



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Escola d'Enginyeria Agroalimentària
i de Biosistemes de Barcelona

RECONSTRUCCIÓ FILOGENÈTICA DE LES ASTERÀCIES DELS PAÏSOS CATALANS

Treball final de grau

Enginyeria de Sistemes Biològics

Autor: Claudia Atienza Jiménez

Tutor: Pol Fernández i Mató

Tutor UPC: Joan Casals Missio

Institució: Institut Botànic de Barcelona (CSIC – Ajuntament de Barcelona)

22 / setembre / 2023

Resum

L'objectiu d'aquest treball és estudiar les relacions evolutives de la família de plantes Asteraceae en els PPCC mitjançant la construcció d'un arbre filogenètic, la qual cosa proporcionarà informació genètica de noves espècies. En el nostre cas es realitzarà un arbre filogenètic a nivell de gènere. Aquest enfocament contribuirà a una millor comprensió de l'evolució de les asteràcies i permetrà l'aplicació d'aquestes dades en camps com la medicina i la cosmètica en el futur.

Lo primer ha sigut completar la llista de gèneres, pel que ha calgut dur a terme un estudi exhaustiu de la sinonímia dels gèneres, tenint en compte la nomenclatura de les diferents obres i bases de dades de referència. Les mostres dels gèneres inclosos a la llista han sigut tractades per obtenir dades de seqüenciació útils utilitzant el *kit Angiosperms353*.

L'anàlisi filogenètica d'aquestes dades genètiques, que constitueix el resultat d'aquest estudi, ha permès explorar les relacions evolutives entre els gèneres d'asteràcies en aquesta regió geogràfica específica. Aquest enfocament molecular proporciona una visió més clara de la diversitat genètica i l'evolució d'aquestes plantes als Països Catalans. Els resultats d'aquest estudi contribueixen a la comprensió de la biodiversitat i de la història evolutiva de les asteràcies en aquesta àrea geogràfica, amb l'objectiu de disposar d'un marc filogenètic adequat que permeti analitzar altres tipus de dades (etnobotàniques i citogenètiques, entre d'altres).

Resumen

El objetivo de este trabajo es estudiar las relaciones evolutivas de la familia de plantas *Asteraceae* en los PPCC mediante la construcción de un árbol filogenético, lo que proporcionará información genética de nuevas especies. En nuestro caso se realizará un árbol filogenético a nivel de género. Este enfoque contribuirá a una mejor comprensión de la evolución de las asteráceas y permitirá la aplicación de estos datos en campos como la medicina y la cosmética en el futuro.

Lo primero ha sido completar la lista de géneros, por lo que ha sido necesario llevar a cabo un estudio exhaustivo de la sinonimia de los géneros, teniendo en cuenta la nomenclatura de las distintas obras y bases de datos de referencia. Las muestras de los géneros incluidos en la lista han sido tratadas para obtener datos de secuenciación útiles utilizando el *kit Angiosperms353*.

El análisis filogenético de estos datos genéticos, que constituye el resultado de este estudio, ha permitido explorar las relaciones evolutivas entre los géneros de asteráceas en esta región geográfica específica. Este enfoque molecular proporciona una visión más clara de la diversidad genética y la evolución de estas plantas en los Países Catalanes. Los resultados de este estudio contribuyen a la comprensión de la biodiversidad y de la historia evolutiva de las asteráceas en esta área geográfica, con el objetivo de disponer de un marco filogenético adecuado que permita analizar otros tipos de datos (etnobotánicos y citogenéticos, entre otros).

Abstract

The aim of this work is to study the evolutionary relationships of the Asteraceae plant family in the PPCCs by constructing a phylogenetic tree, which will provide genetic information for new species. In our case, a phylogenetic tree will be performed at the genus level. This approach will contribute to a better understanding of the evolution of Asteraceae and enable the application of this data in fields such as medicine and cosmetics in the future.

The first thing has been to complete the list of genera, so it has been necessary to carry out an exhaustive study of the synonymy of the genera, taking into account the nomenclature of the different floras and databases. Both officially accepted and non-officially accepted species have been taken to make a comparison and obtain a final result. The samples of the genera have been treated to obtain useful sequencing data using the Angiosperms353 kit.

The phylogenetic analysis of these genetic data, which constitutes the result of this study, has allowed us to explore the evolutionary relationships between the genera of Asteraceae in this specific geographic region. This molecular approach provides a clearer vision of the genetic diversity and evolution of these plants in the Catalan Countries. The results of this study contribute to the understanding of biodiversity and the evolutionary history of Asteraceae in this geographical area, with the aim of having an adequate phylogenetic framework that allows for the analysis of other types of data (ethnobotanical and cytogenetic, among others).

Sumari

ÍNDIX DE FIGURES	6
ÍNDIX DE TAULES	8
SÍMBOLS I ACRÒNIMS	9
AGRAÏMENTS	10
1. INTRODUCCIÓ	11
1.1. <i>Flora dels PPCC: biodiversitat i criteri taxonòmic</i>	12
1.2. <i>Les asteràcies: característiques i diversitat</i>	14
1.3. <i>La família de les asteràcies als Països Catalans: distribució geogràfica</i>	18
1.4. <i>Evolució del sistema de classificació botànica: combinació de criteris clàssics amb tècniques de biologia molecular</i>	19
2. OBJECTIUS	22
2.1. <i>Objectius generals</i>	22
2.2. <i>Objectius específics</i>	22
2.2.1. <i>Objectius científics</i>	22
2.2.2. <i>Objectius d'aprenentatge</i>	23
3. MATERIALS I MÈTODES	24
3.1. <i>Disponibilitat de dades publicades</i>	25
3.2. <i>Obtenció de material per a extracció d'ADN</i>	26
3.2.1. <i>Plecs d'herbari</i>	26
3.2.2. <i>Plantes recol·lectades</i>	26
3.3. <i>Extracció d'ADN</i>	27
3.3.1. <i>Neteja amb columnes</i>	27
3.3.2. <i>Quantificació de la concentració d'ADN de les mostres</i>	27
3.3.3. <i>Electroforesi</i>	28
3.3.4. <i>Sonicació</i>	29
3.4. <i>Construcció de genoteques</i>	29
3.5. <i>Hibridacions a partir del kit Angiosperms353</i>	34
3.6. <i>Anàlisi de les dades</i>	39
4. RESULTATS I DISCUSSIÓ	40
4.1. <i>Resultats relacionats als objectius d'aprenentatge</i>	40

4.2.	<i>Llista definitiva de gèneres</i>	42
4.3.	<i>Obtenció dels gens d'interès</i>	43
4.4.	<i>Resultats de laboratori</i>	47
4.5.	<i>Arbre filogenètic de les asteràcies a nivell de gènere dels PPCC</i>	47
5.	CONCLUSIONS	51
6.	BIBLIOGRAFIA	52
7.	ANNEXOS	58

Índex de figures

Figura 1. Mapa dels PPCC.

(https://ca.m.wikipedia.org/wiki/Fitxer:Mapa_dels_Pa%C3%AFsos_Catalans.svg) _____ 11

Figura 2. Capítols florits de la carxofera (*Cynara scolymus*).

(<https://www.enciclopedia.cat/historia-natural-dels-paisos-catalans/les-asteracies-o-compostes>) _____ 14

Figura 3. Fulla d'*Hypochoeris radicata*.

(https://floracatalana.cat/flora/vasculars/cercador_imatges?field_nom_cientific_value=Hypochoeris&field_sinonims_value=&field_referencia_imatge_value=) _____ 15

Figura 4. Fulla d'*Arctium lappa*.

(https://floracatalana.cat/flora/vasculars/cercador_imatges?field_nom_cientific_value=Arctium&field_sinonims_value=&field_referencia_imatge_value=) _____ 15

Figura 5. Inflorescència terminal d'*Helianthus annuus*.

(https://floracatalana.cat/flora/vasculars/cercador_imatges?field_nom_cientific_value=Helianthus&field_sinonims_value=&field_referencia_imatge_value=) _____ 16

Figura 6. Inflorescència axil·lar de *Prenanthes purpurea*.

(https://floracatalana.cat/flora/vasculars/cercador_imatges?field_nom_cientific_value=Prenanthes&field_sinonims_value=&field_referencia_imatge_value=) _____ 16

Figura 7. Diferents flors d'asteràcies: a) Flor de *Bidens aurea* amb cinc lígules; b) Flor zigomorfa de *Calendula arvensis*; c) Flor actinomorfa de *Conyza bonariensis*.

(<https://floracatalana.cat/flora/vasculars/>) _____ 17

Figura 8. Recol·lecció de l'espècie *Adenostyles alliariae* al municipi de Setcases, Girona. _____ 26

Figura 9. Preparació del gel d'agarosa per a dur a terme una electroforesi.

(<https://www.edvotek.com/site/pdf/101sp.pdf>) _____ 28

- Figura 10.** Procediment per dur a terme l'electroforesi segons el protocol Edvo-kit nº101 d'Edvotek. (<https://www.edvotek.com/site/pdf/101sp.pdf>) _____ 28
- Figura 11.** Procediment a seguir en el pas de *Configuració de la barreja de bloquejadors*. **Font:** Illumina _____ 36
- Figura 12.** Mapa de calor dels 353 gens d'interès presents en els gèneres estudiats. L'escala d'intensitat indica el percentatge de l'expressió gènica de cada gen. Eix X: 353 gens d'interès; Eix Y: Mostres de gèneres d'asteràcies. _____ 45
- Figura 13.** Mapa de calor que indica la quantitat de gens paràlegs presents en els 353 gens d'interès de les mostres estudiades. La barra d'intensitat indica el percentatge de l'expressió gènica dels gens paràlegs. Eix X: 353 gens d'interès; Eix Y: Mostres de gèneres d'asteràcies. _____ 46
- Figura 14.** Arbre filogenètic a nivell de gènere de les asteràcies als PPCC. S'observa a cada node el valor del seu suport. Llegendes: i) clade blau: Cichorioideae; ii) clade rosa: Asteroideae; iii) clade groc: Carduoideae. _____ 50

Índex de taules

Taula 1. Preparació d'un tub lliure de nucleases. Font: Illumina _____	29
Taula 2. Programa del termociclador al pas NEBNext preparació de terminacions . Font: Illumina _____	30
Taula 3. Preparació de la barreja NEBNext preparació de terminacions en el pas Lligament dels adaptadors . Font: Illumina _____	30
Taula 4. Barreja per dur a terme una amplificació PCR al pas Enriquiment per PCR del DNA lligat amb adaptadors . Font: Illumina _____	32
Taula 5. Condicions de la PCR al pas Enriquiment per PCR del DNA lligat amb adaptadors. Font: Illumina _____	32
Taula 6. Paràmetres de la barreja d'hibridació. Font: Illumina _____	34
Taula 7. Paràmetres de la barreja de bloquejadors. Font: Illumina _____	35
Taula 8. Paràmetres del termociclador pel pas de Configuració d'hibridació. Font: Illumina _____	35
Taula 9. Paràmetres de la barreja de PCR Master Mix. Font: Illumina _____	38
Taula 10. Paràmetres del termociclador per la barreja Master Mix. Font: Illumina _____	38
Taula 11. Gèneres de la família <i>Asteraceae</i> presents als PPCC. _____	42
Taula 12. Mostres preparades per a seqüenciar en un futur identificades amb els seus índexs. _	47

Símbols i acrònims

PPCC – Països Catalans

CTAB – Bromur d'hexadeciltrimetilamoni

IAA – Alcohol isoamílic

BME – 2-Mercaptoetanol

NGS – Seqüenciació de Nova Generació

ADN – Àcid desoxiribonucleic

IBB – Institut Botànic de Barcelona

Agraïments

Vull expressar els meus agraïments més sincers a l'Institut Botànic de Barcelona per donar-me l'oportunitat d'introduir-me en el fascinant món de la botànica a través d'aquest Treball de Fi de Grau.

La meva gratitud és profunda i està dirigida de manera especial al meu tutor, Pol Fernández, per tenir la paciència i la bona disposició per a guiar-me en el meu aprenentatge. Pel seu caràcter tan humà i entendre les situacions de cada persona. Per altra banda, agrair a la Dra. Teresa Garnatje per deixar-me descobrir un món totalment nou per a mi. A més, s'agraeix una persona tan entregada a la seva feina.

Ha sigut tot un plaer treballar amb un grup de persones apassionades per la seva feina. La seva dedicació i entusiasme han estat una font constant d'inspiració i motivació en aquest viatge de descobriment i aprenentatge.

Finalment, dedico un agraïment especial als meus amics, família i companys d'estudi, que han estat un suport imprescindible durant aquest projecte.

1. Introducció

La representació cartogràfica dels tàxons vegetals als Països Catalans (PPCC) va ser iniciada per Pius i Quer a partir del 1921 (Bolòs, 1985). L'acumulació de dades al llarg del temps ha permès fer publicacions generals de la distribució de les plantes vasculares dels PPCC. Un important pas en l'evolució dels estudis dels tàxons vegetals va ser la creació de la "Flora dels Països Catalans" d'Oriol de Bolòs i Josep Vigo (Bolòs & Vigo, 2005), actual obra de referència per a conèixer la distribució general dels tàxons en aquest territori. Prèviament, Bolòs (1985), en la seva obra "Corologia de la flora del Països Catalans" ja havia inclòs la distribució geogràfica de les plantes en el territori. Actualment, gràcies a les noves tecnologies, s'està intentant millorar l'anàlisi de totes les dades dels PPCC per tal de poder interpretar-les de manera correcta.

Els PPCC estan situats a la part oriental de la Península Ibèrica i les illes adjacents, incloent-hi les Illes Balears, el País Valencià, la Catalunya del Nord (a l'estat francès), Andorra, la Franja de Ponent, en territori aragonès) i l'Alguer (a l'illa de Sardenya pertanyent a l'estat italià), amb una superfície aproximada de 70.000 km² (Figura 1). Aquest territori presenta una gran diversitat fisiogràfica i biogeogràfica.

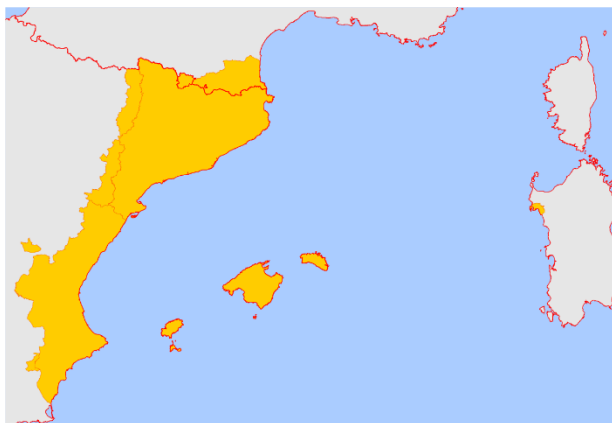


Figura 1. Mapa dels PPCC.

Les representacions cartogràfiques de la distribució son una valuosa eina per a la conservació i l'estudi de les plantes del territori.

(https://ca.m.wikipedia.org/wiki/Fitxer:Mapa_dels_Pa%C3%AFsos_Catalans.svg)

En el nostre cas, aprofundirem en l'estudi de les plantes de la família de les asteràcies, aportant un arbre filogenètic a nivell de gènere. Aquesta família representa el 10% de totes les espècies de plantes amb flor i té una increïble diversitat de formes (Mandel et al., 2019). Consta d'uns 1.100 gèneres i unes 20.000 espècies aproximadament (Aizpuru et al., 2004). Degut a que és una de les famílies de plantes més estudiades avui dia, es pot consultar molta bibliografia, com per exemple alguns articles d'estudis filogenètics de les asteràcies (Mandel et al., 2019; Palazzesi et al., 2022) o

el repositori del *Tree Of Life Explorer* (*Species - Tree of Life Explorer*, <https://treeoflife.kew.org/>), on es pot trobar l'arbre filogenètic de totes les famílies de plantes (*Species - Tree of Life Explorer*, <https://treeoflife.kew.org/>), per a contrastar la informació.

Els arbres filogenètics són estructures matemàtiques que permeten modelitzar les relacions de parentiu entre un grup d'organismes o entre unes determinades seqüències de DNA, RNA o proteïnes (*Els arbres filogenètics* | *enciclopedia.cat*, <https://www.enciclopedia.cat/historia-natural-dels-paisos-catalans/els-arbres-filogenetics>).

En el cas de les asteràcies, els arbres filogenètics amb una representació més àmplia i més resolts es basen en l'ADN del cloroplast (Mandel et al., 2019). Aquesta informació genètica s'ha utilitzat per a comprendre com van sorgir i es van diversificar aquest tipus de plantes. Per altra banda, Bremer al 1994 va realitzar un resum de les característiques morfològiques en un cladograma per a identificar branques basals. Malgrat la rellevància de tots aquests estudis previs, no existeix cap filogènia que inclogui tots els tàxons presents a la Flora dels PPCC.

Amb la qual cosa, el repte i objectiu d'aquest treball és estudiar les relacions evolutives d'aquesta família als PPCC elaborant un arbre filogenètic que permeti donar una visió més completa i aportar dades genètiques de noves espècies. La suma de molts treballs com aquests permetrà en un futur una major comprensió de l'origen i la divergència de les asteràcies al llarg de l'evolució. A més, el fet d'estudiar la filogènia permet seguir les transicions de caràcters d'interès (Panero et al., 2014) per a, posteriorment, utilitzar-los en àmbits com la medicina o la cosmètica, entre d'altres.

1.1. Flora dels PPCC: biodiversitat i criteri taxonòmic

La flora dels PPCC és extremadament diversa i rica, gràcies a que es poden trobar en diferents condicions climàtiques i topogràfiques. Aquesta riquesa es deu a la interacció de diversos factors, com l'orografia, el clima mediterrani, la proximitat del mar i la diversitat dels sòls.

La biodiversitat vegetal és molt notable, amb una gran varietat d'espècies de flora vascular, incloent-hi plantes herbàcies, arbusts i arbres (López-Pujol et al., 2006). A més, es troba quantitat nombrosa de plantes endèmiques, és a dir, aquelles que només es troben en aquesta regió i no en altres parts del món. Aquest grau d'endemisme és una prova de la singularitat i importància de la

flora dels PPCC per a la conservació de la biodiversitat. En part, aquesta biodiversitat és deguda, en part, a la influència humana (Aymerich & Sáez, 2021).

El criteri taxonòmic és fonamental per a l'estudi i la classificació de la flora en aquesta regió. A través dels anys, s'han realitzat nombroses investigacions científiques per intentar esbrinar de manera més precisa l'organització i agrupació de les espècies vegetals segons el seu grau de parentiu i similituds morfològiques, genètiques i ecològiques (Bolòs, 2005; Folch, 1986). Aquestes investigacions han fet que es desenvolupin sistemes de classificació taxonòmica que permeten agrupar les plantes en diferents nivells jeràrquics, com ara ordres, famílies, gèneres i espècies. Gràcies a aquest criteri taxonòmic es pot dur a terme l'elaboració de representacions cartogràfiques de la distribució de les plantes als PPCC. Mitjançant la classificació adequada de les espècies, és possible identificar els territoris on es poden trobar i establir patrons geogràfics. A més, ajuden a identificar patrons de diversitat i endemismes, i contribuir a la conservació de la flora regional. Es poden trobar molts estudis relacionats amb aquest tema, com per exemple l'article d'Otamendi et al. (2020) sobre els principals patrons de distribució de les plantes vasculares dels PPCC.

En aquest treball, el criteri taxonòmic del qual hem partit ha sigut el proporcionat per la Flora dels Països Catalans (Bolòs & Vigo, 2005), el qual és un criteri més aviat sintètic. En aquest, les espècies es consideren àmplies i sovint corresponen a cercles de races més o menys complexes. L'espècie es divideix en subespècies, que a vegades poden arribar a coincidir amb les microespècies utilitzades per diferents botànics. La utilització d'aquest criteri representa certs avantatges (Bolòs & Vigo, 2005): a) l'espècie com a unitat taxonòmica principal acostuma a quedar ben delimitada i és fàcil de reconèixer; b) les afinitats que enllacen els tàxons estretament relacionats es poden observar clarament; c) l'ús de categories infra-específiques permet que, en tots els casos, es pugui discernir les unitats més simples que integren el cercle de formes i que poden tenir una significació geogràfica i ecològica considerable.

Per altra banda, és molt important distingir les categories infraespecífiques fins als nivells més baixos, però cal evitar que la pretensió de separar entre ells tàxons feblement caracteritzats conduïxi a errors de determinació i a confusions que, en ocasions, arriben fins i tot a les grans obres de síntesi florística. Tot i així, cal tenir en compte que aquesta classificació es tracta d'un procés complex i evolutiu, ja que a mesura que es descobreixen noves espècies i s'aclareixen les relacions entre elles, és possible que es produeixin ajustaments en la taxonomia. Això fa que l'estudi

de la flora del PPCC sigui una tasca constant i en evolució, amb l'objectiu de comprendre millor la seva diversitat i conservar el patrimoni natural botànic per a les generacions futures.

1.2. Les asteràcies: característiques i diversitat

Per saber com caracteritzar les asteràcies he seguit el criteri que va seguir el Volum XVI de la *Flora Iberica* (Devesa et al., 2015).

Les asteràcies són plantes que poden ser herbes anuals, biennals o perennes, en rares ocasions com a arbustos, i fins i tot, com a arbres en espècies fora de la Península Ibèrica (Figura 2). A més, algunes de les espècies poden presentar làtex i olis essencials.

Algunes d'aquestes plantes poden ser inermes o espinoses, amb diversitat en la pilositat de les



Figura 2. Capítols florits de la carxofera (*Cynara scolymus*).

(<https://www.enciclopedia.cat/historia-natural-dels-paisos-catalans/les-asteracies-o-compostes>)

seves parts. Acostumen a ser plantes hermafrodites o ginomonoiques, tot i que a vegades poden ser monoiques, dioiques o ginodioiques. Les seves tiges poden ser simples o ramificades, amb secció aproximadament circular, de textura sòlida o a vegades buides en el seu interior, amb o sense espines, foliades o sense fulles. A vegades són alades degut a la decurrència de les fulles o poden ser semblants a tiges sense fulles, i en casos excepcionals, les plantes poden ser acaules (Devesa et al., 2015).

Fulles

es fulles de les asteràcies acostumen a ser alternes o espiralades, amb poca freqüència oposades, i rarament verticil·lades. Són simples i poden estar totes en una roseta (Figura 3) o les primeres poden estar en forma de roseta i la restes caulinars. Per altra banda, la seva consistència pot arribar a ser molt variada. Poden ser herbàcies, carnoses, coriàcies o membranoses. Les fulles poden tenir un pecíol o ser sèssils i, a més, poden tenir una prolongació més enllà del limbe. Les fulles poden ser enteres o estar dividides, i a vegades poden tenir espines o espínules al marge (Figura 3) o, fins i tot, a sobre el limbe. Les fulles tenen predominantment una nervadura pinnada (Figura 4), freqüentment la secundària reticulada, encara que hi ha casos excepcionals amb nervadures paral·lelinèrvies o palmatinèrvies, i no tenen estípules.



Figura 3. Fulla d'*Hypochoeris radicata*.

(https://floracatalana.cat/flora/vasculars/cercador_imatges?field_nom_cientific_value=Hypochoeris&field_sinonims_value=&field_referencia_imatge_value=)



Figura 4. Fulla d'*Arctium lappa*.

(https://floracatalana.cat/flora/vasculars/cercador_imatges?field_nom_cientific_value=Arctium&field_sinonims_value=&field_referencia_imatge_value=)

Inflorescències

Les inflorescències de les asteràcies es presenten en capítols, que poden ser terminals (Figura 5) o axil·lars (Figura 6), solitàries o agrupades en inflorescències complexes. Tenen molta diversitat de formes: globoses, espigades, paniculades o corimboses. A vegades, les últimes fulles caulinars poden envoltar els capítols formant un involucre. Els capítols poden tenir múltiples flors o ser pauciflors, i poques vegades són uniflors.

El receptacle del capítol pot tenir diferents formes: aplanat, convex, cònic, hemisfèric o cilíndric. A més, la seva superfície pot ser llisa, alveolada o rugosa, sòlid o buit, a vegades amb bràctees florals en forma d'escates, setes o pèls. Pot estar envoltat per bràctees ben desenvolupades (involucre), que poden ser herbàcies, membranoses o més o menys coriàcies, pot trobar-se també acompanyades de bràctees més petites que semblen un segon involucre a la base (involucel). Les bràctees poden estar organitzades en una filera única o en diverses sèries, a vegades poc definides, lliures o soldades entre sí. En algunes ocasions amb un apèndix apical de forma molt variada, i poden ser inermes o espinoses.



Figura 5. Inflorescència terminal d'*Helianthus annuus*.

(https://floracatalana.cat/flora/vasculars/cercador_imatges?field_nom_cientific_value=Helianthus&field_sinonims_value=&field_referencia_imatge_value=)



Figura 6. Inflorescència axil·lar de *Prenanthes purpurea*.

(https://floracatalana.cat/flora/vasculars/cercador_imatges?field_nom_cientific_value=Prenanthes&field_sinonims_value=&field_referencia_imatge_value=)

Flors

Les flors de les asteràcies poden ser actinomorfes (Figura 7c) o zigomorfes (Figura 7b), per regla general són pentàmeres (Figura 7a) (excepte el gineceu, el calze, que està molt transformat o falta, i a vegades l'androceu), i són epígines, hermafrodites, unisexuals o, en ocasions, apomíctiques (flors neutres). Algunes vegades totes les flors d'un mateix capítol són sexualment idèntiques (homògames), tant si són hermafrodites com si són unisexuals, però d'altres, les flors poden ser hermafrodites i unisexuals dins d'un mateix capítol (heterògames), amb flors de mida subigual, les de la perifèria més grans o més petites.

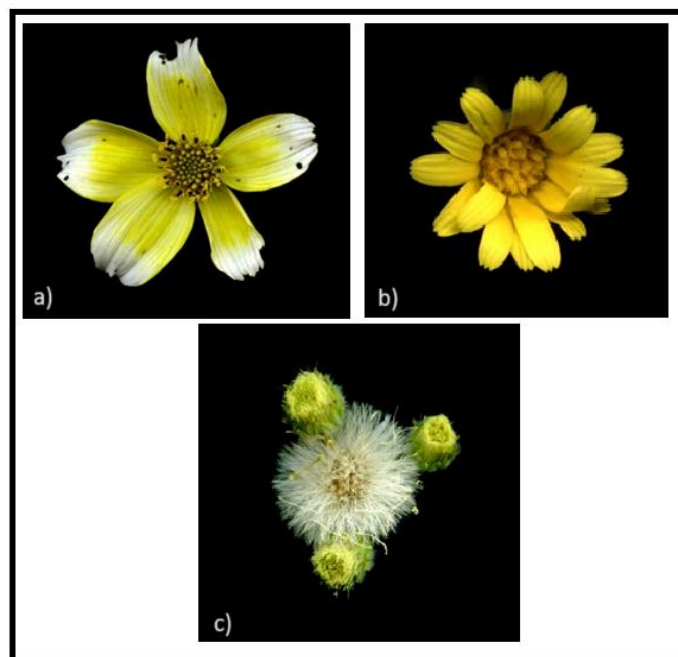


Figura 7. Diferents flors d'asteràcies: a) Flor de *Bidens aurea* amb cinc lígules; b) Flor zigomorfa de *Calendula arvensis*; c) Flor actinomorfa de *Conyza bonariensis*. (<https://floracatalana.cat/flora/vasculars/>)

La corol·la es gamopètala, tetràmera o amb més freqüència pentàmera, de morfologia variable, i rars vegades manca, amb una porció basal estreta, més o menys cilíndrica i poc acolorida (tub), semblant a un peu, i una porció superior més ampla i acolorida (limbe).

El limbe pot ser actinomorfe (Figura 7c), presentant entre 4, 5 o 6 lòbuls o dents de mesura semblant, convergents, erectes, erecto-patents o patents (en les flors tubulars o flòsculs), a vegades poc diferenciat del tub, molt estret (flors filiformes), o zigomorfe (Figura 7b), amb dos lòbuls d'un costat i tres més o menys soldats de l'altre (en les flors bilabiades). També es poden trobar amb tots els lòbuls soldats unilateralment formant una llengüeta (flors ligulades) amb 1, 3 o 5 dents (lígula tridentada o quinquedentada, respectivament).

L'androceu consta de 4 o 5 estams inserits en la part superior del tub de la corol·la, alterns amb els lòbuls o dents en les flors actinomorfes (Figura 7c), soldats per les anteres i, a vegades, també pels filaments, formant un tub pel qual passa l'estil. Els filaments estamentals són cilíndrics, a vegades amb el connectiu ampliat per sota de les anteres (collaret), llisos, glandulars, papil·losos o pilosos. Les anteres són lineals i acostumen a ser calcariformes.

El gineceu de les asteràcies està format per dos carpels units, amb un ovari inferior i unilocular. Presenten un únic òvul anàtrop, amb un tegument i megasporangi prim. La placentació és basal. El pistil té un estil que sovint té un nèctar a l'extrem apical. L'estil està dividit distalment en dues branques que tenen papil·les estigmàtiques a la seva cara adaxial disposades en dues línies separades o bé, en una sola banda contínua.

1.3. La família de les asteràcies als Països Catalans: distribució geogràfica

Les asteràcies es troben distribuïdes àmpliament als PPCC, ocupant diferents hàbitats i regions geogràfiques. La seva diversitat i distribució depèn de factors com el tipus de sòl, el clima, l'altitud i l'ús del territori. Algunes espècies d'asteràcies són molt comunes i es poden trobar en tota la regió, mentre que altres tenen una distribució més restringida.

La distribució geogràfica de les asteràcies als PPCC es pot trobar a l'Atlas Corològic de la Flora Vasculardels Països Catalans, que permet situar cada espècie en els PPCC (Bolòs, 2003).

A partir d'aquesta informació inclosa en l'obra anterior, es poden diferenciar els hàbitats de les asteràcies en diferents zones:

- Zones muntanyoses: com els Pirineus i altres serralades, s'hi poden trobar diverses espècies d'asteràcies adaptades a les condicions alpines i subalpines. Aquestes plantes són especialment diverses en les pastures alpines i prats de muntanya.
- Zones humides: s'hi poden trobar algunes espècies d'asteràcies que són especialment adaptades a aquests ambients, com les que es troben a les vores de rius, llacs i basses.
- Zones costaneres i de planes: es poden trobar en diferents hàbitats, com a les vores del mar, els camps, les vores de camins i els conreus. Algunes espècies són típiques d'aquestes àrees i són comunes a tota la regió.
- Zones alterades: com els marges de carreteres, s'hi poden trobar algunes espècies que són especialment resistents a les perturbacions.

Podem considerar que la seva distribució està influïda per diversos factors ambientals. Aquests factors, a mesura que passa el temps, poden fer que algunes espècies canviïn el seu hàbitat o arribin a extingir-se.

1.4. Evolució del sistema de classificació botànica: combinació de criteris clàssics amb tècniques de biologia molecular

Al llarg de la història, l'evolució del sistema de classificació botànica ha estat influïda per l'ús de tècniques de biologia molecular, aportant una nova dimensió a la comprensió de les relacions evolutives entre les plantes. Aquesta combinació de criteris més clàssics amb dades moleculars ha millorat significativament els sistemes de classificació botànica actuals, aportant precisió i robustesa als resultats obtinguts.

La classificació botànica ha evolucionat al llarg del temps, es va començar basant-se en criteris clàssics que es centraven en característiques morfològiques, com la forma de les fulles, les flors, les arrels i altres parts de les plantes. Aquest enfocament va ser àmpliament acceptat i utilitzat, per tant, va persistir durant molts anys, donant lloc a sistemes de classificació com el de Linné (Benítez de Rojas, 2006). Però, la filogènia molecular, que utilitza totes les dades genètiques per construir arbres filogenètics que representen les relacions evolutives entre espècies, ha fet que es revolucioni el món de la classificació botànica. Avui dia, els sistemes de classificació combinen tant criteris morfològics com dades moleculars per aconseguir classificacions de les plantes, ja que les relacions

evolutives reals poden diferir de les inferides únicament a partir de característiques morfològiques (Benítez de Rojas, 2006). La taxonomia molecular permet una classificació encara més precisa i fonamentada en l'evolució real de les plantes. Amb la qual cosa, la botànica ha avançat des dels criteris morfològics tradicionals fins una combinació d'enfocaments morfològics i moleculars.

En el nostre cas, hem utilitzat el *kit Angiosperms353* (Johnson et al., 2019), un mètode que ha demostrat ser eficient i econòmic (Hale et al., 2020), basat en la tecnologia de seqüenciació massiva, comunament anomenada *Next Generation Sequencing*, la qual ens permet la seqüenciació ràpida i econòmica de regions del genoma, en el nostre cas seqüències d'ADN dels diferents gèneres d'angiospermes (Escalante et al., 2014). Degut a que moltes de les mostres que s'utilitzaven en aquest estudi eren mostres d'herbari, podien estar deteriorades a conseqüència del pas del temps. Per tant, aquesta tecnologia també es pot aplicar a petites quantitats d'ADN molt fragmentat, aprofitant al màxim el material existent (Brewer et al., 2019). A partir d'aquestes dades es pot procedir a fer els estudis bioinformàtics pertinents per a obtenir l'arbre filogenètic.

A partir de les sondes de seqüenciació d'aquest *kit*, les quals són fragments curts d'ADN complementaris a seqüències de gens que es consideren útils per a estudis filogenètics en angiospermes (Johnson et al., 2019). Les sondes capturen les seqüències d'interès, en el nostre cas els 353 gens d'interès que acostumen a ser de còpia única en el genoma de les angiospermes. Aquests van ser escollits gràcies a un disseny previ de sondes (Johnson et al., 2019) que es va centrar en seleccionar els gens ortòlegs. Va ser repte complicat degut a que en les plantes ens trobem amb una gran quantitat de gens paràlegs degut a les duplicacions (Sáenz De Miera Y Carnicer, 2018).

Per a maximitzar el potencial filogenètic de les sondes i reduir els costos de producció s'utilitza un enfocament de *clustering k-medoids* que permet identificar el nombre mínim de seqüències necessàries per representar cada gen d'interès en el conjunt final de sondes. Finalment, a partir de les dades obtingudes, i mitjançant anàlisis bioinformàtiques s'obté l'arbre filogenètic. Es tracta d'un diagrama que representa les relacions evolutives entre organismes. Tot i que la família de les asteràcies és un grup de plantes monofilètic (Mandel et al., 2019), és a dir, tot el grup taxonòmic té un avantpassat original comú (*Plantas y Hongos*, <https://www.plantasyhongos.es/glosario/monofiletico.htm>), està sent un gran repte esbrinar el seu origen i diversificació (Mandel et al., 2019).

A partir de l'arbre filogenètic podem diferenciar diferents termes:

- Clades : grup monofilètic compost per un tàxon ancestral i tots els descendents d'aquest avantpassat.
- Tàxon: es tracta d'un clade al qual se li ha assignat una categoria taxonòmica (Regne, Divisió, Classe, Ordre, Família, Gènere, Espècie)

En conclusió, l'evolució del sistema de classificació botànica ha sigut influenciat per l'ús de tècniques de biologia molecular, que estan aportant una nova dimensió a la comprensió de les relacions evolutives entre les plantes. És molt important doncs seguir obtenint i analitzant dades moleculars per poder complementar significativament la robustesa dels sistemes de classificació botànica actuals.

2. Objectius

2.1. Objectius generals

L'objectiu general de l'estudi consisteix en obtenir un nou marc filogenètic de les asteràcies a nivell de gènere dels PPCC, per a poder analitzar diversos conjunts de dades en el futur. Per exemple *mapejar* altres dades genètiques (incloent les citogenètiques), fitoquímiques i etnobotàniques, entre d'altres, de la família de les asteràcies als PPCC. Aspirem doncs a aprofundir en la comprensió de la complexitat evolutiva i interaccions entre els tàxons d'aquesta important família de plantes en el territori d'estudi.

2.2. Objectius específics

2.2.1. Objectius científics

2.2.1.1. Creació de la llista de gèneres a estudiar

El primer objectiu per a poder començar a fer l'estudi filogenètic de la família *Asteraceae* va ser crear una llista completa dels gèneres que es trobaven als PPCC. A més, s'havia de tenir en compte la sinonímia dels gèneres ja que podia arribar a ocasionar confusions i errors.

2.2.1.2. Obtenció dels gens d'interès

A partir de la llista de gèneres que havíem creat, s'havien d'aconseguir els 353 gens d'interès que buscàvem. Aquest pas podia ser a partir de:

- Dades genòmiques de diferents bases de dades disponibles a d'Internet
- Hibridacions amb el *kit Angiosperms353* al laboratori

2.2.1.3. Construcció d'arbres de gens

L'objectiu era aconseguir un arbre de gènere per cada un dels gens. Aquests tenen informació parcial que cal ser resumida en un arbre final que inclogui tota la informació.

2.2.1.4. Construcció d'un arbre filogenètic de les asteràcies dels PPCC a nivell de gènere

Finalment, després d'haver fet tots els passos anteriors l'objectiu principal del treball va ser elaborar l'arbre filogenètic de les asteràcies dels PPCC a nivell de gènere. El major repte segons el meu criteri va ser adquirir el màxim nombre de dades, ja que és la base per poder a començar a treballar.

2.2.2. Objectius d'aprenentatge

2.2.2.1. Adquisició de coneixements sobre tècniques de laboratori

Són dos els objectius d'aprenentatge sobre tècniques de laboratori:

- Aprendre a treballar en un laboratori de biologia molecular.
- Familiaritzar-me amb les tècniques d'extracció d'ADN, construcció de genoteques i hibridacions.

2.2.2.2. Adquisició de coneixements sobre anàlisis bioinformàtiques

Degut a què no tenia coneixements previs sobre anàlisis bioinformàtiques, els dos primers objectius van ser els següents:

- Aprendre a desenvolupar-me en el llenguatge de programació *Python*.
- Adquirir coneixements per utilitzar programaris d'anàlisis de dades.

2.2.2.3. Adquisició de coneixements sobre recol·lecció de plantes

En aquest cas, l'objectiu principal era aprendre com fer una recol·lecció de les plantes al camp que serien necessàries per a dur a terme la filogènia de les asteràcies. Això incloïa aprendre a identificar les plantes segons la seva morfologia externa i, aprendre a preservar-les, en aquest cas dessecant-les amb gel de sílice. Un altre aspecte important era saber la distribució geogràfica de cada gènere per saber per on buscar.

3. Materials i mètodes

Inicialment es va elaborar un llistat de tots els gèneres de les asteràcies que es poden trobar als PPCC a partir de la Flora Manual dels Països Catalans (Bolòs, 2005). A partir d'aquest llistat es realitzen estudis i anàlisis que explicarem a continuació per tal d'obtenir l'arbre filogenètic a partir d'aquests gèneres.

D'entrada, es va fer una selecció de quines espècies serien les representants de cada gènere. Es va escollir una espècie per gènere seguint una sèrie de criteris. El primer pas que es va dur a terme va ser buscar de quines espècies dels PPCC podíem aconseguir dades genòmiques ja publicades. Es va fer una cerca de dades de seqüenciació dels gèneres en els següents repositoris: Tree Of Life Explorer (*Species - Tree of Life Explorer*, <https://treeoflife.kew.org/>) i NCBI – GenBank (*GenBank Overview*, <https://www.ncbi.nlm.nih.gov/genbank/>). L'objectiu era veure de quants gèneres no teníem la necessitat de fer extraccions, atès que hi havia dades disponibles, i deixar-ho anotat al llistat de gèneres d'asteràcies dels PPCC.

Per tal de veure si les dades que obteníem de l'NCBI – GenBank (*GenBank Overview*, <https://www.ncbi.nlm.nih.gov/genbank/>) eren de bona qualitat vam fer un estudi preliminar a partir d'anàlisis bioinformàtiques. A partir dels resultats obtinguts vam acceptar per bones aquelles dades que tinguessin com a mínim 100 gens dels 353 gens que buscàvem amb una cobertura del 50%, és a dir, que com a mínim recuperéssim la meitat del gen.

A partir d'aquesta llista, vam saber quin eren els gèneres dels quals no teníem cap mena d'informació. Llavors, el següent pas va ser fer una selecció de plecs de l'herbari de l'Institut Botànic de Barcelona (acrònim BC) dels quals es va procedir a extreure'n l'ADN.

Els plecs d'herbari es tracta d'una fulla de paper on es preserva un mostra d'una planta recol·lectada en el seu entorn natural. Aquesta mostra pot ser de fulles, flores, talls..., s'asseca i es fixa al paper. A més, en el plec es posa una etiqueta amb la informació referent a la planta i al seu recol·lector. L'objectiu és mantenir exemplars secs i identificats durant centenars d'anys per tal de poder ser utilitzats en la investigació botànica (Aranzadi, 2023).

Per últim, en el cas dels gèneres que no vam trobar ni dades de seqüenciació a Internet, ni plecs a l'herbari, es va fer una sortida al camp per tal de recol·lectar-los i, posteriorment, fer extraccions d'ADN.

3.1. Disponibilitat de dades publicades

Tal i com s'ha esmentat anteriorment, el primer pas va ser fer una cerca de dades de seqüenciació dels gèneres en dos repositoris:

- A. *Tree Of Life Explorer (Species - Tree of Life Explorer, <https://treeoflife.kew.org/>)*: primer es va mirar de quants gèneres podíem obtenir dades dels gens d'interès a partir d'aquest repositori. Aquesta pàgina web proporciona un arbre filogenètic que els biòlegs, per exemple, utilitzen per a representar les relacions evolutives entre tàxons. És una eina que està bastant actualitzada, amb la qual cosa, podem dir que és fiable. En el nostre cas ens vam fixar en la família *Asteraceae*. En total es van obtenir dades de seqüenciació de 92 gèneres d'un total de 164 gèneres de *Asteraceae* presents als PPCC.
- B. *National Center for Biotechnology Information - GenBank (GenBank Overview, <https://www.ncbi.nlm.nih.gov/genbank/>)*: el segon pas va ser fer una cerca de dades de seqüenciació de la resta de gèneres a la base de dades GenBank, en total 59 gèneres. Es va buscar gènere per gènere en aquesta base de dades i assegurant de que les dades complien algun dels requisits següents: (i) dades directament resultants del *kit Angiosperms353* que és el que es va utilitzar en les nostres extraccions, (ii) dades genòmiques d'ADN que s'hagin seleccionat per hibridació amb altres *kits* d'hibridació que sabem que contenen els nostres gens (p.e., *Compositae*) o (iii) en el cas de no trobar les dades mencionades en els dos apartats anteriors, es van seleccionar dades d'ARN transcriptòmic. Aquestes havien de passar un control posteriorment per veure si eren de bona qualitat. Tal i com he explicat anteriorment, les dades obtingudes per a cada gènere havien de contenir com a mínim 100 gens amb una cobertura del 50%. En total vam poder recuperar dades de seqüenciació de 22 gèneres. Per tant, van quedar 41 gèneres dels quals s'havia de fer les extraccions d'ADN i fer la hibridació amb el *kit Angiosperms353*.

3.2. Obtenció de material per a extracció d'ADN

3.2.1. Plecs d'herbari

En primer lloc, per als gèneres dels quals no s'havien trobat dades de seqüenciació a les bases de dades, es van buscar plecs a l'herbari de l'Institut Botànic de Barcelona. L'objectiu era trobar el màxim nombre de plecs que ens servissin per tal de poder tenir dades genòmiques del màxim nombre de gèneres. Es va fer una cerca prèvia a la base de dades de l'herbari en la qual seguïem uns criteris de selecció: (i) havien de ser plecs que la seva procedència fos dels PPCC i (ii), que no fossin més antics del 1980. A partir d'aquesta informació es va realitzar una llista amb els números de plec de l'herbari BC per tal de facilitar la seva cerca posteriorment.

Un cop teníem la llista de possibles plecs de l'herbari de l'IBB que ens podien servir, es va procedir a la seva cerca. A més dels criteris de selecció esmentats anteriorment, ens havíem de fixar en l'estat del material. Ens interessava material el més verd possible, principalment fulles. Finalment vam obtenir 26 plecs dels quals s'ha extret ADN per a poder afegir-los a les anàlisis posteriors.

3.2.2. Plantes recol·lectades

En el cas dels gèneres que no vam poder obtenir dades de seqüenciació ni als repositoris d'Internet ni amb plecs de l'herbari la idea era anar a recol·lectar-les al camp. Degut a que els temps de floració estan variant per temes com el canvi climàtic o la sequera, només vam poder anar a recol·lectar una espècie representant d'un gènere a la localitat de Setcases, l'*Adenostyles* (Figura 8).



Figura 8. Recol·lecció de l'espècie *Adenostyles alliariae* al municipi de Setcases,

3.3. Extracció d'ADN

A partir del material obtingut dels plecs de l'herbari i el material fresc recol·lectat es va dur a terme el següent protocol per fer les extraccions d'ADN: *96-well plate CTAB extraction by Izzie Fairlie & Lisa Pokorny* (Larridon et al., 2020), amb algunes modificacions. Aquest protocol ens permet treballar de manera simultània amb un nombre elevat de mostres (el protocol complet es troba descrit a l'Annex C).

El procés implica: a) la lisi cel·lular; b) l'eliminació de proteïnes i lípids i; c) la purificació de l'ADN.

3.3.1. Neteja amb columnes

Un cop feta l'extracció d'ADN algunes mostres van quedar molt fosques, segurament per la presència de residus de proteïnes o àcids nucleics, amb la qual cosa vam procedir a netejar-les. Aquest procediment ajuda a obtenir un ADN més pur i concentrat, eliminant contaminants que podrien interferir amb les aplicacions d'investigació posteriors. És un pas molt important per a que tot funcioni correctament després (vegeu el protocol complet a l'Annex C)

3.3.2. Quantificació de la concentració d'ADN de les mostres

Un cop vam tenir totes les extraccions fetes i netejades, vam quantificar la concentració d'ADN de les mostres per tal de veure si les concentracions eren adients per continuar treballant amb elles. Per dur a terme aquest procediment, vam utilitzar un fluoròmetre Qubit™ Flex Fluorometer i vam preparar les mostres que havíem d'introduir a l'aparell. De cada mostra vam preparar en un tub Eppendorf 1 µL de la mostra inicial, és a dir, de l'extracció d'ADN. Seguidament vam preparar una solució que constava d'1 µL del reactiu Quant-iT™ i 199 µL del tampó Quant-iT™. Finalment, amb aquesta solució vam enrasar el tub de la mostra fins arribar a un volum de 200 µL.

Un cop vam preparar la mostra, vam escollir l'assaig *dsDNA High sensitivity* i vam seleccionar els criteris correctes i vam obtenir la concentració de cada mostra. A l'Annex C podem veure el resultats que vam obtenir.

3.3.3. Electroforesi

Es va realitzar l'electroforesi en gel d'agarosa per a separar les molècules segons la seva mida. És un mètode molt eficaç per a les biomolècules carregades elèctricament com l'ADN, que és el que tenim en el nostre cas.

A continuació, explicarem el procediment emprat:

El primer pas a realitzar és preparar el motllo per a omplir-lo amb el gel d'agarosa i poder realitzar l'electroforesi. Seguidament, es va procedir a omplir el motllo amb les mostres i dur a terme l'electroforesi. En la Figura 9 i Figura 10, es pot veure de manera molt clara el procediment que s'ha de seguir:

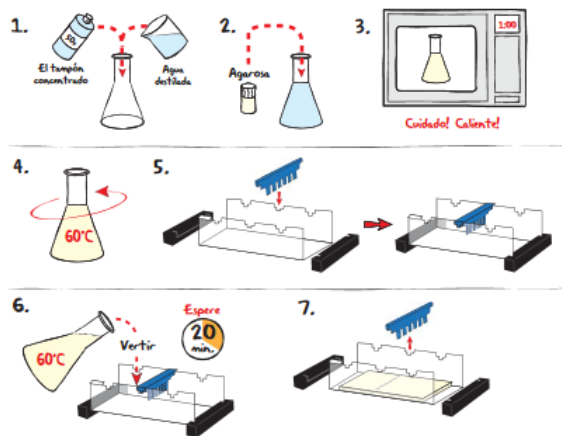


Figura 9. Preparació del gel d'agarosa per a dur a terme una electroforesi.

(<https://www.edvotek.com/site/pdf/101sp.pdf>)

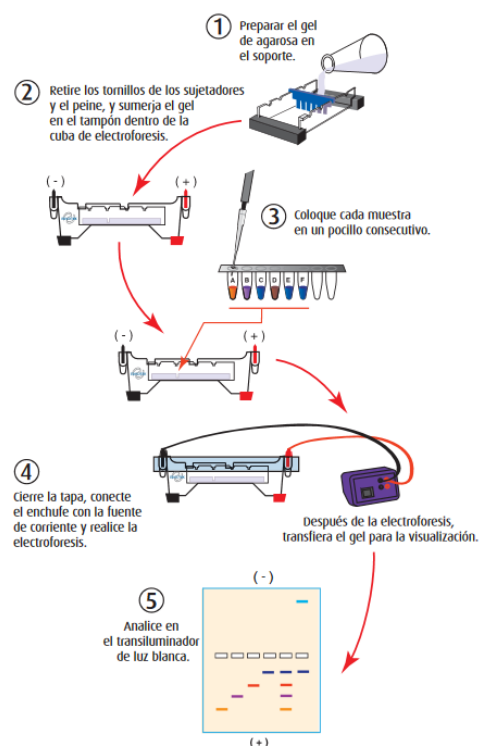


Figura 10. Procediment per dur a terme

l'electroforesi segons el protocol

Edvo-kit n°101 d'Edvotek.

(<https://www.edvotek.com/site/pdf/101sp.pdf>)

3.3.4. Sonicació

El procés de la sonicació ens va permetre tenir mostres d'ADN de mides similars. És a dir, després de fer l'electroforesi, vam observar que hi havia mostres amb més pes molecular que les altres. Amb la qual cosa, vam haver de procedir a la lisi dels fragments d'ADN per a obtenir finalment mostres amb seqüències d'ADN més curtes. Posteriorment, vam fer una comprovació per veure si el procés va ser efectiu realitzant una altra electroforesi. A l'Annex C es pot veure el resultat de les dues electroforesis realitzades, abans i després de la sonicació.

3.4. Construcció de genoteques

El següent pas al laboratori és la construcció de genoteques a partir del *kit NEBNext Ultra II* (Pokorny et al, 2018) per a Illumina, el qual conté els enzims i els tampons necessaris per a convertir una gran quantitat de mostres d'ADN en genoteques d'alta qualitat per a la seqüenciació de nova generació (NGS) a la plataforma Illumina. Aquest *kit* ens permet obtenir genoteques de forma ràpida i fàcil. A continuació explicarem el protocol següent:

Es van preparar 200 ng d'ADN de cada mostra. Es van dissoldre les mostres en aigua. Si el volum d'ADN després de la dissolució era inferior a 25 µL, s'hi va afegir aigua fins a arribar a un volum final de 25 µL.

A. NEBNext preparació de terminacions:

El primer que vam dur a terme va ser preparar un tub lliure de nucleases amb els següents components descrit a la Taula 1.

Taula 1. Preparació d'un tub lliure de nucleases. **Font:** Illumina

NEBNext Ultra II End Prep Enzyme Mix	1,5 µL
NEBNext Ultra II End Prep Reaction Buffer	3,5 µL
ADN fragmentat	25 µL
Volum total	30 µL

Es va pipetejar tot el volum de cada tub uns 10 cops per homogeneïtzar-lo bé. A més, es va fer una breu centrifugació per recollir tot el líquid que es queda enganxat a les parets dels tubs. Una cosa molt important era no contaminar les mostres: per a cada mostra calia utilitzar una punta nova

per a pipetejar. A continuació, vam introduir els tubs en el termociclador amb la tapa escalfada a $\geq 75^{\circ}\text{C}$ i vam executar el següent programa que es pot observar a la Taula 2.

Taula 2. Programa del termociclador al pas NEBNext preparació de terminacions. Font: Illumina

Temps	Temperatura
30 minuts	20°C
30 minuts	65°C
∞	Mantenir a 4°C

B. Lligament dels adaptadors:

A continuació, vam afegir els següents reactius directament a la barreja de preparació de terminacions, és a dir, a la barreja anterior tal i com s'indica a la Taula 3.

Taula 3. Preparació de la barreja NEBNext preparació de terminacions en el pas Lligament dels adaptadors .

End Prep Reaction Mixture	30 μL
NEBNext Ultra II Ligation Master Mix	15 μL
NEBNext Ligation Enhancer	0,5 μL
NEBNext Adaptor for Illumina	1,25 μL
Volum total	46,75 μL

S'ha de tenir en compte que tots els reactius han de ser barrejats abans del seu ús per a un bon funcionament. Per altra banda, el *Ligation Master Mix* i el *Ligation Enhancer* es poden barrejar amb antelació i són estables durant almenys 8 hores a 4°C. No es recomana afegir l'adaptador addicional a una pre-mescla durant el pas de lligament de l'adaptador. Tal i com havíem fet abans, s'ha de fer una breu centrifugació per a recuperar tot el material de les parets dels tubs. Un cop les mostres són homogènies, s'incuben durant 15 minuts a 20°C en el termociclador amb la tapa sense escalfar. Un cop incubades, vam afegir 1,5 μL de l'enzim USER™ a la barreja que teníem de cada mostra. Per finalitzar amb aquest pas, vam homogeneïtzar les mostres correctament i les vam incubar a 37°C durant 15 minuts amb la tapa a $\geq 47^{\circ}\text{C}$.

C. Selecció de mida i neteja del DNA lligat amb adaptadors:

Aquest pas és opcional. Si el material de partida és superior a 50 ng, es pot seguir el protocol de selecció de mida per mantenir la complexitat de la genoteca. Per a una entrada inferior o igual a 50 ng, no es recomana seleccionar la mida. En el nostre cas, sí que vam fer aquest pas.

Amb la qual cosa, vam utilitzar les boles AMPure XP, les quals s'havien de deixar uns 30 minuts aproximadament a una habitació a temperatura ambient abans d'utilitzar-les. A continuació, vam introduir 12,5 µL de boles AMPure XP als 48,25 µL que teníem de la reacció de lligament. Per tal de que la barreja quedés homogènia vam pipetejar uns 10 cops i vam realitzar una breu centrifugació. Seguidament, vam incubar les mostres al gradeta magnètic durant 5 minuts a temperatura ambient. Amb aquest pas el que es vol aconseguir es separar les boles magnètiques del sobrenedant. Després, amb molta cura vam traspasar el sobrenedant, que contenia l'ADN, a tubs nous. Les boles es van descartar en aquest moment.

A continuació, vam afegir un altre cop 5 µL de boles magnètiques a les mostres, les vam homogeneïtzar amb una breu centrifugació i vam deixar-les incubant durant 5 minuts més. En aquest cas, un cop passat el temps establert, vam descartar el sobrenedant que contenia l'ADN que no volíem.

En el següent pas, vam afegir 100 µL d'etanol al 80% a cada tub i el vam deixar incubant durant 30 segons al gradeta magnètic. Després, vam descartar el sobrenedant. Aquest pas es va dur a terme dues vegades. El segon cop que es realitza s'ha de ser curós i treure el màxim sobrenedant possible sense emportar-se les boles magnètiques.

Les boles magnètiques s'han de deixar assecar durant 5 minuts a l'aire. S'ha de tenir cura de que no s'assequin massa ja que això podria disminuir la concentració d'ADN.

A continuació, vam treure els tubs del gradeta magnètic i vam diluir l'ADN que estava a les boles magnètiques en 8,5 µL de 10 mM Tris – HCl or 0,1X TE. Seguidament vam homogeneïtzar les mostres i les vam deixar incubant durant 2 minuts a temperatura ambient.

Per acabar, vam col·locar un altre cop els tubs al gradeta magnètic i, després de 5 minuts, transferim 7,5 µL del líquid a nous tubs de PCR per a amplificar-los posteriorment.

Les mostres poden ser guardades a -20°C.

D. Enriquiment per PCR del DNA lligat amb adaptadors:

Per a dur a terme una amplificació PCR, vam introduir els components descrits a la Taula 4 a tubs estèrils nous.

Taula 4. Barreja per dur a terme una amplificació PCR al pas

Enriquiment per PCR del DNA lligat amb adaptadors . **Font:** Illumina

Adaptor Ligated DNA Fragments	7,5 µL
NEBNext Ultra II Q5 Master Mix	12,5 µL
Index Primer/i7 Primer	2,5 µL
Universal PCR Primer/i5 Primer	2,5 µL
Volum total	25 µL

Vam fer un vòrtex a les mostres fent una centrifugació, les vam introduir al termociclador i vam realitzar una PCR descrita a la Taula 5 les seves condicions.

Taula 5. Condicions de la PCR al pas Enriquiment per PCR del DNA lligat amb adaptadors. **Font:** Illumina

Passos dels cicles	Temperatura	Temps	Cicles
Desnaturalització inicial	98°C	30 segons	1
Desnaturalització	98°C	10 segons	3 - 15
Anellament/extensió	65°C	75 segons	
Extensió final	65°C	5 minuts	1
Emmagatzematge refrigerat	4°C	∞*	

**Quan es posa aquest símbol vol dir que es pot mantenir les mostres el temps que sigui necessari a aquesta temperatura sense que es facin mal bé.*

E. Neteja de la reacció de PCR:

Primer de tot, vam haver de deixar les nostres mostres durant 30 min a temperatura ambient i després vam fer un vòrtex per tal de re-suspendre les boles AMPure XP. A continuació, vam afegir 22,5 µL de les boles resuspeses a la reacció de PCR. Es va barrejar bé pipetejant almenys 10 vegades. S'havia d'anar amb compte d'expulsar tot el líquid de la punta durant l'última barreja.

Després, vam incubar les mostres al gradeta magnètic durant almenys 5 minuts a temperatura ambient. Un cop transcorregut el temps, vam treure amb cura el sobrenedant i el vam descartar.

Per acabar, vam fer el mateix procediment que vam seguir en l'apartat anterior. Vam afegir 100 µL d'etanol al 80% a cada mostra mentre que estaven a la gradeta magnètica. Les vam incubar durant 30 segons i vam descartar el sobrenedant. Aquest pas es va repetir dos cops però al final havíem de treure el màxim sobrenedant possible. A continuació, vam deixar assecar a l'aire les boles durant 5 minuts.

Per acabar, vam diluir l'ADN objectiu afegint 16,5 µL de 0,1X TE. Ho vam barrejar bé i les vam deixar incubar durant 2 minuts a temperatura ambient. Un cop passats els 2 minuts vam introduir les mostres un altre cop al gradeta magnètic i, passats 5 minuts, vam transferir 15 µL de cada mostra a tubs nous de PCR, els vam guardar a -20°C.

3.5. Hibridacions a partir del *kit Angiosperms353*

A partir del sistema d'enriquiment de llibreries NGS en solució, myBaits[®], es va dur a terme la síntesi d'àcids nucleic per produir RNA biotinitades que eren complementàries als nostres objectius de seqüència, és a dir, els nostres gens d'interès. Això ens permet identificar posteriorment les nostres mostres mitjançant una seqüenciació. Vam seguir l'apartat anomenat *High Sensitivity* del protocol *Hybridization Capture for Targeted NGS* (Biosciences, 2023).

A. Configuració d'hibridació

Configuració de la barreja d'hibridació:

El primer que vam fer va ser descongelar els reactius *Hyb*. Els vam fer un vòrtex per a homogeneïtzar-los i després els vam fer una breu centrifugació per a que no quedés material a les parets dels tubs. Tant el reactiu *Hyb N* com el *Hyb S*, es van haver d'escalfar a 60°C per descongelar-los.

Taula 6. Paràmetres de la barreja d'hibridació. **Font:** Illumina

Component	µL / Reacció
Hyb N	9,25
Hyb D	3,5
Hyb S	0,5
Hyb R	1,25
H2O (ronda 1 / ronda 2)	1,10 / 4,40
Esquers	4,40 / 1,10
TOTAL	20

Un cop vam tenir la barreja preparada (Taula 6), vam incubar-la durant 10 minuts a 60°C. A continuació, vam deixar-ho reposar durant 5 minuts abans de continuar. Després, vam repartir 18,5 µL de la barreja per a cada reacció de captura en tubs de 0,2 mL.

Configuració de la barreja de bloquejadors

Degut a que estem treballant amb plantes, vam utilitzar els components descrits a la Taula 7 per fer la barreja de bloquejadors.

Taula 7. Paràmetres de la barreja de bloquejadors. **Font:** Illumina

Component	µL / Reacció
Block O	5
Block X	0,5
TOTAL	5,5

Igual que en el pas anterior, vam repartir 5 mL entre tubs de 0,2 mL. Havíem de tenir dues fileres de tubs preparades. En el nostre cas vam utilitzar un tub només per fer les llibreries. Per tant, vam tenir un tub amb la barreja d'hibridació i un altre tub amb la barreja de bloquejadors.

Després, vam introduir els dos tubs al termociclador seguint el següent programa descrit a la Taula 8.

Taula 8. Paràmetres del termociclador pel pas de Configuració d'hibridació. **Font:** Illumina

Pas	Temperatura	Temps
1	95°C	5 min
2	65°C	5min
3	65°C	∞

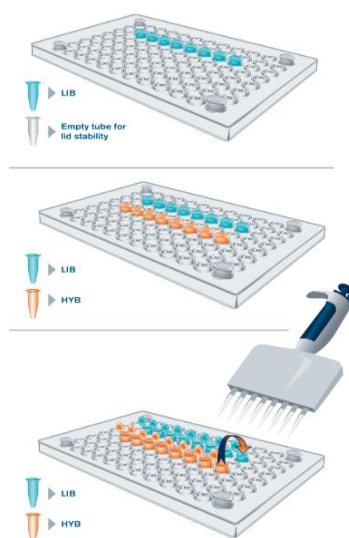


Figura 11. Procediment a seguir en el pas de *Configuració de la barreja de bloquejadors*. **Font:** Illumina

Tal i com es mostra en la Figura 11, vam introduir la barreja d'hibridació al termociclador durant 5 min. Al pas número 2 vam aturar el procés per introduir la barreja de bloquejadors i continuar amb el pas 2.

Per acabar, després del pas 2 vam pipetejar 18 μ L del tub de la barreja de bloquejadors i els vam introduir a l'altre tub. Vam retirar el tub dels bloquejadors i vam fer una breu centrifugació al tub de les hibridacions. Aquest últim el vam deixar al termociclador incubant durant una nit, unes 18 hores aproximadament.

B. Lligament i rentat

Per dur a terme aquest procediment vam decidir que la temperatura de rentat seria 60°C.

Preparació del tampó de rentat X

El primer pas va ser preparar el tampó de rentat X. Vam descongelar el tampó de rentat i el *Hyb S* i el vam homogeneïtzar per tal de que no quedés cap precipitat. A continuació, vam combinar 400 μ L de *Hyb S*, 39,6 mL d'aigua lliure de nucleases i 10 mL de tampó de rentat en un tub de 50 mL. Vam fer un vòrtex i el vam escalfar al bany maria durant 30 minuts abans d'usar-ho.

Preparació de boletes

Només teníem un tub de reacció de rentat, amb la qual cosa només vam treballar aquest tub. Vam afegir a un tub de termociclador 30 μL de boletes. Vam centrifugar-ho per tal de que s'homogeneïtzés.

Seguidament vam introduir el tub amb les boletes a la gradeta magnètica i, un cop van quedar separades, vam retirar el sobrenedant. Després, vam introduir 200 μL del tampó d'unió i vam fer un vòrtex al tub per tal de resuspendre les boletes. A més, vam fer una breu centrifugació per a recuperar el màxim possible de reacció.

A continuació, vam introduir el tub a la gradeta magnètica i vam descartar el sobrenedant. Aquest pas es va repetir dues vegades per a un total de tres rentats. Per acabar, vam resuspendre les boletes ja rentades en 70 μL del tampó d'unió.

Unió de boletes i hibridacions

Vam escalfar el tub amb les boletes fins a 60°C, la temperatura que vam establir des d'un principi, durant 2 minuts. Seguidament vam transvasar la reacció d'hibridació preparada anteriorment al tub que conté les boletes. Per a barrejar-ho tot vam pipetejar diverses vegades.

Després vam introduir el tub anterior, que conté les llibreries més les boletes rentades, al termociclador durant 5 minuts. Al minut 2,5 s'havia d'agitar el tub i fer una breu centrifugació.

Rentat de boletes

Vam introduir el tub a la gradeta magnètica per tal de que les boletes precipitessin fins que la solució fos clara. Un cop separades vam descartar el sobrenedant.

Vam afegir 375 μL de tampó de rentat X escalfat a les boletes i vam posar el tub al bloc de calor durant 15 segons. Seguidament vam fer un vòrtex i una breu centrifugació.

A continuació, vam incubar el tub durant 5 minuts a 60°C al termociclador.

Tots els passos anteriors, és a dir, el rentat que es realitza, es repeteix tres cops. A l'últim rentat intentem treure el màxim líquid possible sense tocar les boletes.

C. Resuspensió i amplificació de les llibreries

Recuperació de la llibreria enriquida

Vam introduir 30 µL de tampó E a les boletes rentades i les vam resuspendre pipetejant.

Amplificació de la llibreria

Vam preparar la següent barreja descrita a la Taula 9 de PCR *Master Mix*.

Taula 9. Paràmetres de la barreja de PCR Master Mix. **Font:** Illumina

Component	Concentració final	µL / Reacció	
Aigua lliure de nucleases	-	5	-
2X KAPA HiFi HotStart Ready Mix	1 X	25	-
P5 library primer (at 10 µM)	500 nM	2,5	AATGATACGGCGACCACCGA
P7 library primer (at 10 µM)	500 nM	2,5	CAAGCAGAAGACGGCATAACGA
Llibreria enriquida	-	15	-
TOTAL		50	-

Un cop feta, vam introduir el tub al termociclador seguint el programa de la Taula 10.

Taula 10. Paràmetres del termociclador per la barreja Master Mix.

Font: Illumina

Pas	Temperatura	Temps	
1	98°C	2 minuts	
2	98°C	20 segons	x14 cicles
3	60°C	30 segons	
4	72°C	45 segons	
5	72°C	5 minuts	
6	8°C	∞	

Amb la qual cosa, ja tenim les nostres reaccions amplificades. Només vam haver de purificar per tal de treure les boletes tal i com hem fet anteriorment.

3.6. Anàlisi de les dades

A partir de les dades que vam obtenir dels repositoris d'Internet esmentats anteriorment, vam seguir el següent procediment:

1. Filtració i control de qualitat de seqüències crues: en aquesta fase, ens centrem en processar les seqüències crues que hem recopilat. Utilitzant eines com FastQC v0.12.1 (Andrews, S. 2010) i Trimmomatic v0.40 (Bolger et al., 2014), realitzem una anàlisi de control de qualitat per assegurar-nos que les seqüències estiguin lliures d'errors i adaptadors no desitjats. Això implica la identificació i la retallada de seqüències de baixa qualitat i adaptadors que podrien afectar la precisió de les nostres dades.
2. Muntatge de seqüències filtrades i aparellades: després de la fase de filtració i control de qualitat, procedim a muntar les seqüències que compleixen els nostres estàndards. Aquest procés ens permet obtenir seqüències més netes i completament aparellades, que seran utilitzades per a anàlisis posteriors.
3. Generació de matrius: amb les seqüències ja processades i aparellades, creem matrius de dades que contenen informació rellevant per a les nostres anàlisis filogenètiques. Aquesta etapa és crucial per a la construcció de models i la inferència de relacions filogenètiques entre les espècies o gens estudiats.
4. Creació d'arbres de gens: a partir de les matrius generades, es realitza un arbre filogenètic per a cada gen d'interès amb el programa Hybpiper (Johnson et al., 2016), és a dir, un total de 353 arbres de gens. Encara que ja tenim un resultat no el podem considerar del tot fiable, per tant, s'han de continuar amb les anàlisis.
5. Unió d'arbres amb ASTRAL-III (Zhang et al., 2018): finalment, per obtenir una representació coherent de les relacions filogenètiques, utilitzem l'eina ASTRAL-III (Zhang et al., 2018) per combinar els arbres filogenètics construïts a partir de les nostres dades. Aquesta etapa ens permet arribar a conclusions més robustes i precises sobre les relacions evolutives entre les espècies o gens objecte d'estudi.

Finalment, es va obtenir un arbre filogenètic a nivell de gènere de les asteràcies als PPCC (vegeu l'Annex C per a més detalls sobre el codi desenvolupat).

4. Resultats i discussió

4.1. Resultats relacionats als objectius d'aprenentatge

Durant la meua estada a l'Institut Botànic de Barcelona, vaig adquirir coneixements en tècniques de laboratori per a la investigació biològica, especialment en extracció d'ADN i construcció de genoteques. L'objectiu principal era completar la filogènia de les asteràcies a nivell de gènere als PPCC, mitjançant l'obtenció de 353 gens de còpia única. Això implicava consolidar els meus coneixements previs sobre extracció d'ADN i perfeccionar-los per a garantir la precisió en tots els passos del procés.

En la següent part vaig aprendre procediments sobre les següents tècniques de laboratori a partir de les extraccions realitzades:

- Purificació d'extraccions d'ADN amb columna.
- Electroforesi.
- Sonicació.
- Construcció de genoteques a partir del *kit* de preparació de llibreries d'ADN NEBNext Ultra II de Illumina®.
- Hibridacions amb el *kit Angiosperms353 (Biosciences, 2023)*.

Per altra banda va ser tot un repte per a mi la part d'anàlisis bioinformàtics, ja que vaig haver d'aprendre i entendre molts conceptes nous. Tot i que durant el meus estudis de grau vaig adquirir unes bases sobre programació, el que vaig aprendre durant aquest treball va anar molt més enllà dels meus coneixements previs.

Per començar, em van ensenyar a comprendre un llenguatge de programació totalment nou per a mi, *Python*. A partir d'aquí, vaig haver d'aprendre el procediment habitual pel tractament d'aquest tipus de dades.

A partir d'aquest llenguatge vaig aprendre les nocions bàsiques de com fer servir els següents programes/*softwares* per analitzar les dades:

- FastQC (Andrews, S. 2010)
- Trimmomatic (Bolger et al., 2014)
- HybPiper (Johnson et al., 2016)
- MAFFT (Katoh & Standley, 2013)
- trimAl (Capella-Gutiérrez et al., 2009)
- FastTree2 (Price et al., 2010)
- TreeShrink (Mai & Mirarab, 2018)
- IQTree (Nguyen et al., 2015)
- ASTRAL (Zhang et al., 2018)
- SLURM (Jette et al., 2003)
- Newick (Junier & Zdobnov, 2010)

4.2. Llista definitiva de gèneres

Després d'haver fet l'estudi de quants gèneres es poden trobar de la família *Asteraceae* als PPCC, vam obtenir un total de 164 gèneres els quals els podem veure a la Taula 11.

Taula 11. Gèneres de la família *Asteraceae* presents als PPCC.

Gèneres <i>Asteraceae</i>					
<i>Achillea</i>	<i>Catananche</i>	<i>Eriocephalus</i>	<i>Jurinea</i>	<i>Pallenis</i>	<i>Solidago</i>
<i>Adenostyles</i>	<i>Centaurea</i>	<i>Eupatorium</i>	<i>Kentrophyllum</i>	<i>Pascalina</i>	<i>Soliva</i>
<i>Aetheorhiza</i>	<i>Chamaemelum</i>	<i>Evax</i>	<i>Lactuca</i>	<i>Petasites</i>	<i>Sonchus</i>
<i>Ambrosia</i>	<i>Chamomilla</i>	<i>Filaginella</i>	<i>Lagoseris</i>	<i>Phagnalon</i>	<i>Soyeria</i>
<i>Anacyclus</i>	<i>Cheirolophus</i>	<i>Filago</i>	<i>Lappa</i>	<i>Picnomon</i>	<i>Staehelina</i>
<i>Andryala</i>	<i>Chiliadenus</i>	<i>Flaveria</i>	<i>Lapsana</i>	<i>Picridium</i>	<i>Symphyotrichum</i>
<i>Antennaria</i>	<i>Chiliadenus</i>	<i>Gaillardia</i>	<i>Lasiopogon</i>	<i>Picris</i>	<i>Tagetes</i>
<i>Anthemis</i>	<i>Chondrilla</i>	<i>Galactites</i>	<i>Launaea</i>	<i>Pilosella</i>	<i>Tanacetum</i>
<i>Arctium</i>	<i>Chrysanthemoides</i>	<i>Galatella</i>	<i>Leontodon</i>	<i>Podospermum</i>	<i>Taraxacum</i>
<i>Arctotheca</i>	<i>Chrysanthemum</i>	<i>Galinsoga</i>	<i>Leontopodium</i>	<i>Prenanthes</i>	<i>Tephrosieris</i>
<i>Arnica</i>	<i>Cicerbita</i>	<i>Gamochaeta</i>	<i>Leucanthemopsis</i>	<i>Pseudognaphalium</i>	<i>Thrinicia</i>
<i>Arnoseris</i>	<i>Cichorium</i>	<i>Gazania</i>	<i>Leucanthemum</i>	<i>Pterotheca</i>	<i>Tolpis</i>
<i>Artemisia</i>	<i>Cirsium</i>	<i>Geropogon</i>	<i>Leuzea</i>	<i>Pulicaria</i>	<i>Tragopogon</i>
<i>Aster</i>	<i>Cnicus</i>	<i>Gnaphalium</i>	<i>Ligularia</i>	<i>Reichardia</i>	<i>Tripleurospermum</i>
<i>Asteriscus</i>	<i>Coleostephus</i>	<i>Guizotia</i>	<i>Limbarda</i>	<i>Rhagadiolus</i>	<i>Tussilago</i>
<i>Atractylis</i>	<i>Conyza</i>	<i>Gymnostyles</i>	<i>Logfia</i>	<i>Rhaponticum</i>	<i>Tyrimnus</i>
<i>Balsamita</i>	<i>Cosmos</i>	<i>Hedypnois</i>	<i>Mantiscalca</i>	<i>Rudbeckia</i>	<i>Urospermum</i>
<i>Bellis</i>	<i>Cota</i>	<i>Helianthus</i>	<i>Matricaria</i>	<i>Santolina</i>	<i>Verbesina</i>
<i>Bellium</i>	<i>Cotula</i>	<i>Helichrysum</i>	<i>Melitella</i>	<i>Saussurea</i>	<i>Volutaria</i>
<i>Bidens</i>	<i>Crepis</i>	<i>Helminthia</i>	<i>Microlonchus</i>	<i>Schkuhria</i>	<i>Wedelia</i>
<i>Bombycilaena</i>	<i>Crupina</i>	<i>Helminthotheca</i>	<i>Micropus</i>	<i>Schlagintweitia</i>	<i>Willemetia</i>
<i>Calendula</i>	<i>Cynara</i>	<i>Hieracium</i>	<i>Mulgedium</i>	<i>Scolymus</i>	<i>Xanthium</i>
<i>Calycocorsus</i>	<i>Diotis</i>	<i>Homogyne</i>	<i>Mycelis</i>	<i>Scorzonera</i>	<i>Xeranthemum</i>
<i>Carduncellus</i>	<i>Dittrichia</i>	<i>Hyoseris</i>	<i>Notobasis</i>	<i>Scorzoneroides</i>	<i>Zacintha</i>
<i>Carduus</i>	<i>Doronicum</i>	<i>Hypochoeris</i>	<i>Omalotheca</i>	<i>Senecio</i>	
<i>Carlina</i>	<i>Echinops</i>	<i>Inula</i>	<i>Onopordum</i>	<i>Seriola</i>	
<i>Carpesium</i>	<i>Eclipta</i>	<i>Iva</i>	<i>Osteospermum</i>	<i>Serratula</i>	
<i>Carthamus</i>	<i>Erigeron</i>	<i>Jasonia</i>	<i>Otanthus</i>	<i>Silybum</i>	

L'obtenció del nombre de gèneres d'asteràcies total als PPCC va ser el primer pas. Primer de tot es van buscar els gèneres que es trobaven als PPCC a partir de la Flora Manual dels Països Catalans (Bolòs et al., 2005). Addicionalment, es va fer un estudi sobre la sinonímia dels gèneres. A l'haver-hi incongruències entre el criteri de la Flora manual dels Països Catalans i altres recursos com *Plants of the World Online* (*Plants of the World Online* | Kew Science, <https://powo.science.kew.org/>) es va consultar a experts en la família com la Dra. Oriane Hidalgo i en cas de dubte es van incloure tots

els possibles gèneres per a comprovar-ne la sinonímia. Els gèneres poden tenir sinònims que s'han utilitzat en el passat o en diferents contextos. A més, la nomenclatura botànica pot estar canviant degut a canvis basats en noves investigacions i descobriments (p.e., Rouhan & Gaudeul (2014)).

Per tant, realitzar aquest estudi i verificar quins gèneres estan oficialment acceptats és essencial per assegurar-se de que s'està treballant amb la terminologia i la classificació botànica més actualitzada i precisa. A més, les nostres anàlisis permeten saber si les possibles sinonímies són reals o no (vegeu l'apartat "Arbre filogenètic a nivell de gènere de les asteràcies dels PPCC" per a més detalls). Això ajuda a garantir la consistència i precisió en la identificació i l'anàlisi de les plantes estudiades.

4.3. Obtenció dels gens d'interès

Degut a la falta de temps, només es van poder incloure a l'anàlisi les dades obtingudes dels repositoris d'Internet. És a dir, l'arbre filogenètic final no inclourà les mostres obtingudes mitjançant les extraccions d'ADN al laboratori. Aquestes mostres van ser processades i es va aconseguir material de bona qualitat per a 19 gèneres addicionals que seran seqüenciats posteriorment. Per tant, finalment es va treballar amb un grup de 119 gèneres que representen un 72,5% del total de gèneres presents en l'àrea d'estudi.

Per a obtenir una representació visual de la qualitat de dades utilitzades es va utilitzar un mapa de calor (Figura 12) en el qual es veu quines mostres passaven els lliminars de qualitat i quines no. Es van haver de descartar dades de 21 gèneres per falta de recuperació de gens. Aquests resultats posen èmfasi en la importància de comprovar la qualitat de les dades publicades ja que la seva qualitat varia enormement.

Per altra banda, també vam obtenir un mapa de calor però en aquest cas dels gens paràlegs (Figura 13). Els gens paràlegs són gens d'una mateixa espècie que tenen seqüències d'ADN semblants i codifiquen proteïnes amb funcions semblants o relacionades. Són gens que tenen més còpies de les esperades, normalment per duplicacions, i que poden aportar un senyal filogenètic erroni (Tekle et al., 2008). Aquest gràfic va permetre saber quins gens s'havien de descartar per la presència de paràlegs.

En el nostre cas, es va decidir eliminar els gens que presentaven més de 2 còpies. Un total de 15 gens es van eliminar amb aquest criteri, quedant un 95,70% de gens que es van considerar fiables.

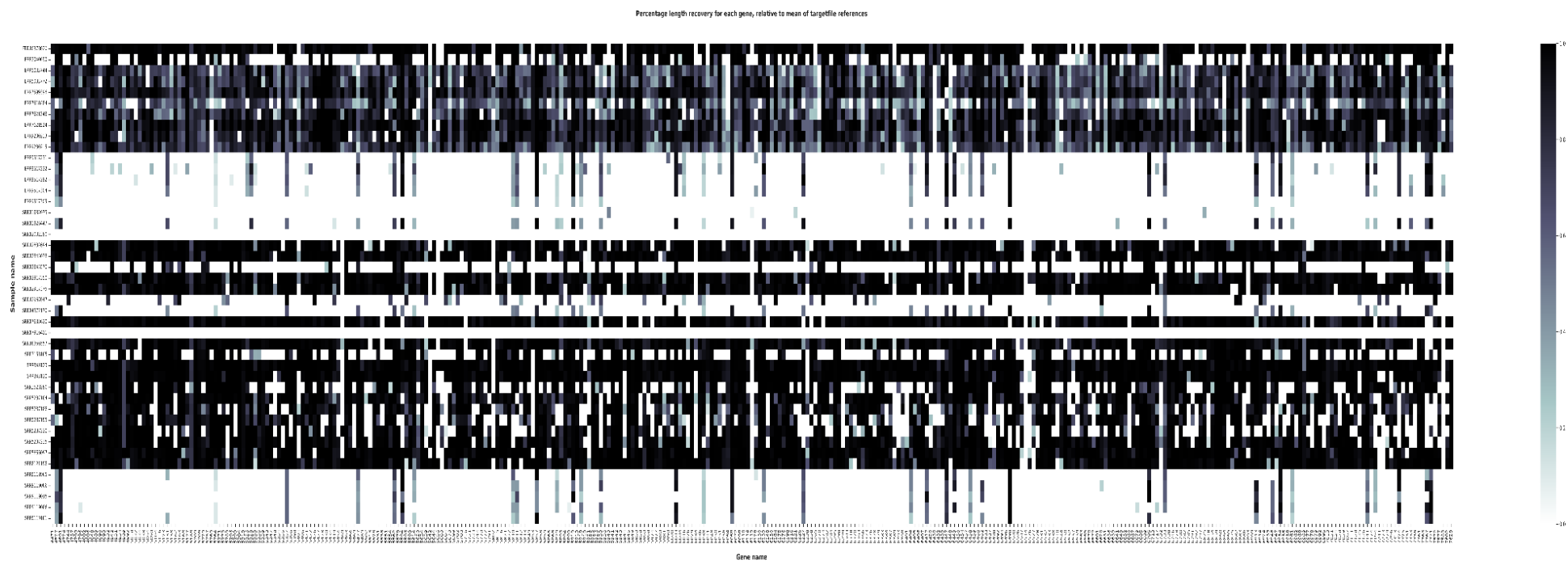


Figura 12. Mapa de calor dels 353 gens d'interès presents en els gèneres estudiats. L'escala d'intensitat indica el percentatge de l'expressió gènica de cada gen. Eix X: 353 gens d'interès; Eix Y: Mostres de gèneres d'asteràcies.

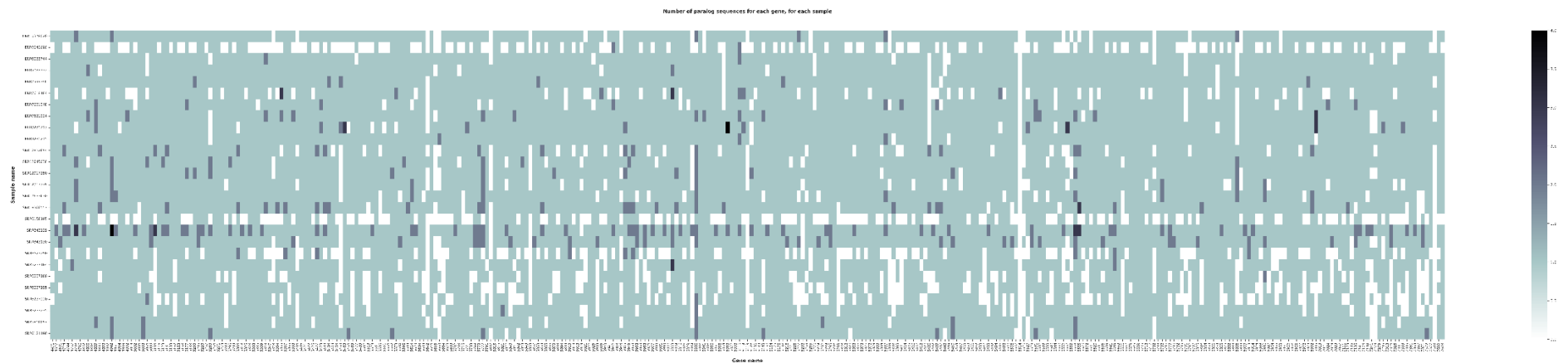


Figura 13. Mapa de calor que indica la quantitat de gens paràlegs presents en els 353 gens d'interès de les mostres estudiades. La barra d'intensitat indica el percentatge de l'expressió gènica dels gens paràlages. Eix X: 353 gens d'interès; Eix Y: Mostres de gèneres d'asteràcies.

4.4. Resultats de laboratori

Pel que fa a la part de laboratori, vam realitzar llibreries d'un total de 29 mostres. D'aquestes van sortir bé, és a dir, que tenien una concentració d'ADN major de 7ng/mL, 24 mostres. En la Taula 12 es pot observar un resum dels índexs que tenen cada mostra (i7XX, i5XX). Gràcies a aquests identificadors es podran diferenciar en un futur quin gènere és cada mostra a l'hora de seqüenciar-les.

Taula 12. Mostres preparades per a seqüenciar en un futur identificades amb els seus índexs.

	i723	i724	i715	i722
i509	CP0006	CP0002	CP0021	
i510	CP0009	CP0003		CP0010
i511	CP0012	CP0004		CP0024
i512	CP0013	CP0005		CP0025
i513	CP0014	CP0008		CP0026
i514	CP0015	CP0016		CP0007
i515	CP0022	CP0017		CP0011

4.5. Arbre filogenètic de les asteràcies a nivell de gènere dels PPCC

En primer lloc hem elaborat un arbre filogenètic per cada un dels 338 gens d'interès que buscàvem i que van passar el filtre de paralogia. Aquests arbres tenen una resolució menor en comparació a la de l'arbre final. En aquest cas, no és una bona eina per a resoldre les relacions evolutives dels gèneres. Per tant, per tal d'obtenir el resultat final, s'ha utilitzat el mètode de coalescència a partir del programa ASTRAL (Zhang et al., 2018) per a unificar-los i obtenir un arbre amb una major qualitat i resolució (Liu et al., 2019).

Es pot observar que l'arbre filogenètic obtingut (Figura 14) té una topologia bastant similar a, per exemple, l'arbre que podem trobar al *Tree Of Life Explorer* (*Species - Tree of Life Explorer*, <https://treeoflife.kew.org/>) o en alguns articles sobre *Asteraceae* (Mandel et al., 2019). Això dona

suport al nostre resultat. Es veuen tres grans clades diferenciats i amb bon suport que corresponen a les tres principals sub-famílies d'asteràcies tal i com es descriu a l'article de Mandel et al. (2019). Hem decidit utilitzar el mateix criteri de colors per tal de facilitar la comparació. Addicionalment, dins del clade *Asteroideae* trobem clades que corresponen a les tribus que també es corresponen als trobats a la literatura.

Primer de tot s'han afegit un parell de *outgroups*, *Mutisia spinosa* i *Cyclolepis genistoides*, els quals s'utilitzen com a referència per a determinar les relacions evolutives entre els gèneres. Aquests *outgroups* no estan relacionats amb el grup en el qual està centrat el nostre arbre filogenètic però estan relacionats de manera més distant en termes evolutius. La inclusió d'aquests gèneres permet l'arrelament de l'arbre, ja que proporcionen una referència externa per establir la direcció de les branques i determinar l'ancestre comú. Per tant, a l'afegir aquestes espècies s'està construint un marc de referència més sòlid.

A continuació, ens podem fixar en que hi ha diferents mostres dels mateix gènere, com per exemple al clade *Asteroideae* tenim dues mostres del gènere *Pallenis*. Això és degut a aquelles mostres que tenien incongruències de sinonímia i que volíem comprovar. El fet de que en el nostre arbre hagin sortit com a grups germans en la majoria dels casos ens indica que en molts casos la sinonímia entre els dos gèneres és real i, per tant, caldria actualitzar el criteri de la Flora Manual dels Països Catalans (Bolòs et al., 2005) ja que és en aquesta obra que la sinonímia no ha estat considerada. Un altre exemple podria ser el gènere *Centaurea* del clade *Carduoideae*. Del clade *Cichorioideae* també tenim un parell de mostres juntes del gènere *Cicerbita*. En contrast, trobem dos casos en els que les mostres identificades com el mateix gènere no són germanes. És el cas dels gèneres *Achillea* i *Filago*. En primer lloc veiem que les dues espècies d'*Achillea* es troben distants indicant que possiblement no són part del mateix gènere. Aquest resultat concorda amb els últims estudis que identifiquen *Achillea maritima* com a *Otanthus maritimus* (David Criado, no publicat). El cas de *Filago* és diferent ja que els dos individus són propers i veiem que *Bombycilaena* es col·loca com a germana de *Filago pyramidata*. Segons POWO (*Plants of the World Online | Kew Science*, <https://powo.science.kew.org/>), *Bombycilaena erecta* és una espècie acceptada que té com a sinònim *Filago erecta*. Segons les nostres anàlisis i resultats aquesta sinonímia hauria de ser revisada i *Filago erecta* hauria de ser el nom acceptat.

A més, un altre factor que es pot observar per veure la fiabilitat del nostre arbre són els suports dels nodes. Tot i que la majoria dels nodes tenen un suport d'1, tenim casos on aquest valor no és tan elevat. En el cas de que el node tingui un suport major a 0,95, es pot considerar que la posició filogenètica dels clades està consolidada i és fiable (Minh et al., 2013). En canvi, nodes amb un suport inferior poden indicar que la posició filogenètica no és massa segura. Cal recordar que els valors dels suports dels nodes poden variar segons la metodologia emprada per a construir l'arbre filogenètic (Mandel et al., 2019). Dins la subfamília *Carduoideae* trobem dos suports baixos, dins les *Cichorioideae* no en trobem cap però dins les *Asteroideae* en trobem 11. Un exemple més concret és el baix suport de *Doronicum* que ens indicaria que la seva posició filogenètica és dubtosa. Segons Mandel et al. (2019) i altres estudis, aquest gènere l'hauríem de trobar dins de la tribu *Senecioneae* que correspon al clade que inclou *Senecio*. En futurs estudis caldrà revisar la qualitat de les dades i afegir altres membres del gènere per a assegurar la seva posició en la filogènia.

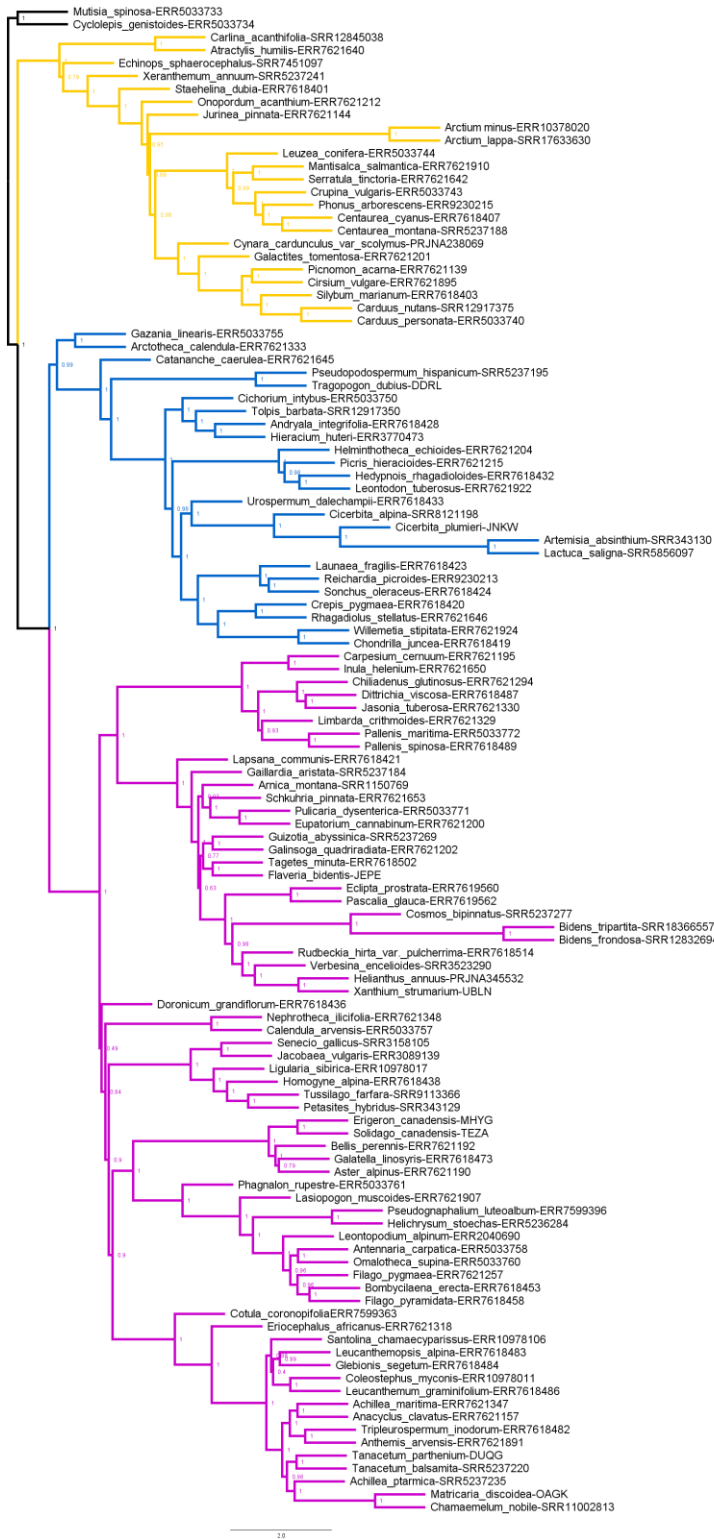


Figura 14. Arbre filogenètic a nivell de gènere de les asteràcies als PPCC. S'observa a cada node el valor del seu suport. Llegenda: i) clade blau: Cichorioideae; ii) clade rosa: Asteroideae; iii) clade groc: Carduoideae.

5. Conclusions

Primer de tot, s'han identificat un total de 164 gèneres de la família *Asteraceae* als PPCC. Degut a que no s'han inclòs al resultat final les mostres fresques o de l'herbari (un total de 19 mostres), sinó que només vam fer servir les dels repositoris d'Internet, queda pendent completar l'arbre filogenètic final amb els gèneres que falten. Finalment l'estudi s'ha fet amb un total de 119 gèneres que representen un 72,5% del total de gèneres presents en aquest projecte.

A continuació, a partir de la qualitat de les seqüències publicades vam considerar que contenien fiables un 95,70% dels gens.

Els resultats obtinguts en l'arbre filogenètic final són concordants amb altres estudis i fonts d'informació. Es poden diferenciar 3 clades: *Cichorioideae*, *Asteroideae* i *Gymnarrhenoideae*. A més, es va observar que algunes mostres d'un mateix gènere estaven agrupades estretament, amb la qual cosa, això proporciona un nivell de confiança en els resultats.

L'ús d'*outgroup* com *Mutisia spinosa* i *Cyclolepis genistoides* va ser essencial per a arrelar l'arbre filogenètic i proporcionar una referència externa per a establir les direccions de les branques i determinar l'ancestre comú. Això va ajudar a construir un marc de referència més sòlid per a la nostra recerca.

La representació visual de l'arbre filogenètic i l'anàlisi dels suports dels nodes ens va permetre avaluar la fiabilitat de les relacions filogenètiques. Tot i que la gran part de l'arbre el suport dels seus nodes és alt, alguns nodes s'han de revisar ja que el suport és massa baix. Seria de suma importància continuar investigant els casos on la resolució de l'arbre no és clara.

En resum, l'estudi proporciona una visió detallada de la construcció de l'arbre filogenètic a nivell de gènere de les asteràcies als PPCC, des de la identificació dels gèneres fins l'avaluació de la veracitat de l'arbre resultant passant per la identificació de possibles nous gèneres en els PPCC no considerants en la Flora Manual dels Països Catalans i l'eliminació d'altres que passarien a ser sinònims.

6. Bibliografia

Aizpuru, Carretero, & Devesa. (2004). *Compositae (Asteraceae)*.

<https://www.unavarra.es/herbario/htm/Compositae.htm>

Aranzadi. (s. f.). *El origen e importancia de los herbarios*. Aranzadi. Recuperado 16 de octubre de

2023, de <https://www.aranzadi.eus/origen-e-importancia-de-los-herbarios>

Benítez de Rojas, C. (2006). *BOTÁNICA SISTEMÁTICA*.

Biosciences, A. (2023). *Hybridization Capture for Targeted NGS*.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence

data. *Bioinformatics*, 30(15), 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>

Bolòs, O. D. (1985). *Corologia de la flora dels Països Catalans = Chorology of the flora of catalan countries: Volum introductori / Oriol de Bolòs: Chorology of the Flora of Catalan Countries: Information and Comments*. Institut d'Estudis Catalans.

Brewer, G. E., Clarkson, J. J., Maurin, O., Zuntini, A. R., Barber, V., Bellot, S., Biggs, N., Cowan, R. S., Davies, N. M. J., Dodsworth, S., Edwards, S. L., Eiserhardt, W. L., Epiawalage, N., Frisby, S., Grall, A., Kersey, P. J., Pokorny, L., Leitch, I. J., Forest, F., & Baker, W. J. (2019). Factors Affecting Targeted Sequencing of 353 Nuclear Genes From Herbarium Specimens Spanning the Diversity of Angiosperms. *Frontiers in Plant Science*, 10. <https://www.frontiersin.org/articles/10.3389/fpls.2019.01102>

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972-1973. <https://doi.org/10.1093/bioinformatics/btp348>

de Bolòs, O. (2003). *Atlas corològic de la flora vascular dels països catalans*. ORCA. https://books.google.es/books?id=i-cwE7_NCZQC

de Bolòs, O. (2005). *Flora manual dels Països Catalans*. Pòrtic. <https://books.google.es/books?id=ZTJDAAAACAAJ>

Devesa, J. A., Quintanar, A., & García, M. Á. (2015). *Flora iberica*. <https://bibdigital.rjb.csic.es/records/item/9895-flora-iberica?offset=1>

Els arbres filogenètics | enciclopedia.cat. (s. f.). Recuperado 16 de octubre de 2023, de <https://www.enciclopedia.cat/historia-natural-dels-paisos-catalans/els-arbres-filogenetics>

Escalante, A. E., Jardón Barbolla, L., Ramírez-Barahona, S., & Eguiarte, L. E. (2014). The study of biodiversity in the era of massive sequencing. *Revista Mexicana de Biodiversidad*, 85(4), 1249-1264. <https://doi.org/10.7550/rmb.43498>

Fairlie, I., & Pokorny, L. (s. f.). *96-well plate CTAB extraction*.

Folch, R. (1986). *La vegetació dels Països Catalans – Institució Catalana d'Història Natural*. <https://blogs.iec.cat/ichn/la-vegetacio-dels-paisos-catalans/>

GenBank Overview. (s. f.). Recuperado 12 de julio de 2023, de <https://www.ncbi.nlm.nih.gov/genbank/>

- Hale, H., Gardner, E. M., Viruel, J., Pokorny, L., & Johnson, M. G. (2020). Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in plants. *Applications in Plant Sciences*, 8(4), e11337. <https://doi.org/10.1002/aps3.11337>
- Jette, M., Yoo, A., & Grondona, M. (2003). SLURM: Simple linux utility for resource management. En *LECTURE NOTES IN COMPUTER SCIENCE*. https://doi.org/10.1007/10968987_3
- Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., Zerega, N. J. C., & Wickett, N. J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment1. *Applications in Plant Sciences*, 4(7), apps.1600016. <https://doi.org/10.3732/apps.1600016>
- Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., Eiserhardt, W. L., Epitawalage, N., Forest, F., Kim, J. T., Leebens-Mack, J. H., Leitch, I. J., Maurin, O., Soltis, D. E., Soltis, P. S., Wong, G. K., Baker, W. J., & Wickett, N. J. (2019). A Universal Probe Set for Targeted Sequencing of 353 Nuclear Genes from Any Flowering Plant Designed Using k-Medoids Clustering. *Systematic Biology*, 68(4), 594-606. <https://doi.org/10.1093/sysbio/syy086>
- Junier, T., & Zdobnov, E. M. (2010). The Newick utilities: High-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics*, 26(13), 1669-1670. <https://doi.org/10.1093/bioinformatics/btq243>
- Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30(4), 772-780. <https://doi.org/10.1093/molbev/mst010>

- Larridon, I., Villaverde, T., Zuntini, A. R., Pokorny, L., Brewer, G. E., Epiawalage, N., Fairlie, I., Hahn, M., Kim, J., Maguilla, E., Maurin, O., Xanthos, M., Hipp, A. L., Forest, F., & Baker, W. J. (2020). Tackling Rapid Radiations With Targeted Sequencing. *Frontiers in Plant Science*, *10*, 1655. <https://doi.org/10.3389/fpls.2019.01655>
- Liu, L., Anderson, C., Pearl, D., & Edwards, S. V. (2019). Modern Phylogenomics: Building Phylogenetic Trees Using the Multispecies Coalescent Model. En M. Anisimova (Ed.), *Evolutionary Genomics: Statistical and Computational Methods* (pp. 211-239). Springer. https://doi.org/10.1007/978-1-4939-9074-0_7
- López-Pujol, J., Bosch, M., Simon, J., & Blanché, C. (2006). *Diversitat isoenzimàtica de la flora vascular silvestre dels Països Catalans*.
- Mai, U., & Mirarab, S. (2018). TreeShrink: Fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics*, *19*(Suppl 5), 272. <https://doi.org/10.1186/s12864-018-4620-2>
- Mandel, J., Dikow, R., Siniscalchi, C., Thapa, R., Watson, L., & Funk, V. (2019). A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. *Proceedings of the National Academy of Sciences*, *116*, 201903871. <https://doi.org/10.1073/pnas.1903871116>
- Mandel, J. R., Dikow, R. B., Siniscalchi, C. M., Thapa, R., Watson, L. E., & Funk, V. A. (2019). A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. *Proceedings of the National Academy of Sciences*, *116*(28), 14083-14088. <https://doi.org/10.1073/pnas.1903871116>

- Minh, B. Q., Nguyen, M. A. T., & Von Haeseler, A. (2013). Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution*, 30(5), 1188-1195. <https://doi.org/10.1093/molbev/mst024>
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1), 268-274. <https://doi.org/10.1093/molbev/msu300>
- Palazzesi, L., Pellicer, J., Barreda, V. D., Loeuille, B., Mandel, J. R., Pokorný, L., Siniscalchi, C. M., Tellería, M. C., Leitch, I. J., & Hidalgo, O. (2022). Asteraceae as a model system for evolutionary studies: From fossils to genomes. *Botanical Journal of the Linnean Society*, 200(2), 143-164. <https://doi.org/10.1093/botlinnean/boac032>
- Panero, J. L., Freire, S. E., Ariza Espinar, L., Crozier, B. S., Barboza, G. E., & Cantero, J. J. (2014). Resolution of deep nodes yields an improved backbone phylogeny and a new basal lineage to study early evolution of Asteraceae. *Molecular Phylogenetics and Evolution*, 80, 43-53. <https://doi.org/10.1016/j.ympev.2014.07.012>
- Plantas y Hongos*. (s. f.). Recuperado 4 de septiembre de 2023, de <https://www.plantasyhongos.es/glosario/monofiletico.htm>
- Plants of the World Online | Kew Science*. (s. f.). Plants of the World Online. Recuperado 10 de septiembre de 2023, de <https://powo.science.kew.org/>
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*, 5(3), e9490. <https://doi.org/10.1371/journal.pone.0009490>

- Rouhan, G., & Gaudeul, M. (2014). Plant Taxonomy: A Historical Perspective, Current Challenges, and Perspectives. En P. Besse (Ed.), *Molecular Plant Taxonomy: Methods and Protocols* (pp. 1-37). Humana Press. https://doi.org/10.1007/978-1-62703-767-9_1
- Sáenz De Miera Y Carnicer, L. (2018). Las vicilinas, una lección sobre evolución. *Ambiociencias*, 44-55. <https://doi.org/10.18002/ambioc.v0i6.4910>
- Species—Tree of Life Explorer*. (s. f.). Recuperado 12 de julio de 2023, de <https://treeoflife.kew.org/specimen-viewer>
- Tekle, Y. I., Grant, J., Anderson, O. R., Nerad, T. A., Cole, J. C., Patterson, D. J., & Katz, L. A. (2008). Phylogenetic placement of diverse amoebae inferred from multigene analyses and assessment of clade stability within 'Amoebozoa' upon removal of varying rate classes of SSU-rDNA. *Molecular Phylogenetics and Evolution*, 47(1), 339-352. <https://doi.org/10.1016/j.ympev.2007.11.015>
- Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19(6), 153. <https://doi.org/10.1186/s12859-018-2129-y>
- Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19(6), 153. <https://doi.org/10.1186/s12859-018-2129-y>

7. Annexos

Annex A. Informació relativa als plecs d'herbari utilitzats

En la Taula A1 es troba informació sobre els plecs d'herbari utilitzats per dur a terme l'arbre filogenètic, com la localitat de la mostra, qui la va recol·lectar, etc.

Taula A1. Informació relativa als plecs d'herbari utilitzats.

Gènere	Recol·lector	Localitat	Any	Nº herbari	Codi mostra
<i>Notobasis syriaca</i>	T. Garnatje & R. Vilatersana	Campillos, Málaga	2000	BC 871211	CP0001
<i>Anthemis triumffeti</i>	J. Molero & S. Pyke	Vilanova de Prades, Osca	2014	BC 940323	CP0002
<i>Scorzonera laciniata</i>	T. Garnatje	Caldes de Montbui, Girona	2017	BC 971447	CP0003
<i>Omalothea sylvatica</i>	Ivan	Setcases, Girona			CP0004
<i>Prenanthes purpurea</i>	Ivan	Setcases, Girona			CP0005
<i>Mycelis muralis</i>	Josep Nuet Badia	La Vall d'Aran, Lleida	2019	BC 985046	CP0006
<i>Asteriscus aquaticus</i>	Abel Boldú	Les Borges Blanques, Lleida	1973	BC 631532	CP0007
<i>Gnaphalium uliginosum</i>	S. Arrabal & M. Galvany	Vosges, França	2008	BC 843320	CP0008
<i>Klasea flavescens</i>	P. M. Uribe-Echebarria	Agoncillo, Osca	2010	BC 874578	CP0009
<i>Scolymus hispanicus</i>	S. Pyke	L'Escala, Girona	2005	BC 907027	CP0010
<i>Cota altissima</i>	C. Benedí	Sant Sadurní de l'Heura, Girona	1983	BC 984540	CP0011
<i>Soliva stolonifera</i>	U. de Bolós	Sa Marina de Llucmajor, Balears	1986	BC 652447	CP0012
<i>Ambrosia psilostachya</i>	A. Rosell i J. Nuet Badia	Les Planes, Barcelona	1983	BC 656830	CP0013
<i>Scorzonera hirsuta</i>	M. Casanovas & M. C. Martinell	Montsec d'Ares, Lleida	2005	BC 872145	CP0014
<i>Chamaemelum mixtum</i>	S. Pyke	Port de La Selva, Girona	2002	BC 866136	CP0015
<i>Symphotrichum pilosum</i>	S. Pyke	Hospitalet de Llobregat, Barcelona	2019	BC 978975	CP0016
<i>Saussurea alpina</i>	Guardiola, M. & Petit, A.	Erill, Lleida	2006	BC 871819	CP0017
<i>Chrysanthemoides monilifera</i>	ipse	Le Napoule, França	1981	BC 645092	CP0018
<i>Cheirolephus intybaceus</i>	J. C. Cristóbal	Jijona, València	2021	BC 984398	CP0019
<i>Logfia gallica</i>	S. Pyke	Barcelona, Barcelona	2011	BC 915156	CP0020
<i>Pilosella officinarum</i>	D. Pérez Prieto	Vallcebre, Barcelona	2021	BC 984212	CP0021
<i>Leontodon autumnalis</i>	Moench	Alp, Girona	2006	BC 864114	CP0023
<i>Bellium bellidioides</i>	M. A. Cardona, J. Contandriopoulos i L. Llorens	Mallorca, Balears	1981	BC 653564	CP0024
<i>Cladanthus mixtus</i>	Fco. Górniz	Valdefresno, León	2006	BC 866431	CP0025
<i>Gamochaeta subfalcata</i>	S. Pyke	Barcelona, Barcelona	2001	BC 866584	CP0026

Annex B. Dades relatives a la part de laboratori

Taula B1. Quantitat de mostra utilitzada per dur a terme les extraccions d'ADN.

Espècie	Codi mostra	Pes (g)
<i>Notobasis syriaca</i>	CP0001	0,02
<i>Anthemis triumfetti</i>	CP0002	0,022
<i>Scorzonera laciniata</i>	CP0003	0,019
<i>Omalotheca sylvatica</i>	CP0004	0,027
<i>Prenanthes purpurea</i>	CP0005	0,023
<i>Mycelis muralis</i>	CP0006	0,008
<i>Asteriscus aquaticus</i>	CP0007	0,015
<i>Gnaphalium uliginosum</i>	CP0008	0,013
<i>Klasea flavescens</i>	CP0009	0,032
<i>Scolymus hispanicus</i>	CP0010	0,032
<i>Cota altissima</i>	CP0011	0,023
<i>Soliva stolonifera</i>	CP0012	0,027
<i>Ambrosia psilostachya</i>	CP0013	0,024
<i>Scorzonera hirsuta</i>	CP0014	0,006
<i>Chamaemelum mixtum</i>	CP0015	0,022
<i>Symphotrichum pilosum</i>	CP0016	0,019
<i>Saussurea alpina</i>	CP0017	0,014
<i>Chrysanthemoides monilifera</i>	CP0018	0,028
<i>Cheirolophus intybaceus</i>	CP0019	0,029
<i>Logfia gallica</i>	CP0020	0,043
<i>Pilosella officinarum</i>	CP0021	0,022
<i>Leontodon autumnalis</i>	CP0023	0,025
<i>Bellium bellidioides</i>	CP0024	0,012
<i>Cladanthus mixtus</i>	CP0025	0,03
<i>Gamochoaeta subfalcata</i>	CP0026	0,024

Per fer les extraccions d'ADN es van agafar les següents quantitats de cada plec que es mostren en la Taula B1:

Protocol per dur a terme extraccions d'ADN:

A. Trituració de cèl·lules vegetals i extracció d'ADN amb tampó CTAB+BME

En primer lloc, es van pesar entre 0,01 i 0,03 g de cada mostra i es van introduir en tubs Eppendorf de 2 mL amb 2 boles metàl·liques a cada tub.

Un pas molt important per dur a terme correctament tot el protocol és la identificació de les mostres. Es van anomenar les mostres de la següent forma: CPXXXX.

Els tubs amb les mostres i les boles metàl·liques es van congelar durant

aproximadament 1 hora a -20°C. Aquest pas facilita la trituració del material vegetal.

Seguidament, les mostres es trituraren en una màquina anomenada Geno/Grinder® durant un minut a 1500 cops/minut. Es pot repetir aquest procediment uns minuts més en el cas de que les mostres no quedin triturades totalment i, en el cas de que les mostres siguin molt dures, es poden triturar manualment. Es convenient fer una breu centrifugació als tubs per a què no quedin restes de material als taps.

Un cop es van tenir les mostres totalment triturades, es va afegir a cada tub 400 µL del tampó CTAB+BME amb una pipeta multicanal p1000. Es va barrejar pipetejant per tal de que quedés el més homogeni possible.

Per acabar, es van tancar tots els tubs i es van incubar al bany d'aigua durant 24 hores a 65°C. És necessari posar alguna cosa de pes a sobre per tal de que els tubs no s'obrin i s'evapori la solució.

B. Eliminació de proteïnes amb cloroform:IAA i precipitació de l'ADN

Un cop van passar les 24 hores d'incubació, mantenint el pes sobre els tubs, es van retirar les mostres del bany d'aigua i es van deixar refredar durant 15 minuts.

A continuació, a l'interior de la campana de gasos, es va abocar el volum necessari de (24:1) cloroform:isoamil alcohol (IAA) en un recipient de vidre que ens permetés pipetejar-lo. En el cas que no s'utilitzi immediatament s'ha de tancar perquè no s'evapori. Es van retirar els tancs dels tubs i hi es va afegir un volum de 400 µL de cloroform:IAA amb una pipeta multicanal p1000 als tubs que contenien l'extracte vegetal.

Després, es va barrejar les mostres invertint manualment els tubs de 25 a 50 vegades per a que s'emulsióni el contingut. I, per acabar, es van centrifugar les mostres durant 10 minuts a 3700 rpm. Una vegada centrifugades, es va transferir la part aquosa de cada tub a uns tubs nous amb una pipeta multicanal. S'ha d'evitar tocar o pipetejar la. Es van eliminar els tubs, les boles metàl·liques, l'extracte vegetal i els residus de cloroform:IAA seguint la normativa.

Per acabar, es va afegir un volum d'isopropanol a -20°C igual al de la fase aquosa que contenen els tubs. Un cop va ser afegit, es veu la precipitació d'un sediment (*pellet*) blanc. Es van tancar els tubs assegurant-nos que estan ben tancats i es van barrejar un altre cop invertint-los unes 10 vegades. En el cas de que siguin mostres d'herbari, han d'estar precipitant durant de tres a set nits a -20°C. Si és material fresc, només cal deixar-lo precipitar durant una nit.

C. Rentat, assecat i resuspensió de l'ADN

Es van retirar les mostres del congelador i es van centrifugar durant 20 minuts a 3700 rpm. Un cop acabada la centrifugació, s'havia de visualitzar clarament un sediment (*pellet*).

Es van treure els tancs i es van dipositar invertits sobre un tros de paper. Després, es van abocar l'isopropanol i, a continuació, es va treure l'última gota d'isopropanol que quedava al tub amb un paper net.

Es va preparar etanol al 70% en un recipient de plàstic en el qual s'havia de poder pipetejar després. Se'n varen afegir 500 µL a cada tub per a rentar els sediments (*pellets*). Es van tancar de manera segura i es va realitzar fer un vòrtex per tal de que els sediments (*pellets*) es desenganxessin i quedessin suspesos en l'etanol.

Seguidament, es van introduir les mostres a la centrifugadora durant 15 minuts a 3700 rpm. Un cop centrifugades, es van retirar l'etanol amb cura de que no caigués el sediment (*pellet*). Es va realitzar

un segon rentat amb etanol tal i com l'havíem fet anteriorment. Es va retirar l'etanol de les mostres i es van deixar assecar els sediments (*pellets*) a l'aire durant una nit o a l'estufa a 40-60°C durant una hora.

Per finalitzar, es van afegir 50-100 µL d'aigua ultrapura Milli-Q, es van tancar els tubs i es van incubar durant 30 min a 60°C a l'estufa.

Un cop fetes les extraccions d'ADN es va fer una electroforesi per tal de veure la grandària de les mostres. A la Figura B1 es pot veure el gel d'agarosa carregat.

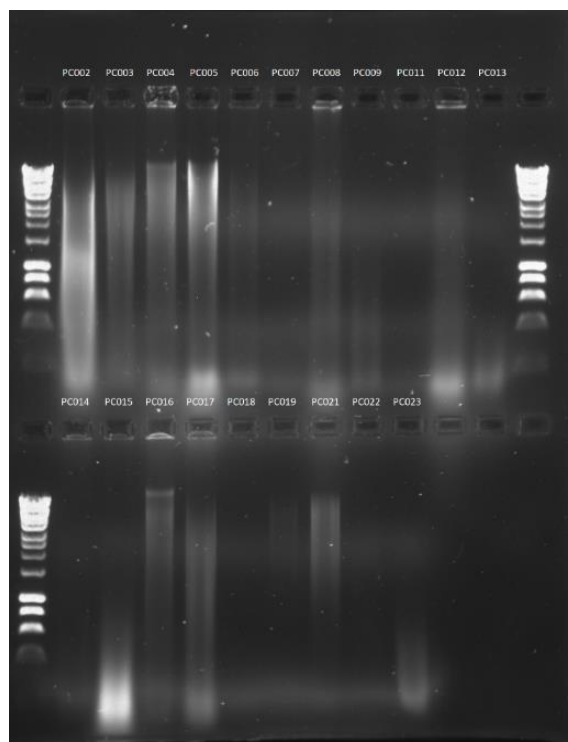


Figura B1. Electroforesi de les extraccions d'ADN.

En la Taula B3 es mostren els gèneres els quals van ser sotmesos a la sonicació per tal de tenir seqüències curtes d'ADN:

Espècie	Codi mostra
<i>Klasea flavescens</i>	CP0009
<i>Cota altíssima</i>	CP0011
<i>Ambrosia psilostachya</i>	CP0013
<i>Chrysanthemoides monilifera</i>	CP0018
<i>Cheirolophus intybaceus</i>	CP0019
<i>Pilosella officinarum</i>	CP0021
<i>Leontodon autumnalis</i>	CP0023

Taula B2. Mostres sotmeses a sonicació.

Posteriorment es va fer una altre electroforesi per tal de comprovar que tot havia anat correctament i les mostres tenien un pes molecular inferior, Figura B2.

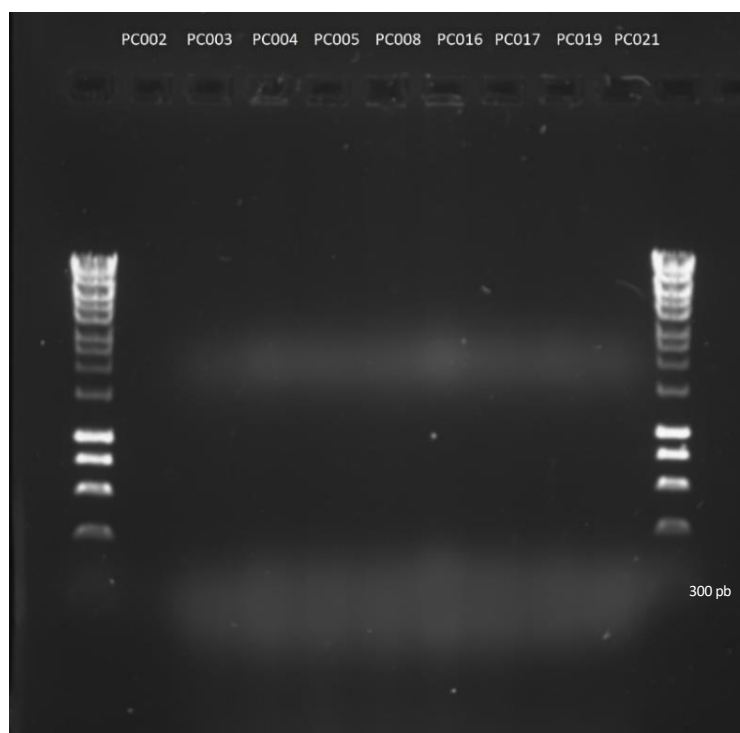


Figura B2. Electroforesi de les mostres un cop sonicades

Protocol per dur a terme la neteja amb columnes de les extraccions d'ADN:

1. Primer de tot es va preparar 5 volums de la solució tampó PB (500 μ L per a 100 μ L d'ADN)
2. A continuació, es van carregar les mostres a les columnes i, aquestes, es van col·locar en tubs sense tapa de 2 mL. Es van centrifugar durant 1-2 minuts a 13.000 rpm.
3. Seguidament es van rentar afegint 700 μ L per cada mostra de la solució tampó PE i es va centrifugar un altre cop durant 1-2 minuts a 13.000 rpm.
4. Es van buidar els tubs col·lectors i es va tornar a afegir 700 μ L de la solució tampó PE. En aquest cas es va tornar a centrifugar durant 1-2 minuts però a 13.300 rpm.
5. Es va tornar a centrifugar durant 1 minut per tal de descartar que quedés PE o etanol.
6. Després, es va traspasar el que hi havia a la columna a tubs d'1,5 mL, es va afegir 50 μ L d'aigua a cada tub i es deixaren reposar tres minuts.
7. Per acabar, es van centrifugar les mostres durant 1 min a 13.000 rpm.

Annex C. Dades relatives a les anàlisis bioinformàtics

Amb la primera cerca de dades al *GenBank* (*GenBank Overview*, <https://www.ncbi.nlm.nih.gov/genbank/>) vam obtenir el primer llistat de mostres observat a la Taula C1 que s'havien d'analitzar per comprovar si eren de bona qualitat o no. A partir dels anàlisis bioinformàtics pertinents vam obtenir els següents estadístics:

Taula C1. Resultats obtinguts de les primeres mostres que vam cercar al GenBank. *NumReads*: número de seqüències del gen d'interès; *ReadsMapped*: número de seqüències mapejades exitosament; *PctOnTarget*: proporció de seqüències alineades; *GenesMapped*: gens mapejats; *GenesWithContigs*: gens mapejats amb contigs; *GenesWithSeqs*: gens seqüenciats completament; *GenesAt50pct*: gens amb un 50% de cobertura del gen d'interès.

Name	NumReads	ReadsMapped	PctOnTarget	GenesMapped	GenesWithContigs	GenesWithSeqs	GenesAt50pct
ERR10378020	51322460	1075204	21	352	352	327	325
ERR2040690	22177772	102134	5	344	194	189	176
ERR5033744	1106038	263693	238	352	339	336	281
ERR5033772	988338	273489	277	352	341	339	285
ERR7599396	12681356	3653901	288	352	352	346	323
ERR7618484	1499422	331835	221	350	332	291	207
ERR7621348	6149350	1487284	242	351	350	323	291
ERR7621924	2180334	646930	297	351	344	342	326
ERR9230213	2268798	1053629	464	351	344	331	312
ERR9230215	3883620	1579586	407	353	350	343	296
ERR9517231	3632608	289791	80	352	301	46	32
ERR9517232	19764104	1141545	58	353	344	66	35
ERR9517292	5060762	233089	46	351	306	46	30
ERR9517304	4453616	372498	84	352	323	46	27
ERR9517305	88392	4109	46	228	34	15	2
SRR11832075	2155304	145237	67	347	278	5	0
SRR11926447	15822562	331832	21	352	134	44	26
SRR12002390	1419986	71362	50	353	74	0	0
SRR12832694	22058456	446188	20	353	344	314	305
SRR12845038	22831972	1062321	47	352	352	331	328
SRR12845570	13952234	62042	4	352	104	71	67
SRR12917350	22454234	1443722	64	353	353	334	326
SRR12917375	29108960	745989	26	353	341	321	313
SRR12967047	3110880	187615	60	349	316	49	39
SRR14572170	1821608	54885	30	353	212	40	26
SRR17633630	64829750	2273533	35	352	352	328	324
SRR18366557	41132090	1785039	43	353	352	336	326
SRR3158105	14774226	149375	10	352	235	184	175
SRR343129	56343010	950756	17	352	352	343	343
SRR343130	52880072	612948	12	352	351	333	330
SRR3523290	17587734	50491	3	352	236	229	221
SRR5237184	40820201	142150	3	352	305	304	298
SRR5237188	30414531	97762	3	348	265	265	249
SRR5237195	37374881	90219	2	348	252	251	241
SRR5237220	23094047	84925	4	348	248	244	234
SRR5237235	52558260	185799	4	350	304	303	294
SRR5856097	37209310	2243389	60	352	352	329	324
SRR8121198	35402708	2315310	65	352	352	339	327
SRR9118965	4680714	118358	25	328	96	35	20
SRR9119006	4516496	163082	36	341	83	39	23
SRR9119015	499382	23003	46	350	135	33	19
SRR9119086	7529948	141708	19	327	81	37	21
SRR9119101	569194	50522	89	352	228	35	18

En la Taula C1 podem observar diversos paràmetres:

- *NumReads*: fa referència al número de seqüències que s'han seqüenciat d'un dels 353 gens d'interès.
- *ReadsMapped*: es refereix al número de seqüències que s'han *mapejat* exitosament a una regió específica del gen d'interès.
- *PctOnTarget*: indica la proporció de seqüències que s'han alineat amb el gen d'interès en relació amb el total de seqüències generades.
- *GenesMapped*: en aquesta columna es pot observar quants gens han sigut exitosament *mapejats*.
- *GenesWithContigs*: en aquest cas aquesta columna també indica quants gens han sigut exitosament *mapejats*, però en fragments d'ADN anomenats *contigs*, els quals són fragments d'ADN més complets del gens d'interès.
- *GenesWithSeqs*: és un altre manera de mesurar quants gens s'han seqüenciat de manera quasi completa degut a que es miren seqüències del gen més llargues.
- *GenesAt50pct*: en aquesta columna s'observa el número de gens que presenten un 50% o més de les seqüències del gens d'interès. És a dir, indica quants d'aquests gens estan presents al menys a la meitat de les seqüències de la mostra. Aquesta mesura es pot fer en diferents percentatges.

A partir de la columna de *GenesAt50pct* vam decidir que les mostres que presentessin menys de 100 gens les descartaríem per no arribar a conclusions errònies. Per tant, d'aquesta llista vam haver de descartar 16 mostres. Aquesta va ser una primera neteja de dades per posteriorment seguir buscant mostres que ens poguessin servir.

A partir de les dades que vam obtenir dels repositoris d'Internet esmentats anteriorment, vam seguir el següent procediment:

1. Filtració i control de qualitat de seqüències crues: *FastQC i Trimmomatic*

Un cop descarregades totes les dades que tenim de cada gènere, es va procedir a fer el següent:

- Dins de la carpeta on teníem les dades originals (brutes), es van descomprimir les seqüències:

```
gunzip -k PAFTOL_0*.fastq.gz
```

- Un cop descomprimides les seqüències, es van moure des de la carpeta de *raw* fins a la carpeta *trim_qc*.

```
mv *.fastq ../../work/trim_qc/.
```

- A partir del programa *Fastqc* es va fer un primer control de qualitat amb l'script *00_fastqc_pre_trim.sh*.

```
bash 00_fastqc_pre_trim.sh
```

- Seguidament, es va fer un filtrat amb *Trimmomatic* per eliminar adaptadors i seqüències de baixa qualitat amb l'script *00_trim_qc.sh*.

```
bash 00_trim_qc.sh
```

- Per acabar, es va realitzar un control de qualitat de les seqüències filtrades *Paired End* amb l'script *00_fastqc_post_trim.sh*.

```
bash 00_fastqc_post_trim.sh
```

Ensamblatge de seqüències filtrades i aparellades

- Es van copiar els *reads* aparellats de *trim_qc* a *hybpbip_nuc*.

```
cp trim_qc/*P.fastq ./hybpbip_nuc/
```

- Es van descarregar i descomprimir l'arxiu de dianes nuclears *mega353*

```
wget https://github.com/chrisjackson-pellicle/NewTargets/raw/master/mega353.fasta.zip
```

```
unzip mega353.fasta.zip
```

- Seguidament, es va crear una llista de noms sense duplicats:

```
basename -s _1P.fastq *_1P.fastq > names.txt
```

- Es va executar el programa *Hybpiper* per recuperar dianes nuclears amb l'*script* *01_hybpip_nuc.sh*.

```
bash ../../scripts/01_hybpip_nuc.sh
```

- Per últim, es va calcular els estadístics per comprovar que *Hybpiper* ha funcionat:
Càlcul de les longituds per les seqüències mapejades i ensamblades:

```
python /home/csic/dve/pfm/bin/HybPiper/get_seq_lengths.py mega353.fasta names.txt dna > nuc_seq_len.txt
```

Càlcul de les estadístiques resum de *Hybpiper*:

```
python /home/csic/dve/pfm/bin/HybPiper/hybpiper_stats.py nuc_seq_len.txt names.txt > nuc_stats.txt
```

Comprovació dels gens paràlegs trobats per *Hybpiper*:

```
while read name;
do echo $name
python /home/csic/dve/pfm/bin/HybPiper/paralog_investigator.py $name 2>> nuc_para.log
done < names.txt
```

Per últim, vam recuperar les seqüències:

```
python /home/csic/dve/pfm/bin/HybPiper/retrieve_sequences.py mega353.fasta . dna supercontig 2>> nuc_retrseqs.log
```

Generació de matrius

En aquesta part utilitzem els següents programes:

- MAFFT: per alinear les seqüències
- trimAL: per filtrar les seqüències
- FastTree2: per inferir fileres ràpides
- TreeShrink: per eliminar taxes fora de rang
- Nw_ed: per filtrar per última vegada

A continuació, explico els passos que vam realitzar per dur a terme aquest procediment:

- Primer de tot, es va moure els arxius FNA amb les seqüències *supcontig* des de la carpeta de *hybpip_nuc*

```
mv work/hybpip_nuc/*.FNA work/aln_nuc/.
```

- Es van posar en quarantena els gens amb tres o més paràlegs.

Primer es va crear una llista i una carpeta dels paràlegs:

```
nano para_3-n-up.txt  
  
mkdir para_3-n-up
```

A continuació, es va fer un bucle a partir de la comanda *while* per traslladar els gens que tenen paràlegs:

```
while read para;  
do mv $para".FNA" para_3-n-up/.;  
done < para_3-n-up.txt
```

- Seguidament es va posar en quarantena els gens amb un valor inferior a dos terços de la mitjana de la matriu.

Primer vam contar les seqüències:

```
for gene in *.FNA;  
do grep --with-filename --count ">" $gene;  
done > seqs_per_gene_no-para_3-n-up.txt
```

Es va crear una llista de gens amb un valor inferior a dos terços de la matriu i es va crear una carpeta respectivament:

```
nano less_two_thirds_median_seqs.txt  
  
mkdir less_two_thirds_median_seqs
```

Per últim, es va traslladar els gens de la llista a la carpeta utilitzant un bucle a partir de la comanda *while*:

```
while read gene;  
do mv $gene".FNA" less_two_thirds_median_seqs/.;  
done < less_two_thirds_median_seqs.txt
```

- Alinear les seqüències amb MAFFT:

Es va crear una carpeta pels resultats:

```
mkdir aln_nuc_325
```

Es va haver d'obtenir directoris d'arxius FNA:

```
find . -maxdepth 1 -name "*.FNA" > paths_to_fna_senecio.txt
```

Es va executar l'script *bash 02_mafft_nuc.sh*:

```
find . -maxdepth 1 -name "*.aln" paths_to_aln.txt
## paths to FNA files
aln_file=$(cat paths_to_aln.txt | awk -v line=$SLURM_ARRAY_TASK_ID '{if (NR == line) print $0}')
## get the subdir (000*/) in which each fasta was saved and file name
trim_file=$(basename $aln_file | sed 's/\.aln/\.trim.aln/g')
tree_file=$(basename $aln_file | sed 's/\.aln/\.trim.tree/g')
## trim aln files
trimal -in $aln_file -out ./trimAL_222/$trim_file -automated1

## fasttree
FastTree ./trimAL_222/$trim_file > ./fasttree_222/$tree_file
## THE END
```

Filtració amb TrimAL

- Es va crear els següents directoris:

```
mkdir trimAL_222
```

- Es va executar l'script *bash 06_TrimAl_nuc.sh*:

```
find . -maxdepth 1 -name "*.aln" paths_to_aln.txt
## paths to FNA files
aln_file=$(cat paths_to_aln.txt | awk -v line=$SLURM_ARRAY_TASK_ID '{if (NR == line) print $0}')
## get the subdir (000*/) in which each fasta was saved and file name
trim_file=$(basename $aln_file | sed 's/\.aln/\.trim.aln/g')
## trim aln files
trimal -in $aln_file -out ./trimAL_222/$trim_file -automated1
```

Creació d'arbres de gens a partir del programa IQTree

- Es va executar l'script *bash 07_IQTree_nuc.sh*:

```
...
## iqtree files
for i in *.aln;
do
iqtree -s $i -m MFP
done
...
```

Unió dels arbres a partir del programa ASTRAL

- Es va crear una carpeta de treball i es va concatenar els fitxers de l'arbre gènic reduït i retallat:

```
mkdir sptre_nuc

cd sptre_nuc

cat ../gtre_nuc/gtre_316_nuc/*.shrnk.trm.treefile > poppy.nuc.316.shrnk.trm.gene.trees
```

- Es va crear un fitxer especificant els llindars i es va col·lapsar les biparticions sota els llindars esmentats:

```
nano bs_thresholds.txt

while read bs;
do
    /PATH_TO_TOOL/bin/nw_ed poppy.nuc.316.shrnk.trm.gene.trees 'i & b' '$bs' 'o' > poppy.nuc.316.bs$bs.shrnk.trm.gene.trees
done < bs_thresholds.txt
```

- Es va enviar l'script de lot a SLURM:

```
sbatch 10-0_ASTRAL_sptr_bs_nuc_poppy.sh
```

Aquest script executa l'ASTRAL dins d'un bucle mentre iterem sobre tots els llindars de col·lapse BS (amb o sense la bandera -2 t):

```
while read bs;
do
    java -jar /PATH_TO_TOOL/bin/ASTRAL/astral.5.6.3.jar -i poppy.nuc.316.bs$bs.shrnk.trm.gene.trees -t 2 -o poppy.nuc.316.bs$bs.t2.shrnk.trm.sp.tre 2> poppy.nuc.316.bs$bs.t2.shrnk.trm.sp.log

    #java -jar /PATH_TO_TOOL/bin/ASTRAL/astral.5.6.3.jar -i poppy.nuc.316.bs$bs.shrnk.trm.gene.trees -o poppy.nuc.316.bs$bs.shrnk.trm.sp.tre 2> poppy.nuc.316.bs$bs.shrnk.trm.sp.log
done < bs_thresholds.txt
```

Finalment, es va obtenir l'arbre filogenètic a nivell de gènere de les asteràcies als PPCC.