Current Status of Propagation, Use, Legalization, and Haploid-Inducer-Mediated Genetic Engineering', *Plants (Basel)*, vol. 11, no. 9, p. 1236, May 2022, doi: 10.3390/plants11091236.

[2] H. Carliner, Q. L. Brown, A. L. Sarvet, and D. S. Hasin, 'Cannabis use, attitudes, and legal status in the U.S.: A review', *Preventive Medicine*, vol. 104, pp. 13–23, Nov. 2017, doi: 10.1016/j.ypmed.2017.07.008.

[3] O. Aizpurua-Olaizola *et al.*, 'Evolution of the Cannabinoid and Terpene Content during the Growth of Cannabis sativa Plants from Different Chemotypes', *Journal of Natural Products*, vol. 79, no. 2, pp. 324–331, 2016, doi: 10.1021/acs.jnatprod.5b00949.

[4] W. A. Devane, F. A. Dysarz III, M. R. Johnson, L. S. Melvin, and A. C. Howlett, 'Determination and characterization of a cannabinoid receptor in rat brain', *Molecular Pharmacology*, vol. 34, no. 5, pp. 605–613, 1988.

[5] P. F. Whiting *et al.*, 'Cannabinoids for medical use: A systematic review and meta-analysis', *JAMA - Journal of the American Medical Association*, vol. 313, no. 24, pp. 2456–2473, 2015, doi: 10.1001/jama.2015.6358.

[6] J. T. Fischedick, A. Hazekamp, T. Erkelens, Y. H. Choi, and R. Verpoorte, 'Metabolic fingerprinting of Cannabis sativa L., cannabinoids and terpenoids for chemotaxonomic and drug standardization purposes', *Phytochemistry*, vol. 71, no. 17–18, pp. 2058–2073, 2010, doi: 10.1016/j.phytochem.2010.10.001.

[7] I. Galve-Roperh, C. Sánchez, M. L. Cortés, T. G. Del Pulgar, M. Izquierdo, and M. Guzmán, 'Anti-tumoral action of cannabinoids: Involvement of sustained ceramide accumulation and extracellular signal-regulated kinase activation', *Nature Medicine*, vol. 6, no. 3, pp. 313–319, 2000, doi: 10.1038/73171.

[8] S. Blasco-Benito *et al.*, 'Appraising the "entourage effect": Antitumor action of a pure cannabinoid versus a botanical drug preparation in preclinical models of breast cancer', *Biochemical Pharmacology*, vol. 157, pp. 285–293, 2018, doi: 10.1016/j.bcp.2018.06.025.

[9] S. Zou and U. Kumar, 'Cannabinoid receptors and the endocannabinoid system: Signaling and function in the central nervous system', *International Journal of Molecular Sciences*, vol. 19, no. 3, 2018, doi: 10.3390/ijms19030833.

[10] M. A. ElSohly, M. M. Radwan, W. Gul, S. Chandra, and A. Galal, 'Phytochemistry of Cannabis sativa L', *Progress in the chemistry of organic natural products*, vol. 103, pp. 1–36, 2017, doi: 10.1007/978-3-319-45541-9_1.

[11] O. Hardiman *et al.*, 'Amyotrophic lateral sclerosis', *Nature Reviews Disease Primers*, vol. 3, 2017, doi: 10.1038/nrdp.2017.71.

[12] A. Cerrato *et al.*, 'Phytocannabinomics: Untargeted metabolomics as a tool for cannabis chemovar differentiation', *Talanta*, vol. 230, 2021, doi: 10.1016/j.talanta.2021.122313.

[13] J. Zheng, M. Johnson, R. Mandal, and D. S. Wishart, 'A comprehensive targeted metabolomics assay for crop plant sample analysis', *Metabolites*, vol. 11, no. 5, 2021, doi: 10.3390/metabo11050303.

[14] P. G. Vásquez-Ocmín *et al.*, 'Cannabinoids vs. whole metabolome: Relevance of cannabinomics in analyzing Cannabis varieties', *Analytica Chimica Acta*, vol. 1184, 2021, doi: 10.1016/j.aca.2021.339020.

[15] T. M. Sirangelo, R. A. Ludlow, and N. D. Spadafora, 'Multi-Omics Approaches to Study Molecular Mechanisms in Cannabis sativa', *Plants*, vol. 11, no. 16, 2022, doi: 10.3390/plants11162182.

[16] A. Shapira, P. Berman, K. Futoran, O. Guberman, and D. Meiri, 'Tandem Mass Spectrometric Quantification of 93 Terpenoids in Cannabis Using Static Headspace Injections', *Analytical Chemistry*, vol. 91, no. 17, pp. 11425–11432, 2019, doi: 10.1021/acs.analchem.9b02844.

[17] M. C. Monti, P. Frei, S. Weber, E. Scheurer, and K. Mercer-Chalmers-Bender, 'Beyond Δ9-tetrahydrocannabinol and cannabidiol: chemical differentiation of cannabis varieties applying targeted and untargeted analysis', *Analytical and Bioanalytical Chemistry*, vol. 414, no. 13, pp. 3847–3862, 2022, doi: 10.1007/s00216-022-04026-2.

[18] C.-R. Li *et al.*, 'LC-MS-based untargeted metabolomics reveals chemical differences of Cannabis leaves from different regions of China', *Industrial Crops and Products*, vol. 176, 2022, doi: 10.1016/j.indcrop.2021.114411.

[19] F. Thoma, A. Somborn-Schulz, D. Schlehuber, V. Keuter, and G. Deerberg, 'Effects of Light on Secondary Metabolites in Selected Leafy Greens: A Review', *Frontiers in Plant Science*, vol. 11, 2020, doi: 10.3389/fpls.2020.00497.

[20] K. A. Aliferis and D. Bernard-Perron, 'Cannabinomics: Application of Metabolomics in Cannabis (Cannabis sativa L.) Research and Development', *Frontiers in Plant Science*, vol. 11, 2020, doi: 10.3389/fpls.2020.00554.

[21] A. Rashid, V. Ali, M. Khajuria, S. Faiz, S. Gairola, and D. Vyas, 'GC–MS based metabolomic approach to understand nutraceutical potential of Cannabis seeds from two different environments', *Food Chemistry*, vol. 339, p. 128076, Mar. 2021, doi: 10.1016/j.foodchem.2020.128076.

[22] Y. Zhou *et al.*, 'Metascape provides a biologist-oriented resource for the analysis of systems-level datasets', *Nat Commun*, vol. 10, no. 1, Art. no. 1, Apr. 2019, doi: 10.1038/s41467-019-09234-6.

[23] K. Dettmer, P. A. Aronov, and B. D. Hammock, 'Mass spectrometry-based metabolomics', *Mass Spectrometry Reviews*, vol. 26, no. 1, pp. 51–78, 2007, doi: 10.1002/mas.20108.

[24] A. Villate *et al.*, 'Review: Metabolomics as a prediction tool for plants performance under environmental stress', *Plant Science*, vol. 303, 2021, doi: 10.1016/j.plantsci.2020.110789.

[25] C. H. Johnson, J. Ivanisevic, and G. Siuzdak, 'Metabolomics: Beyond biomarkers and towards mechanisms', *Nature Reviews Molecular Cell Biology*, vol. 17, no. 7, pp. 451–459, 2016, doi: 10.1038/nrm.2016.25.

[26] E. T. Wurtzel and T. M. Kutchan, 'Plant metabolism, the diverse chemistry set of the future', *Science*, vol. 353, no. 6305, pp. 1232–1236, 2016, doi: 10.1126/science.aad2062.

[27] A. Raza *et al.*, 'Impact of climate change on crops adaptation and strategies to tackle its outcome: A review', *Plants*, vol. 8, no. 2, 2019, doi: 10.3390/plants8020034.

[28] R. Kumar, A. Bohra, A. K. Pandey, M. K. Pandey, and A. Kumar, 'Metabolomics for plant improvement: Status and prospects', *Frontiers in Plant Science*, vol. 8, 2017, doi: 10.3389/fpls.2017.01302.

[29] T. Cajka and O. Fiehn, 'Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics', *Anal. Chem.*, vol. 88, no. 1, pp. 524–545, Jan. 2016, doi: 10.1021/acs.analchem.5b04491.

[30] A. Ribbenstedt, H. Ziarrusta, and J. P. Benskin, 'Development, characterization and comparisons of targeted and non-targeted metabolomics methods', *PLOS ONE*, vol. 13, no. 11, p. e0207082, Nov. 2018, doi: 10.1371/journal.pone.0207082.

[31] J.-L. Wolfender, G. Marti, A. Thomas, and S. Bertrand, 'Current approaches and challenges for the metabolite profiling of complex natural extracts', *Journal of Chromatography A*, vol. 1382, pp. 136–164, 2015, doi: 10.1016/j.chroma.2014.10.091.

[32] C. Martias *et al.*, 'Optimization of Sample Preparation for Metabolomics Exploration of Urine, Feces, Blood and Saliva in Humans Using Combined NMR and UHPLC-HRMS Platforms', *Molecules*, vol. 26, no. 14, Art. no. 14, Jan. 2021, doi: 10.3390/molecules26144111.

[33] M.-E. P. Papadimitropoulos, C. G. Vasilopoulou, C. Maga-Nteve, and M. I. Klapa, 'Untargeted GC-MS metabolomics', *Methods in Molecular Biology*, vol. 1738, pp. 133–147, 2018, doi: 10.1007/978-1-4939-7643-0_9.

[34] G. Tooulakou *et al.*, 'Alarm Photosynthesis: Calcium Oxalate Crystals as an Internal CO2 Source in Plants', *Plant Physiology*, vol. 171, no. 4, pp. 2577–2585, Aug. 2016, doi: 10.1104/pp.16.00111.

[35] R. C. H. De Vos, S. Moco, A. Lommen, J. J. B. Keurentjes, R. J. Bino, and R. D. Hall, 'Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry', *Nature Protocols*, vol. 2, no. 4, pp. 778–791, 2007, doi: 10.1038/nprot.2007.95.

[36] H. K. Kim, Y. H. Choi, and R. Verpoorte, 'NMR-based metabolomic analysis of plants', *Nature Protocols*, vol. 5, no. 3, pp. 536–549, 2010, doi: 10.1038/nprot.2009.237.

[37] T. Fujii *et al.*, 'Direct metabolomics for plant cells by live single-cell mass spectrometry', *Nature Protocols*, vol. 10, no. 9, pp. 1445–1456, 2015, doi: 10.1038/nprot.2015.084.

[38] J. Lisec, N. Schauer, J. Kopka, L. Willmitzer, and A. R. Fernie, 'Gas chromatography mass spectrometry–based metabolite profiling in plants', *Nat Protoc*, vol. 1, no. 1, pp. 387–396, Jun. 2006, doi: 10.1038/nprot.2006.59.

[39] J. M. Amigo, 'Data mining, machine learning, deep learning, chemometrics: Definitions, common points and trends (Spoiler Alert: VALIDATE your models!)', *Brazilian Journal of Analytical Chemistry*, vol. 8, no. 32, pp. 22–38, 2021, doi: 10.30744/BRJAC.2179-3425.AR-38-2021.

[40] G. E. Hinton and R. R. Salakhutdinov, 'Reducing the dimensionality of data with neural networks', *Science*, vol. 313, no. 5786, pp. 504–507, 2006, doi: 10.1126/science.1127647.

[41] E. Small and H. D. Beckstead, 'Cannabinoid Phenotypes in Cannabis sativa', *Nature*, vol. 245, no. 5421, Art. no. 5421, Sep. 1973, doi: 10.1038/245147a0.

[42] A. Rivas-Ubach *et al.*, 'Topsoil depth substantially influences the responses to drought of the foliar metabolomes of Mediterranean forests', *Perspectives in Plant Ecology, Evolution and Systematics*, vol. 21, pp. 41–54, 2016, doi: 10.1016/j.ppees.2016.06.001.

[43] H. K. Kim and R. Verpoorte, 'Sample preparation for plant metabolomics', *Phytochemical Analysis*, vol. 21, no. 1, pp. 4–13, 2010, doi: 10.1002/pca.1188.

[44] S. Wu *et al.*, 'Mapping the Arabidopsis Metabolic Landscape by Untargeted Metabolomics at Different Environmental Conditions', *Molecular Plant*, vol. 11, no. 1, pp. 118–134, 2018, doi: 10.1016/j.molp.2017.08.012.

[45] K. Chetnik, L. Petrick, and G. Pandey, 'MetaClean: a machine learning-based classifier for reduced false positive peak detection in untargeted LC–MS metabolomics data', *Metabolomics*, vol. 16, no. 11, 2020, doi: 10.1007/s11306-020-01738-3.

[46] A. C. Schrimpe-Rutledge, S. G. Codreanu, S. D. Sherrod, and J. A. McLean, 'Untargeted Metabolomics Strategies—Challenges and Emerging Directions', *J. Am. Soc. Mass Spectrom.*, vol. 27, no. 12, pp. 1897–1905, Dec. 2016, doi: 10.1007/s13361-016-1469-y.

[47] Z. Pang *et al.*, 'MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights', *Nucleic Acids Research*, vol. 49, no. W1, pp. W388–W396, Jul. 2021, doi: 10.1093/nar/gkab382.

[48] Z. Pang *et al.*, 'Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data', *Nat Protoc*, vol. 17, no. 8, Art. no. 8, Aug. 2022, doi: 10.1038/s41596-022-00710-w.

[49] Y. Lu, Z. Pang, and J. Xia, 'Comprehensive investigation of pathway enrichment methods for functional interpretation of LC–MS global metabolomics data', *Briefings in Bioinformatics*, vol. 24, no. 1, p. bbac553, Jan. 2023, doi: 10.1093/bib/bbac553.

[50] R. G. Sadygov, F. M. Maroto, and A. F. R. Hühmer, 'ChromAlign: A two-step algorithmic procedure for time alignment of three-dimensional LC-MS chromatographic surfaces', *Analytical Chemistry*, vol. 78, no. 24, pp. 8207–8217, 2006, doi: 10.1021/ac060923y.

[51] S. Fan *et al.*, 'Systematic Error Removal Using Random Forest for Normalizing Large-Scale Untargeted Lipidomics Data', *Analytical Chemistry*, vol. 91, no. 5, pp. 3590–3596, 2019, doi: 10.1021/acs.analchem.8b05592.

[52] T. Vogt, 'Phenylpropanoid biosynthesis', *Molecular Plant*, vol. 3, no. 1, pp. 2–20, 2010, doi: 10.1093/mp/ssp106.

[53] Y. Yang and C. Benning, 'Functions of triacylglycerols during plant development and stress', *Current Opinion in Biotechnology*, vol. 49, pp. 191–198, Feb. 2018, doi: 10.1016/j.copbio.2017.09.003.

[54] H. U. Kim, 'Lipid Metabolism in Plants', *Plants*, vol. 9, no. 7, Art. no. 7, Jul. 2020, doi: 10.3390/plants9070871.

[55] R. A. Dixon, 'Natural products and plant disease resistance', *Nature*, vol. 411, no. 6839, pp. 843–847, 2001, doi: 10.1038/35081178.

[56] M. Chadwick, H. Trewin, F. Gawthrop, and C. Wagstaff, 'Sesquiterpenoids lactones: Benefits to plants and people', *International Journal of Molecular Sciences*, vol. 14, no. 6, pp. 12780–12805, 2013, doi: 10.3390/ijms140612780.

[57] E. R. Radwanski and R. L. Last, 'Tryptophan biosynthesis and metabolism: Biochemical and molecular genetics', *Plant Cell*, vol. 7, no. 7, pp. 921–934, 1995.

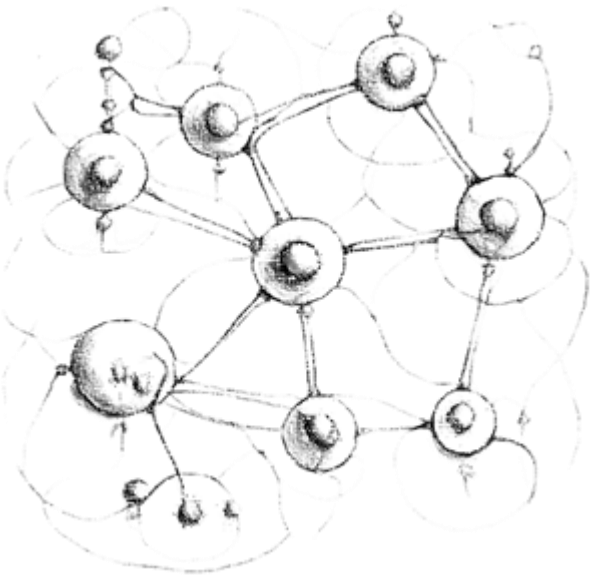
[58] T. M. Kutchan, 'Alkaloid biosynthesis - The basis for metabolic engineering of medicinal plants', *Plant Cell*, vol. 7, no. 7, pp. 1059–1070, 1995.

[59] R. A. Razal, S. Ellis, S. Singh, N. G. Lewis, and G. H. N. Towers, 'Nitrogen recycling in phenylpropanoid metabolism', *Phytochemistry*, vol. 41, no. 1, pp. 31–35, 1996, doi: 10.1016/0031-9422(95)00628-1.

[60] T. Tohge, M. Watanabe, R. Hoefgen, and A. R. Fernie, 'Shikimate and phenylalanine biosynthesis in the green lineage', *Frontiers in Plant Science*, vol. 4, no. MAR, 2013, doi: 10.3389/fpls.2013.00062.

[61] H. Maeda and N. Dudareva, 'The shikimate pathway and aromatic amino acid biosynthesis in plants', *Annual Review of Plant Biology*, vol. 63, pp. 73–105, 2012, doi: 10.1146/annurev-arplant-042811-105439.

[62] R. Leonardi, Y.-M. Zhang, C. O. Rock, and S. Jackowski, 'Coenzyme A: Back in action', *Progress in Lipid Research*, vol. 44, no. 2–3, pp. 125–153, 2005, doi: 10.1016/j.plipres.2005.04.001.



CHAPTER 2:

DETERMINATION OF PREDICTIVE CONSTITUTIVE METABOLIC MARKERS FOR MEDICAL CANNABIS CULTIVAR ADAPTABILITY TO TROPICAL CLIMATES THROUGH LC-HRMS BASED UNTARGETED METABOLOMICS

San Nicolas M., Villate A., Walther D., Olivares M., Aizpurua-Olaizola O., Usobiaga A.

Abstract

Currently, the potential of cannabis as treatment for diverse symptoms, pathologies or illnesses of neurological regard is widely acknowledged, therefore, a resultant increase in social and institutional acceptance towards the introduction of cannabis as a natural and alternative healthcare source has led diverse countries across the planet to adopt frontline positions concerning the research and development (R&D) in this emerging discipline. Nevertheless, the native weather of certain geographical regions could be a constraint for the quality assurance of this naturally grown medicinal product, such as the tropical climate to which various of those countries are subjected. Its distinctive high temperature and humidity may induce an undesired stress in the plants, which, consequently, could reduce, or even halt, the biosynthetic yield of its bioactive compounds. Hence, for medical cannabis cultivations in such environments, the adaptability of cannabis cultivars should primarily be assessed to ensure the adequate active pharmaceutical ingredient (API) natural production. Thus, the objective of this work was to identify constitutive metabolic markers at early vegetative stages to predictively determine the adaptability of cannabis cultivars to tropical environments through High Performance Liquid Chromatography coupled to qOrbitrap High Resolution Mass Spectrometer (HPLC-qOrbitrap) based untargeted metabolomics. The chromatographic data were used to train and externally validate Partial Least Squares Discriminant Analysis (PLS-DA) predictive models and calculate the subsequent VIPs, from which the potential biomarkers were determined. The obtained results showed that 6 metabolites, 1 primary (3-[8,13-bis(ethenyl)-18-(1-hydroxy-3-methoxy-3-oxopropyl)-3,7,12,17-tetramethyl-22,23-dihydroporphyrin-2-yl]propanoic acid) and 5 secondaries (Piperochromenoic acid, Mammea C/BB, Dalbergichromene, Dihydrocordoin and Hedychilactone D), all of them detected in the non-polar extract of the leaf tissue, could be claimed as such, as each of them triggered an adaptive metabolic mechanism to face the uneasy scenario. This work thereby paved the way for the development of a targeted analytical approach for the identified metabolic markers to facilitate the quality assurance at early stages for medical cannabis cultivation in countries with tropical climates.

Keywords: plant metabolomics, biomarker, prediction, machine learning, adaptive mechanism, cultivar adaptability

1. Introduction

Over the last few years, glimpses of diverse changes and progresses have been witnessed towards the application of medical cannabis as a natural substitute to dogmatically stablished synthetic drugs [1], [2]. Vivid evidence of the current systematic acceptance rate is the surging tendency in the US states that have already authorised cannabis for medical purposes, as the current number stands at 38 out of 50 [3]. Moreover, in 2021 the removal of cannabis and cannabis resin from Schedule IV of the Single Convention on narcotic drugs (1961) took place [4], where the World Health Organisation (WHO) recommended its elimination from the list of narcotic drugs and recognised its therapeutic potential [5]. Such measures have led to a burgeoning trend of institutional acceptance, because of which several countries around the globe have decided to start implementing bills for the regulation or legalisation of cannabis for medical purposes. In Europe, for instance, countries such as Switzerland and Germany have already set up government programmes to allow clinical trials to document and monitor the medical effects experienced by subjects, whether they are neurological patients or controls [6], [7]. Legislative changes in these countries opened new medical horizons through the introduction and application of cannabinoids, the naturally occurring active pharmaceutical ingredients (API) in *Cannabis sativa L.* plant species. Currently, major cannabinoids in cannabis, Δ^9-tetrahydrocannabinol (THC) and cannabidiol (CBD), are widely known for their therapeutic potential, due to their agonist status to the receptors of the human endocannabinoid system (ECS) [8]–[17]. However, in these recent years, the legislation towards medical cannabis has also undergone major changes in various countries apart from the US and Europe, for instance, Argentina [18], Uruguay [19], Colombia [20], Saint Vincent and Grenadines [21] and Jamaica [22] in America; Australia [23] in Oceania; Lesotho and South Africa in Africa [24], [25]; and Thailand [26] in Asia.

In these countries the current medical cannabis R&D situation is diverse, as each is taking its own step towards different horizon aims; nevertheless, they may have to deal with some common factors for controlled quality medical cannabis productions, as, for instance, climatic or environmental challenges. Even though cannabis is documented to be native from western Asia [27]–[29], it has been cultivated around the world for millennia [30], and despite some years ago the cannabis taxonomy was widely known as *Sativa* or *Indica* [28], plenty of different cultivars currently exist due to Mendelian genetics [31] based breeding during several ages, consequently diluting the morphological and physiological distinction threshold between both subspecies. As a result, many cannabis cultivars may now exhibit limited adaptability to certain climatic conditions, which may turn the autochthonous atmospheres of some countries into a stressful cultivation environment. That is the case of some of the previously stated countries, which are distinguished for their tropical climate, featuring high temperature and humidity. These abiotic stressors can lead to unpleasant environmental conditions for the cultivation of controlled medical cannabis, as the inner growth and development metabolism of the plant may result disturbed or even harmed [32]–[37], therefore supressing the biosynthetic pathways of cannabinoids [38]–[41].

Hence, the metabolism of a plant is integrated by to classes of metabolites. On the one hand, primary metabolites are essential compounds that are involved in fundamental biochemical pathways necessary for the growth, development and basic functions of a plant, such as amino acids, sugars, lipids, nucleic acids or organic acids [42]–[44]. On the other hand, secondary metabolites are compounds produced by plants that are not directly involved in the essential

processes of growth or development, but often provide key roles in plant defence, signalling and interactions with the surrounding environment, as phenolic compounds, flavonoids, alkaloids or glycosides [45]–[50]. Currently, it is widely known that a plant individual can trigger adaptive metabolic mechanisms modifying these metabolite content levels or related metabolic pathways in its cells to deal with uneasy scenarios, such as an stressful environment, therefore, untargeted metabolomics could be an excellent resource for the early detection of predictive metabolic markers that determine the adaptability of different cannabis cultivars to such stressful environments [51]. Thus, following this research line, in this work the determination of predictive constitutive metabolic markers for cannabis cultivars adaptability to tropical climates was aimed by means of untargeted metabolomic analysis by HPLC-qOrbitrap, coupled with advanced multivariate data-analysis approaches.

2. Materials and methods

2.1. Reagents and solvents

Solvents for chromatographic analysis, acetonitrile (ACN), methanol (MeOH) and water (H_2O) of HPLC grade, were acquired in Panreac/AppliChem (ITW Reagents, S.R.L., Italy). Formic acid (FA) for LC-MS was purchased from Fischer Chemicals (Thermo Fisher Scientific Inc., USA) and ammonium acetate (NH_4Ac, for molecular biology ≥ 98 %) was acquired in Sigma Aldrich (Merck Group, USA). Calibration of the qOrbitrap was performed with Pierce™ LTQ Velos ESI Positive Ion Calibration Solution and Pierce™ ESI Negative Ion Calibration Solution by Thermo Scientific (Thermo Fisher Scientific Inc., USA).

For sample extraction, MeOH (anhydrous) and trichloromethane ($CHCl_3$, synthesis grade) were obtained in Macron Fine (Avantor, Inc., USA) and Scharlau (Scharlab S.L., Spain), respectively. Water was filtered to a resistivity of 18.2 MΩ·cm @ 25 °C and a total organic content (TOC) less than 3 ng·mL^{-1} using a Q-POD water dispenser and a Millipak express 40 (0.22 µm filter) by Merck (Merck Group, USA).

2.2. Plant samples

Three indoor cultivations were conducted at the facilities of Sovereign Fields (Sovereign Fields S.L., Spain) for the untargeted metabolomics study of cannabis. Each cultivation was carried out in two isolated 24 m^2 growth rooms, with 6 Ceramic Metal Halide (CMH) lamps (315 W/37000 lm) by Lumatek (Lumatek Ltd., UK) in the roof of each one, at 2.10 m height and uniformly distributed through the room, coupled to Adjust-A-Wings large enforcer reflectors.

Plantlet clones/biological replicates were always obtained through branch cutting from mother-plants 6 weeks before the cultivation starting time (t_0). The plantlets spent two weeks in incubators, during which they developed the roots. Then, the plantlets were taken out the incubators and transplanted to 0.5 L pots containing a soil/hummus/nutrient mixture. Specifically, the mixture was composed of 80 % of Light mix soil of Biobizz (Biobizz Worldwide S.L., Spain), 20 % of hummus, and 10 g/L of farmer mix nutrient solution by Lurpe (Lurpe Natural Solutions, Spain), which is composed of bat guano, bone meal, kelp meal, Azomite®, organic alfalfa, insect frass, blood meal, dolomite, langbeinite humic and fulvic acids, and a complex blend of rhizobacteria and Trichoderma. Plants stayed the next two weeks at room temperature conditions in the greenhouse and at last, transplanted to 11 L pots with the same mixture for two more weeks in acclimatisation before the starting point (t_0).

Samples were cultivated for 12 weeks (84 days) starting from t_0 time, divided in vegetative stage (first 4 weeks/28 days) and flowering stage (next 8 weeks/56 days). These stages were defined by the photoperiod regime (vegetative 18h light with lamps at 50 % intensity, flowering 12h light with lamps at 80 % intensity).

Each growth room was subjected to specific environmental conditions, based on temperature and humidity. One of the rooms, called the *Control* room, tried to simulate the growth conditions of indoor cannabis cultivations, so the temperature was kept between 23-25 °C during the daytime and was prevented from dropping below 20 °C during the night, while the relative humidity was kept at the overall of 60 %. The environmental conditions in the other room, called the *Tropical Stress* room, were set at an average temperature and humidity of 29-31 °C and 80 %, respectively.

Plant samples were periodically harvested and instantly frozen in liquid nitrogen, following the quenching method. A representative leaf and inflorescence tissue sample was collected from each biological replicate at each corresponding harvesting time. For doing so, the sampling design proposed in *M. San Nicolas et al.* [52] was followed.

In each harvesting time-point, leaf and flower samples were harvested from 6 biological replicates per growth-condition in the 1st and 2nd cultivations for adequate sample size for untargeted metabolomic study [53]–[56], and from 3 biological replicates in the case of the 3rd cultivation. The samples were kept at -40 °C temperature prior to analysis.

2.2.1. 1st cultivation

Plant clones of 2 different cultivars were grown, one of high THC and low CBD concentrations (SFT1, a chemotype I cultivar), and one of low THC and high CBD concentrations (SFC1, a chemotype III cultivar) [57]. A total of 78 plant clones as biological replicates per variety were produced. Of the 78 replicates per variety, 6 were harvested at t_0 time. The other 72 clones would be divided in both growth rooms, 36 per room. Sample collection strategy is illustrated in *Table 1*. Sampled plants were removed from the rooms, so the density of plants in the rooms decreased over time. The monitored data of the environmental conditions of both growth rooms is appended in *Figure S1* and *S2* of Supplementary Material (SM).

2.2.2. 2nd cultivation

The experimental design of the second cultivation was optimised according to the exploratory results obtained in the untargeted metabolomic study of the first cultivation. Thus, sample harvesting times were reduced by half, allowing the introduction of a greater number of cultivars into the growth rooms, therefore enhancing the representative coverage of the subsequent metabolomic study. Moreover, the samplings were performed on the same plant individuals, so the plant amount per room was kept constant during the whole cultivation, the same as the initial amount of the first cultivation. This way, plant clones of 12 cultivars were grown: 8 of chemotype I (SFT1, SFT2, SFT3, SFT4, SFT5, SFT6, SFT7 and SFT8), 2 of balanced concentrations in THC and CBD levels (SFB1 and SFB2, chemotype II cultivars) and 2 of chemotype III (SFC1 and SFC2) [57]. The 2 varieties grown in the first cultivation (SFT1 and SFC1) were also cultivated in the second cultivation.

A total of 18 plant clones as biological replicates per variety were produced. Of the 18 replicates per variety, 6 were harvested at t_0 time, and the other 12 clones were divided in both growth rooms. Sample collection strategy is shown in *Table 1*.

2.2.3. 3rd cultivation

The third cultivation was carried out for the qualitative validation of the results obtained through the first and second cultivation. Hence, plant clones from a total of 15 different cultivars were grown for diversity coverage, from which two were the ones cultivated in both the first and second cultivations (SFT1 and SFC1), and another two were also cultivated in the second cultivation (SFB1 and SFC2). From the cultivated varieties, 5 belonged to chemotype I (SFT1, SFT9, SFT10, SFT11 and SFT12), 4 to chemotype II (SFB1, SFB3, SFB4, SFB5) and 6 to chemotype III (SFC1, SFC2, SFC3, SFC4, SFC5 and SFC6) [57]. The monitored data of the environmental conditions of both growth rooms is appended in *Figure S3* and *S4* of SM, alongside the data of the 2nd cultivation.

As previously stated, 3 biological replicates per-growth condition, per variety, were grown in the cultivation, therefore, a total of 9 plant clones were produced. From the 9 replicates per cultivar, 3 were harvested at t_0 time, and the other 6 were divided in both growth rooms.

In accordance with the defined objective, only leaf tissue samples were harvested at the cultivation starting time (t_0), in order to search for the potential constitutive metabolic markers, and also flower tissue samples at the end of cultivation, for the determination of corresponding cannabinoid contents in the cannabis cultivars (*Table 1*).

Table 1: Sample-harvesting scheme of the three experimental cultivation cycles

	Sample harvesting time-point	**Experimental cultivation**		
		1st cultivation	2nd cultivation	3rd cultivation
Vegetative stage	t_0 (Day 0/Week 0)	Leaves	Leaves	Leaves
	t_1 (Day 14/Week 2)	Leaves	Ø	Ø
	t_2 (Day 28/Week 4)	Leaves	Leaves	Ø
Flowering stage	t_3 (Day 42/Week 6)	Leaves + Inflorescences	Ø	Ø
	t_4 (Day 56/Week 8)	Leaves + Inflorescences	Leaves + Inflorescences	Ø
	t_5 (Day 70/Week 10)	Leaves + Inflorescences	Ø	Ø
	t_6 (Day 84/Week 12)	Leaves + Inflorescences	Leaves + Inflorescences	Inflorescences

2.3. Analysis of cannabinoids

Δ^9-Tetrahydrocannabinol (THC), Δ^9-Tetrahydrocannabinolic acid (THCA), Cannabidiol (CBD), Cannabidiolic acid (CBDA), Cannabichromene (CBC), Cannabichromencic acid (CBCA), Cannabinol (CBN), Cannabinolic acid (CBNA), Cannabigerol (CBG), Cannabigerolic acid (CBGA), Tetrahydrocannabivarin (THCV), Tetrahydrocannabivarinic acid (THCVA), Cannabidivarin (CBDV) and Cannabidivarinic acid (CBDVA) cannabinoids were analysed in the inflorescences of the cannabis cultivars grown in the Control and Tropical Stress environments at the end of the cultivations. This was performed using a Shimadzu Prominence-i LC-2030C 3D Plus liquid chromatographic system with an integrated UV detector, with a NexLeaf CBX Potency 2.7 µm (150 x 4.6 mm) chromatographic column coupled with a NexLeaf CBX Potency 2.7 µm (5 x 4.6 mm) pre-column (after a liquid-liquid extraction using ethanol as dissolvent), according to method developed in *Aizpurua-Olaizola et al.* [57], in the Sovereign Fields S.L. facilities. Cannabinoids were quantified at 220 nm wavelength.

2.4. Experimental procedure for untargeted metabolomics analysis

Extraction and subsequent dilution of frozen plant samples was performed prior to analysis as defined in the *Experimental procedure* section of *M. San Nicolas et al.* [52] using a trichloromethane-water-methanol ($CHCl_3$:H_2O:CH_3OH (2:1:1, v/v)) solvent mix. To correct the possible chromatographic drift and signal fluctuation across the analysis sequences in the untargeted research of the cannabis samples, one quality control (QC) sample was prepared per plant tissue type and extract phase. This QC sample consisted of the pooled sum of 10 µL aliquots of the final extract of each sample and it was periodically analysed through the corresponding analysis sequences.

Untargeted metabolomics analysis of all the extracts was performed by a Thermo Scientific Dionex Ultimate 3000 liquid chromatograph coupled to a Thermo Scientific Q Exactive Focus quadrupole-Orbitrap mass spectrometer (UHPLC-q-Orbitrap), equipped with a heated electrospray ionisation source (HESI, Thermo-Fisher Scientific, CA, USA). A reverse phase Kinetex 1.7 µm C18 100 Å (150x3 mm) with AJ0-8782 C-18 pre-column (Phenomenex, USA) was used for compound retention. The chromatographic analysis were also performed exactly as described in the *Chromatographic methods and general MS settings* section of *M. San Nicolas et al.* [52]. Based on empirical evidence shared in that work, analysis were only performed in positive ionisation mode, since way more chromatographic features were annotated in comparison with the negative ionisation. Thus, the search for potential constitutive biomarkers was focused on the ionisation mode that offered greater chromatographic feature population and higher In-Silico feature identification probability. The chromatographic feature detection in the qOrbitrap was performed in the MS2 data-dependant (DDA) mode.

2.5. Data-processing

Data processing of the raw chromatograms for the untargeted metabolomic study of cannabis was performed using Compound Discoverer 3.3 software by Thermo Fischer.

A minimum signal-to-noise ratio (S/N) of 3 was stablished in order to consider a chromatographic peak feasible and subsequently integrate its corresponding under-peak area. In the detected chromatographic features, MS1 was used as precursor for compound detection, with a mass tolerance of 5 ppm after performing 5 scans per feasible peak. Retention times of chromatographic features were corrected towards the pooled QC samples. Compounds were

grouped across the samples with a mass tolerance of 5 ppm and a retention time (RT) tolerance of 0.2 minutes based on the preferred ion, i.e. $[M+H]^+$. Moreover, a maximum threshold of 1 minute peak width at half peak height was stablished in every detected feature.

In addition, the features were also filtered on the basis of the quality of the chromatographic peaks, which was defined in terms of four criteria: the jaggedness, the zig-zag index, the FWHM2base (Feature width at half max to base) and the modality of the peaks [58]. Each of these metrics was measured at 5/10, being this value, the representation of its contribution compared to the other parameters contributors. Each measured metric contributed equally to the overall peak quality. Hence, the overall peak quality filter threshold was set at 6/10 for a minimum number of 4 samples, and only features whose peaks exceeded this filter would be further accepted. An area filter was also applied, so that the maximum area of every detected feature would have to be greater than, at least, 1,000,000 intensity counts in one sample.

Metabolite annotation was performed using the databases specified below. At least a partial putative spectral match with a minimum single candidate compound among the provided spectral fragmentation libraries would be needed for metabolite identification. The referred libraries were Cannabis Sativa L. endogen suspect list, retrieved from Plant Metabolic Network database [59]–[62], LipidMaps structure database [63], Endogenous metabolites database of 4,400 compounds provided by Thermo Scientific, 6549 flavonoids structure database from Arita laboratory [64], Carotenoids database [65], Chemical Entities of Biological Interest (ChEBI) database [66], The Food Database (FooDB) [67], Human Metabolome Database (HMDB) [68], Kyoto Encyclopedia of Genes and Genomes (KEGG) [69] and Small Molecule Pathway Database (SMPDB) [70]. For structural elucidation of the detected features through *in-silico* fragmentation mzLogic data analysis algorithm and Mass Frontier 7.0 spectral interpretation software were used, both from Thermo Fisher Scientific. The Fragment Ion Search (FISh) score of the metabolite candidates was calculated with 2.5 mmu high accuracy mass tolerance, 0.5 Da low accuracy mass tolerance and a threefold S/N threshold. Metabolite identification was performed according to the confidence levels defined in *Schrimpe-Rutledge et al.* [71]. No pure reference standard was used in this work; hence, the highest annotated identification confidence level was 2 (putative compound identification due to MS2 match through *in-silico* fragmentation).

Detected features were filtered according to the specified filters and criteria, therefore, supervised multivariate data-analysis approaches were applied with the remaining features, maximising the covariance between the sample-classes and the detected features to determine the Variables of Importance in Projection (VIPs) for the defined outcome, to identify potential biomarkers of cannabis cultivars adaptability to tropical environments. The data analysis approaches were performed using Metaboanalyst 5.0 [72], [73] and the PLS_Toolbox 9.0 in MATLAB (Eigenvector Research, Inc., USA).

3. Results and discussion

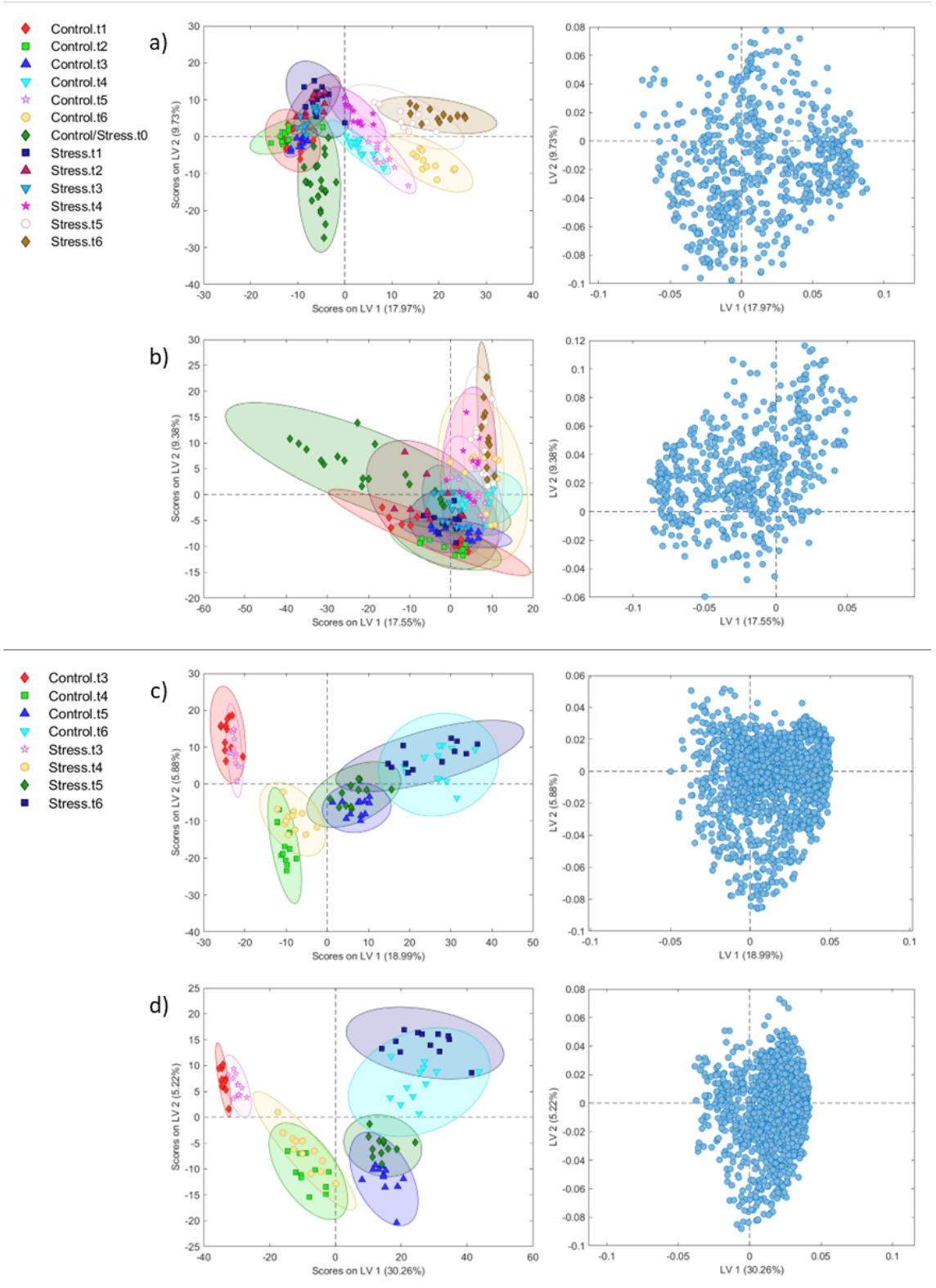
3.1. 1st Cultivation data exploration for 2nd cultivation optimisation

An exploratory observational study of the 1st cultivation results was done for the determination of the optimal sample-harvesting times in the 2nd cultivation, which was the one that would provide the later model training data, since there was the willing of embracing greater cultivar diversity coverage. Obtained data is present in *Tables S1, S2, S3 and S4* of SM.

Thus, both plant-tissue-wise and phase-wise supervised Partial Least Squares Discriminant Analysis (PLS-DA) classification models were built with each's autoscaled results for explorative purposes, whereby the major behaviour trends of the different condition and harvesting time sample classes were observed through the cultivation time. To accomplish it, hypothetical missing data and chromatographic drift were beforehand corrected by Compound Discoverer 3.3. software.

Given that this first approach consisted of a observational study, in-depth identification of the detected features through *in-silico* fragmentation was not performed, therefore every feature was identified with a 3rd identification confidence level according to *Schrimpe-Rutledge et al.* [71], i.e. tentative identification of the compounds by MS1 data and monoisotopic mass.

At first glance, *Figure 1* shows that all the sample classes are clustered around the centre of the variable space; however, taking a closer look, a time-wise evolution tendency can be distinguished during the cultivation, where the variance between different harvesting times happened to be significantly higher than the difference among coetaneous condition classes. Indeed, a clearer appreciation of this trend can be seen in the results of the flower samples, due to the presence of fewer sample-classes. Furthermore, differences can be observed between the two condition classes at every sample harvesting time, which appear to remain roughly constant throughout the cultivation; hence, a glimpse of a potential difference between the metabolomes of the plant individuals grown in the different environmental conditions could be deciphered from these results, implying that the cannabis cultivars have indeed adapted and survived to the tropical stress environment due to a metabolic adaptive mechanism. Being this the study case, the identification of potential constitutive biomarkers of the adaptability of cannabis cultivars to topical environments seemed a feasible task.



<u>Figure 1:</u> PLS-DA scores and loadings in the 1st and 2nd latent variables of the 1st cultivation data a) Leaf tissue, polar phase b) Leaf tissue, non-polar phase c) Flower tissue, polar phase d) Flower tissue, non-polar phase

An apparent conclusion from these results is that, particularly in the case of leaf samples, the consecutive harvesting-times classes happen to overlap. This meant that some successive harvesting time-points could be negligible while maintaining the significance of the total explained information. This observation allowed the experimental design of the 2nd cultivation to be defined with, at least, half the sample harvesting times of the 1st cultivation. It is true that in the case of the flower samples the referred overlap was considerably less than in the case of the leaf samples, meaning that the difference between successive flower sample classes was of greater significance. Nevertheless, since the final aim of the work was defined as the determination of potential constitutive markers (i.e. markers that are implicitly present from an early moment of the cultivation), emphasis should be focused on the leaf tissue, as this sample matrix was the one present throughout all the cultivation time. It was therefore concluded that the sample harvesting time-points of the 2nd cultivation could be limited to t_0, t_2, t_4 and t_6 times, in the case of leaf samples, and to t_4 and t_6 in the case of inflorescences, without loss of significance in the untargeted metabolomics study of the plants. This meant that samples would be collected every 4 weeks instead of every 2 weeks, which was considered a reasonable time between harvesting time-points for plant individuals to cope with the stress induced in the sample-cuttings, so that this induced stress would not be discernible in the analytical results. Hence, it was decided that plant samples could be collected from the same individuals throughout the 2nd cultivation, thus reducing the number of biological replicates per cultivar in each growth room from 36 to 6, allowing the introduction of more varieties, from 2 to 12, and increasing the biological variability. This was an outcome of great interest, as the subsequent data derived from the 2nd cultivation would have the purpose of training the corresponding data-analysis models for potential biomarkers determination. Therefore, covering representative biological variability turned to be an issue of great significance.

3.2. Cultivar adaptability determination

Following the conclusions drawn from the resultant data from the 1st cultivation, the 2nd cultivation was carried out, from which the obtained results were addressed for the determination of potential biomarkers of the adaptability of cannabis cultivars to environments of tropical stress, but prior to this, the adaptability of each variety had to be defined. Hence, the adaptability was defined on the basis of cannabinoid production. Specifically, the adaptability of each cultivar was defined by the relative difference in the total cannabinoid concentration in inflorescences at the end of the cultivation between its biological replicates grown in the Control and Tropical Stress environments (*Eq. 1*).

$$Eq.1: Relative\ cannabinoid\ content\ difference\ (\%) = \frac{\Sigma[Cannabinoids]Control - \Sigma[Cannabinoids]Stress}{\Sigma[Cannabinoids]Control} * 100$$

Figure 2 shows the relative cannabinoid content difference for each cultivar. Considering the severe abiotic factors that plants grown in the Tropical Stress room underwent, it was expected that the plants grown in this environment would produce significantly lower overall cannabinoids contents, which was subsequently confirmed, only excluding one of the varieties (SFC1), which showed higher total cannabinoid concentration in the plants grown under abiotic

stress. Taking this into account, a good adaptability was defined in a resultant relative difference smaller than 33.3 %. As it can be there observed, 3 out of 12 cultivars were classified as varieties with Good adaptability while the rest corresponded to the Bad adaptability class.

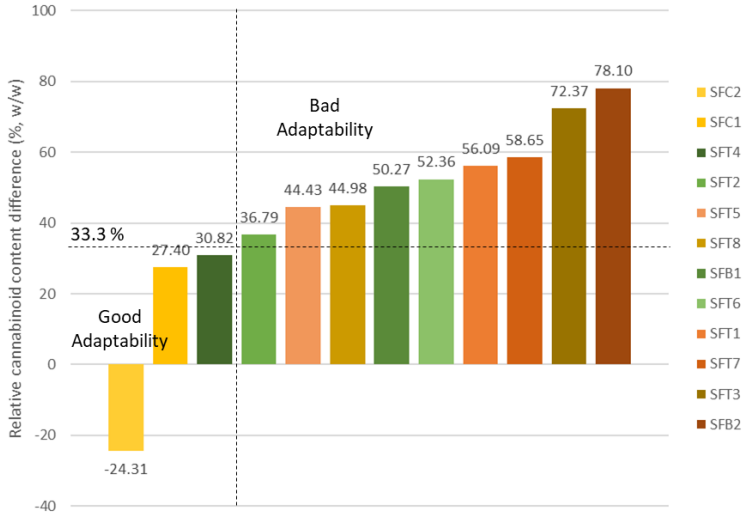


Figure 2: Relative cannabinoid difference between Control and Tropical Stress inflorescences at the end of the 2nd cultivation, weighed towards the cannabinoid production of each cultivar in the Control environment

Two chemotype III cultivars (SFC1 and SFC2) and one chemotype I cultivar (SFT4) were determined as likely to adapt to the studied stress environment. It was of undeniable interest that, among the cultivars of Good adaptability, chemotype diversity existed, since every cannabinoid profile is interesting in the medical or pharmaceutical field, depending on the targeted therapeutic treatment. Thus, for the determination of potential constitutive markers, a PLS-DA model was trained with the resultant data of the 2nd cultivation for each plant tissue and polarity nature of the extracts with the cultivars belonging to the Good and Bad adaptability classes, with which an exhaustive search would be carried out to ascertain the most relevant metabolites for the significant classification of the two classes.

3.3. Biomarker determination for cultivar adaptability (2nd and 1st cultivations)

Chromatographic analysis results of each plant tissue and extract phase from the 2nd cultivation were processed by Compound Discoverer 3.3 as stated in *section 2.5.*, alongside the corresponding harvesting time-points samples from the 1st cultivation (t_0, t_2, t_4 and t_6 in the 1st cultivation). Thus, a single dataset would be obtained that would unify the integrated data from the 2nd and 1st cultivations, which would be later divided into the training and testing sets, respectively, in the subsequent multivariate models. In this case, the hypothetical missing data was imputed through *k*-Nearest Neighbours (KNN, n = 5) and empirical Bayes methods (EBM) was applied for the eventual chromatographic drift and batch effect correction [74], since plant samples were analysed in different analysis sequences. In the case of leaf samples, 2nd and 1st cultivation samples were analysed in 3 and 2 different batches (*See Figure 3*), respectively; in the case of flower samples, both 2nd and 1st cultivation samples were analysed in 2 batches each. Obtained data is shown in *Tables S5, S6, S7* and *S8* of SM.

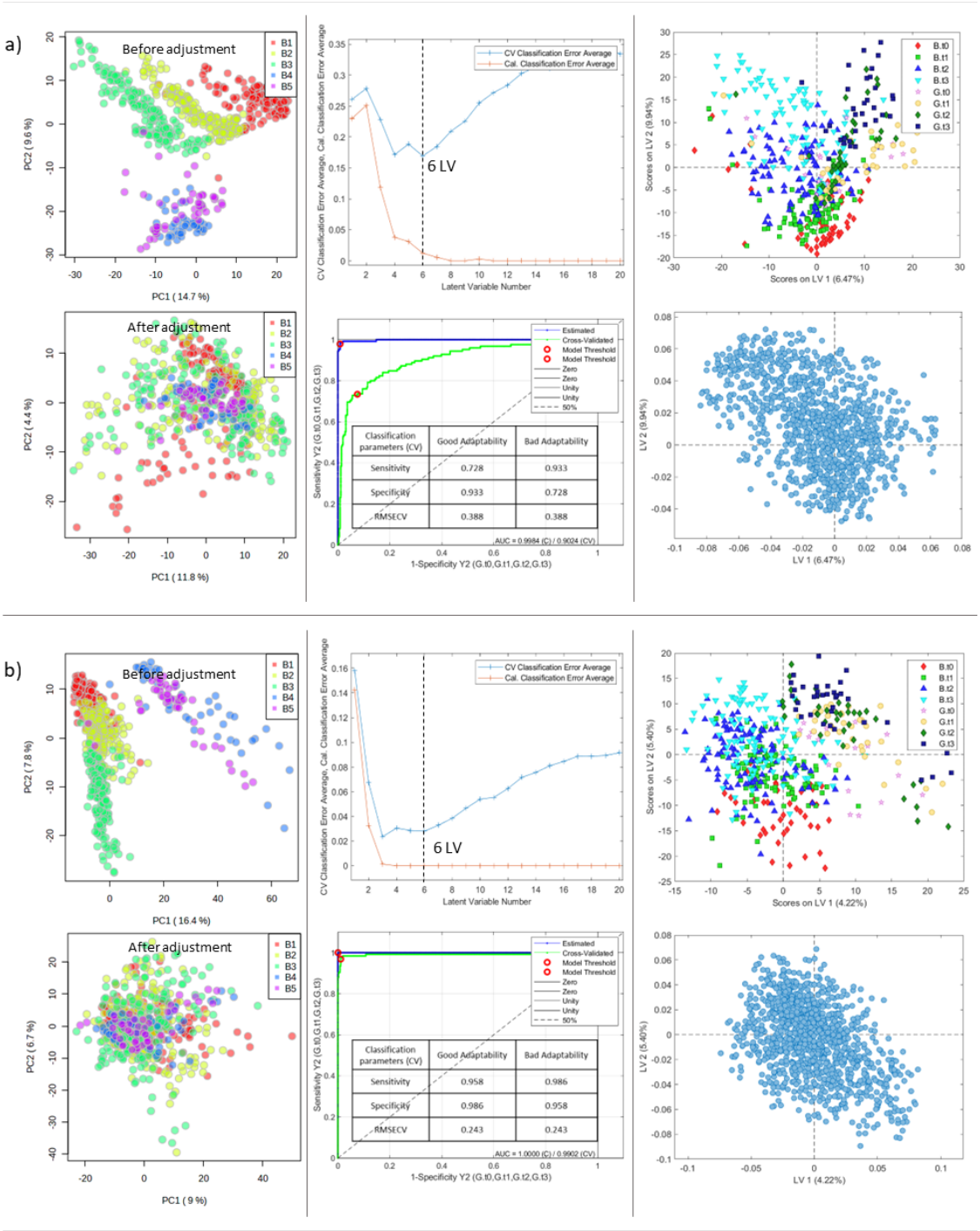
Hence, a PLS-DA model was trained with the autoscaled data of each plant tissue and extract phase following the defined adaptability classes. Cross validation was performed by random subsampling means, with 10 data splits permutated in 5 iterations. The latent variable (LV) number in each corresponding classification model was determined by the minimisation cross-validated average classification error in conjunction with the maximisation of the predictive capacity of the model (*See Figure 3*). Cross-validated classification performance parameters of the Good adaptability class are summed up in *Table 2* for each of the study cases.

Table 2: Cross-validated classification performance parameters of the Good adaptability class in each of the trained models

	Leaf tissue		**Flower tissue**	
	Polar extract	Non-polar extract	Polar extract	Non-polar extract
LV number	6	6	4	4
Sensitivity (CV)	0.728	0.958	0.992	0.870
Specificity (CV)	0.933	0.986	0.969	0.955
RMSECV	0.388	0.243	0.252	0.283

Since the final objective was to identify constitutive markers, the multivariate data-analysis approach was focused from the leaf tissue perspective, and the resultant data from the flower tissue was used to validate the later results. Therefore, the classification parameters of both PLS-DA models trained with the resultant data from the leaf tissue can be observed in *Figure 3*.

As shown in *Figure 3*, it turned out that there was a significant improvement in the cross-validated classification performance parameters from the polar to the non-polar extract of the leaf tissue. This could be visually appreciated in both Receiving Operating Characteristic (ROC) curves, which reflect the predictive power of the models, where their optimum cross-validated sensitivity and specificity parameters were 0.958 and 0.986, respectively, for the Good adaptability class, in the non-polar phase, compared to the corresponding values of 0.728 and 0.933 in the polar phase. Nevertheless, the search for the biomarkers was carried out in both phases.



<u>Figure 3:</u> Followed data-analysis approach for the PLS-DA model training for the determination of constitutive biomarkers of cannabis cultivars adaptability towards tropical stress in the leaf tissue. From left to right: Batch effect correction by EBM; PLS-DA LV number determination, alongside the corresponding ROC curve; and PLS-DA scores and loadings of the Good (G) and Bad (B) adaptability classes in the 1st and 2nd LV. a) Polar phase b) Non-polar phase

Therefore, the potential markers were determined through VIP calculation and, for the unequivocal identification of the markers through all the cultivation time, a minimal VIP value of 2 was set as a rigorous exclusion criterion. In addition, as previously stated, perspective was

focused into the identification of the corresponding constitutive markers, hence, p-values between the raw signals of each detected feature were calculated by ANOVA between the samples composing the Good and Bad adaptability classes in the t_0 harvesting time, for both 2nd and 1st cultivations, and a 0.05 p-value threshold was set as a strict criterion for the verifiability of the corresponding markers at 95 % confidence at the starting time of both cultivations. Thus, detected features that were successful in surpassing all three knockout criteria were considered as potential biomarkers of the cannabis cultivars adaptability to tropical stress, which are showed in *Table S9* of SM. 5 potential biomarker candidates were determined in the polar phase extract of the leaf tissue, and 14, in the corresponding non-polar case. A higher number of candidates were expected to be identified in the non-polar phase extract, as the previous cross-validated classification performance parameters were significantly better than in the polar extract, fact that turned out to be true. Moreover, 2 out of the 5 candidate metabolites from the polar phase extract, Nicotinic acid riboside and D-Ribitol-5-phosphate, could not be identified as putative metabolites (2nd identification confidence level according to *Schrimpe-Rutledge et al.* [71]) by *in-silico* fragmentation, so they were annotated as tentative metabolites (3rd identification confidence level). On the contrary, the 14 candidates determined in the non-polar phase extract could all be identified as putative metabolites at 2nd identification confidence level.

The significance of the candidate metabolic markers identified in the leaf tissue was put into perspective with their corresponding relevance among both adaptability classes in the flower tissue, in order to further validate their assessment as potential metabolic markers throughout the whole cultivation time, regardless of the studied tissue. This way, an additional exclusion filter was applied to the marker candidate metabolites during the flowering stage. Thus, focusing on *Figure 4*, a clear assessment could be made stating that the 5 marker candidates detected in the polar phase extract of the leaf tissue have an utter lack of relevance for the distinction between the Good and Bad adaptability classes in the floral tissue, which consequently lead them to removal from the candidate list for the determination of biomarkers of the adaptability of cultivars to tropical environments. As a result, the polar phase extract of the plant samples was declared irrelevant for the defined objective.

Once the polar phase extract was excluded from further research, attention was focused on the non-polar phase extract, since various of the present candidates did indeed fulfil the previously defined criteria for being deemed potential markers. However, to ascertain the total generalizability and transferability of the subsequent results, metabolites consisting of just carbon and hydrogen (C_XH_Y molecular formula) were also removed from the potential markers list of the non-polar phase extract, due to their limited ionisability through HESI, related to their hypothetical restricted detectability [75]. This way, metabolite candidates that did not fulfil the minimum compulsory significance for the classification between the Good and Bad adaptability classes in the flower tissue (4-methoxy-3-(3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraenyl)-benzene-1,2-diol; N-[(2S,3S,6R)-2-(hydroxymethyl)-6-(2-oxo-2-piperidin-1-ylethyl)-3,6-dihydro-2H-pyran-3-yl]-1,3-benzodioxole-5-carboxamide and Xanthatin), and reported hydrocarbon based metabolites (3-Allylcyclohexene; γ-Curcumene; Ectocarpene; Curcumene and 1,3,5-Heptatriene) were consequently eliminated from the potential marker list, narrowing the list of potential biomarkers of the cannabis cultivars adaptability to tropical environments the remaining 6: Piperochromenoic acid; Mammea C/BB; Dalbergichromene; Dihydrocordoin; Hedychilactone D and 3-[8,13-bis(ethenyl)-18-(1-hydroxy-3-methoxy-3-oxopropyl)-3,7,12,17-tetramethyl-22,23-dihydroporphyrin-2-yl]propanoic acid.

<u>Figure 4:</u> Exclusion parameters of the determined potential biomarker candidates for the adaptability of cannabis cultivars to tropical environments in a) Polar extracts b) Non-polar extracts

In this line, to ensure that the resultant 6 potential biomarkers were correctly identified, their chromatographic confirmation was assessed by the retention time index (RTI) calculation through the OTrAMS uncertainty measurement methodology developed by the University of Athens (UOA-RTI platform) [76], [77]. Therefore, in essence, a two-factor verification was performed for the identification of the candidate metabolites under consideration. As a result, as stated in *Table 3*, from the 6 potential biomarker candidates, Piperochromenoic acid, Mammea C/BB and Dalbergichromene surpassed the RTI at 1st level, meaning the uncertainty between their experimental and predicted retention times (RT) was completely accepted, resulting in highly likely identification; the RT of Dihydrocordoin was acknowledged at 2nd level, which meant that it was accepted although there was an error between the experimental and predicted RT; and the RT uncertainty of Hedychilactone D and 3-[8,13-bis(ethenyl)-18-(1-hydroxy-3-methoxy-3-oxopropyl)-3,7,12,17-tetramethyl-22,23-dihydroporphyrin-2-yl]propanoic acid turned out to classify as not reliable at 3rd level, fact that could derive in two conclusions. On one hand, it could happen that these metabolites were not exactly the annotated metabolites, which is an unlikely fact since the experimental MS2 fragmentation spectra of both candidates perfectly matched with their theoretical respective (*Figure S5* of SM). Nonetheless, on the other hand, the chromatographic method could have not been the most suitable for these two compounds, and this could be the reason why the uncertainty between their experimental and predicted RT resulted to be considerable, as their RT were the highest among the 6 candidates. Therefore, in possible future works related to this research, it is

suggested to improve the chromatographic analysis method for the analysis of very low polarity compounds in the non-polar extracts.

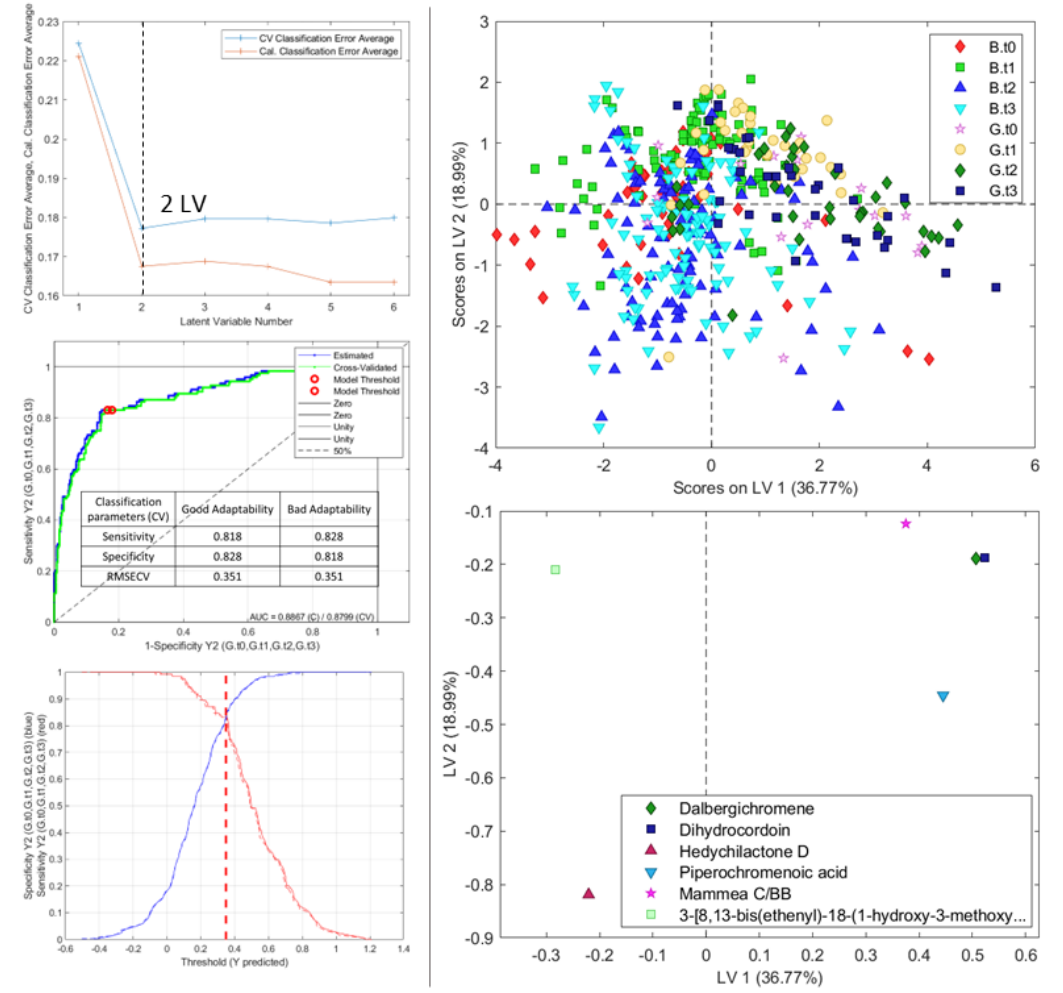
Table 3: Analytical features of the determined potential constitutive biomarkers of the adaptability of cannabis cultivars to tropical stress

Metabolite	Calculated molecular weight	m/z	Mass error (ΔDa)	ICL*	RT (min)	RTI	Metabolite class	
1) Piperochromenoic acid	340.20350	341.21078	-0.00035	2	14.42	box1	Bicyclic monoterpenoid	Secondary
2) Mammea C/BB	400.22462	401.23189	-0.00036	2	14.31	box1	Coumarin	
3) Dalbergichromene	254.09406	255.10134	-0.00023	2	11.01	box1	Neoflavonoid	
4) Dihydrocordoin	310.15653	311.16381	-0.00036	2	11.01	box2	Dihydrochalcone	
11) Hedychilactone D	330.18284	331.19011	-0.00027	2	15.40	box3	Labdane diterpenoid	
14) 3-[8,13-bis(ethenyl)-18-(1-hydroxy-3-methoxy-3-oxopropyl)-3,7,12,17-tetramethyl-22,23-dihydroporphyrin-2-yl]propanoic acid	592.26848	593.27576	-0.00009	2	16.89	box3	Porphyrin	Primary

*ICL: Identification Confidence Level

3.4. External classification performance validation of the potential biomarkers (2nd and 1st cultivations)

Having annotated the potential biomarkers metabolites through the two-factor identification methodology, a 2 LV PLS-DA model was trained with just the autoscaled data of the 6 potential biomarkers across the leaf samples of the 2nd cultivation, in order to assess the trueness of their subsequent performance through external validation for the Good and Bad adaptability classes discrimination (*Figure 6*).



<u>Figure 6:</u> Figures of merit of the PLS-DA model trained with only the potential biomarkers data. From left to right: PLS-DA LV number determination; ROC curve and estimated sensitivity and specificity responses of the Good adaptability class; and PLS-DA scores and loadings of the two adaptability classes in the 1st and 2nd LV

Hence, the predictive classification performance of the trained PLS-DA models was compared through external validation, carrying out the prediction of the non-polar extracts of the 1st cultivation leaf samples of the corresponding harvesting time-points. Therefore, the external validation results of the previously trained PLS-DA model with all the detected chromatographic features (*See section 3.3.*), and the model trained with just the determined potential biomarker data, were contrasted. The prediction results are summed up in *Table 4*, and total prediction results are present in *Tables S10* and *S11* of SM.

Table 4: Prediction/External validation results of the 1st cultivation non-polar phase extract of the leaf samples through the PLS-DA classification models trained with the corresponding 2nd cultivation data

		SFT1 (Bad adaptability)	SFC1 (Good adaptability)	Overall
All features	Correct predictions	45/46	0/48	45/94
	Trueness (%)	97.83	0.00	47.87
Potential biomarkers	Correct predictions	39/46	14/48	53/94
	Trueness (%)	84.78	29.17	56.38

The prediction outcomes shown in *Table 5* suggested that the metabolites detected in the non-polar phase extract of the leaf, selected as potential biomarkers of the adaptability of cannabis cultivars to tropical stress surroundings, correctly performed as such. It is true that in the case of the PLS-DA model trained with the information of all the detected chromatographic features, the correct predictions for the cultivar belonging to the Bad adaptability (SFT1) class resulted to be almost perfect (45/46), to the detriment of the model trained with only the potential biomarker data (39/46). Nevertheless, the improvement of the model trained with just the potential biomarker data in the prediction of the cultivar belonging to the Good adaptability class is undeniable (14/48), as it resulted impossible to correctly predict any sample of the cultivar belonging to the Good adaptability class through the PLS-DA including all the data (0/48). Thus, there were two main reasons for this to happen. On the one hand, a key issue to consider was the statistical weigh of each class present in the model training. The Good adaptability class was only composed with the samples of 3 cultivars (SFC2, SFC1 and SFT4), while the Bad adaptability class consisted of the rest 9 cultivars (SFT2, SFT5, SFT8, SFB1, SFT6, SFT1, SFT7, SFT3 and SFB2), as defined in section 3.2. This meant that the Bad adaptability class possessed the triple statistical weight than the corresponding to the Good adaptability class, which, a priori, favours the balance of the predictive classification performance towards the Bad adaptability class. However, the results of the external validation showed the opposite, from where it could be assessed that all the rest of the detected chromatographic features interfered in the prediction of the 1st cultivation samples, worsening the overall trueness, meaning they acted as noise towards the predictive power of the model including very chromatographic feature. On the other hand, the number of LVs composing each predictive model should also be considered. The model involving every detected feature was cumbersome, that is why it needed of 6 LV to reach the minimum cross-validated classification average error, in comparison with the potential biomarkers based one, which with just 2 LV, achieved it. Therefore, the model based on just the potential biomarkers turned out to be simpler in both terms of chromatographic feature variables and latent variables, which made it a considerably simpler model and, consequently, significantly reduced the probability of misprediction. Considering these facts, the improvement with just the potential biomarker data is evident, so the stated metabolites were validated as such.

Hence, considering the data used for training the referred PLS-DA classification models, these results suggest that the prediction results in the external validation would have be even better, if the 2nd cultivation had been carried out with more cultivars and the statistical weight difference between the Good and Bad adaptability classes had been lower or null, which would further confirm the status of Piperochromenoic acid, Mammea C/BB, Dalbergichromene,

Dihydrocordoin, Hedychilactone D and 3-[8,13-bis(ethenyl)-18-(1-hydroxy-3-methoxy-3-oxopropyl)-3,7,12,17-tetramethyl-22,23-dihydroporphyrin-2-yl]propanoic acid as biomarkers of the adaptability of cannabis cultivars to tropical stress.

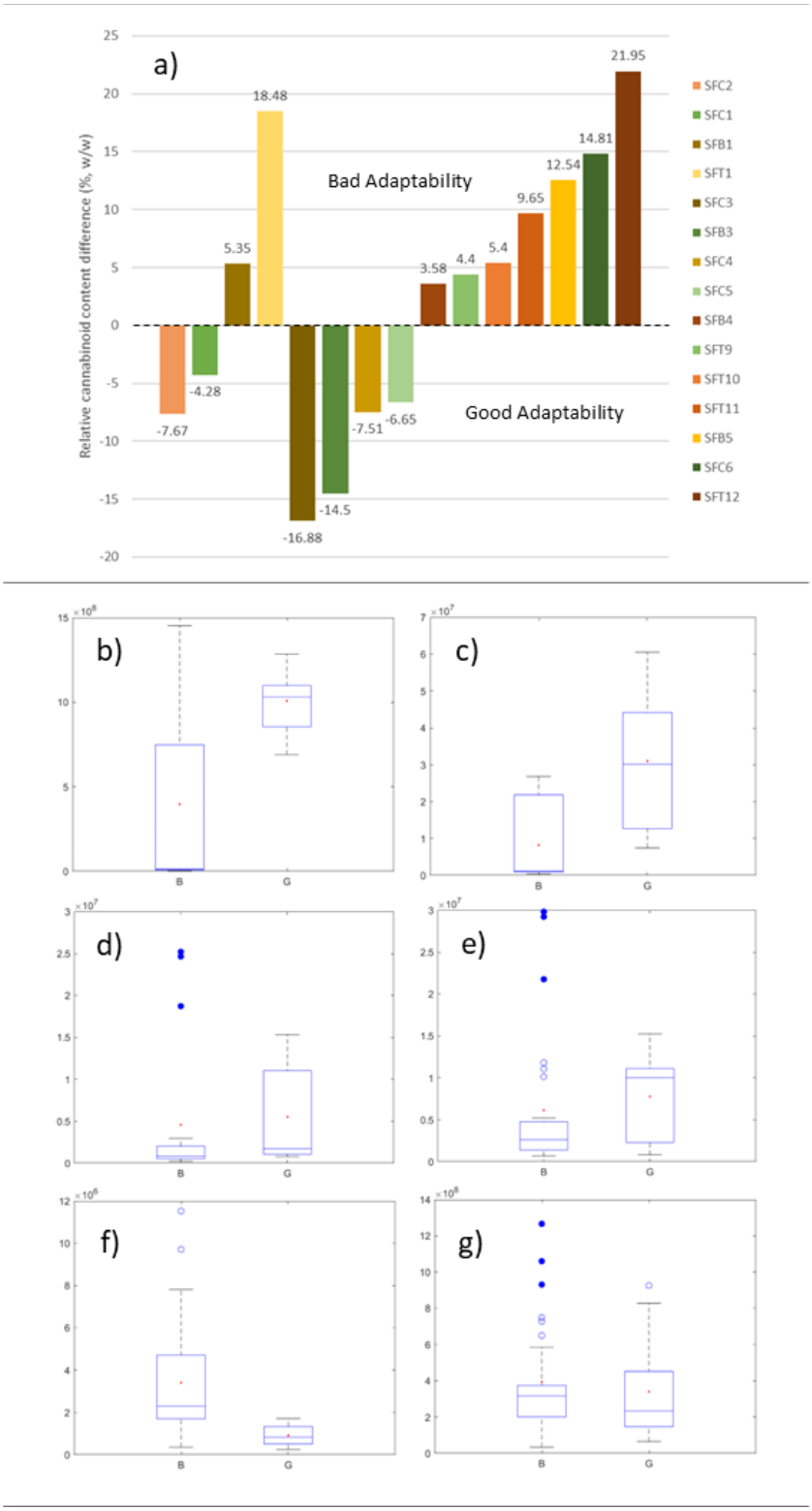
3.5. Qualitative validation of metabolic markers (3rd cultivation)

The cannabis cultivars grown in the 3rd cultivation were used for qualitative validation of the obtained results, with the aim of acknowledging their total transferability to unknown cultivars. Hence, untargeted metabolomic analysis of the cultivated varieties was performed in the leaf tissue of the plantlets at t_0 stage and the significance of the signal difference of the constitutive metabolic markers determined between the Good and Bad adaptability cultivars was assessed (raw data is shown in *Table S12* of SM). As previously stated, new cultivars were introduced in this cultivation, therefore their adaptability also needed to be determined by means of cannabinoid analysis in the inflorescences of their corresponding plant individuals grown in Control and Tropical Stress environments at the end of the cultivation time.

Figure 7 shows the relative cannabinoid content difference of each cultivar. SFC1 and SFT1 cultivars were previously studied in the 1st and 2nd cultivations, and SFC2 and SB1, in the 2nd. As it can there be observed, the content differences of the previously studied cultivars turned out to be different from those obtained in the 2nd cultivation (*See section 3.2.*). The reason that may lie behind this fact is that the overall difference in environmental temperature and humidity between the Control and Tropical Stress rooms was not as severe as in the 2nd Cultivation, as it can be observed in *Figures S3* and *S4* of SM. It was for this reason that the relative differences in cannabinoids also resulted to decrease. At first glance, this might seem a deviation in the results; nevertheless, since the determined metabolic markers were deemed as constitutive, the relative difference in the total cannabinoid content did not affect the reported conclusions, since the biomarkers were ascertained in a prior stage to the abiotic Control/Stress induction. Therefore, taking as a reference the relative cannabinoid content difference of the previously studied cultivars, the Good/Bad adaptability distinction threshold was reduced to 0%. According to this, previously studied SFC2 and SFC1 cultivars, and newly introduced SFC3, SFB3, SFC4 and SFC5 were assessed as Good adaptability cultivars, while the rest were not considered to adapt correctly to the simulated tropical environment.

Considering this, content levels of markers identified in both the Good and Bad adaptability cultivars, as depicted in Figure 7, exhibited a similar behaviour to what was observed in the 2nd and 1st Cultivations (See *Figure S5* of SM). It should be noted that in this last external validation, 3-[8,13-bis(ethenyl)-18-(1-hydroxy-3-methoxy-3-oxopropyl)-3,7,12,17-tetramethyl-22,23-dihydroporphyrin-2-yl]propanoic acid did not show such a significant content difference between the cultivars of Good and Bad adaptability. Bearing in mind that the difference of this biomarker slightly reduced from the 1st to the 2nd cultivation among both cultivar classes, it might result that, as the variety diversity increased and differed from the training and testing data sets, this metabolic marker moved away from the spotlight in this qualitative validation cultivation. Nevertheless, as the results in the 2nd and 1st cultivations confirmed its significance, its role as a metabolic marker can neither be denied, although further research should be carried out to utterly confirm it.

So, this 3rd cultivation, ultimately validated the role of these metabolites as constitutive metabolic markers for assessing the adaptability of cannabis cultivars to tropical environments, even in unrelated cultivation scenarios, at least in the case of the secondary metabolites.



<u>Figure 7:</u> a) Relative cannabinoid difference between Control and Tropical Stress inflorescences at the end of the 3rd cultivation, weighed towards the cannabinoid production of each cultivar in the Control environment. Raw signal of each determined biomarker in both G and B classes at t_0 harvesting time: b) Piperochromenoic acid c) Mammea C/BB d) Dalbergichromene e) DIhydrocordoin f) Hedychilactone D g) 3-[8,13-bis(ethenyl)-18-(1-hydroxy-3-methoxy-3-oxopropyl)-3,7,12,17-tetramethyl-22,23-dihydroporphyrin-2-yl]propanoic acid

3.6. Comprehension of the physiological demeanour of the determined metabolic markers

The 6 potential metabolic markers assessed in this study had one chemical factor in common, which was their low polarity. Nevertheless, the physiological role of each of them in the adaptive metabolism of the cannabis plant species towards the tropical stress turned out to be fundamentally diverse, judging by their respective metabolite class. A brief glance to the PLS-DA loadings plot in section 3.4. is enough to infer behavioural differences among the determined biomarkers, given that two of the six selected compounds, Hedychilactone D and the porphyrinic metabolite specifically, appeared in opposition to the other 4 biomarkers over the 1st LV axis, therefore them resulting in inverse correlation towards the Good adaptability class. In other words, the two stated metabolites appeared in greater overall contents in the samples corresponding to the Bad adaptability class, while the levels of the other four metabolites were higher in the cultivars of Good adaptability (see *Figure S5* of SM). Nevertheless, those two compounds did not belong to the same metabolite classes. In fact, of the 6 determined biomarkers, 5 were pondered as secondary metabolites, including Hedichilactone D, while the porphyrinic compound was claimed as a primary metabolite (*see Table 3, section 3.3.*) [45]. This meant that the porphyrinic biomarker possesses a metabolic role which is directly related to the development and growth of the plant [78], unlike the rest of secondary metabolites, whose presence and corresponding contents are inherently adjusted in the metabolism of the plant with the purpose of forcing an adaptive response towards the simulated uneasy scenario [45].

3.6.1. The porphyrinic compound, a primary metabolite

This metabolite is the demetallised derivative of 13^{1}-hydroxy-magnesium-protoporphyrin IX 13-monomethyl ester metabolite. This protoporphyrin is a primary metabolite which actively participates in the porphyrin metabolism of plants, being an intermediate in the biosynthesis of chlorophyll (*Figure S6* and *S7* of SM), therefore, it is an essential compound for the correct functioning of photosynthesis, which is one of the fundamental biological processes for the necessary growth and development of the plants. Photosynthesis is known to be a highly sensitive metabolic process to high temperature, resulting in reduced metabolic yield, increased reactive oxygen species (ROS) release and augmented production of secondary metabolites [79], [80]. Indeed, high temperatures can trigger the degradation of chlorophyll, leading to a lower light-capturing efficiency [81], [82]. In addition, high humidity alone is not believed to trigger an adaptive mechanism related to chlorophyll, as this abiotic factor does not directly affect its stability and its related cellular processes; nonetheless, it could exacerbate the impact of high temperature in the photosynthetic process, since high humidity can certainly affect the stomatal closure rate in leaves due to direct affection to endogenous abscisic acid (ABA) levels [83]–[85], which is the phytohormone responsible of regulating this adaptive physiological process (*Figure S8* of SM). Therefore, this side effect could end up in a possible increase in the cellular oxidative stress caused by the high temperature.

Thus, considering the stated facts, and that greater constitutive levels of this metabolite derivative were found in the cultivars belonging to the Bad adaptability class, it could be deduced that lower inherent contents of this compound could favour the corresponding cultivars' adaptability towards high temperature and humidity. This could happen due to a higher demetallisation rate of the 13^{1}-hydroxy-magnesium-protoporphyrin IX 13-monomethyl ester metabolite in the cultivars belonging to the Bad adaptability class, therefore its metabolic function would result further disrupted towards the biosynthesis of chlorophyll. Consequently,

lower photosynthetic rate would occur in the Bad adaptability cultivars, hence yielding lower amounts of organic matter through this metabolic process (*Eq. 2*).

$$Eq.\,2\colon 6\,CO_2 + 6\,H_2O \rightarrow C_6H_{12}O_6 + 6\,O_2$$

Hence, as this organic matter is the raw ingredient for the biosynthesis of many of the metabolites and compounds present in the living being, a greater chemical yield in *Eq. 2* could be the source of an enhanced cannabinoid production in the cultivars of Good adaptability to tropical stress.

3.6.2. The secondary metabolites

The trend of the five secondary metabolites determined as biomarkers turned to be heterogeneous, as each of them belonged to a distinct metabolite class (*see Table 3, section 3.3.*). Moreover, as previously stated, Hedichilactone D showed higher overall levels in the cultivars of Bad adaptability, suggesting that the subsequent adaptive mechanism triggered by this secondary metabolite was different from the other four.

In this regard, piperochromenoic acid, mammea C/BB metabolite, and dalbergichromene were all chromene-derivatives. More specifically, they belonged to the bicyclic monoterpenoid, coumarin and neoflavonoid metabolite classes, respectively (*Figure S9* of SM). Dihydrocordoin, on the other hand, was deemed a dihydrochalcone. Regardless of the specific metabolite class to which each compound belonged, these biomarkers surely would play an important metabolic role for the successful adaptability of the plants to high temperature and humidity stress, behaving as defensive metabolic agents. As constitutive secondary metabolites present in leaf tissue, they might inhibit the generation of ROS and further reduce their presence in the biological tissue once they appear, as many antioxidant flavonoids are believed to behave in the chloroplasts [86], [87]. Moreover, similarly to phenolic metabolites, these might activate antimicrobial and antioxidant countermeasures as well, to protect the plant from pathogenic infections and toxic effects of ROS, as consequent up-regulation of genes of the phenylpropanoid pathway occurs for increased phenolic accumulation in response to abiotic stress (*Figure S10* of SM) [88], [89]. Bearing in mind that high temperature and humidity negatively affects the photosynthesis, leading to a consequence such as an increased amount of ROS, a higher in-built content of these secondary metabolites would reasonably enhance the adaptability of cultivars to such scenario, as its adaptive capacity would be inherently strengthened to cope with it.

In the opposite way, as a labdane diterpenoid, a greater presence of Hedychilactone D in the cultivars of Bad adaptability breaks the observed pattern of the identified secondary metabolites. However, this fact may make sense if a deeper look is taken into their metabolic role. Indeed, diterpenoid labdanes are source compounds for the biosynthesis of Geranyl pyrophosphate (GPP) and *ent*-Kaurene through the MEP/DOXP and Mevalonate pathways (*Figure S11, S12* and *S13* of SM) [90], [91], which are the main intermediates for the Gibberellin hormone biosynthesis (*Figure S14 of SM*). Gibberellin is an essential primary metabolite for the correct stem growth, i.e. adecuate plant growth and development (*Figure S8* of SM). In this line, it was observed that, in the t_0 leaf tissues of either the 2nd or 1st cultivations, the content of gibberellin A15 (in the closed lactone form) was higher in the varieties of Good adaptability (*Figure S15* of SM), therefore, it could be deduced that the content of the labdane diterpenoid biomarker was lower in the Good adaptability cultivars due to a higher constitutive yield in the

biosynthesis of this hormone, fact that would mean that the cultivars of Good adaptability were inherently better prepared to face the unfavourable growth scenario.

4. Conclusions

This research work showed that, through three experimental cultivations of different cannabis cultivars grown in control and tropical stress environments, and untargeted metabolomics analysis of strategically harvested plant samples by HPLC-qOrbitrap, it was possible to determine 6 constitutive metabolites with predictive capacity for determining the adaptability of cannabis cultivars to tropical environments, i.e. high temperature and humidity. This objective was achieved by exploring the suitability of different multivariate data-analysis approaches, defining adequate predictive variable settings of the PLS-DA models and identifying potential markers by strict exclusion criteria set by VIPs and p-values calculations.

Hence, of the 6 metabolites identified as constitutive metabolic markers with predictive capacity, one resulted to be primary and the other five, secondary. According to the obtained untargeted results, it was deducted that lower inherent content of the demetallised porphyrinic primary metabolite (3-[8,13-bis(ethenyl)-18-(1-hydroxy-3-methoxy-3-oxopropyl)-3,7,12,17-tetramethyl-22,23-dihydroporphyrin-2-yl]propanoic acid) in the cultivars of Good adaptability triggered a primary metabolic adaptive mechanism towards the abiotic stress provoked by the high temperature and humidity, involving the chlorophyll biosynthetic pathway through porphyrin metabolism and subsequent photosynthetic process. The other 5 secondary metabolites triggered adaptive metabolic mechanisms based on minimising cellular damage of the plant, mainly based on the scavenging of the increased ROS amount released due to the stressful synergetic effect caused by high environmental temperature and humidity. The inherent contents of 4 of them (Piperochromenoic acid, Mammea C/BB, Dalbergichromene y Dihydrocordoin) resulted to be higher in the cultivars of Good adaptability, while the content of Hedychilactone D was lower due to a higher constitutive biosynthetic yield of Gibberellin A15, thereby improving the readiness of the cultivars of Good adaptability to face the severe abiotic conditions.

Therefore, by means of statistically based strict exclusion criteria, and predictive and qualitative external validations of the results, it was empirically demonstrated that the metabolites identified as potential constitutive markers with predictive capacity to determine the adaptability of medical cannabis cultivars to tropical environments did indeed possess such attribute, thereby fulfilling the primarily targeted goal of this research work.

Thus, the future perspective of this research is based on the premise of developing a targeted analysis method of the identified metabolic markers for their early quantification in the leaf tissue, in order to determine their later adaptability to tropical environments before the plants are even introduced into the flowering stage, therefore assessing, by quantitative decision-making criteria, the adequate quality control assurance in medical cannabis cultivations in countries with tropical climates.

Supplementary Material

The Supplementary Material of Chapter 2 is available in the link or in the QR below:

Table S1: 1st Cultivation Metabolomics Data. Leaf, Polar

Table S2: 1st Cultivation Metabolomics Data. Leaf, Non-polar

Table S3: 1st Cultivation Metabolomics Data. Flower, Polar

Table S4: 1st Cultivation Metabolomics Data. Flower, Non-polar

Table S5: 2nd&1st Cultivations Metabolomics Data. Leaf, Polar

Table S6: 2nd&1st Cultivations Metabolomics Data. Leaf, Non-polar

Table S7: 2nd&1st Cultivations Metabolomics Data. Flower, Polar

Table S8: 2nd&1st Cultivations Metabolomics Data. Flower, Non-polar

Table S9: Candidate metabolic markers determined in both phase extracts of leat tissue

Table S10: PLS-DA predictions. External validation results. All metabolites including model

Table S11: PLS-DA predictions. External validation results. Only potential markers model

Table S12: 3rd Cultivation Metabolomics Data. Leaf, Non-polar

Figure S1: 1st Cultivation Temperature and Humidity data (Control room)

Figure S2: 1st Cultivation Temperature and Humidity data (Tropical stress room)

Figure S3: Temperature monitorisation in the 2nd and 3rd cultivations

Figure S4: Humidity monitorisation in the 2nd and 3rd cultivations

Figure S5: Analytical features of the determined potential markers

Figure S6: Porphyrin metabolism

Figure S7: Biosynthesis of plant secondary metabolites

Figure S8: Plant hormone signal transduction

Figure S9: Biosynthesis of various plant secondary metabolites

Figure S10: Pheylpropanoid biosynthesis

Figure S11: Terpenoid backbone biosynthesis

Figure S12: Diterpenoid biosynthesis

Figure S13: Biosynthesis of terpenoids and steroids

Figure S14: Biosynthesis of plant hormones

Figure S15: Gibberelin A15 (closed lactone form) analytical features

Bibliography

[1] A. I. Fraguas-Sánchez and A. I. Torres-Suárez, 'Medical Use of Cannabinoids', *Drugs*, vol. 78, no. 16, pp. 1665–1703, Nov. 2018, doi: 10.1007/s40265-018-0996-1.

[2] P. F. Whiting *et al.*, 'Cannabinoids for medical use: A systematic review and meta-analysis', *JAMA - Journal of the American Medical Association*, vol. 313, no. 24, pp. 2456–2473, 2015, doi: 10.1001/jama.2015.6358.

[3] 'State Medical Cannabis Laws'. Accessed: Jun. 02, 2023. [Online]. Available: https://www.ncsl.org/health/state-medical-cannabis-laws

[4] 'Recommendation of removal for cannabis and cannabis resin from Schedule IV of 1961 Single Convention on Narcotic Drugs', presented at the United Nations Comission on Narcotic Drugs, Vienna, Dec. 2020. Accessed: Jun. 02, 2023. [Online]. Available: https://www.who.int/news/item/04-12-2020-un-commission-on-narcotic-drugs-reclassifies-cannabis-to-recognize-its-therapeutic-uses

[5] K. Riboulet-Zemouli and M. A. Krawitz, 'WHO's first scientific review of medicinal Cannabis: from global struggle to patient implications', *Drugs, Habits and Social Policy*, vol. 23, no. 1, pp. 5–21, Jan. 2022, doi: 10.1108/DHS-11-2021-0060.

[6] *Kontrollierte Abgabe von Cannabis: Eckpunktepapier der Bundesregierung liegt vor*. 2022. Accessed: Jun. 02, 2023. [Online]. Available: https://www.bundesgesundheitsministerium.de/ministerium/meldungen/kontrollierte-abgabe-von-cannabis-eckpunktepapier-der-bundesregierung-liegt-vor.html

[7] *Federal Act on Narcotics and Psychotropic Substances (Narcotics Act, NarcA)*. 2022. Accessed: Jun. 02, 2023. [Online]. Available: https://www.fedlex.admin.ch/eli/cc/1952/241_241_245/en

[8] O. Aizpurua-Olaizola, I. Elezgarai, I. Rico-Barrio, I. Zarandona, N. Etxebarria, and A. Usobiaga, 'Targeting the endocannabinoid system: future therapeutic strategies', *Drug Discovery Today*, vol. 22, no. 1, pp. 105–110, 2017, doi: 10.1016/j.drudis.2016.08.005.

[9] V. Di Marzo, 'New approaches and challenges to targeting the endocannabinoid system', *Nature Reviews Drug Discovery*, vol. 17, no. 9, pp. 623–639, 2018, doi: 10.1038/nrd.2018.115.

[10] C. Muller, P. Morales, and P. H. Reggio, 'Cannabinoid ligands targeting TRP channels', *Frontiers in Molecular Neuroscience*, vol. 11, 2019, doi: 10.3389/fnmol.2018.00487.

[11] V. Reddy *et al.*, 'Targeting the endocannabinoid system: a predictive, preventive, and personalized medicine-directed approach to the management of brain pathologies', *EPMA Journal*, vol. 11, no. 2, pp. 217–250, 2020, doi: 10.1007/s13167-020-00203-4.

[12] J. Tam, L. Hinden, A. Drori, S. Udi, S. Azar, and S. Baraghithy, 'The therapeutic potential of targeting the peripheral endocannabinoid/CB1 receptor system', *European Journal of Internal Medicine*, vol. 49, pp. 23–29, 2018, doi: 10.1016/j.ejim.2018.01.009.

[13] L. Cristino, T. Bisogno, and V. Di Marzo, 'Cannabinoids and the expanded endocannabinoid system in neurological disorders', *Nature Reviews Neurology*, vol. 16, no. 1, pp. 9–29, 2020, doi: 10.1038/s41582-019-0284-z.

[14] M. Maccarrone *et al.*, 'Endocannabinoid signaling at the periphery: 50 years after THC', *Trends in Pharmacological Sciences*, vol. 36, no. 5, pp. 277–296, 2015, doi: 10.1016/j.tips.2015.02.008.

[15] P. Berman, L. Sulimani, A. Gelfand, K. Amsalem, G. M. Lewitus, and D. Meiri, 'Cannabinoidomics – An analytical approach to understand the effect of medical Cannabis treatment on the endocannabinoid metabolome', *Talanta*, vol. 219, 2020, doi: 10.1016/j.talanta.2020.121336.

[16] S. Zou and U. Kumar, 'Cannabinoid receptors and the endocannabinoid system: Signaling and function in the central nervous system', *International Journal of Molecular Sciences*, vol. 19, no. 3, 2018, doi: 10.3390/ijms19030833.

[17] A. Villate, M. San Nicolas, O. Aizpurua-Olaizola, M. Olivares, A. Usobiaga, and N. Etxebarria, 'Quantification of Endocannabinoids in Human Plasma', *Methods in Molecular Biology*, vol. 2687, pp. 107–126, 2023, doi: 10.1007/978-1-0716-3307-6_9.

[18] 'Cannabis medicinal', Argentina.gob.ar. Accessed: Jan. 08, 2024. [Online]. Available: https://www.argentina.gob.ar/justicia/derechofacil/leysimple/cannabis-medicinal

[19] C. J. C. F. Fijnaut and B. de Ruyver, *The Third Way: A Plea for a Balanced Cannabis Policy*. BRILL, 2015.

[20] 'Cannabis law and legislation in Colombia | CMS Expert Guides'. Accessed: Jan. 07, 2024. [Online]. Available: https://cms.law/en/int/expert-guides/cms-expert-guide-to-a-legal-roadmap-to-cannabis/colombia

[21] 'Medicinal Cannabis Industry Given Legal Footing - Ministry of Agriculture, Forestry, Fisheries and Rural Transformation'. Accessed: Jan. 07, 2024. [Online]. Available: https://agriculture.gov.vc/agriculture/index.php?option=com_content&view=article&id=419:medicinal-cannabis-industry-given-legal-footing-3&catid=104&Itemid=288

[22] C. M. Lindsay *et al.*, 'Potency trends of cannabis in Jamaica during the period of 2014 to 2020', *Drug Test Anal*, Jun. 2023, doi: 10.1002/dta.3527.

[23] T. G. Administration (TGA), 'Guidance for the use of medicinal cannabis in Australia: Overview | Therapeutic Goods Administration (TGA)'. Accessed: Jan. 08, 2024. [Online]. Available: https://www.tga.gov.au/resources/resource/guidance/guidance-use-medicinal-cannabis-australia-overview

[24] C. Rusenga, G. Klantschnig, N. Carrier, and S. Howell, 'Cannabis policy changes in Africa are welcome. But small producers are the losers', The Conversation. Accessed: Jan. 08, 2024. [Online]. Available: http://theconversation.com/cannabis-policy-changes-in-africa-are-welcome-but-small-producers-are-the-losers-179681

[25] AfricaNews, 'Morocco joins growing list of African countries to legalize cannabis', Africanews. Accessed: Jan. 08, 2024. [Online]. Available: https://www.africanews.com/2021/05/27/morocco-joins-growing-list-of-african-countries-to-legalize-cannabis/

[26] S. Assanangkornchai, K. Thaikla, M. Talek, and D. Saingam, 'Medical cannabis use in Thailand after its legalization: a respondent-driven sample survey', *PeerJ*, vol. 10, p. e12809, Jan. 2022, doi: 10.7717/peerj.12809.

[27] S. A. Bonini *et al.*, 'Cannabis sativa: A comprehensive ethnopharmacological review of a medicinal plant with a long history', *Journal of Ethnopharmacology*, vol. 227, pp. 300–315, 2018, doi: 10.1016/j.jep.2018.09.004.

[28] J. M. McPartland, 'Cannabis Systematics at the Levels of Family, Genus, and Species', *Cannabis and Cannabinoid Research*, vol. 3, no. 1, pp. 203–212, 2018, doi: 10.1089/can.2018.0039.

[29] M. Ren *et al.*, 'The origins of cannabis smoking: Chemical residue evidence from the first millennium BCE in the Pamirs', *Science Advances*, vol. 5, no. 6, 2019, doi: 10.1126/sciadv.aaw1391.

[30] 'History of cannabis', The University of Sydney. Accessed: Jan. 08, 2024. [Online]. Available: https://www.sydney.edu.au/lambert/medicinal-cannabis/history-of-cannabis.html

[31] W. Bateson, *Mendel's Principles of Heredity - A Defence, with a Translation of Mendel's Original Papers on Hybridisation*. Cambridge University Press, 2009.

[32] H. Zhang, J. Zhu, Z. Gong, and J.-K. Zhu, 'Abiotic stress responses in plants', *Nat Rev Genet*, vol. 23, no. 2, Art. no. 2, Feb. 2022, doi: 10.1038/s41576-021-00413-0.

[33] M. Hasanuzzaman *et al.*, 'Reactive oxygen species and antioxidant defense in plants under abiotic stress: Revisiting the crucial role of a universal defense regulator', *Antioxidants*, vol. 9, no. 8, pp. 1–52, 2020, doi: 10.3390/antiox9080681.

[34] A. Raza *et al.*, 'Impact of climate change on crops adaptation and strategies to tackle its outcome: A review', *Plants*, vol. 8, no. 2, 2019, doi: 10.3390/plants8020034.

[35] C. Waszczak, M. Carmody, and J. Kangasjärvi, 'Reactive Oxygen Species in Plant Signaling', *Annual Review of Plant Biology*, vol. 69, pp. 209–236, 2018, doi: 10.1146/annurev-arplant-042817-040322.

[36] T. Isah, 'Stress and defense responses in plant secondary metabolites production', *Biological research*, vol. 52, no. 1, p. 39, 2019, doi: 10.1186/s40659-019-0246-3.

[37] K. Vahdati and C. Leslie, *Abiotic Stress: Plant Responses and Applications in Agriculture*. BoD – Books on Demand, 2013.

[38] D. Jin, K. Dai, Z. Xie, and J. Chen, 'Secondary Metabolites Profiled in Cannabis Inflorescences, Leaves, Stem Barks, and Roots for Medicinal Purposes', *Scientific Reports*, vol. 10, no. 1, 2020, doi: 10.1038/s41598-020-60172-6.

[39] T. Gülck and B. L. Møller, 'Phytocannabinoids: Origins and Biosynthesis', *Trends in Plant Science*, vol. 25, no. 10, pp. 985–1004, 2020, doi: 10.1016/j.tplants.2020.05.005.

[40] M. E. Bergman, B. Davis, and M. A. Phillips, 'Medically useful plant terpenoids: Biosynthesis, occurrence, and mechanism of action', *Molecules*, vol. 24, no. 21, 2019, doi: 10.3390/molecules24213961.

[41] X. Luo *et al.*, 'Complete biosynthesis of cannabinoids and their unnatural analogues in yeast', *Nature*, vol. 567, no. 7746, pp. 123–126, 2019, doi: 10.1038/s41586-019-0978-9.

[42] A. Canarini, C. Kaiser, A. Merchant, A. Richter, and W. Wanek, 'Root exudation of primary metabolites: Mechanisms and their roles in plant responses to environmental stimuli', *Frontiers in Plant Science*, vol. 10, 2019, doi: 10.3389/fpls.2019.00157.

[43] O. Stein and D. Granot, 'An overview of sucrose synthases in plants', *Frontiers in Plant Science*, vol. 10, 2019, doi: 10.3389/fpls.2019.00095.

[44] M. Yin, Y. Zhang, and H. Li, 'Advances in research on immunoregulation of macrophages by plant polysaccharides', *Frontiers in Immunology*, vol. 10, no. FEB, 2019, doi: 10.3389/fimmu.2019.00145.

[45] M. Erb and D. J. Kliebenstein, 'Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites: The Blurred Functional Trichotomy1[OPEN]', *Plant Physiology*, vol. 184, no. 1, pp. 39–52, 2020, doi: 10.1104/PP.20.00433.

[46] A. Crozier, I. B. Jaganath, and M. N. Clifford, 'Phenols, Polyphenols and Tannins: An Overview', in *Plant Secondary Metabolites*, John Wiley & Sons, Ltd, 2006, pp. 1–24. doi: 10.1002/9780470988558.ch1.

[47] F. Thoma, A. Somborn-Schulz, D. Schlehuber, V. Keuter, and G. Deerberg, 'Effects of Light on Secondary Metabolites in Selected Leafy Greens: A Review', *Frontiers in Plant Science*, vol. 11, 2020, doi: 10.3389/fpls.2020.00497.

[48] J. L. Berini *et al.*, 'Combinations of Abiotic Factors Differentially Alter Production of Plant Secondary Metabolites in Five Woody Plant Species in the Boreal-Temperate Transition Zone', *Front. Plant Sci.*, vol. 9, 2018, doi: 10.3389/fpls.2018.01257.

[49] B. Labarrere, A. Prinzing, T. Dorey, E. Chesneau, and F. Hennion, 'Variations of secondary metabolites among natural populations of sub-antarctic ranunculus species suggest functional redundancy and versatility', *Plants*, vol. 8, no. 7, 2019, doi: 10.3390/plants8070234.

[50] A. Piasecka *et al.*, 'Drought-related secondary metabolites of barley (Hordeum vulgare L) leaves and their metabolomic quantitative trait loci', *The Plant Journal*, vol. 89, no. 5, pp. 898–913, 2017, doi: 10.1111/tpj.13430.

[51] A. Villate *et al.*, 'Review: Metabolomics as a prediction tool for plants performance under environmental stress', *Plant Science*, vol. 303, 2021, doi: 10.1016/j.plantsci.2020.110789.

[52] M. San Nicolas *et al.*, 'Exploratory optimisation of a LC-HRMS based analytical method for untargeted metabolomic screening of Cannabis Sativa L. through Data Mining', *Analytica Chimica Acta*, vol. 1279, p. 341848, Oct. 2023, doi: 10.1016/j.aca.2023.341848.

[53] A. M. Rodrigues, A. I. Ribeiro-Barros, and C. António, 'Experimental design and sample preparation in forest tree metabolomics', *Metabolites*, vol. 9, no. 12, 2019, doi: 10.3390/metabo9120285.

[54] T. F. Jorge *et al.*, 'Mass spectrometry-based plant metabolomics: Metabolite responses to abiotic stress: MASS SPECTROMETRY-BASED PLANT METABOLOMICS', *Mass Spec Rev*, vol. 35, no. 5, pp. 620–649, Sep. 2016, doi: 10.1002/mas.21449.

[55] S. Subramaniam, 'Challenges in Biological Data Integration in the Post-genome Sequence Era', in *Data Integration in the Life Sciences*, B. Ludäscher and L. Raschid, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005, pp. 1–1. doi: 10.1007/11530084_1.

[56] J. Lisec, N. Schauer, J. Kopka, L. Willmitzer, and A. R. Fernie, 'Gas chromatography mass spectrometry–based metabolite profiling in plants', *Nat Protoc*, vol. 1, no. 1, pp. 387–396, Jun. 2006, doi: 10.1038/nprot.2006.59.

[57] O. Aizpurua-Olaizola *et al.*, 'Evolution of the Cannabinoid and Terpene Content during the Growth of Cannabis sativa Plants from Different Chemotypes', *Journal of Natural Products*, vol. 79, no. 2, pp. 324–331, 2016, doi: 10.1021/acs.jnatprod.5b00949.

[58] K. Chetnik, L. Petrick, and G. Pandey, 'MetaClean: a machine learning-based classifier for reduced false positive peak detection in untargeted LC–MS metabolomics data', *Metabolomics*, vol. 16, no. 11, 2020, doi: 10.1007/s11306-020-01738-3.

[59] C. Hawkins *et al.*, 'Plant Metabolic Network 15: A resource of genome-wide metabolism databases for 126 plants and algae', *Journal of Integrative Plant Biology*, vol. 63, no. 11, pp. 1888–1905, 2021, doi: 10.1111/jipb.13163.

[60] L. Chae, T. Kim, R. Nilo-Poyanco, and S. Y. Rhee, 'Genomic signatures of specialized metabolism in plants', *Science*, vol. 344, no. 6183, pp. 510–513, 2014, doi: 10.1126/science.1252076.

[61] K. Dreher, 'Putting the plant metabolic network pathway databases to work: Going offline to gain new capabilities', *Methods in Molecular Biology*, vol. 1083, pp. 151–171, 2014, doi: 10.1007/978-1-62703-661-0_10.

[62] L. Chae, I. Lee, J. Shin, and S. Y. Rhee, 'Towards understanding how molecular networks evolve in plants', *Current Opinion in Plant Biology*, vol. 15, no. 2, pp. 177–184, 2012, doi: 10.1016/j.pbi.2012.01.006.

[63] M. J. Conroy *et al.*, 'LIPID MAPS: update to databases and tools for the lipidomics community', *Nucleic Acids Research*, p. gkad896, Oct. 2023, doi: 10.1093/nar/gkad896.

[64] M. Arita and K. Suwa, 'Search extension transforms Wiki into a relational system: A case for flavonoid metabolite database', *BioData Mining*, vol. 1, no. 1, p. 7, Sep. 2008, doi: 10.1186/1756-0381-1-7.

[65] J. Yabuzaki, 'Carotenoids Database: Structures, chemical fingerprints and distribution among organisms', *Database*, vol. 2017, no. 1, 2017, doi: 10.1093/database/bax004.

[66] J. Hastings *et al.*, 'ChEBI in 2016: Improved services and an expanding collection of metabolites', *Nucleic Acids Research*, vol. 44, no. D1, pp. D1214–D1219, 2016, doi: 10.1093/nar/gkv1031.

[67] 'FooDB'. Accessed: Nov. 27, 2023. [Online]. Available: https://foodb.ca/

[68] D. S. Wishart *et al.*, 'HMDB 5.0: The Human Metabolome Database for 2022', *Nucleic Acids Research*, vol. 50, no. D1, pp. D622–D631, 2022, doi: 10.1093/nar/gkab1062.

[69] M. Kanehisa and S. Goto, 'KEGG: Kyoto Encyclopedia of Genes and Genomes', *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000, doi: 10.1093/nar/28.1.27.

[70] T. Jewison *et al.*, 'SMPDB 2.0: Big improvements to the small molecule pathway database', *Nucleic Acids Research*, vol. 42, no. D1, pp. D478–D484, 2014, doi: 10.1093/nar/gkt1067.

[71] A. C. Schrimpe-Rutledge, S. G. Codreanu, S. D. Sherrod, and J. A. McLean, 'Untargeted Metabolomics Strategies—Challenges and Emerging Directions', *J. Am. Soc. Mass Spectrom.*, vol. 27, no. 12, pp. 1897–1905, Dec. 2016, doi: 10.1007/s13361-016-1469-y.

[72] Z. Pang *et al.*, 'Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data', *Nat Protoc*, vol. 17, no. 8, Art. no. 8, Aug. 2022, doi: 10.1038/s41596-022-00710-w.

[73] Z. Pang *et al.*, 'MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights', *Nucleic Acids Research*, vol. 49, no. W1, pp. W388–W396, Jul. 2021, doi: 10.1093/nar/gkab382.

[74] W. E. Johnson, C. Li, and A. Rabinovic, 'Adjusting batch effects in microarray expression data using empirical Bayes methods', *Biostatistics*, vol. 8, no. 1, pp. 118–127, 2007, doi: 10.1093/biostatistics/kxj037.

[75] E. M. Thurman, I. Ferrer, and D. Barceló, 'Choosing between atmospheric pressure chemical ionization and electrospray ionization interfaces for the HPLC/MS analysis of pesticides', *Analytical Chemistry*, vol. 73, no. 22, pp. 5441–5449, 2001, doi: 10.1021/ac010506f.

[76] R. Aalizadeh *et al.*, 'Development and Application of Liquid Chromatographic Retention Time Indices in HRMS-Based Suspect and Nontarget Screening', *Analytical Chemistry*, vol. 93, no. 33, pp. 11601–11611, 2021, doi: 10.1021/acs.analchem.1c02348.

[77] R. Aalizadeh, N. S. Thomaidis, A. A. Bletsou, and P. Gago-Ferrero, 'Quantitative Structure–Retention Relationship Models To Support Nontarget High-Resolution Mass Spectrometric Screening of Emerging Contaminants in Environmental Samples', *J. Chem. Inf. Model.*, vol. 56, no. 7, pp. 1384–1398, Jul. 2016, doi: 10.1021/acs.jcim.5b00752.

[78] L. Taiz, E. Zeiger, I. M. Møller, and A. Murphy, 'Plant physiology and development.', *Plant physiology and development.*, no. Ed. 6, 2015, Accessed: Dec. 23, 2023. [Online]. Available: https://www.cabdirect.org/cabdirect/abstract/20173165866

[79] J. Berry and O. Bjorkman, 'Photosynthetic Response and Adaptation to Temperature in Higher Plants', *Annual Review of Plant Physiology*, vol. 31, no. 1, pp. 491–543, 1980, doi: 10.1146/annurev.pp.31.060180.002423.

[80] S. Mathur, D. Agrawal, and A. Jajoo, 'Photosynthesis: Response to high temperature stress', *Journal of Photochemistry and Photobiology B: Biology*, vol. 137, pp. 116–126, 2014, doi: 10.1016/j.jphotobiol.2014.01.010.

[81] A. Sharma *et al.*, 'Phytohormones regulate accumulation of osmolytes under abiotic stress', *Biomolecules*, vol. 9, no. 7, 2019, doi: 10.3390/biom9070285.

[82] A. K. Tewari and B. C. Tripathy, 'Temperature-stress-induced impairment of chlorophyll biosynthetic reactions in cucumber and wheat', *Plant Physiology*, vol. 117, no. 3, pp. 851–858, 1998, doi: 10.1104/pp.117.3.851.

[83] M. Okamoto, Y. Tanaka, S. R. Abrams, Y. Kamiya, M. Seki, and E. Nambara, 'High humidity induces abscisic acid 8'-hydroxylase in stomata and vasculature to regulate local and systemic abscisic acid responses in Arabidopsis', *Plant Physiology*, vol. 149, no. 2, pp. 825–834, 2009, doi: 10.1104/pp.108.130823.

[84] H. Bauer *et al.*, 'The Stomatal Response to Reduced Relative Humidity Requires Guard Cell-Autonomous ABA Synthesis', *Current Biology*, vol. 23, no. 1, pp. 53–57, Jan. 2013, doi: 10.1016/j.cub.2012.11.022.

[85] T. Kuromori, E. Sugimoto, and K. Shinozaki, 'Intertissue Signal Transfer of Abscisic Acid from Vascular Cells to Guard Cells', *Plant Physiology*, vol. 164, no. 4, pp. 1587–1592, Apr. 2014, doi: 10.1104/pp.114.235556.

[86] G. Agati, E. Azzarello, S. Pollastri, and M. Tattini, 'Flavonoids as antioxidants in plants: Location and functional significance', *Plant Science*, vol. 196, pp. 67–76, Nov. 2012, doi: 10.1016/j.plantsci.2012.07.014.

[87] G. Agati, P. Matteini, A. Goti, and M. Tattini, 'Chloroplast-located flavonoids can scavenge singlet oxygen', *New Phytologist*, vol. 174, no. 1, pp. 77–89, 2007, doi: 10.1111/j.1469-8137.2007.01986.x.

[88] S. Kumar, Md. M. Abedin, A. K. Singh, and S. Das, 'Role of Phenolic Compounds in Plant-Defensive Mechanisms', in *Plant Phenolics in Sustainable Agriculture : Volume 1*, R. Lone, R. Shuab, and A. N. Kamili, Eds., Singapore: Springer, 2020, pp. 517–532. doi: 10.1007/978-981-15-4890-1_22.

[89] A. Sharma, B. Shahzad, A. Rehman, R. Bhardwaj, M. Landi, and B. Zheng, 'Response of phenylpropanoid pathway and the role of polyphenols in plants under abiotic stress', *Molecules*, vol. 24, no. 13, 2019, doi: 10.3390/molecules24132452.

[90] R. J. Peters, 'Two rings in them all: The labdane-related diterpenoids', *Natural Product Reports*, vol. 27, no. 11, pp. 1521–1530, 2010, doi: 10.1039/c0np00019a.

[91] T. Toyomasu, M. R. Shenton, and K. Okada, 'Evolution of Labdane-Related Diterpene Synthases in Cereals', *Plant and Cell Physiology*, vol. 61, no. 11, pp. 1850–1859, Nov. 2020, doi: 10.1093/pcp/pcaa106.

3. KAPITULUA:

NIR-IRUDI HIPERESPEKTRAL ETA IKASKUNTZA AUTOMATIKO BIDEZKO *CANNABIS SATIVA L.*-REN KEMOTIPOEN SAILKAPEN EZ-INBADITZAILEA

3. CHAPTER:

NIR-HYPERSPECTRAL IMAGING AND MACHINE LEARNING FOR NON-INVASIVE CHEMOTYPE CLASSIFICATION IN *CANNABIS SATIVA L.*

Computers and Electronics in Agriculture, vol. 217, no. 108551 (2024).

DOI: 10.1016/j.compag.2023.108551

San Nicolas M., Villate A., Alvarez-Mora I., Olivares M., Aizpurua-Olaizola O., Usobiaga A., Amigo J.M.

Laburpena

Erabilpen medizinalerako kannabisak gaur egun jaso duen onarpen-tasa publikoen igoeraren ondorioz, landare espezie honen ekoizpena handitu egin da mundu osoan. Hala ere, eraldatze-prozesuan murgilduta dagoen lege esparruak ez du alde batera uzten kannabisaren kultiborako metodologien ezagutza legez kanpoko jatorriduna izatea, urtetan zehar kalitate-kontroleko arau eta produkzio-prozesu zuzenetako gidalerro gabe emana izan dena. Urtetako kultibo ohitura hauek, gerora, kontrol gabeko produktuak eman litzake emaitza bezala. Ildo horretan, lan honen helburua kannabis barietate desberdinen hazkuntzan zehar haien kemotipoaren sailkapen ez-inbaditzailerako metodologia bat garatzea izan da, kannabisaren produkzio prozesuan duda-gabeko ekoizpen-kontrol garbirako metodologiak aplikatzeko. Horrela, irudi hiperespektralak (HSI), aldagai anitzeko datu analisirako ikuspegi desberdinekin batera, hala nola, osagai nagusien analisia (PCA) eta karratu partzial minimoen analisi diskriminatzailea (PLS-DA), landareen *in-situ* analisi ez-inbaditzailea ahalbidetu zuen. Hori posible egiteko, hiru kimiotioentzako bi PLS-DA sailkapen eredu trebatu ziren landareen datu espektralekin, landare loreen kannabinoideen edukian oinarritutakoak, bi ereduen arteko diferentzia landarearen zurtoin zatia partzialtasun moduan aintzat hartzea izanda. Horrela, loreen datu espektralen bidez lortutako sentikortasun eta espezifikotasun balioak hurrengoak izan ziren: 0.845/0.845 I kemotipoaren kasuan, 0.954/0.920 II kemotipoarentzako, eta 0.888/0.925 III kemotipoarentzako. Azkenik, PLS-DA hierarkiko baten bidez (HPLS-DA), zurtoina interferentzia moduan kontuan hartzen zuena, % 94.7-ko egiazkotasun orokorra lortu zen 57 landare indibiduoren kanpo-balidazioan, hurrengo eran banatuta: % 92.3-ko egiazkotasuna I kemotipoarentzat, % 100.0-koa II kemotipoarentzat eta % 88.9-koa III kemotipoarentzat. Emaitza hauetan oinarrituta, kannabis kultiboen nekazaritza-kontrol integralerako teknika analitiko ez-inbaditzaile baten bitarteko kontzeptu froga gauzatu zen, ordura arte frogatu gabeko jazoera zena. Ondorioz, lan honek erabilpen medizinalerako kannabis laboreen kalitate kontrolerako teknologia ez-inbaditzailearen garapenerako bidea are gehiago erraztu lezake. Izan ere, azaleratzen dabilen industria honek kannabisaren kemotipoak era zorrotzean kontrolatu beharko ditu eta abantaila handia suposatuko luke denbora luzea behar duten teknika analitiko suntsitzaileak saihestea, hala nola, kromatografia.

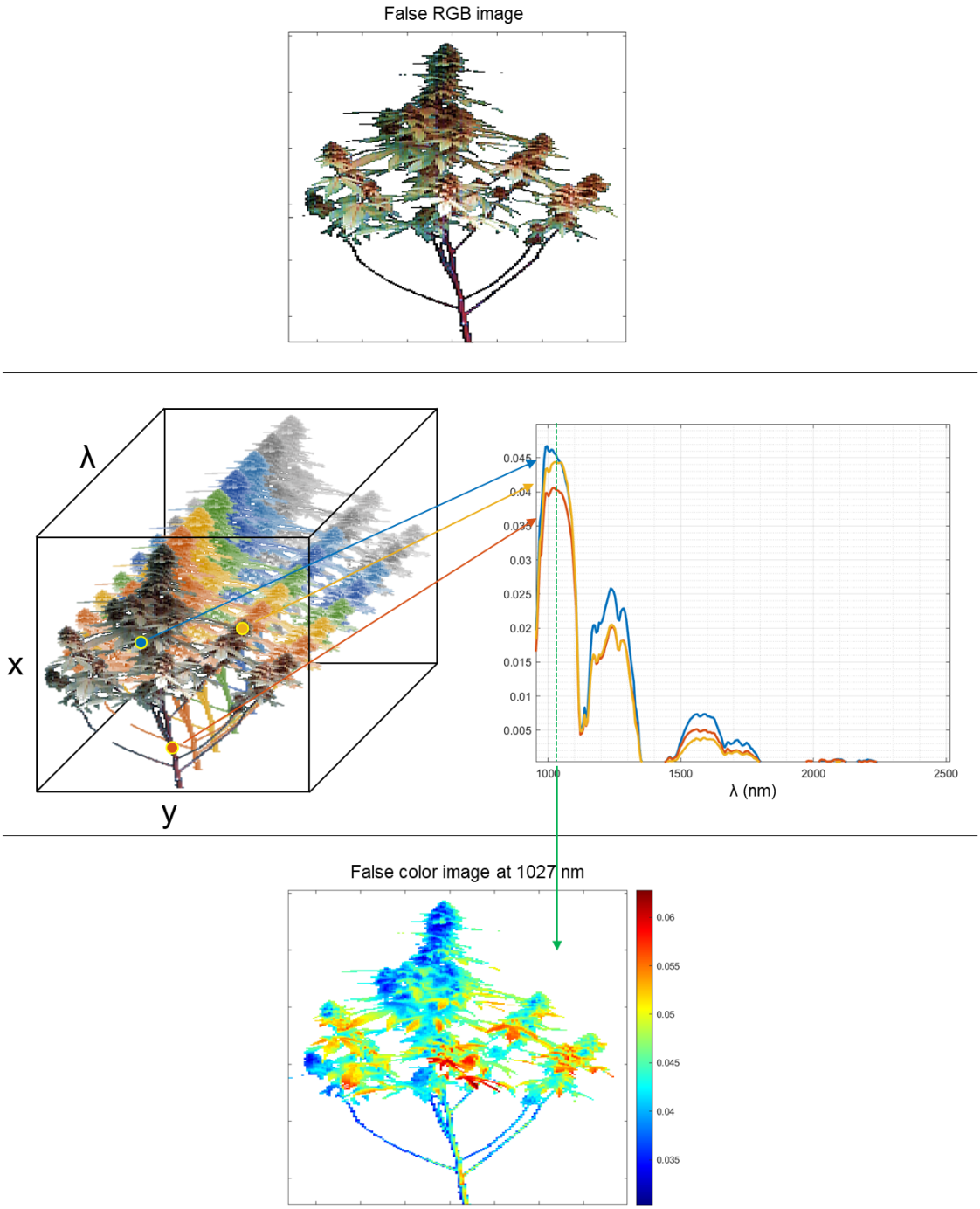
Hitz gakoak: analisi ez-suntsitzailea, kalamua, kannabinoideak, kalitate kontrola, PLS-DA, kanpo-berrespena

1. Sarrera

Azken urteotan, erabilpen medizinaleko kannabisarenganako onarpen instituzionalaren joerak, gaixotasun anitzen edo min aringarrien tratamendurako medikamentu sintetikoen alternatiba mediko gisa, igoera harrigarria izan du [1]–[4]. Honen adibide dira Amerikako Estatu Batuetako (AEB) 38 estatutan erabilpen medikurako baimenaren onarpena, Marihuanaren Arau Medikuen (MML) baiespenaren bitartez edo Droga Narkotikoen Konbentzio Bakarreko IV Agendan kannabis eta kannabis-erretxinaren desklazifikazioa [6], horrela bere potentzial terapeutikoa aitortuz. Jazoera horien ondorioz, Europako hainbat herrialdek kannabisaren erabilera terapeutikoaren aldeko jarrera hartu dute, hala nola, Alemaniak edo Suitzak [7] [8]. Medikuntzan kannabisetik eratorritako produktuak bereganatzeko interes hori kannabinoideetatik dator, batez ere, giza sistema endokannabinoidearekin (ECS) elkarreragiten duten landare-espezie honetako konposatu bioaktibo terpenoideen klasea [9]–[11]. ECS-a giza nerbio-sistemaren parte da, eta funtzio fisiologiko garrantzitsua betetzen du, hainbat prozesutan parte hartzen baitu, hala nola, garunaren plastikotasunaren erregulazioan, garapen neuronalean edo energia-oreka edo apetituaren erregulazioan, besteak beste [4]. Hala ere, kannabinoideen klasearen barruan, konposatu desberdinak biosintetizatzen dira kannabis-landarean, eta horietako bakoitzak modu desberdinean elkarreragiten du ECSren hartzaileekin, ondorio mediko desberdinak lortuz [12]–[14]. Hori dela eta, erabilpen medizinalerako kannabisaren gaineko interesaren hazkundeak agerian utzi du metodo zehatz eta eraginkorren beharrizana landare-ekoizpeneko prozesuan kalitate-kontrola bermatzeko.

Gaur egun, bi kannabinoide dira orokorrean ezagunenak farmazeutikoki-aktibo diren osagai potentzial moduan (API moduan), eta hauek dira kannabis-loreetan kontzentrazio handienetan agertzen direnak ere: Δ^9-tetrahidrokannabinol-a (THC) eta kannabidiol-a (CBD) [16], [17]. Hori dela, eta, kannabinoide hauek aurkitu zirenetik, kannabis barietateak Mendeliar-genetika legeetan oinarrituta elkargurutzatu izan dira, efektu terapeutikoa hobetzeko eta kontsumo-segurtasuna handitzeko THC/CBD ratio onenen bila. Horren ondorioz, gaur egun barietate zenbatezinak aurki daitezke kannabis espezie barruan, baina horietako gehienak hiru kemotipotan sailka litezke, kannabinoide nagusi horien edukiaren arabera [18], [19]; I kemotipoa, THC-a kannabinoide nagusi gisa duena; II kemotipoa, THC eta CBD eduki konparagarriak dituena; eta III kemotipoa, CBD-a kannabinoide nagusi gisa duena [15]. Hala, kannabis medizinalean, barietateak behar bezala sailkatu behar dira kannabinoideen kuantifikazio zehatzaren bidez, non, orokorrean, diodo detektagailuari lotutako kromatografia likidoa (LC-DAD), masa-espektrometriari lotutako kromatografia likidoa (LC-MS) edo sugar-ionizazio bidezko detektagailuari akoplatutako gas kromatografia (GC-FID) erabili izan den ohiko erreminta moduan [20]–[26]. Analisi hauetan prozesuak lagin-tratamendu suntsitzailea dakar, baita laginen prestaketa neketsua edo denbora luze eta baliabide ugari behar duten teknika analitiko garestiak ere. Horiez gain, azken urteotan, infragorri-hurbileko espektroskopiak (NIRS) interes handia erakarri du kannabinoideen analisiari dagokionez, zehaztasun eta doitasun galerarik gabeko lan kostu baxuagoko prozedurei esker [27]–[32]. Hala eta guztiz ere, landare-laginetako hezetasunak muga prozedural bat suposa lezake NIRSa erabiltzeko orduan, ur-molekularen lotura kimikoek NIR eskualdean seinale esanguratsuak aurkezten baitituzte. Beraz, esan liteke, oro har, hezetasunak interferentzia kuantitatiboak edo kualitatiboak eragin ditzakeela metodo honen bidezko kannabinoideen analisian. Hori dela eta, Europar Batasunak (EB) batzorde-araudi bat (2017/1155 zenbakiduna) eskuordetu zuen. Bertan, prozedura esperimental bat deskribatzen zen, kannabinoideen analisirako aurretratamendu moduan kannabis-laginen lehortzea barne hartzen zuena [33], eta, oro har, prozedura hori jarraitzen da

[27], [28], [31], [32], [34], [35]. Ondorioz, nahiz eta NIRS-ak analisi kromatografikoak baino lagin-tratamendu errazagoa behar izan, prozedura esperimental inbaditzailea eskatzen du kannabis-landarean, kannabinoideen analisi adierazgarria egiteko.



<u>1. Irudia:</u> Kannabis-landare baten 930-2500 nm-ko uhin-luzera tarteko NIR-irudi hiperespektralaren bistaratzea

Ildo honetan, irudi hiperespektrala (HSI) kannabisaren kemotipoa zehazteko teknika analitiko ez-inbaditzaile potentzial gisa agertzen da, irudien informazio espazialaren eta espektroskopiaren informazio espekrala batzen dituena, bien abantailak bateratuz [36]. Irudi optikoen bidez, HSI-ak objektuen bistaratze bidimentsionala ahalbidetzen du, pixel bakoitzean espektro elektromagnetiko zabala eskuratzen den bitartean. Horrela, pixel bakoitzean kolore primarioak (gorria, berdea, urdina) irudikatu ordez, pixeletako espektro zabala banda espektraletan banatzen da, eta lehen begirada batean ikus daitekeena baino informazio gehiago eskura daiteke [37]. Horregatik, NIR-HSI-ak, espektroskopiaren bidezko lagin baten konposizio kimikoa zehaztea ahalbidetzeaz gain, laginaren bolumenean zehar duen banaketaren bistaratze bidimentsionala ahalbidetzen du.

Hori dela eta, neurtutako landarearen informazio kimiko guztia jasotzen duen irudi hiperespektral bat hiru dimentsioko matrize gisa ikus daiteke, bi dimentsio espazial dituena, pixeletan (x eta y) banatua, eta dimentsio espektral bat (λ) (*1 Irudia*) [37]. Horregatik, irudi hiperespektralek, normalean, nahaste multikonposatuaren mendekotasuna izaten dute, eta oso gutxitan izaten dituzte osagai espezifikoekin erlazionatutako aldagai espektral selektiboak [37]. Hala ere, arrazoi zehatz horregatik, HSI-ak NIRS analisian interferentziak eragiten dituzten hezetasun-bandekin erlazionatutako aldagai selektiboei aurre egin diezaiekete, landare bizi batean analisi ez-inbaditzailea ahalbidetuz. Baina irudi hiperespektralek zarata espaziala edo espektrala, edo datu erredundanteak ere izan ditzakete, beraz, datuak behar bezala tratatu behar dira lortu nahi diren helburuak behar bezala ondorioztatzeko. Ildo horretan, ikasketa automatikoko planteamenduek hutsune hori bete dezakete eta aldagai anitzeko erronkak kudeatu, datu hiperespektraletatik eredu esanguratsuak atera eta kemotipoen sailkapen eraginkor automatizatua eskaintzeko, horrela landareen ekoizpenean kalitate-kontrola berma dadin kannabis landareeen analisi azkar eta ez-inbaditzailea ahalbidetuz.

Horrez gain, NIR-HSI eta ikasketa automatiko bidezko kannabinoide nagusien kuantifikazioan zentratutako lan berri batzuen *ad hoc* ekarpena aipagarria da. Adibidez, *S. K. Abeysekera et al.*-en lanean [38] NIR-HSI teknologia kannabisaren THCA edukia zehatz balioesteko tresna baliogarria zela frogatu zen. Hau *W. S. Holmes et al.*-ek [39] lortutako ondorioa izan zen ere, baina kasu honetan CBDAri zegokionez, bai loreen, bai hostoen analisian. *Y. Lu et al.*-en [40] ere NIR-HSI-a, ohiko metodo analitikoen aurka, kannabisaren lore-materialeko kannabinoide nagusiak azkar kuantifikatzeko tresna baliagarria izan daitekeela zehaztu zen. Horrela, berriki egindako ikerketa enpirikoek teknologia honek kannabisean hainbat helburutarako duen potentziala erakusten ari dira, parametro analitikoak zalantzan jarri gabe prozedurak sinplifikatuz. Hori dela eta, NIR-HSI bezalako teknika ez-inbaditzaile berritzaileak maizago sartu beharko lirateke kultiboen kalitate kontrolean, landako eta laborategiko lan-kargak murrizte aldera.

Beraz, ikerketa honen helburua zen HSI-ren bideragarritasuna zehaztea izan zen, ikasketa automatikoko planteamendu desberdinekin batera, kannabisaren kemotipoen sailkapen ez-inbaditzailerako. Lan honen irismena kannabisaren kemotipoak sailkatzeko edozein lagin-aurretratamenduren beharra baztertuz metodo zehatz eta fidagarri baten garapena izan da, analisi-prozesua *in-situ* arintzeko, eta horrela, kannabisaren industria medikoan kalitate-kontrola erraztuz. Proposatutako metodologiak kultibo-praktiken hobekuntzarako erabilgarritasun handia dakar kannabisarekin lotutako ikerketan eta baita haren ekoizpenean erabaki egokiak ahalbidetzean ere.

2. Material eta metodoak

2.1. Kannabis landareen kultibazioa

Kannabis landareak Sovereign Fields S.L.-ren instalazioetan kultibatu ziren (Larramendi 3 kalea, Donostia-San Sebastian 20006, Espainia). Landareak 11 L-ko loreontzi beltzetan hazi ziren, lur/hummus/elikagai nahasturatekin. Zehazki, nahasturak Biobizz Worldwide S.L. -ko (Lezama-Leguizamon parke industriala, Gorbeia 11 kalea, Etxebarri 48450, Espainia) "Light mix soil" -a izan zuen % 80ean, eta hummusa % 20ean. Horrez gain, nahastura honi Lurpe Natural Solutions-eko (Zubiate 3 kalea, Lemoa 48330, Espainia) 10 $g \cdot L^{-1}$ "Farmer mix nutrient solution" ere gehitu zitzaion, saguzar guanoz, hezur-irinez, alga-irinez, Azomitea®z, alpalpa organikoaz, intsektu-zaborraz, odol-irinez, dolomitaz, langbeinita azido humikoz eta fulbikoz, eta, errizobakterioz eta trikodermaz osatuta dagoena. Kultibo denbora, guztira, 12 aste izan ziren: lehen 4 asteak landareen etapa begetatiboari zegozkion, landareen hazkuntza-aldia izan zena, eta hurrengo 8 asteak, lorakuntza-etapari. Etapa hauek argialdi erregulazioak definitu zituen: 18 ordu argi/6 ordu ilun etapa begetatiborako, eta 12 ordu argi/12 ordu ilun, lorakuntza-etaparako. Landareen NIR-irudi hiperespektralak kultibo-denborako hamargarren astean eskuratu ziren.

57 landare-banakoen NIR-irudi hiperespektralak eskuratu ziren lan honetan, hurrengo barietateetan banatuta: 10 *Dairy Queen* (DQ), 9 *Futura 75* (FT), 10 *Remedy* (RE), 8 *Roma* (RO), 10 *Tel Aviv* (TA) eta 10 *White Widow* (WW). Landare-banako bakoitza zenbakitua izan zen.

2.2. Kannabinoideen analisia

Kannabis barietate bakoitzaren kemotipoa definitzeko kannabinoideen kontzentrazioa Sovereign Fields S.L.-ren instalazioetan neurtu zen ultramore detektagailu bati akoplatutako kromatografia likido (LC-UV) bidez, *Aizpurua-Olaizola et al.* laneko metodoa jarraituz [15]. Lortutako emaitzak kemotipoa definitzeko erreferentzia gisa erabili ziren eta hurrengo sailkapena zehaztu zen: TA, RO eta DQ barietateak I kemotipo taldean definitu ziren ($[THC_{Totala}]/[CBD_{Totala}] > 10$); RE eta WW barietateak, II kemotipo taldean ($0.3 < [THC_{Totala}]/[CBD_{Totala}] < 3$); eta FT barietatea, III kemotipoan ($[THC_{Totala}]/[CBD_{Totala}] < 0.1$) [18].

2.3. NIR-irudi hiperespektralen eskurapena

Irudiak *in-situ* hartu ziren, eguzki argipean, landa-muntaidun HySpex SWIR 384 kamera hiperespektrala erabiliz (HySpex by NEO, Østensjøveien 34, N-0667 Oslo, Norvegia). Irudiak udako egun berean (2021eko uztailaren 2-an) eskuratu ziren 09:00 eta 14:00 arteko ordu tartean.

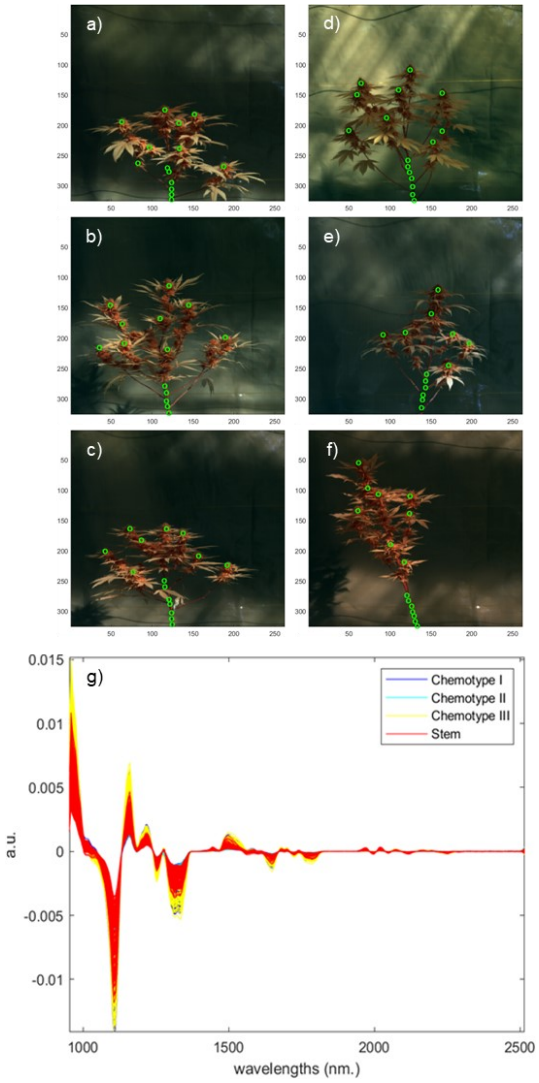
Kamera hiperespektralaren espektro-eremua 930-2500 nm tartean definituta egon zen, 288 kanal espektraletan banatuta 5.45 nm-ko espektro bereizmenarekin, eta bereizmen espaziala 384 x 286 pixelekoa izan zen. Kamera 3 metroko distantziako lentearekin hornituta egon zen eta zelai-muntaiaren tripodearen errotorearen errotazio-angelua 16°-koa izan zen, 841 mm-ko ikuseremu lineala (FOV) lortuz. Ondorioz, irudien pixelek 2.19/2.19 mm-ko tamaina izan zuten.

Irudi hiperespektralak eskuratzeko, landare-banakoak Sovereign Fields S.L.-ren negutegi barruko pareta zuri baten aurrean kokatu ziren. Honen aurrean, paretarekiko perpendikularki, kameraren tripodea loreontzitik 3 metrora kokatu zen. Kameraren 16°-ko biraketa angelua kontuan izanda, landarea FOV-aren zentruan lerrokatu zen. Parametro guztiak ezarrita, irudiak ENVI ® irudi hiperespektralak prozesatzeko softwarearekin eskuratu eta erradiometrikoki kalibratu ziren (NV5 Geospatial Solutions, Inc., Broomfields, Colorado, AEB). Aldi bakoitzean landare-banako baten irudiak hartu ziren.

2.4. Irudi hiperespektralen analisia

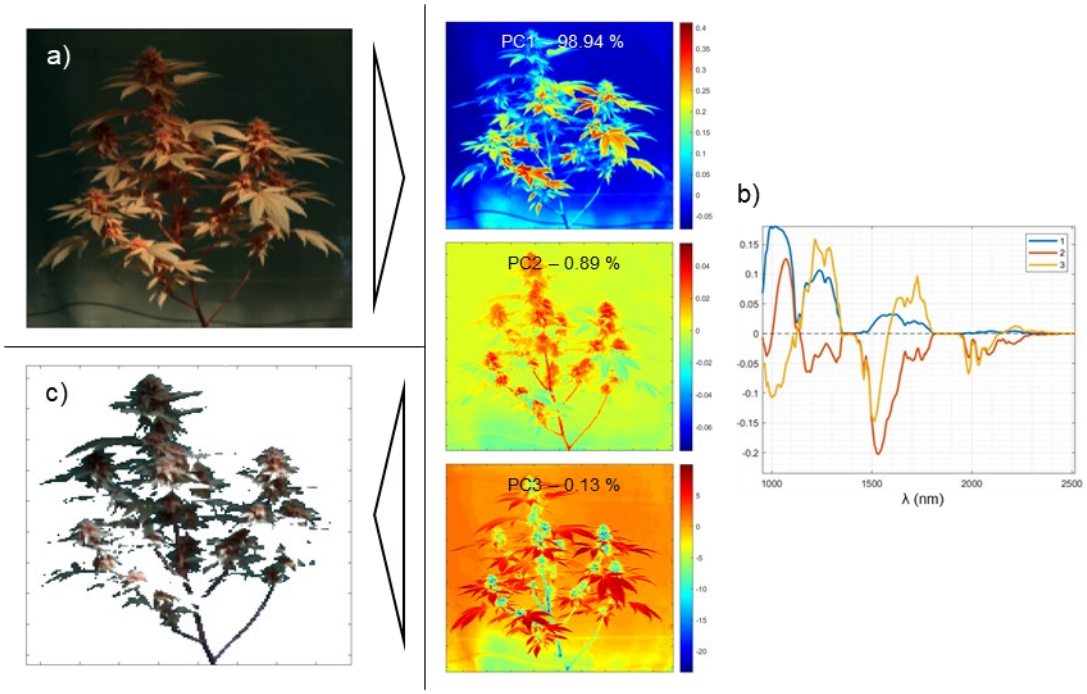
NIR-irudi hiperespektralak MATLAB ingurunera (The MathWorks Inc., USA) inportatu eta HYPER-Tools 3.0 (libreki eskuragarri www.hypertools.org-en) erabiliz maneiatu ziren [41].

NIR-irudi hiperespektraletatik 502 lore-espektro atera ziren guztira, eskuzko lore-pixel aukeraketaren bitartez: 249 I kemotipoari zegozkienak, eta, II eta III kemotipoko 178 eta 75, hurrenez hurren. Horiez gain, irudi guztietatik, 219 zurtoin espektru aukeratu ziren guztira, landareen zurtoinetako pixel-ak eskuz aukeratuz, *2. Irudian* ikus daitekeen bezala. Aukeratutako espektruei Savitzky-Golay deribatua (7-ko leiho zabalera; bigarren polinomio ordena; lehen deribatu ordena) eta batezbesteko zentratua aplikatu zitzaien datuen aurretratamendu moduan, gerorako sailkapen ereduak trebatzeko.



<u>2. Irudia:</u> Zegozkien sailkapen ereduen trebakuntzarako espektruak ateratzeko lore- eta zurtoin-pixelen eskuzko aukeraketa landare banako desberdinen irudi hiperespektraletan a) DQ5 banakoa b) TA6 banakoa c) RO6 banakoa d) RE2 banakoa e) WW5 banakoa f) FT9 banakoa

Kemotipoen iragarpen zuzenerako eta datuen bistaratze egokirako, irudi bakoitzaren atzealdea ezabatu zen. Hori egiteko, PCA aplikatu zen irudi bakoitzean, batezbesteko zentratuaren ostean [42], [43]. Bigarren osagai nagusiko *score*-ek landare-banakoen egokitasun morfologikoa ahalbideratu zuten, batez ere loreetan fokatutakoa, horrela irudien atzealde irregularra ezabatzeko aukera emanez, *3. Irudian* ikus daitekeen bezala.



<u>3. Irudia:</u> PCA bidezko maskaratze morfologikoa DQ4 landare-banakoan a) Landarearen RGB irudi faltsua b) 1, 2 eta 3. osagai nagusietako laginen-proiekzioak (scores) eta aldagaien-proiekzioak (loadings) c) Landarearen RGB irudi faltsu maskaratua

Lan honetan PLS-DA ereduan oinarritutako bi sailkapen estrategia proposatu ziren [44]. Lehen hurbilketa, kalibrazio-eredua hiru kemotipoen 502 lore-espektru sortarekin zuzenean trebatuz gauzatu zen. Bestalde, bigarren estrategia bi mailako eredu hierarkikoan oinarritu zen: lehen mailan, 502 lore-espektruen eta 219 zurtoin-espektruen bidez trebatutako sailkapen ereduarekin, zurtoina landarearen gainerakotik sailkatzeko hurbilketa proposatu zen; bigarren mailan, landarearen zurtoinari ez zegokion zatia hiru kemotipoen artean sailkatu zen aurretiaz 502 lore-espektruen bidez trebatutako sailkapen eredua erabiliz. Sailkapen ereduak PLS_Toolbox 9.0 (Eigenvector Reseach, USA) erremintaren bidez garatu ziren eta ausazko azpimultzoen zeharkako berrespena aplikatu zitzaien, 10 datu-banaketarekin eta 5 errepikapenekin.

Sailkapen-ereduen trebakuntzarako erabili ez ziren pixelak gerora iragarri ziren hauen bidez irudi hiperespektral guztietan.

3. Emaitzak eta eztabaida

3.1. Kannabinoide nagusiek karakterizazioa

Barietate bakoitzaren kemotipoa haien lore lehor lagin bateratuaren kromatografia bidezko analisi suntsitzaile bidez definitu zen. Zegozkien kannabinoide nagusien kontzentrazioak *1. Taulan* agertzen dira, non DQ, TA eta RO I kemotipoko barietateak izan zirela ikus daitekeen, WW eta RE barietateei II kemotipoa egokitu zitzaien eta FT III kemotipoko barietate bakar moduan sailkatu zen.

1. Taula: LC-UV bidez zehaztutako THC eta CBD kannabinoideen kontzentrazioak kannabis barietate desberdinen lore lehorretan, eta barietate bakoitzari dagozkien kemotipoak

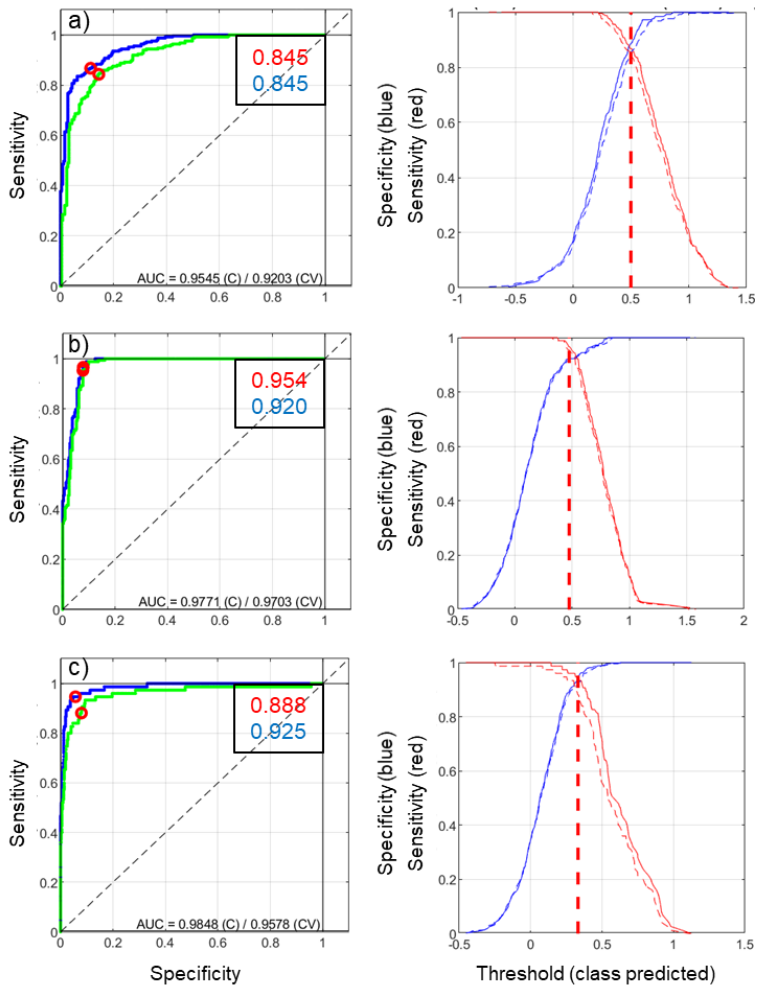
	Kannabinoide kontzentrazioa (% m/m)			
Barietatea	**THC totala**	**CBD totala**	**[THC_{Totala}]/[CBD_{Totala}]**	**Klasea**
White Widow	5.6	7.5	0.75	II Kemotipoa
Dairy Queen	20.2	0.7	28.86	I Kemotipoa
Tel Aviv	16.9	0.2	84.50	I Kemotipoa
Roma	17.3	0.2	86.50	I Kemotipoa
Futura 75	0.7	15.6	0.04	III Kemotipoa
Remedy	6.8	8.8	0.77	II Kemotipoa

3.2. PLS-DA ereduaren errendimendua

3.2.1. PLS-DA ereduaren trebakuntza

Hiru kemotipoak sailkatzeko eredua 8 aldagai latente (LV) erabiliz trebatu zen. Ereduaren zeharkako berrespena ausazko azpimultzoketa bidez egin zen, hiru klaseek pisu estatistiko desberdinak baitzituzten, 5 errepikapenetan permutatutako 10 datu banaketekin. Zeharkako berrespenak eman zituen sentikortasun- eta espezifikotasun-balio optimoak honako hauek izan ziren klase bakoitzerako: 0.845/0.845 I kemotiporako, 0.954/0.920 II kemotiporako eta 0.888/0.925 III kemotiporako. Klase bakoitzari zegozkien jasotako funtzionamendu-ezaugarrien (ROC) kurbak *4. Irudian* agertzen dira.

Sailkapen-ereduaren fidagarritasuna bermatzeko, 100 iterazioko permutazio-froga bat egin zen trebakuntzarako erabilitako datuekin, ereduaren esangura estatistikoa egiaztatzeko. Honen arabera, ereduaren iragarpen-gaitasuna gehiegi egokitutakoa den ala ez zehazten da, hau zorizko emaitza ala dagokien lagin-aldagaien arteko korrelazioaren emaitza dela egiaztatuz [45]. Ereduaren ausazko probabilitatea Wilcoxon probaren bidez frogatu zen [46], non, bai bere kabuz iragarritako emaitza, bai zeharka-berretsitako emaitza, 0 izan zen kemotipo klase bakoitzerako. Beraz, eredua estatistikoki irmoa eta esanguratsua zela egiaztatu zen, hau da, ereduaren zeharka-berretsitako parametroak eta haren iragarpen-gaitasuna ez zeudela ausazko faktoreen eraginpean, eta, ondorioz, iragarpen emaitza oro aldagaien eta klaseen arteko korrelazioaren egiazko emaitza izango litzatekeela.



4. Irudia: Kalkulatutako (urdina) eta zeharka berrestutako (berdea) ROC kurbak eta kalkulatutako (lerro jarraia) eta zeharka berrestutako (lerro etena) erantzun kurbak a) I kemotipoa b) II kemotipoa c) III kemotipoa

3.2.2. PLS-DA ereduaren kanpo-berrespena

Landare bakoitzean iragarritako kemotipoen probabilitateak Material Osagarriaren (SM) *S1 Taulan* aurkezten dira. Irudien pixel bakoitzean dagokion kemotipoa iragartzeko gutxieneko atalasea % 50eko probabilitatean ezarri zen. Honekin batera, landare-banako bakoitzaren kemotipoaren kanpo-iragarpenerako araua ezarri zen, gutxienez irudi bat osatzen zuten pixelen % 50a kemotipoetako batean sailkatu behar zirela ezartzen zuena. Aitorpen honen arabera egiazko klase bakoitzaren kasuan iragarritako kemotipoak *2. Taulan* erakusten dira.

2. Taula: Trebatutako PLS-DA sailkapen ereduaren bidez 3 kemotipoei zegozkien 6 kannabis barietateetako 57 landare-banakoen egiazko eta iragarritako kemotipo klaseak

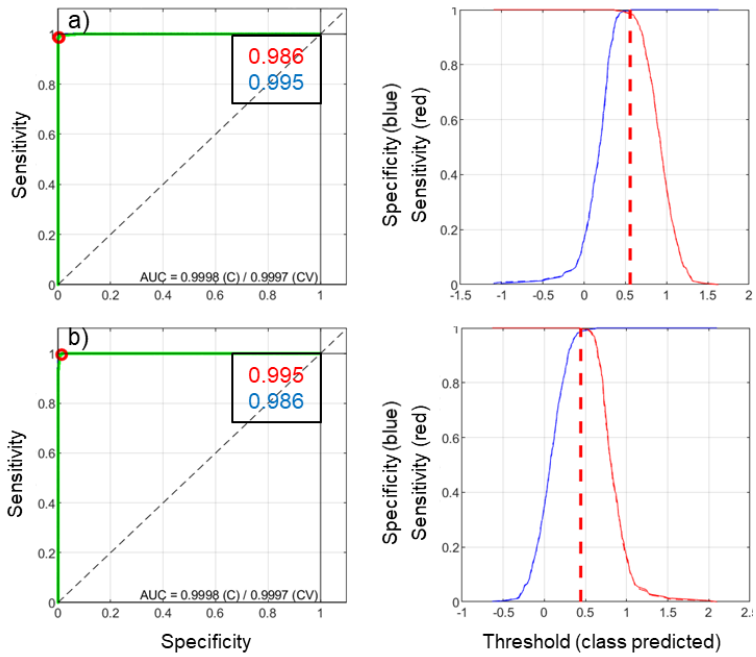
		Iragarritako klasea (Kemotipo probabilitatea > 50 %)				
		I Kem.	II Kem.	III Kem.	Ez-sailkatua	
Egiazko klasea	I Kem. (n=28)	25	2	0	1	28
	II Kem. (n=20)	0	20	0	0	20
	III Kem. (n=9)	2	0	6	1	9
		27	22	6		Guztira=57

2. Taulan aurkeztutako emaitzen arabera 4 landare-banako gaizki sailkatu zirela ondorioztatu daiteke (% 7.02 sailkapen okerreko proportzioa). Beste bi landare-banakoak ez ziren klaseetako batean ere ez sailkatu (% 3.52 ez-sailkapen proportzioa), beraz, 57 landare-banakotik, 51 izan ziren zuzen sailkatutakoak (% 89.47ko egiazkotasuna). Zentzu honetan, II kemotipoko landare guztiak zuzen sailkatu zirela (% 100 egiazkotasuna) aipatzea merezi du. Emaitza onenak klase honi egokitzea espero zen, sailkapen-ereduaren trebakuntzan sentikortasun eta espezifikotasun balio altuenak aurkeztu zituen eta (0.954 eta 0.920, hurrenez hurren). Honetaz aparte, I kemotipoko landareen zenbatekoa handiena izan zen hiru klaseen artean (n=28). Klase hau hiru barietate desberdinek osatu zuten, sailkapen okerreko proportzio handiago batera gida zezakeen faktorea, landare-banakoen arteko aldakortasun biologiko handiagoagatik. Hau izan liteke klase hau osatzen duten 28 banakoetatik 25 soilik zuzenki iragarri izanaren arrazoia (%89.29ko egiazkotasuna). Azkenik, sailkapen-ereduak emaitza okerrenak eman zituen III kemotipoko landareetan. Kasu honetan, 9 landare-banakoetatik 6 bakarrik sailkatu ziren modu egokian (% 66.67ko egiazkotasuna). Orokorrean, emaitza hauek ez lirateke egokiak izango sailkapen-eredu adierazgarri baten garapenerako. Hala ere, jazoera hau sailkapen-ereduaren partzialtasunaren ondorioa izan liteke. Bada, iragarpen emaitzei begirada sakonago bat botaz, ikus liteke, orokorrean, landareen zurtoina I kemotipoan sailkatzen zela, landare-banako bakoitzaren kemotipoa edozein zela ere. Hain zuzen ere, partzialtasun hau nabariagoa da III kemotipoko landareetan, non I kemotipoan sailkatutako pixel kopuruak irudien proportzio esanguratsua barne hartzen zuen. Ondorioz, sailkapen-eredua bi sailkapen mailatan hierarkizatu zen, irudietan landareen zurtoina aparteko klase moduan kontuan hartzeko. Planteamendu aldaketa honek iragarpen partzialtasun hau deuseztea espero zen, horrela iragarpen-gaitasun hobea edukiko zuen sailkapen-eredu bat lortuz.

3.3. HPLS-DA ereduaren errendimendua

3.3.1. HPLS-DA ereduaren trebakuntza

HPLS-DA ereduaren lehen sailkapen-mailak landareen zurtoinen eta zurtoina ez ziren zatien pixelen arteko sailkapenean zetzan. Eredu honek, zurtoinari ez zegokion landare zatian, zeharka-berretsitako 0.986ko eta 0.995eko sentikortasun- eta espezifikotasun-balioak aurkeztu zituen hurrenez hurren, eta 0.995ekoak eta 0.986koak, zurtoinaren kasuan, 4 LV-rekin (*5. Irudia*). HPLS-DA-ren bigarren sailkapen-maila aurretiaz trebatutako PLS-DA ereduari zegokion (3.2.1. atala).



5. Irudia: Kalkulatutako (urdina) eta zeharka-berretsitako (berdea) ROC kurbak eta kalkulatutako (lerro jarraia) eta zeharka-berretsitako (lerro etena) erantzun kurbak a) Zurtoinetik aparteko klasea b) Zurtoin klasea

3.3.2. HPLS-DA ereduaren kanpo-berrespena

HPLS-DA ereduaren bidez landare-banako bakoitzarentzat iragarritako kemotipoen probabilitateak SM-ko *S2 Taulan* aurkezten dira. Pixel bakoitzean dagokion kemotipoa iragartzeko gutxieneko atalasea % 50eko probabilitatean finkatu zen ere, aurreko PLS-DA-ren bidezko iragarpenen kasuaren moduan. Hala eta guztiz ere, kasu honetan, landareen kemotipoaren iragarpenerako, irudi bakoitza osatzen zuten pixel guztien artean zurtoinari zegozkien pixelak kontuan hartu ziren ere, bai banakoen artean, bai barietateen artean, landarearen tamaina eta morfologia aldatzen baitzen. Horrela ere, sailkapen-arau bat ezarri zen bigarren sailkapen mailan, gutxienez landarea osatzen zuten zurtoinetik aparteko pixelen % 50a kemotipoetako batean sailkatu behar zirela adieraziz. Adierazpen honen arabera, *3. Taulan* landareetan iragarritako kemotipo klaseak benetan zegozkien klaseekin bateratuta aurkeztu dira.

3. Taulan aurkeztutako emaitzen arabera, ez zen sailkatu gabeko landare-banakorik gelditu HPLS-DA-ren bidezko kemotipoaren iragarpenean, beraz, landare-banako guztietan % 50a baino probabilitate handiagoko sailkapena gauzatu zen kemotipoetako batean (% 0.00 ez-sailkapen proportzioa). Hala ere, 57 landare-banakotik 5 ez ziren zuzenki sailkatu (% 91.23ko egiazkotasuna, % 8.77ko sailkapen okerreko proportzioa). PLS-DA-ren kasuan bezala, II kemotipoko banako guztiak zuzenki sailkatu ziren (% 100eko egiazkotasuna), baina III kemotipoko landareen kasuan, emaitzak hobetu egin ziren, 9 landare-banakotik 8tan sailkapen zuzena lortuz (% 88.89ko egiazkotasuna). III kemotipoko landare kopurua kontuan hartuz, emaitza hau hobekuntza nabarmena izan zen PLS-DA-ren bidez lortutako emaitzekiko (% 22.22ko hobekuntza). Hala ere, I kemotipoko egiazkotasun proportzioa jaitsi egin zen (% 3.58ko jaitsiera), 28 landare-banakotik 24 zuzenki sailkatu baitziren (%85.71eko egiazkotasuna).

3. Taula: Trebatutako HPLS-DA sailkapen ereduaren bidez 3 kemotipoei zegozkien 6 kannabis barietateetako 57 landare-banakoen egiazko eta iragarritako kemotipo klaseak

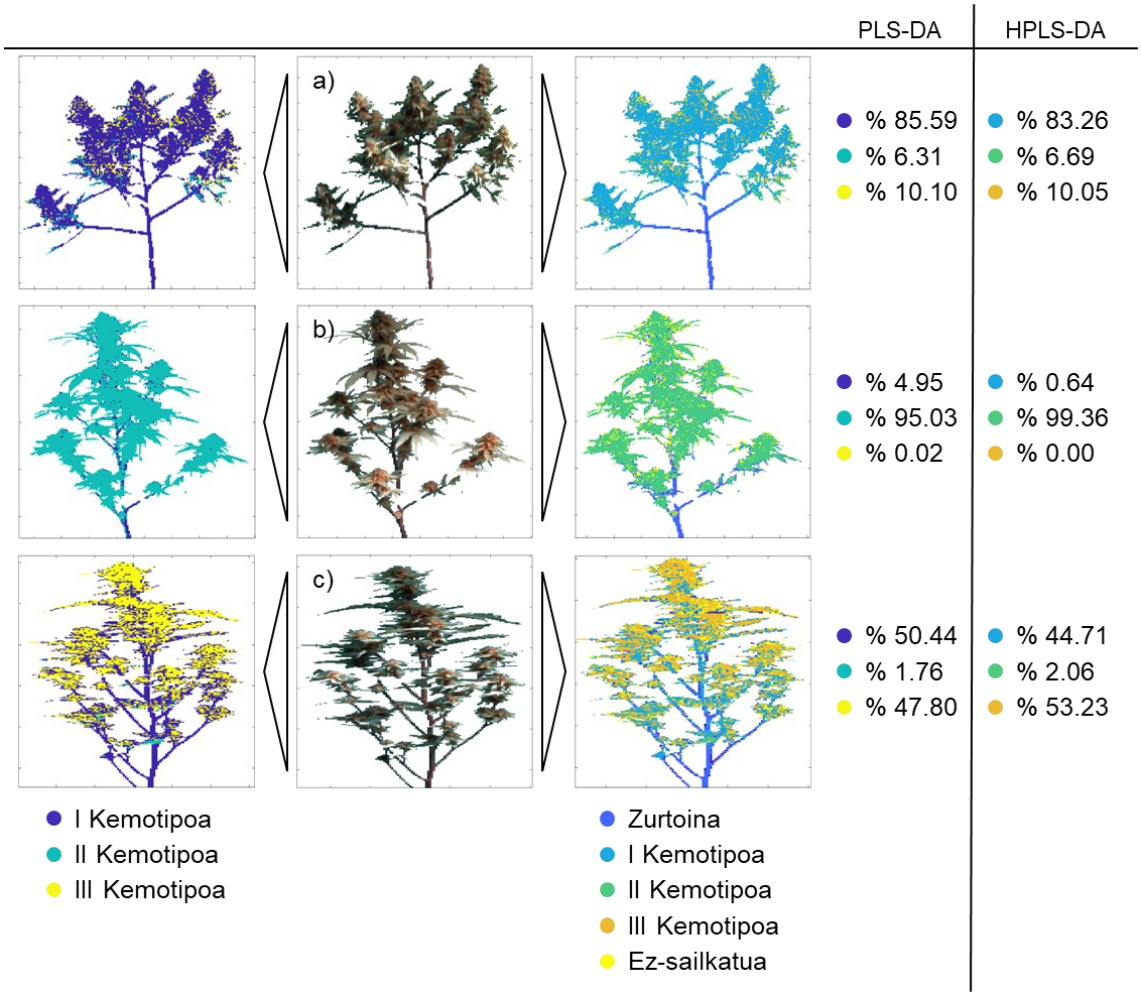
		Iragarritako klasea (Kemotipo probabilitatea > 50 %)				
		I Kem.	II Kem.	III Kem.	Ez-sailkatua	
Egiazko klasea	I Kem. (n=28)	24	4	0	0	28
	II Kem. (n=20)	0	20	0	0	20
	III Kem. (n=9)	1	0	8	0	9
		25	24	8		Guztira=57

3.4. Sailkapen ereduen arteko konparaketa

Orokorrean, *4. Taulan* aurkezten diren emaitzak kontutan hartuz eta PLS-DA eta HPLS-DA ereduak konparatuz, esan liteke eredu hierarkikoak PLS-DA ereduak baino emaitza hobeak lortu zituela sailkatu gabeko landarerik ez baitzen gelditu. *6. Irudian* iragarpen-adibide batzuen bistaratzea ikus daitezke, non, adibidez, III kemotipokoa den FT2 landare-banakoa, PLS-DA ereduaren bidez I kemotipoan sailkatu zela ikus daitekeen, eta aldiz, HPLS-DA ereduaren bidez zuzenki sailkatu zen. Azken honek iragarpen-gaitasun orokorra hobetu zuen, III kemotipoaren egiazkotasuna hobetu baitzuen zurtoinaren partizaltasunaren deuseztatzea zela eta. Ondorioz, lortutako emaitzak egiazkoenganako hurbilagoak izan ziren.

4. Taula: Sailkapen-ereduen iragarpen-gaitasunaren konparaketa

	PLS-DA	**HPLS-DA**
Egiazkotasun orokorra	% 89.47	% 91.23
I Kemotipoa	% 89. 29	% 85.71
II Kemotipoa	% 100.00	% 100.00
III Kemotipoa	% 66.67	% 88.89
Sailkapen okerren proportzioa	% 7.02	% 8.77
Ez-sailkapen proportzioa	% 3.51	% 0.00



6. Irudia: PLS-DA (ezkerra) eta HPLS-DA (eskuina) bidezko kemotipoen iragarpen adibideak landare desberdinetan: a) RO3 landare-banakoa (I kemotipoa) b) WW10 landare-banakoa (II kemotipoa) c) FT2 landare-banakoa (III kemotipoa)

3.5. Kontzeptu frogaren baieztapena

Bi ereduen kasuan, bi landare-banako zehatzetan dudarik gabeko sailkapen okerra gertatu zen. Bi banako horiek, Tel Aviv 3 eta Tel Aviv 4 izan ziren, teorikoki I kemotipoari zegozkienak, baina probabilitate ia osoz II kemotipoan sailkatu zirenak. Tel Aviv 3 banakoaren kasuan, PLS-DA ereduaren bidez lortutako probabilitate emaitzak % 8.05koa izan zen I kemotipoarentzat, % 91.95koa II kemotipoarentzat eta % 0.00koa III kemotipoarentzat; HPLS-DA-ren bidez, ordea, iragarritako probabilitatea % 1.33koa izan zen I kemotipoarentzat, % 98.67koa II kemotipoarentzat eta % 0.00koa III kemotipoarentzat. Tel Aviv 4 banakoaren kasuan, iragarpen emaitzak antzekoak izan ziren. PLS-DA-k % 9.02ko probabilitatea eman zuen I kemotipoarentzat, % 90.98koa II kemotipoarentzat eta % 0.00koa III kemotipoarentzat; HPLS-DA-ren bitartez, berriz, balio hori % 0.09koa izan zen I kemotipoarentzat, % 99.91koa II kemotipoarentzat eta % 0.00koa III kemotipoarentzat. Bi banakoak era esanguratsuan sailkatu ziren II kemotipoan eta, nahiz eta emaitza hauek nekez sinesgarriak izan, kultibazio instalazioetatik eratorritako landareen etiketatze akats baten ondorioz baliteke zuzenak izatea. Jazoera hau gerora egiaztatu zen, zeren eta, irudien eskuratze momentuan, WW (II kemotipoa) landare batzuk gaizki

etiketatuta egon baitziren eta haien trazabilitatearen kontrola galdu baitzen. Hortaz, bi banako zehatz hauen beste TA eta WW landareenganako konparagarritasuna aztertzeko *Student t*-test bat egin zen talde horien iragarpen-emaitzen artean [47]. *t* froga estatistikoak bi taldeen arteko bariantza berdinen sinesmenaren bitartez kalkulatu ziren. Alde batetik, TA3 eta TA4 banakoak Tel Aviv barietateei egokitzearen suposizioa proposatu zen hipotesi nulu moduan, bi banako horietan I kemotipoarekiko lortutako probabilitate emaitzak klase honetako banakoen emaitzekin konparatuz. Beste aldetik, hipotesi nulu moduan bi landare-banako horien WW barietatearekiko antzekotasuna aztertu zen, horretarako bi banako horietan II kemotipoarekiko lortutako probabilitateak azken barietate honetako landareetan lortutako probabilitatearekin konparatuz. Bi kalkuluak, *5. Taulan* erakusten direnak, HPLS-DA ereduaren bidez lortutako probabilitate emaitzekin egin ziren.

5. Taula: TA3 eta TA4 landare-banakoen eta TA barietateko beste landareen arteko, eta, TA3 eta TA4 landare-banakoen eta WW barietateko landareen arteko t-test estatistikoa

	***t*-Testa: Bi talde, bariantza berdinak suposatuz (α=0.05)**			
	I Kemotipoa		II Kemotipoa	
	TA3 eta TA4	Gainontzeko TA-ak	TA3 eta TA4	WW-ak
Batez bestekoa	0.71	69.66	99.29	92.80
Bariantza	0.77	103.10	0.77	38.01
Behaketak	2	8	2	10
Bariantza taldekatua	90.31		34.29	
t estatistikoa	-9.18		1.43	
P(T≤t) buztan batentzako	8.03E-06		0.09	
t Kritikoa buztan batentzako	1.86		1.81	

t-test estatistikoaren emaitzek TA3 eta TA4 banakoen TA barietatearekiko egokitzapenaren hipotesi nulua ezeztatu zuten, kalkulatutako *t* -9.18 izan baitzen (Buztan batentzako *t* kritikoa = 1.86), WW barietatearekiko antzekotasuna onartu zuen bitartean (kalkulatutako *t* = 1.43, buztan batentzako *t* kritikoa = 1.81). Hau horrela izanda, landare-banako hauek WW moduan zuzenki etiketatu izan balira, HPLS-DA ereduaren bitarteko iragarpenak nabarmenki hobetuko lirateke, I kemotipoaren kasuan % 92.31ko egiazkotasuna lortuz, % 100.00ekoa II kemotipoaren kasuan, eta III kemotipoarentzako % 88.89ko egiazkotsauna ezarriz. Era berean, sailkapen okerreko proportzioa % 5.26ra jaitsi da eta soilik 3 landare-banako geldituko baitziren gaizki iragarrita. Horrela, HPLS-DA ereduaren iragarpenen egiazkotasun orokorra % 94.74an geldituko litzateke, sailkapen ereduarentzako iragarpen-gaitasun zehatza izango zena. Hipotesi hau onartuz, klase bakoitzaren kasurako iragarritako batezbesteko probabilitateak *6. Taulan* aurkeztu dira. Emaitza hauen arabera, II kemotipoko landareentzako sailkapen errendimendua bikaina izan da, % 5.59-

ko doitasuna duen batezbesteko iragarritako probabilitatea % 96.63koa izan baita. Beste kasuetan, iragarritako batezbesteko probabilitate- eta doitasun-emaitzak % 67.85 eta % 15.10 izan dira I kemotipoko landareentzat, eta, % 61.75 eta % 17.82 III kemotipoaren kasuan, hurrenez hurren. Jazoera hau II kemotipoko landareek bi kannabinoide nagusiak kontzentrazio esanguratsuetan izatearen ondorioz gerta liteke. Kontrara, beste bi klaseek kannabinoide nagusietako bat soilik zeukaten kontzentrazio esanguratsuan, beraz, ondorioztatu liteke bi kannabinoide nagusien presentziak nabarmenki ahalbidetzen duela NIR-irudi hiperespektralaren bidezko sailkapen adierazgarria kannabis barietateetan.

6. Taula: Kemotipo bakoitzean iragarritako batezbesteko probabilitatea, desbiderapen estandarra eta doitasuna, TA3 eta TA4 banakoak WW barietatekoak zirela onartuz

	I Kemotipoa (%)	**II Kemotipoa (%)**	**III Kemotipoa (%)**
Egiazkotasuna	92.31	100.00	88.89
Iragarritako batez besteko probabilitatea	67.85	96.63	61.75
Desbiderapen estandarra	10.24	5.40	11.01
Doitasuna (DEE)	15.10	5.59	17.82

*DEE: Desbiderapen estandar erlatiboa

4. Ondorioak

Lan honetan, *Cannabis sativa L.*-ren NIR-irudi hiperespektralen bidez I, II eta III kemotipoetako barietateak era esanguratsuan sailka daitezkeela ondorioztatu zen, horrela teknika analitiko inbaditzaileak ekidinez. Hau posiblea izan zen bi mailan hierarkizatutako sailkapen eredu baten bitartez: lehen sailkapen-mailan landarearen zurtoina eta zurtoinari ez zegozkion atalak sailkatu ziren, eta bigarren mailan zurtoina ez zen landare zatia hiru kemotipoetako batean sailkatu zen, NIR informazio espektralean oinarrituta. HPLS-DA ereduari dagokionez, iragarpen emaitzek erakutsi zuten II kemotipoko landareak bikain sailkatzen zirela, I eta II kemotipoen kasuan 2 eta 1 landare-banakok sailkapen okerra izan zuten bitartean, hurrenez hurren. Hau, % 94.74ko sailkapen orokorreko egiazkotasunean itzultzen da, sailkapen errendimenduaren kontzeptu froga egokia izan zena. Gainera, aurretiaz aipatu den bezala, TA3 eta TA4 landare-banakoak, teorikoki I kemotipokoak zirenak, era esanguratsuan sailkatu ziren II kemotipoan. Nahiz eta hau anomalia izan, gerora konfirmatu zen etiketatze-akats bat izan zela kultibazio-instalazioetan landare hauen kultibo denboran zehar, White Widow (II kemotipoa) barietateko landare batzuk gaizki etiketatu ziren eta. Emaitzak okertu beharrean, akats honek ikerketa lan honen helburua enpirikoki frogatzen lagundu zuen, *in-situ* analisiak landareen gaineko kontrol prozedurala ahalbidetu baitzuen, haien loreetan kannabinoideen karakterizazio inbaditzailea saihestuz. Hala ere, esan beharra dago jazoera hau akats iturri potentziala izan ahalko litzateke ela datuetan beste edozein egoeratan, hortaz, derrigorrezkoa da trazabilitatea egokia mantentzea landare-banakoen identifikazioan lagin bolumen handiagoko etorkizuneko landa-lanetako ikerketetan,

landareen izendatze planteamendu zorrotzekin. Gainera, onuragarria izango litzateke sailkapen ereduaren errendimendua kultibazio inguru desberdinetan frogatzea, Sovereign Fields S.L.-ko instalazioetakoez aparte, beste hazkuntza toki batzuetarako emaitzen transferigarritasuna ziurtatzeko.

Hala bada, NIR-HSI-an oinarritutako metodologiak, HPLS-DA-ren bidezko kemotipo sailkapenaren eskutik, NIRS bidezko kannabinoideen analisiaren zailtasun handienari aurre egin zion, hau da, landare biziko hezetasuna, horrela landare-banako bizi baten gaineko analisi adierazgarria ahalbidetuz. Hala ere, berebizikoa da lan honen helburua azpimarratzea, landare bakoitzari zegokion kemotipoaren iragarpen-sailkapena metodologia erraz eta ez-inbaditzaile baten bitartez produkzioko kalitate kontroleko neurri moduan, eta kannabinoideen kuantifikazioarekin ez nahastea, non ohiko metodologia landare-materiala lehortzea eta NIRS ala kromatografia bezalako beste teknika batzuen bidezko analisia izango zena.

Ondorioz, lan honen helburu nagusia lortu egin zen, hau da, labore-kultiboen hazkuntzaren kontrol integralerako kontzeptu-froga, eta are zehatzago, kannabisarena, teknika analitiko ez-inbaditzaile baten bitartez frogatzea. Honek bidea erraztu lezake nekazaritza-sektoreko kalitate-kontrolerako teknologia ez-inbaditzailearen garapenerako, zehaztutako helburua lortu egin baitzen ohiko teknika analitikoak saihestuz, hala nola, kromatografia edo NIR espektroskopia konbentzionala. Hala ere, etorkizunean hau posiblea izateko, sailkapen ereduak trebakuntza datu adierazgarriagoak beharko lituzke. Emaitzen orokortasuna areagotzeko, lagin-kopuru handiagoa beharko litzateke, barietate-aniztasun estaldura zabalagoarekin eta kultibazio-denbora desberdinetako laginekin. Hortaz, ikerketa ildo honetatik aurrera egiteko proposamen egokia sailkapen eredua hiru kemotipoetako barietate kopuru handiagoarekin trebatzea izango litzateke, idealki, kemotipoen arteko pisu estatistikoen desberdintasuna orekatuz, eta hazkuntza-denbora desberdinetan landareen hostoetatik eratorritako hosto-espektruekin konbinatuz. Era honetan, kultibo-denbora desberdinen araberako erregresio joera bat lortuko litzateke, landare-banakoen kemotipoen iragarpena hazkuntza etapa goiztiarretan ahalbidetuz.

Material Osagarria

3. Kapituluko material osagarria esteka honetan ala beheko QR kodean dago eskuragarri:

S1 Taula: PLS-DA ereduaren bidezko kemotipo iragarpenak

S2 Taula: HPLS-DA ereduaren bidezko kemotipo iragarpenak

Bibliografia

[1] L. Cristino, T. Bisogno, and V. Di Marzo, 'Cannabinoids and the expanded endocannabinoid system in neurological disorders', *Nature Reviews Neurology*, vol. 16, no. 1, pp. 9–29, 2020, doi: 10.1038/s41582-019-0284-z.

[2] N. Black *et al.*, 'Cannabinoids for the treatment of mental disorders and symptoms of mental disorders: a systematic review and meta-analysis', *The Lancet Psychiatry*, vol. 6, no. 12, pp. 995–1010, 2019, doi: 10.1016/S2215-0366(19)30401-8.

[3] D. Cox-Georgian, N. Ramadoss, C. Dona, and C. Basu, 'Therapeutic and medicinal uses of terpenes', in *Medicinal Plants: From Farm to Pharmacy*, 2019, pp. 333–359. doi: 10.1007/978-3-030-31269-5_15.

[4] O. Aizpurua-Olaizola, I. Elezgarai, I. Rico-Barrio, I. Zarandona, N. Etxebarria, and A. Usobiaga, 'Targeting the endocannabinoid system: future therapeutic strategies', *Drug Discovery Today*, vol. 22, no. 1, pp. 105–110, 2017, doi: 10.1016/j.drudis.2016.08.005.

[5] 'State Medical Cannabis Laws'. Accessed: Jun. 02, 2023. [Online]. Available: https://www.ncsl.org/health/state-medical-cannabis-laws

[6] 'Recommendation of removal for cannabis and cannabis resin from Schedule IV of 1961 Single Convention on Narcotic Drugs', presented at the United Nations Comission on Narcotic Drugs, Vienna, Dec. 2020. Accessed: Jun. 02, 2023. [Online]. Available: https://www.who.int/news/item/04-12-2020-un-commission-on-narcotic-drugs-reclassifies-cannabis-to-recognize-its-therapeutic-uses

[7] *Kontrollierte Abgabe von Cannabis: Eckpunktepapier der Bundesregierung liegt vor*. 2022. Accessed: Jun. 02, 2023. [Online]. Available: https://www.bundesgesundheitsministerium.de/ministerium/meldungen/kontrollierte-abgabe-von-cannabis-eckpunktepapier-der-bundesregierung-liegt-vor.html

[8] *Federal Act on Narcotics and Psychotropic Substances (Narcotics Act, NarcA)*. 2022. Accessed: Jun. 02, 2023. [Online]. Available: https://www.fedlex.admin.ch/eli/cc/1952/241_241_245/en

[9] J. K. Booth and J. Bohlmann, 'Terpenes in Cannabis sativa – From plant genome to humans', *Plant Science*, vol. 284, pp. 67–72, 2019, doi: 10.1016/j.plantsci.2019.03.022.

[10] M. Maccarrone *et al.*, 'Endocannabinoid signaling at the periphery: 50 years after THC', *Trends in Pharmacological Sciences*, vol. 36, no. 5, pp. 277–296, 2015, doi: 10.1016/j.tips.2015.02.008.

[11] V. Di Marzo, 'New approaches and challenges to targeting the endocannabinoid system', *Nature Reviews Drug Discovery*, vol. 17, no. 9, pp. 623–639, 2018, doi: 10.1038/nrd.2018.115.

[12] C. Muller, P. Morales, and P. H. Reggio, 'Cannabinoid ligands targeting TRP channels', *Frontiers in Molecular Neuroscience*, vol. 11, 2019, doi: 10.3389/fnmol.2018.00487.

[13] S. A. Bonini *et al.*, 'Cannabis sativa: A comprehensive ethnopharmacological review of a medicinal plant with a long history', *Journal of Ethnopharmacology*, vol. 227, pp. 300–315, 2018, doi: 10.1016/j.jep.2018.09.004.

[14] S. Zou and U. Kumar, 'Cannabinoid receptors and the endocannabinoid system: Signaling and function in the central nervous system', *International Journal of Molecular Sciences*, vol. 19, no. 3, 2018, doi: 10.3390/ijms19030833.

[15] O. Aizpurua-Olaizola *et al.*, 'Evolution of the Cannabinoid and Terpene Content during the Growth of Cannabis sativa Plants from Different Chemotypes', *Journal of Natural Products*, vol. 79, no. 2, pp. 324–331, 2016, doi: 10.1021/acs.jnatprod.5b00949.

[16] T. Gülck and B. L. Møller, 'Phytocannabinoids: Origins and Biosynthesis', *Trends in Plant Science*, vol. 25, no. 10, pp. 985–1004, 2020, doi: 10.1016/j.tplants.2020.05.005.

[17] A. I. Fraguas-Sánchez and A. I. Torres-Suárez, 'Medical Use of Cannabinoids', *Drugs*, vol. 78, no. 16, pp. 1665–1703, Nov. 2018, doi: 10.1007/s40265-018-0996-1.

[18] E. Small and H. D. Beckstead, 'Cannabinoid Phenotypes in Cannabis sativa', *Nature*, vol. 245, no. 5421, Art. no. 5421, Sep. 1973, doi: 10.1038/245147a0.

[19] M. A. Lewis, E. B. Russo, and K. M. Smith, 'Pharmacological Foundations of Cannabis Chemovars', *Planta Medica*, vol. 84, no. 4, pp. 225–233, 2018, doi: 10.1055/s-0043-122240.

[20] O. Aizpurua-Olaizola, J. Omar, P. Navarro, M. Olivares, N. Etxebarria, and A. Usobiaga, 'Identification and quantification of cannabinoids in Cannabis sativa L. plants by high performance liquid chromatography-mass spectrometry', *Analytical and Bioanalytical Chemistry*, vol. 406, no. 29, pp. 7549–7560, 2014, doi: 10.1007/s00216-014-8177-x.

[21] M. San Nicolas *et al.*, 'Analysis of cannabinoids in plants, marijuana products and biological tissues', *Comprehensive Analytical Chemistry*, vol. 90, pp. 65–102, 2020, doi: 10.1016/bs.coac.2020.04.002.

[22] S. Chandra, M. M. Radwan, C. G. Majumdar, J. C. Church, T. P. Freeman, and M. A. ElSohly, 'New trends in cannabis potency in USA and Europe during the last decade (2008–2017)', *European Archives of Psychiatry and Clinical Neuroscience*, vol. 269, no. 1, pp. 5–15, 2019, doi: 10.1007/s00406-019-00983-5.

[23] C. Citti, D. Braghiroli, M. A. Vandelli, and G. Cannazza, 'Pharmaceutical and biomedical analysis of cannabinoids: A critical review', *Journal of Pharmaceutical and Biomedical Analysis*, vol. 147, pp. 565–579, 2018, doi: 10.1016/j.jpba.2017.06.003.

[24] P. Berman *et al.*, 'A new ESI-LC/MS approach for comprehensive metabolic profiling of phytocannabinoids in Cannabis', *Scientific Reports*, vol. 8, no. 1, 2018, doi: 10.1038/s41598-018-32651-4.

[25] C. Citti, B. Pacchetti, M. A. Vandelli, F. Forni, and G. Cannazza, 'Analysis of cannabinoids in commercial hemp seed oil and decarboxylation kinetics studies of cannabidiolic acid (CBDA)', *Journal of Pharmaceutical and Biomedical Analysis*, vol. 149, pp. 532–540, 2018, doi: 10.1016/j.jpba.2017.11.044.

[26] M. San Nicolas *et al.*, 'Exploratory optimisation of a LC-HRMS based analytical method for untargeted metabolomic screening of Cannabis Sativa L. through Data Mining', *Analytica Chimica Acta*, vol. 1279, p. 341848, Oct. 2023, doi: 10.1016/j.aca.2023.341848.

[27] C. Sánchez-Carnerero Callado, N. Núñez-Sánchez, S. Casano, and C. Ferreiro-Vera, 'The potential of near infrared spectroscopy to estimate the content of cannabinoids in Cannabis

sativa L.: A comparative study', *Talanta*, vol. 190, pp. 147–157, 2018, doi: 10.1016/j.talanta.2018.07.085.

[28] C. Duchateau, J.-M. Kauffmann, M. Canfyn, C. Stévigny, K. De Braekeleer, and E. Deconinck, 'Discrimination of legal and illegal Cannabis spp. according to European legislation using near infrared spectroscopy and chemometrics', *Drug Testing and Analysis*, vol. 12, no. 9, pp. 1309–1319, 2020, doi: 10.1002/dta.2865.

[29] R. Deidda *et al.*, 'New perspective for the in-field analysis of cannabis samples using handheld near-infrared spectroscopy: A case study focusing on the determination of Δ9-tetrahydrocannabinol', *Journal of Pharmaceutical and Biomedical Analysis*, vol. 202, 2021, doi: 10.1016/j.jpba.2021.114150.

[30] J. Espel Grekopoulos, 'Construction and Validation of Quantification Methods for Determining the Cannabidiol Content in Liquid Pharma-Grade Formulations by Means of Near-Infrared Spectroscopy and Partial Least Squares Regression', *Medical Cannabis and Cannabinoids*, vol. 2, no. 1, pp. 43–55, 2019, doi: 10.1159/000500266.

[31] K. Su *et al.*, 'NIR spectroscopy for rapid measurement of moisture and cannabinoid contents of industrial hemp (Cannabis sativa)', *Industrial Crops and Products*, vol. 184, 2022, doi: 10.1016/j.indcrop.2022.115007.

[32] S. Yao, C. Ball, G. Miyagusuku-Cruzado, M. M. Giusti, D. P. Aykas, and L. E. Rodriguez-Saona, 'A novel handheld FT-NIR spectroscopic approach for real-time screening of major cannabinoids content in hemp', *Talanta*, vol. 247, 2022, doi: 10.1016/j.talanta.2022.123559.

[33] *Commission Delegated Regulation (EU) 2017/1155 of 15 February 2017 amending Delegated Regulation (EU) No 639/2014 as regards the control measures relating to the cultivation of hemp, certain provisions on the greening payment, the payment for young farmers in control of a legal person, the calculation of the per unit amount in the framework of voluntary coupled support, the fractions of payment entitlements and certain notification requirements relating to the single area payment scheme and the voluntary coupled support, and amending Annex X to Regulation (EU) No 1307/2013 of the European Parliament and of the Council*, vol. 167. 2017. Accessed: Nov. 08, 2023. [Online]. Available: http://data.europa.eu/eli/reg_del/2017/1155/oj/eng

[34] C. Jarén, P. C. Zambrana, C. Pérez-Roncal, A. López-Maestresalas, A. Ábrego, and S. Arazuri, 'Potential of NIRS Technology for the Determination of Cannabinoid Content in Industrial Hemp (Cannabis sativa L.)', *Agronomy*, vol. 12, no. 4, 2022, doi: 10.3390/agronomy12040938.

[35] D. Valinger *et al.*, 'Development of ANN models based on combined UV-vis-NIR spectra for rapid quantification of physical and chemical properties of industrial hemp extracts', *Phytochemical Analysis*, vol. 32, no. 3, pp. 326–338, 2021, doi: 10.1002/pca.2979.

[36] J. M. Amigo, 'Chapter 1.1 - Hyperspectral and multispectral imaging: setting the scene', in *Data Handling in Science and Technology*, vol. 32, J. M. Amigo, Ed., in Hyperspectral Imaging, vol. 32. , Elsevier, 2019, pp. 3–16. doi: 10.1016/B978-0-444-63977-6.00001-8.

[37] J. M. Amigo, H. Babamoradi, and S. Elcoroaristizabal, 'Hyperspectral image analysis. A tutorial', *Analytica Chimica Acta*, vol. 896, pp. 34–51, 2015, doi: 10.1016/j.aca.2015.09.030.

[38] S. K. Abeysekera *et al.*, 'Sparse reproducible machine learning for near infrared hyperspectral imaging: Estimating the tetrahydrocannabinolic acid concentration in Cannabis sativa L.', *Industrial Crops and Products*, vol. 192, 2023, doi: 10.1016/j.indcrop.2022.116137.

[39] W. S. Holmes *et al.*, 'On machine learning methods to estimate cannabidiolic acid content of Cannabis sativa L. from near-infrared hyperspectral imaging', presented at the Conference Record - IEEE Instrumentation and Measurement Technology Conference, 2023. doi: 10.1109/I2MTC53148.2023.10175994.

[40] Y. Lu, X. Li, S. Young, X. Li, E. Linder, and D. Suchoff, 'Hyperspectral imaging with chemometrics for non-destructive determination of cannabinoids in floral and leaf materials of industrial hemp (Cannabis sativa L.)', *Computers and Electronics in Agriculture*, vol. 202, 2022, doi: 10.1016/j.compag.2022.107387.

[41] N. Mobaraki and J. M. Amigo, 'HYPER-Tools. A graphical user-friendly interface for hyperspectral image analysis', *Chemometrics and Intelligent Laboratory Systems*, vol. 172, pp. 174–187, 2018, doi: 10.1016/j.chemolab.2017.11.003.

[42] K. Pearson, 'LIII. On lines and planes of closest fit to systems of points in space', *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, Nov. 1901, doi: 10.1080/14786440109462720.

[43] H. Hotelling, 'Analysis of a complex of statistical variables into principal components', *Journal of Educational Psychology*, vol. 24, pp. 417–441, 1933, doi: 10.1037/h0071325.

[44] L. Ståhle and S. Wold, 'Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study', *Journal of Chemometrics*, vol. 1, no. 3, pp. 185–196, 1987, doi: 10.1002/cem.1180010306.

[45] E. Lopez, J. Etxebarria-Elezgarai, J. M. Amigo, and A. Seifert, 'The importance of choosing a proper validation strategy in predictive models. A tutorial with real examples', *Analytica Chimica Acta*, vol. 1275, p. 341532, Sep. 2023, doi: 10.1016/j.aca.2023.341532.

[46] F. Wilcoxon, 'Individual Comparisons by Ranking Methods', *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[47] Student, 'The probable error of a mean', *Biometrika*, vol. 6, no. 1, pp. 1–25, 1908, doi: 10.1093/biomet/6.1.1.

GENERAL CONCLUSIONS AND FUTURE PERSPECTIVES

Overall, it could be stated that the objectives and sub-objectives defined for this research project were achieved.

On the one hand, the intended outcomes regarding the metabolomic study of cannabis were successfully reached. First, a streamlined methodology was optimised based on previous plant metabolomics studies in the literature. By empirical comparison of the experimental variables described in these previous studies, it was concluded that, in order to obtain the broadest possible metabolic coverage in the representative metabolomics study of the cannabis plant, an experimental methodology based on a two-phase chemical extraction through $CHCl_3$:CH_3OH:H_2O (2:1:1, v/v) solvent combination, followed by the use of a C18 reversed-phase chromatographic column for the untargeted analysis by HPLC-qOrbitrap, and the analysis of both inflorescence and leaf as complementary biological tissues for representative screening of the whole plant individual, would have to be adopted. Following this approach, this developed methodology was applied in the untargeted analysis for the search of predictive constitutive metabolic markers for the determination of the adaptability of cannabis cultivars to tropical climate environments. From this investigation it was concluded that the non-polar phase was the extract of interest, as this contained the metabolites that fulfilled the exclusion filters based on the applied statistical criteria, thereby totally excluding the polar-phase extract. Hence, the metabolites Piperochromenoic acid, Mammea C/BB, Dalbergichromene, Dihydrocordoin, Hedichylactone D and 3-[8,13-bis(ethenyl)-18-(1-hydroxy-3-methoxy-3-oxopropyl)-3,7,12,17-tetramethyl-22,23-dihydroporphyrin-2-yl]propanoic acid were statistically determined as the referring markers, and their physiological behaviour as such was confirmed by their metabolic adaptive effect on the cannabis plant.

On the other hand, the feasibility of the NIR-HSI for the traceability and quality control in medical cannabis cultivation sites was confirmed through a proof of concept for the non-invasive chemotype classification of cannabis cultivars. In this case, it was concluded that the key determinant for the accomplishment of the defined objective relied in the multivariate data-analysis approach to be adopted, which entirely depended on the obtained data, since, as it could be observed, the predictive capability towards the classification of chemotype III plants did not result as effective as for chemotype I and II plants. This resulted so due to the lower statistical weight of the model training spectra corresponding to the chemotype III in comparison to the others, as its population was smaller than chemotype I and II's.

Hence, these conclusions set the direction of future perspectives for these research lines. Regarding the metabolomic study of cannabis, the next step would be to transition from an untargeted screening to the development of a targeted analytical method for the quantification of the determined metabolic markers in immature stages of the cultivars, in order to enable the correct decision-making when sending novel cultivars to regions of the world with tropical climates or to identify suitable varieties in the breeding operations of such geographies. To achieve this, the concentration levels of these metabolites in the immature stage of the cultivars of good and bad adaptability would have to be determined, in order to establish a concentration threshold by which to classify the novel varieties of cannabis to be developed in the future through the Mendelian genetics-based breeding process.

On the other hand, regarding the assessment of the feasibility of NIR-HSI for the classification of cannabis chemotypes by means of an easy and non-invasive analysis, the next stage would be to extend the classification model with more cultivars of chemotypes I, II and III, in order to increase the variability between cultivars and therefore render a more representative and

transferable model. Moreover, additional spectroscopic data of chemotype III should be integrated in order to narrow the gap in the statistical weight between the different chemotype classes so as to obtain better prediction results. Furthermore, it could be further considered to extend the scope of this approach to chemotype IV and V cultivars that are currently being developed in the medical cannabis industry, as they are also attracting a great deal of interest for medical purposes. In addition, it would also be of great interest to complement the corresponding multivariate data-analysis approach with spectroscopic data at different stages of cultivation, so as to be able to monitor the plants during the entire cultivation process and thus ensure their traceability without mistake by means of an automated mechanism integrated in the cultivation site itself.