

Títol del treball:

**Cerca i optimització de marcadors genètics per a la
diferenciació de les dues espècies de sonso**

Estudiant: Àlex Martínez Guillem

Grau en Biologia

Correu electrònic: alexmartinezuguillem@gmail.com

Tutor: Dr. Jordi Viñas de Puig

Empresa / institució: Universitat de Girona

Vistiplau tutor:

Nom del tutor: Dr. Jordi Viñas de Puig

Empresa / institució: Universitat de Girona

Correu(s) electrònic(s): jordi.vinas@udg.edu

Abstract

The fish known as sand eel includes two species known as *Gymnammodytes cicereus* and *Gymnammodytes semisquamatus*. Both are present in the Mediterranean Sea although it's practically impossible to differentiate them morphologically due to their high similarity, leading to an indiscriminate fishing of both. This is the reason why it's necessary to be able to differentiate them at molecular level with the aim of determine how many populations of each specie exist in order to avoid the loss of viability of any of them.

The Genetics Ichthyology Laboratory of the University of Girona (LIG-UdG) has designed a mitochondrial marker for this purpose, which has allowed to differentiate the two species and determine that there is only one population for each of them. Following this objective, this paper proposes confirming the molecular differentiation of the two species using nuclear markers.

To perform this, an exhaustive bibliographic search has been carried out in order to find nuclear markers that have been used in previous studies to establish the phylogeny of species genetically related to those of this study. Subsequently, the search of primers of the selected markers has been proceeded and the corresponding DNA extraction of individuals from both species, previously differentiated mitochondrially, has been performed. The optimal PCR conditions have been designed and the amplicons replication delimited by the pairs of selected primers has been performed, in addition, the quality of the PCR products has been verified by agarose gel electrophoresis. Next, the cleaning and purification of the resulting amplicons has been carried out with an *ExoSAP* protocol and has been followed by a *BigDye* sequencing protocol to sequence them for subsequent bioinformatic analyses.

With the results obtained it can be concluded, through the construction of phylogenetic trees using the neighbour joining method, that *Gymnammodytes cicereus* and *Gymnammodytes semisquamatus* are two different species within the same genus. At the same time, the distance between these two species has been calculated, along with an outgroup conformed by genetically close species, which has allowed to corroborate the final conclusion of this study.

Resum

El peix conegut com a sonso comprèn dues espècies conegudes com a *Gymnammodytes cicereus* i *Gymnammodytes semisquamatus*. Ambdues es troben presents al Mar Mediterrani tot i que és pràcticament impossible diferenciar-les morfològicament degut a la seva elevada similitud, donant lloc a una pesca indiscriminada de totes dues. Per aquest motiu sorgeix la necessitat de poder-les diferenciar a nivell molecular i així poder determinar quantes poblacions de cada espècie existeixen per evitar la pèrdua de viabilitat d'alguna d'elles.

El Laboratori d'Ictiologia Genètica de la Universitat de Girona (LIG-UdG) ha dissenyat un marcador mitocondrial amb aquesta finalitat, el qual ha permès diferenciar les dues espècies i determinar que només existeix una població per a cada una d'elles. Seguint amb aquest objectiu, aquest treball proposa confirmar la diferenciació molecular de les dues espècies utilitzant marcadors nuclears.

Per dur-ho terme s'ha fet una exhaustiva cerca bibliogràfica, amb l'objectiu de trobar marcadors nuclears que s'hagin fet servir en estudis anteriors per establir la filogènia d'espècies genèticament vinculades a les d'aquest estudi. Posteriorment s'ha procedit a la cerca de *primers* dels marcadors seleccionats i s'han realitzat les corresponents extraccions de DNA d'individus d'ambdues espècies, prèviament diferenciats mitocondrialment. S'han dissenyat les condicions òptimes de PCR i s'ha procedit a la replicació dels amplicons delimitats pels parells de *primers* seleccionats, a més, la qualitat dels productes de PCR s'ha comprovat mitjançant electroforesis en gel d'agarosa. A continuació, s'ha procedit a la neteja i purificació dels amplicons resultants amb un protocol d'*ExoSAP* i s'ha seguit amb un protocol de seqüenciació *BigDye* per tal de seqüenciar-los per a posteriors anàlisis bioinformàtiques.

A nivell de resultats s'ha pogut concloure, mitjançant la construcció d'arbres filogenètics pel mètode *neighbour joining*, que *Gymnammodytes cicereus* i *Gymnammodytes semisquamatus* són dues espècies diferents dins del mateix gènere. Paral·lelament també s'ha calculat la distància entre aquestes dues espècies, juntament amb un *outgroup* conformat per espècies genèticament properes, que ha permès corroborar la conclusió final d'aquest estudi.

Resumen

El pez conocido como saltón comprende dos especies conocidas como *Gymnammodytes cicereus* y *Gymnammodytes semisquamatus*. Ambas se encuentran presentes en el Mar Mediterráneo aunque es prácticamente imposible diferenciarlas morfológicamente debido a su elevada similitud, dando lugar a una pesca indiscriminada de ambas. De ahí surge la necesidad de poderlas diferenciar a nivel molecular y así poder determinar cuántas poblaciones de cada especie existen para evitar la pérdida de viabilidad de alguna de ellas.

El Laboratorio de Ictiología Genética de la Universidad de Girona (LIG-UdG) ha diseñado un marcador mitocondrial con esta finalidad, el cual ha permitido diferenciar las dos especies y determinar que sólo existe una población para cada una de ellas. Siguiendo con este objetivo, este trabajo propone confirmar la diferenciación molecular de las dos especies utilizando marcadores nucleares.

Para llevarlo a cabo se ha hecho una exhaustiva búsqueda bibliográfica, con el objetivo de encontrar marcadores nucleares que hayan sido utilizados en estudios anteriores para establecer la filogenia de especies genéticamente vinculadas a las de este estudio. Posteriormente se ha procedido a la búsqueda de *primers* de los marcadores seleccionados y se han realizado las correspondientes extracciones de ADN de individuos de ambas especies, previamente diferenciados mitocondrialmente. Se han diseñado las condiciones óptimas de PCR y se ha procedido a la replicación de los amplicones delimitados por los pares de *primers* seleccionados, además, la calidad de los productos de PCR se ha comprobado mediante electroforesis en gel de agarosa. A continuación, se ha procedido a la limpieza y purificación de los amplicones resultantes con un protocolo de *ExoSAP* y se ha seguido con un protocolo de secuenciación *BigDye* con el objetivo de secuenciarlos para posteriores análisis bioinformáticos.

A nivel de resultados se ha podido concluir, mediante la construcción de árboles filogenéticos por el método *neighbour joining*, que *Gymnammodytes cicereus* y *Gymnammodytes semisquamatus* son dos especies diferentes dentro del mismo género. Paralelamente también se ha calculado la distancia entre estas dos especies, junto con un *outgroup* conformado por especies genéticamente cercanas, que ha permitido corroborar la conclusión final de este estudio.

Índex

1. Introducció	1
2. Objectives	4
3. Metodologia	6
3.1. Cerca bibliogràfica de marcadors nuclears	6
3.2. Cerca bioinformàtica de <i>primers</i> i disseny de les condicions de PCR.....	6
3.3. Preparació i conservació dels <i>primers</i>	8
3.4. Extraccions de DNA	9
3.5. Amplificació de DNA mitjançant PCR i comprovació electroforètica.....	9
3.6. Seqüenciació, precipitació i lectura de seqüències.....	12
3.7. Tractament bioinformàtic de dades	14
3.8. Comentari ètic	17
4. Resultats i discussió	17
4.1. Selecció dels marcadors nuclears	18
4.2. Cerca bioinformàtica de <i>primers</i> i disseny de les condicions de PCR.....	19
4.3. Resultat de les PCR i comprovacions electroforètiques	21
4.4. Anàlisis bioinformàtiques.....	24
5. Conclusions	27
6. Bibliografia	28

1. Introducció

El peix conegut com a sonso comprèn dues espècies del gènere *Gymnammodytes*, de la família *Ammodytidae* i de l'ordre dels *Perciformes* (Froese et al., 2018): la primera d'elles coneguda amb el nom científic de *Gymnammodytes cicereus* (Rafinesque, 1810) (Figura 1A) i la segona amb el nom de *Gymnammodytes semisquamatus* (Jourdain, 1879) (Figura 1B). Se les coneix amb els noms comuns de sonso blau i sonso ros respectivament.

Ambdues espècies presenten unes característiques morfològiques molt similars dificultant-ne per tant la seva diferenciació a simple vista, especialment en individus en estadi juvenil, de la mateixa manera que succeeix amb altres espècies d'aquesta mateixa família (Mitchell et al., 1998). Es caracteritzen per ser peixos de talla petita que arriben a un màxim de 15 cm de longitud en la seva edat adulta. De la mateixa manera que la resta d'ammodítids es caracteritzen per presentar un cos allargat i cilíndric amb un cap punxegut i una coloració lleugerament blavosa. Mostren la particularitat de no presentar ni bufeta natatòria ni aletes pèlviques. Presenten una boca protractil amb una mandíbula inferior que predomina per sobre de la superior i es caracteritza per la seva absència de dents. Pel que fa a les aletes aquestes són curtes en el cas de les anals i pectorals però, en canvi, presenten una dorsal llarga, la qual es troba molt a prop d'una línia lateral molt ramificada (Froese et al., 2018; Mas et al., 2000).



Figura 1A: Exemplars de *Gymnammodytes cicereus* (sonso blau). Figura 1B: Exemplars de *Gymnammodytes semisquamatus* (sonso ros).

Ambdues espècies són bentòniques i viuen de forma gregària a zones amb poca profunditat i que continguin sorra gruixuda. Durant la nit es refugien enterrant-se en caus que construeixen a la sorra i es reactiven amb la sortida del Sol, motiu pel qual se'ls coneix com a peixos demersals. Formen bancs d'individus de la mateixa espècie

en zones costaneres i s'alimenten principalment de zooplàncton (Froese et al., 2018; The IUCN Red List, 2018).

Pel que fa a la distribució geogràfica s'observen certes diferències en les zones principals que habiten cadascuna de les dues espècies d'estudi. Per una banda, *G. cicereus* es troba principalment a l'Atlàntic oriental i al Mar Mediterrani: incloent el Mar Adriàtic, el Mar Egea i el Mar Negre, Senegal i Portugal. Tot i així, també s'ha pogut observar a Mauritània, Angola i al sud del Marroc (Froese et al., 2018). Per altra banda, *G. semisquamatus* es troba des de la costa meridional de Noruega i les illes Shetland fins a la península Ibèrica, incloent-hi les Illes Britàniques i el Mar del Nord, però no el Mar Bàltic (Froese et al., 2018). També es troba, en menor mesura, al Mar Mediterrani. Tenint en compte això és coherent observar que a Catalunya la major part de les captures corresponen a l'espècie *G. cicereus*, on representa el 98% del total de les captures que es realitzen (Lleonart et al., 2014).

La pesca comercial del sonso a la costa catalana es troba gestionada pel comitè de co-gestió del sonso, el qual es troba regulat per l'administració, pescadors, científics i conservacionistes i que té com a objectiu l'establiment d'un pla de gestió per a la pesqueria del sonso. Degut a això les zones de pesca autoritzades es troben reduïdes a cinc confraries pesqueres, concretament a Arenys de Mar, Blanes, Sant Feliu de Guíxols, Palamós i l'Estartit (Figura 2), que comprenen una línia de costa de 120 km. L'època en què s'autoritza la pesca del sonso és la situada entre els mesos de març i desembre, amb un període de veda fixat entre el 15 de desembre i el 15 de febrer aproximadament (Lleonart et al., 2014).

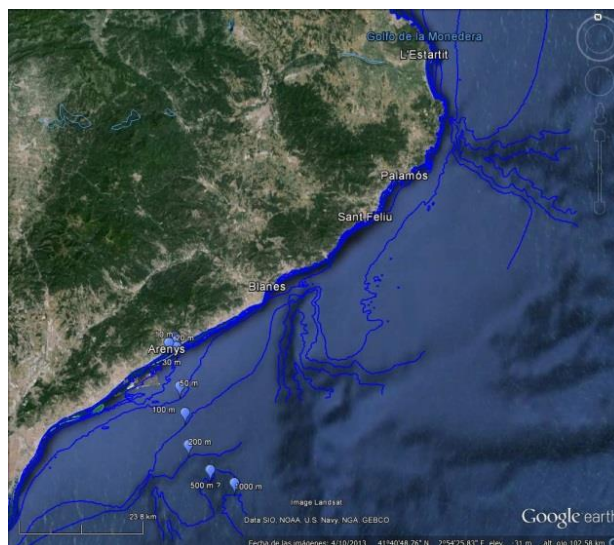


Figura 2: Localització geogràfica de les cinc confraries pesqueres on s'autoritza la pesca del sonso.

En base a tot això, es considera necessari estudiar un seguit de factors que poden comprometre la conservació de les dues espècies així com la correcta gestió dels recursos. Un d'aquests factors és el fet que ambdues espècies són pescades de manera simultània on, com ja s'ha dit anteriorment, l'espècie *G. cicereus* forma part de pràcticament la totalitat de les captures. Aquesta gran diferència pot ser un indicador que les poblacions de *G. semisquamatus*, l'espècie minoritària, són molt més reduïdes que les de *G. cicereus*, per tant, una pesca indiscriminada d'ambdues espècies pot comprometre la futura viabilitat i conservació de l'espècie menys abundant (Waples et al., 2008; Hauser et al., 2002).

Per altra banda, es té un coneixement molt reduït de l'estructura poblacional de cadascuna de les dues espècies, tot i així, la informació pesquera així com l'existència d'un període de reproducció simultani entre les dues espècies sembla indicar la presència d'una única població. Per tant, conèixer de manera acurada l'estructura d'aquesta hipotètica població permetria una millor gestió i conservació del recurs.

Per tal de dur-ho a terme és necessari esbrinar quantes poblacions existeixen per a cada una de les espècies, de manera que sorgeix la necessitat de poder diferenciar els individus en funció de a quina espècie pertanyen, la qual cosa permetrà, alhora, estudiar si hi ha un impacte diferencial entre les dues espècies provocat per l'activitat pesquera (Waples et al., 2008).

Considerant el que ja s'ha dit amb anterioritat, és molt complicat diferenciar aquestes espècies a nivell morfològic, per tant, l'ús de marcadors moleculars podria ser el mètode més adequat per dur-ho a terme (Punt, A., 2006; Waples et al., 2008). D'aquesta manera, prèviament a l'inici d'aquest treball, el Laboratori d'Ictiologia Genètica de la Universitat de Girona (LIG-UdG) ha utilitzat un marcador de DNA mitocondrial (mtDNA) per tal d'obtenir la primera evidència que indica que realment es tracta de dues espècies diferents (Dades no publicades) (Figura 3). Concretament, s'ha utilitzat un marcador basat en la regió control del DNA mitocondrial (mtDNA CR), el qual és àmpliament conegut en el camp de la biologia molecular per haver estat utilitzat en múltiples estudis d'aquesta índole (Azuma et al., 2013).

Per tal de confirmar que dues espècies són completament diferents són necessaris un mínim de dos marcadors moleculars no lligats que ho corroborin. En aquest cas se n'ha utilitzat un de mitocondrial que ha permès diferenciar-les i es procedirà a fer el mateix amb un o més marcadors nuclears (Awise et al., 1990).

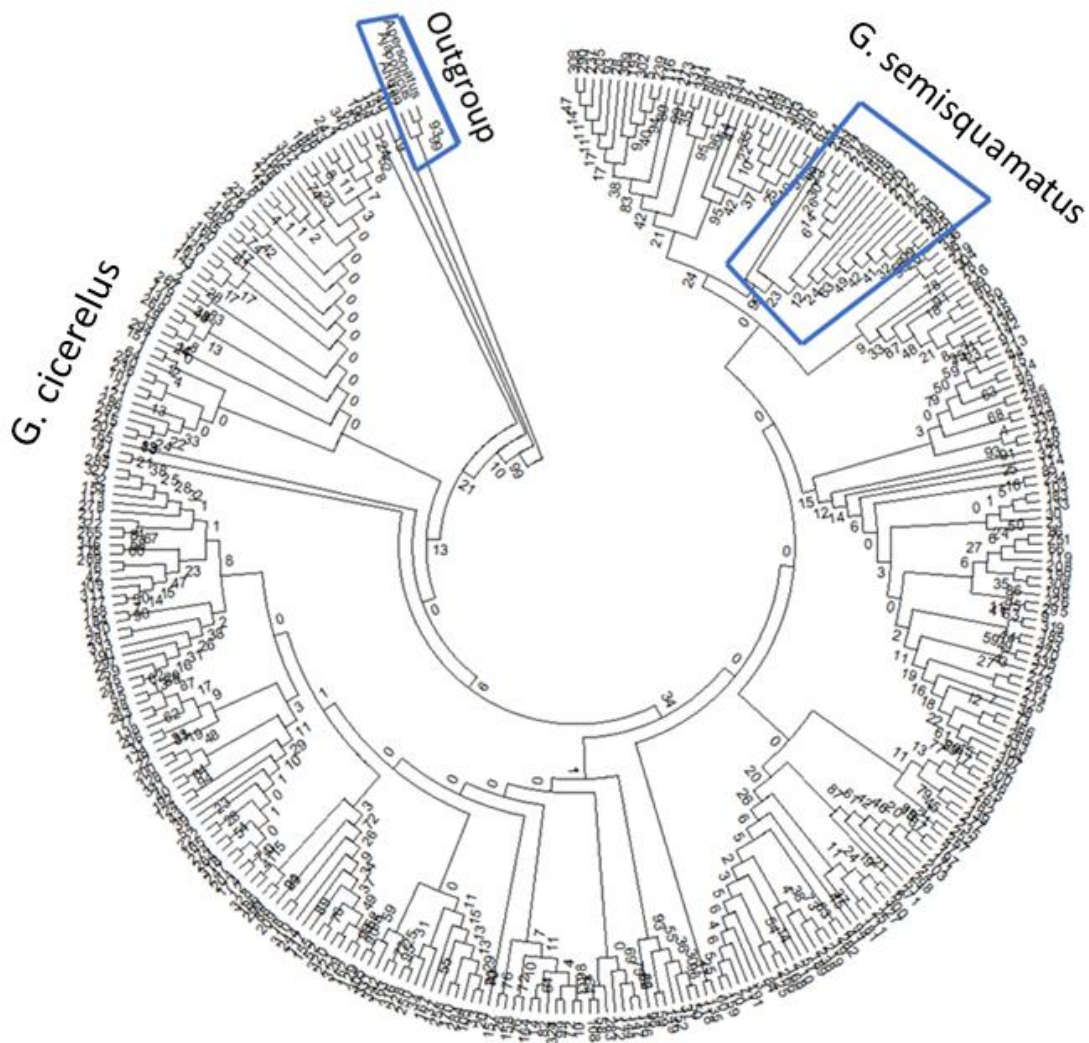


Figura 3: Arbre filogenètic obtingut a partir de l'estudi amb el marcador de la regió control del DNA mitocondrial. Es mostren els diferents haplotips mitocondrials detectats juntament amb l'espècie a la qual pertanyen. En requadres s'indiquen els haplotips que corresponen a l'*outgroup*, que són espècies que actuen com a referència, i els que corresponen a l'espècie *G. semisquamatus*. Tots els haplotips no inclosos dins dels requadres són els corresponents a l'espècie *G. cicerelus*.

2. Objectives

In the introduction it has been mentioned the difficult to differentiate the two species raised. From here emerges the project developed by the LIG-UdG in order to be able to differentiate them molecularly through the use of markers.

In order to verify if two individuals belong to a single species or to two different species is necessary to confirm it using at least two non-bound molecular markers. The LIG-UdG has done the test using one mitochondrial marker, while this work is based in the results obtained using nuclear markers, in other words, it's the second part of this project.

Basing on this, this work presents the following objectives which will be carried out during the course of the work in order to solve the questions that arise:

- Bibliographic search of nuclear markers that are viable for this study and allow to contrast the results obtained with the mitochondrial marker. Similar studies published previously will carry the information to know the best nuclear markers that can be used in this case with a high rate of success.
- Bioinformatic search of primers and design of PCR conditions for the new markers. It will be necessary to obtain the highest quality pair of primers of each selected nuclear marker in order to increase the probability of obtaining good results, as well as consider the size of the desired fragment (amplicon) and the conditions under which the different PCRs will be carried out.
- Verify if the selected primers work correctly, with individuals from the two mitochondrial differentiated species, for subsequent sequencing and bioinformatic data treatment. Despite searching for the primers with the greatest probability of functioning, and selecting the first ones of the best quality possible, there is no guarantee that ensure that they will work correctly for the two studied species. Therefore, it's necessary to check its efficiency by electrophoresis. Subsequently, it's necessary to sequence the samples in order to carry out a series of bioinformatic analyses that allow reaching the final conclusions of this work.

Considering all the antecedents related in the introduction and the objectives raised for the resolution of the questions that arise the main hypothesis of the work is proposed, which dictates that *Gymnamodytes cicereus* and *Gymnamodytes semisquamatus* are two different species differentiable with nuclear markers. Therefore, it will be necessary to comply with each one of the objectives to be able to confirm or reject this hypothesis.

3. Metodologia

En aquest apartat es procedirà a descriure les diferents metodologies i protocols que s'han seguit durant la realització d'aquest treball. Aquestes es troben descrites en diferents subapartats, els quals fan referència a cadascun dels objectius plantejats per tal de complir-los.

3.1. Cerca bibliogràfica de marcadors nuclears

El primer pas en la realització d'aquest treball ha consistit en dur a terme una cerca bibliogràfica exhaustiva de la literatura científica, per tal de descobrir si estudis previs ja han tractat amb anterioritat el tema principal d'aquest projecte. Per tal de fer-ho s'han utilitzat bases de dades que contenen informació seria, fiable i contrastada, com per exemple Google Scholar, PubMed o Genbank.

S'han fet servir diferents paraules clau per tal de poder trobar la informació desitjada, entre elles s'hi troben algunes com: *Gymnamodytes cicereus*, *Gymnamodytes semisquamatus*, *Gymnamodytes*, *Ammodytidae* i la combinació de totes les anteriors amb les paraules *genetics*, *taxonomy*, *phylogeny* i *nuclear markers*.

L'objectiu principal ha estat trobar estudis que hagin utilitzat marcadors nuclears en espècies genèticament vinculades a *G. cicereus* i a *G. semisquamatus*, ja que d'aquesta manera s'espera que les seqüències dels marcadors en cada una d'elles siguin bastant similars i permetin la cerca de *primers* compatibles.

3.2. Cerca bioinformàtica de primers i disseny de les condicions de PCR

Un *primer* és una seqüència curta d'àcid nucleic (d'aproximadament 20 parells de bases) que serveix com a punt de partida per a la replicació del DNA. La seva necessitat radica en el fet que la DNA polimerasa no pot sintetitzar una nova cadena de DNA si no hi ha una cadena preexistent. En condicions naturals aquesta cadena està constituïda per RNA que es genera per tal de dur a terme aquesta funció, per això, en condicions experimentals on es vol replicar una seqüència de DNA es requereix de *primers* que supleixin aquesta funció. S'utilitzen dos *primers* diferents, el *primer forward* i el *primer reverse* (Ye et al., 2012).

Per tal d'iniciar amb la cerca dels *primers*, el primer que s'ha dut a terme ha estat descarregar les seqüències corresponents als marcadors nuclears escollits de cada una de les espècies, genèticament vinculades a les d'aquest estudi, que s'hagin seleccionat. La majoria d'articles faciliten els *accession number* de cada marcador i

espècie, els quals serveixen com a codis que al introduir-los al buscador de NCBI permeten la redirecció cap a la pàgina de GenBank on s'hi troba tota la informació relacionada.

Aquí es pot descarregar la seqüència corresponent en diferents formats, tot i que en aquest cas s'ha descarregat en format FASTA degut a la seva compatibilitat amb multitud de programes que permeten l'edició de seqüències. Per tant, s'han obtingut les seqüències corresponents a cada marcador nuclear per a cada una de les espècies genèticament properes que s'han seleccionat i s'han classificat en funció del marcador nuclear al qual pertanyen.

Seguidament, s'ha utilitzat el programa BioEdit versió 7.0.5 (Hall, T.A., 1999) per tal de realitzar un alineament múltiple de les espècies properes a cada un dels marcadors separatament.

Per tal de fer això, s'ha executat el programa BioEdit i s'han carregat les seqüències de les espècies seleccionades d'un dels marcadors nuclears. Acte seguit s'ha procedit a realitzar un alineament múltiple mitjançant l'eina ClustalW Multiple alignment (Larkin et al., 2007). Gràcies a això, s'ha pogut obtenir la seqüència consens representativa de les seqüències de cada una de les espècies alineades. S'ha repetit el procés per a cada marcador nuclear, obtenint així tantes seqüències consens com marcadors s'hagin seleccionat.

L'objectiu d'això és que al realitzar els alineaments amb espècies genèticament vinculades a les que formen part d'aquest l'estudi és bastant probable que els marcadors mantinguin un nivell de conservació suficient com perquè els *primers* seleccionats a partir de les seqüències consens funcionin correctament.

Un cop s'han realitzat els alineaments per a cada marcador i s'han obtingut les diferents seqüències consens s'ha procedit al disseny de les condicions de PCR. Per tal de realitzar-ho s'han utilitzat tant el programa BioEdit com el programa Primer3 (Rozen et al., 2012).

El primer pas ha consistit en analitzar el resultat dels alineaments per assegurar-se que aquests siguin de qualitat, és a dir, que les diferents seqüències alineades presentin una longitud similar i no continguin gaps entremig de les seqüències. Seguidament, amb el programa BioEdit s'ha buscat una zona de l'inici de l'alineament amb un alt nivell de conservació entre les espècies utilitzades i s'ha fet el mateix amb una zona del final, així el programa Primer3 ha cercat els corresponents *primers forward* i *reverse*, respectivament, de millor qualitat inclosos en aquestes dues zones.

La seqüència compresa entre un i altre *primer* correspon al fragment de la seqüència que serà replicat en les futures PCR (amplicó).

Com ja s'ha dit, s'ha introduït la seqüència al programa Primer3 per tal que aquest pugui cercar el parell de *primers* de millor qualitat possible, en funció de certs paràmetres que presenta el programa, per tal de garantir les millors condicions de PCR. Les premisses que s'han tingut en compte són les següents (Stock, S., 2009; Yu-Huei, C., 2015):

- Longitud dels *primers* d'entre 18 i 24 nucleòtids.
- Mida de l'amplicó de 500 parells de bases aproximadament.
- Contingut de citosina (C) i guanina (G) d'entre un 40 i un 60%.
- *Melting temperature* d'entre 50 i 62 °C.
- Evitar *primers* que puguin dimeritzar.
- Evitar *primers* que siguin autocomplementaris.

Aquestes premisses s'han tingut en consideració per tal d'evitar problemes durant la realització de les corresponents PCR. La importància de la longitud del *primer* radica en el fet que si aquesta fos molt reduïda el *primer* seria molt inespecífic i podria hibridar a múltiples zones de la seqüència. Per altra banda, la *melting temperature* correspon a la temperatura a la qual un 50% de les molècules de DNA es troben encara formant una doble hèlix estable i l'altre 50% es separa en molècules d'un sol filament. A més, varia en funció de la composició de bases nitrogenades que conforma el *primer*, on un contingut excessiu o deficient en G i C donaria lloc a *melting temperatures* a les quals la DNA polimerasa no pot treballar eficientment. Finalment, també s'han considerat *primers* que no presentin regions complementàries entre ells per tal d'evitar que dimeritzin, així com *primers* que no presentin zones complementàries amb si mateixos ja que d'aquesta manera es replegarien formant un *loop*, donant lloc així a *primers* de baixa eficiència que no serien capaços d'hibridar correctament amb la seqüència per tal de replicar l'amplicó (Stock, S., 2009; Yu-Huei, C., 2015).

Considerant tots aquests factors s'ha seleccionat com a prioritària la primera opció donada pel programa Primer3. Tot i així també s'ha anotat la segona parella de *primers* de millor qualitat considerant l'existència de possibles errors amb la primera.

3.3. Preparació i conservació dels *primers*

Els *primers* són distribuïts dins de petits vials correctament etiquetats juntament amb un full que recull tota la informació corresponent a cada un d'ells, és a dir, informació

tal com el nom que se l'hi va assignar al *primer* al moment de demanar-lo, la seva longitud, la seva seqüència o la seva concentració en nanomols (nmols).

Després de comprovar que tot estigui en ordre, s'ha procedit a preparar els *primers* per a la seva conservació i posterior ús. Per fer això ha estat necessari preparar la que es coneix com a *working solution* o solució a la qual el *primer* presenta activitat. El mètode per dur-ho a terme es troba explicat detalladament al full d'informació que lliura la casa comercial juntament amb els *primers*.

3.4. Extraccions de DNA

Com ja s'ha mencionat a la introducció, aquest treball correspon a la segona part del projecte dut a terme pel LIG-UdG, per tant, s'han fet servir els mateixos individus que es van recol·lectar i classificar mitjançant l'ús del marcador mitocondrial prèviament mencionat. Aquests, van ser conservats individualitzats en alcohol al 96% per a futurs estudis genètics, indicant si formen part de *G. cicereus* (espècie 1) o *G. semisquamatus* (espècie 2).

L'objectiu principal és aconseguir extreure DNA de diferents individus de cada espècie per tal de comprovar si els marcadors nuclears que s'han seleccionat funcionen i, en última instància, verificar si la classificació en espècies que se n'obtingui correspon amb la determinada pel marcador mitocondrial. D'aquesta manera, es podria acabar determinant el nombre exacte de poblacions presents de cada espècie per així posar en pràctica nous criteris ètics i de sostenibilitat, ja que això permetria una millor gestió i conservació del recurs especialment per a l'espècie menys abundant.

L'extracció de DNA s'ha dut a terme pel mètode que fa ús de la columna *MicroSpin* i del conjunt de tampons conformat pel de lisi, lisi/unió, desinhibició i rentat. Addicionalment també es fa ús de la proteïnasa K i de l'isopropanol. Gràcies a l'actuació de tots aquests reactius, juntament amb un seguit de centrifugacions, s'ha pogut acabar obtenint DNA genòmic completament purificat sense presència de proteïnes o qualsevol altre component cel·lular.

3.5. Amplificació de DNA mitjançant PCR i comprovació electroforètica

La reacció en cadena de la polimerasa (PCR) és una tècnica utilitzada en el camp de la biologia molecular. Permet obtenir un elevat nombre de còpies d'un fragment concret de DNA, en aquest cas el que es troba delimitat pel parell de *primers* seleccionat per a un marcador nuclear concret.

L'objectiu de realitzar aquesta tècnica és, com ja s'ha dit, obtenir una gran quantitat de còpies del fragment de DNA desitjat per tal de poder-lo acabar seqüenciant i obtenir els resultats finals d'aquest estudi.

Per tal de dur a terme una PCR són necessaris diferents reactius així com les extraccions de DNA realitzades prèviament i els *primers* seleccionats. Els reactius utilitzats així com la quantitat que se n'ha fet servir es basen en els utilitzats a l'estudi amb el marcador mitocondrial (Taula 1).

La preparació de la PCR s'ha dut a terme sota campana per tal d'evitar la contaminació de les mostres amb DNA forà. Per aquest motiu s'han distribuït els diferents estris necessaris (pipetes, puntes de pipeta, tubs eppendorf i reactius) dins la campana i s'han irradiat amb llum ultraviolada, cosa que ha permès eliminar el DNA que poguessin tenir impregnat ja sigui d'origen humà o bacterià, entre d'altres. És important no irradiar amb llum ultraviolada ni les mostres de DNA purificat, ni la polimerasa, ni els *primers* ni els dNTPs, ja que la capacitat mutagènica d'aquesta els degradaria.

Taula 1: Llistat de reactius necessaris per dur terme les diferents PCR. Es mostra la concentració en què es troben tant en *stock* com en reacció, juntament amb el volum necessari de cada un suposant una PCR per a una única mostra. Els *primers forward* i *reverse* corresponen als del marcador nuclear que s'estigui testant.

Reactiu	Concentració <i>stock</i>	Concentració reacció	Volum
H ₂ O			15,9 µl
<i>Buffer</i>	10x	1x	2,5 µl
MgCl ₂	25 mM	1 mM	1 µl
<i>Primer forward</i>	10 µM	0,4 µM	1 µl
<i>Primer reverse</i>	10 µM	0,4 µM	1 µl
dNTPs	10 mM	200 µM	0,5 µl
Taq. Polimerasa		0,5u	0,1 µl
Total			22 µl <i>mix</i> (+ 3 µl mostra)

Els volums mostrats (Taula 1) són els de referència considerant que només es realitzés la PCR per a una mostra, però com és evident, cada marcador nuclear ha estat testat amb un número de mostres més elevat, normalment situat entre quatre i cinc mostres de cada espècie. Per això, el que s'ha realitzat cada vegada és un *mix* de PCR, és a dir, les quantitats enllistades anteriorment han estat multiplicades pel número de mostres a amplificar. Per exemple, suposant una PCR que contingui quatre individus de l'espècie 1, quatre de l'espècie 2 i un control negatiu (on no haurà d'amplificar res, sinó els resultats obtinguts en aquesta PCR no seran fiables perquè segurament s'hi haurà produït contaminació) es farà un *mix* que contingui nou vegades

els volums prèviament enllistats. També és correcte preparar un *mix* que contingui un volum més dels necessaris per si es produeix algun error en el moment de pipetejar.

Seguidament, s'han col·locat 22 µl del *mix* a tants tubs eppendorf com mostres a processar per posteriorment introduir-hi els 3 µl de l'extracció de DNA corresponent. Fet això s'han col·locat els diferents tubs dins del termociclador per començar la reacció en cadena de la polimerasa.

Per tal que la reacció es dugui a terme és necessari que el termociclador es trobi configurat en el programa corresponent. En aquest cas s'ha fet servir el mateix que es va utilitzar en les PCR realitzades amb el marcador mitocondrial (Figura 4).

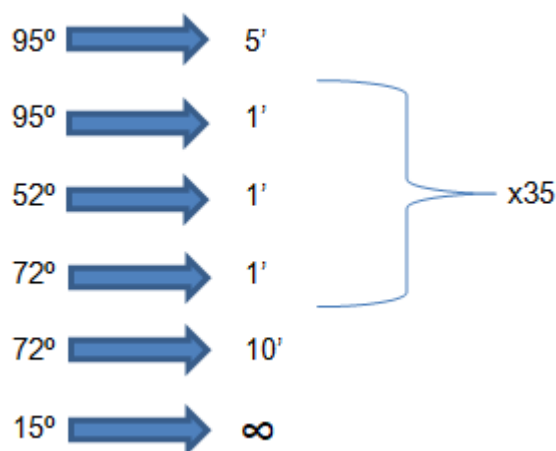


Figura 4: Programa configurat en el termociclador per dur a terme les PCR.

Com es pot observar, el programa del termociclador que duu a terme la reacció d'amplificació es basa en aprofitar la capacitat de la DNA polimerasa per replicar el fragment desitjat comprès entre el parell de *primers* utilitzat. Perquè això sigui possible el termociclador alterna cicles d'alta temperatura amb cicles de temperatura més baixa. El procés inicia amb un xoc tèrmic a 95 °C durant 5 minuts per tal de provocar la desnaturalització i corresponent separació de les dues cadenes que conformen el DNA. Es segueix amb un cicle que es repeteix 35 vegades, el qual alterna 1 minut a 95 °C que permet la separació de les dues cadenes de DNA, 1 minut a 52 °C que permet la unió del *primer* a la seva regió complementària del DNA i 1 minut a 72 °C on la DNA polimerasa replicarà el fragment de DNA comprès entre els dos *primers* utilitzats, usant-los com a punt de partida per a la síntesi del nou DNA mitjançant els dNTPs que s'han afegit prèviament. Finalment, un cop finalitzats els 35 cicles, es manté a 72 °C durant 10 minuts per assegurar una correcta amplificació dels nous fragments de DNA que s'han generat (Sambrook et al., 2001). Un cop la reacció ha

finalitzat el termociclador manté les mostres a una temperatura de 15 °C per a la seva conservació.

Tot aquest procés s'ha realitzat per a cada marcador nuclear seleccionat, fent servir entre quatre i cinc mostres de cada espècie i un control negatiu en cada un d'ells. Prèviament, però, s'ha volgut comprovar tant la qualitat de les extraccions de DNA realitzades com l'expertesa al moment de muntar una PCR. Amb aquesta finalitat s'han dut a terme diverses PCR compostes per quatre individus de cada espècie, un control negatiu i un control positiu (extracció de DNA d'un individu de l'espècie *Auxis rochei*) utilitzant el marcador mitocondrial de la regió control. Això ha permès comprovar la qualitat de les extraccions de DNA així com l'habilitat al moment de realitzar les diferents PCR, completament necessari abans de començar a experimentar amb els marcadors nuclears.

Posteriorment, s'ha procedit a la realització de les corresponents electroforesis per tal de comprovar si el parell de *primers* seleccionat per a cada marcador ha hibridat amb la seva regió complementària del DNA, és a dir, si s'ha aconseguit l'amplificació de l'amplicó per tal de poder procedir a la seqüenciació d'aquest.

Per dur-ho a terme s'han realitzat un seguit d'electroforesis en gel d'agarosa a l'1% tenyint amb *GelRed*, el qual és un agent intercalant que produeix certa fluorescència al ser il·luminat amb llum ultraviolada. D'aquesta manera s'ha pogut visualitzar el DNA en forma de bandes que al comparar-les amb un marcador de pesos moleculars, i recordant la mida que ha de tenir l'amplicó de cada marcador nuclear, ha estat possible determinar si les PCR han funcionat correctament.

Un cop s'ha aconseguit que el parell de *primers* d'un marcador nuclear concret funcioni per a ambdues espècies ha estat necessari netejar els productes de PCR mitjançant un protocol d'*ExoSAP*, per tal de procedir-ne a la seva seqüenciació.

L'*ExoSAP* rep el seu nom del conjunt de reactius que s'utilitzen en aquest procés, els quals són una exonucleasa I que elimina els *primers* sobrants que no han hibridat amb el DNA i una *Shrimp Alkaline Phosphatase (SAP)* que elimina els dNTPs sobrants. Per tant, la funció principal d'aquest procés és la de purificar els productes de PCR obtinguts prèviament, obtenint així els productes d'*ExoSAP*.

3.6. Seqüenciació, precipitació i lectura de seqüències

Amb l'objectiu de seqüenciar el DNA s'ha seguit un protocol de seqüenciació per *BigDye*, el qual es realitza a partir d'un kit comercial que permet la seqüenciació de *novo*, la reseqüenciació i l'acabat amb plantilles de productes de PCR o plasmidis.

La qualitat del DNA present en el producte de PCR pot variar a causa de diversos factors, entre els que s'inclouen la longitud del fragment de DNA amplificat, l'excés de *primers* i dNTPs i la presència de proteïnes i sals residuals. Això es tradueix en una seqüenciació de menor qualitat que presentarà molt més soroll de fons ocasionat per les reaccions de seqüenciació. Per tant, la utilització del *BigDye* ha permès una millor seqüenciació i una visualització dels resultats molt més clara i fiable.

Per dur a terme la seqüenciació de les mostres amb *BigDye* s'ha generat un *mix* amb els reactius necessaris. Els valors de referència suposant una única mostra han estat els següents:

- 4 µl d'aigua miliQ.
- 1 µl de *BigDye 3.1 Premix*.
- 2,5 µl de *Buffer BigDye*.
- 0,5 µl de *primer forward* (del marcador nuclear que s'estigui utilitzant).

Volum final: 8 µl.

El número de volums del *mix* ha variat en funció del número de mostres que s'estiguessin utilitzant, de la mateixa manera que amb el *mix* de PCR. Finalment s'han disposat, en nous tubs eppendorf, 8 µl de *mix* i 2 µl del producte d'*ExoSAP* corresponent.

Fet això s'han col·locat tots els tubs generats al termociclador per tal de posar en marxa el programa de seqüenciació (Figura 5).

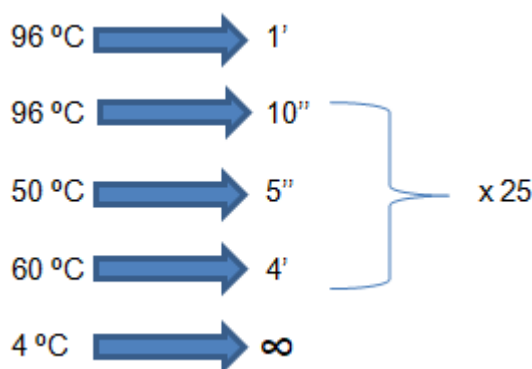


Figura 5: Programa configurat en el termociclador per dur a terme les reaccions de seqüenciació.

El protocol referit en aquest apartat s'ha realitzat per a la totalitat de productes de PCR obtinguts de cada marcador nuclear que prèviament ha donat bons resultats. Cal mencionar que si la seqüenciació fallés es repetiria el protocol de seqüenciació per *BigDye* utilitzant el *primer reverse* en comptes del *forward*.

Arribats en aquest punt ja es disposa de totes les mostres seqüenciades per a cada marcador nuclear que ha funcionat. El problema, però, radica en el fet que és necessari concentrar i dessalinitzar els productes obtinguts de les reaccions de seqüenciació per tal que el seqüenciador pugui llegir la seqüència de cada mostra correctament. Si no es realitzés aquesta precipitació el DNA no es trobaria en la concentració suficient com perquè el seqüenciador el pogués llegir, alhora que les diferents sals presents en la solució interferirien en els resultats obtinguts (Sambrook et al., 1989).

En aquest cas s'ha optat per una reacció de precipitació amb etanol en comptes d'isopropanol. Això és així, bàsicament, pel fet que el DNA és menys soluble en isopropanol fent que tendeixi a precipitar a una menor concentració juntament amb les sals presents en la solució. En canvi, el DNA és més soluble en etanol fent que sigui necessària una major quantitat d'aquest per tal que precipiti, amb l'avantatge que les sals seguiran en solució sense precipitar. Per tant, s'opta per la precipitació amb etanol en casos com el d'aquest estudi, on el volum de la mostra és reduït i amb una elevada concentració de DNA que es conservarà en fred (Sambrook et al., 1989).

Finalment, les mostres s'han introduït al seqüenciador (seqüenciador automàtic *applied ABI310*), el qual ha permès la lectura de la seqüència de DNA de cada mostra analitzada. Aquest aparell dona els resultats en forma d'una seqüència de bases nitrogenades amb pics d'una altura concreta que varien en funció de la intensitat del senyal detectat. També genera un arxiu que conté la seqüència per a futures anàlisis bioinformàtiques.

3.7. Tractament bioinformàtic de dades

Després de seqüenciar els diferents individus per a cada marcador nuclear utilitzat s'han obtingut un seguit d'arxius en format AB1 que poden ser oberts amb el programa BioEdit per tal de visualitzar les seqüències. Tot i així aquest programa no permet l'edició d'aquestes seqüències, motiu pel qual s'ha utilitzat el programa Chromas (Technelysium, 1996) per tal de complir amb aquest propòsit.

El programa Chromas permet visualitzar la seqüència confinada dins d'un arxiu AB1, representant-la en funció de la intensitat i presència de cada base dins la seqüència (Figura 6).

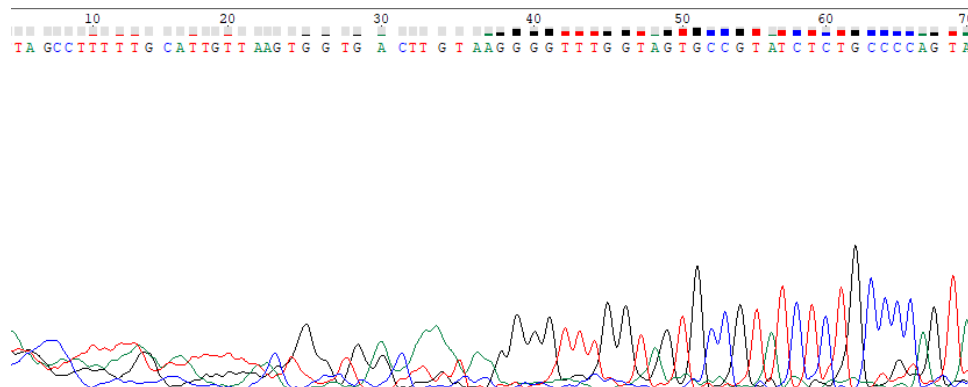


Figura 6: Exemple d'un cromatograma visualitzable mitjançant el programa Chromas. Es pot apreciar el soroll a l'inici de la seqüència, on no hi ha cap pic clarament definit.

L'objectiu d'utilitzar aquest programa ha estat el de suprimir el soroll present en les seqüències, és a dir, principalment els fragments que apareixen a banda i banda de les seqüències amb una qualitat bastant reduïda. Com ja s'ha dit, aquests fragments corresponen al soroll generat per la pròpia reacció de seqüenciació així com a petits fragments amplificats i seqüenciats que no corresponen al fragment delimitat pels primers utilitzats en el procés de PCR.

Fet això, s'ha procedit a realitzar un alineament d'aquestes seqüències amb aquelles que es van alinear per poder seleccionar els primers de cada marcador, és a dir, les seqüències corresponents a les espècies genèticament vinculades a *G. cicereus* i *G. semisquamatus*. Com és evident, s'ha realitzat un alineament independent per a cada marcador nuclear seqüenciat, mitjançant el programa BioEdit. Si pel motiu que sigui durant la seqüenciació amb *BigDye* s'ha utilitzat el primer reverse perquè el forward no ha funcionat caldrà obtenir el reverse complement de la seqüència, cosa que el mateix programa pot fer sense cap tipus de complicació.

Per tal de poder determinar si realment són dues espècies és necessari construir un arbre filogenètic mitjançant el programa MEGA (Tamura et al., 2013). Aquest programa no pot treballar amb els arxius en format FASTA que utilitza BioEdit, tot i així, el mateix MEGA disposa d'una eina que ha permès fer la conversió de format FASTA a format MEGA. Un cop s'han convertit els arxius al format adequat s'han carregat en el programa per tal de construir un arbre filogenètic diferent per a cada marcador nuclear.

A través de la interfície del programa MEGA s'ha procedit a la construcció dels arbres filogenètics pel mètode conegut com a *neighbor joining*, el qual agrupa les mostres mitjançant la generació informatitzada de matrius de distància entre cada parell de taxons que permeten anar-los unint entre ells. Aquest tipus d'arbre té en consideració les diferències presents en la velocitat evolutiva entre les diferents branques que el

formen. Normalment s'utilitza al sospitar de poca similitud entre les seqüències (Saitou et al., 1987).

Els paràmetres que s'han utilitzat per a la seva construcció es poden observar a la Figura 7. S'ha utilitzat el test filogenètic basat en el mètode del *bootstrap*, el qual determina el número de rèpliques que es realitzen, per tant, a major bootstrap major robustesa de la anàlisi. Per altra banda, s'ha utilitzat el model *Kimura de 2 paràmetres*, el qual es basa en considerar que les transicions són més freqüents que les transversions. Per tant, aquest model depèn de dos paràmetres: La probabilitat d'observar una transició i la probabilitat d'observar una transversió. En base a tot això, l'arbre s'ha construït basant-se en aquests dos paràmetres (Saitou et al., 1987). Alhora, s'ha configurat una deleció completa dels gaps o zones amb buits d'informació.

Option	Selection
Analysis	Phylogeny Reconstruction
Scope	All Selected Taxa
Statistical Method	Neighbor-joining
Phylogeny Test	
Test of Phylogeny	Bootstrap method
<i>No. of Bootstrap Replications</i>	500
Substitution Model	
Substitutions Type	Nucleotide
Model/Method	Kimura 2-parameter model
Substitutions to Include	d: Transitions + Transversions
Rates and Patterns	
Rates among Sites	Uniform rates
<i>Gamma Parameter</i>	<i>Not Applicable</i>
Pattern among Lineages	Same (Homogeneous)
Data Subset to Use	
Gaps/Missing Data Treatment	Complete deletion
<i>Site Coverage Cutoff (%)</i>	<i>Not Applicable</i>

Figura 7: Paràmetres utilitzats per a la realització de l'arbre filogenètic mitjançant el programa MEGA.

Finalment, amb el mateix programa MEGA s'ha dissenyat una matriu de distàncies per a cada marcador nuclear amb l'objectiu de calcular la distància present entre les dues espècies i les de referència (espècies genèticament vinculades a les d'aquest estudi). S'ha calculat la distància agrupant cada mostra en grups diferents en funció dels resultats obtinguts en l'arbre filogenètic i s'ha procedit a calcular la distància entre els diferents grups, utilitzant els mateixos paràmetres que s'han utilitzat en la construcció de l'arbre (Figura 7).

Per acabar, es mostra un esquema general de tot el protocol seguit per dur terme aquest estudi (Figura 8).

Esquema del Protocol

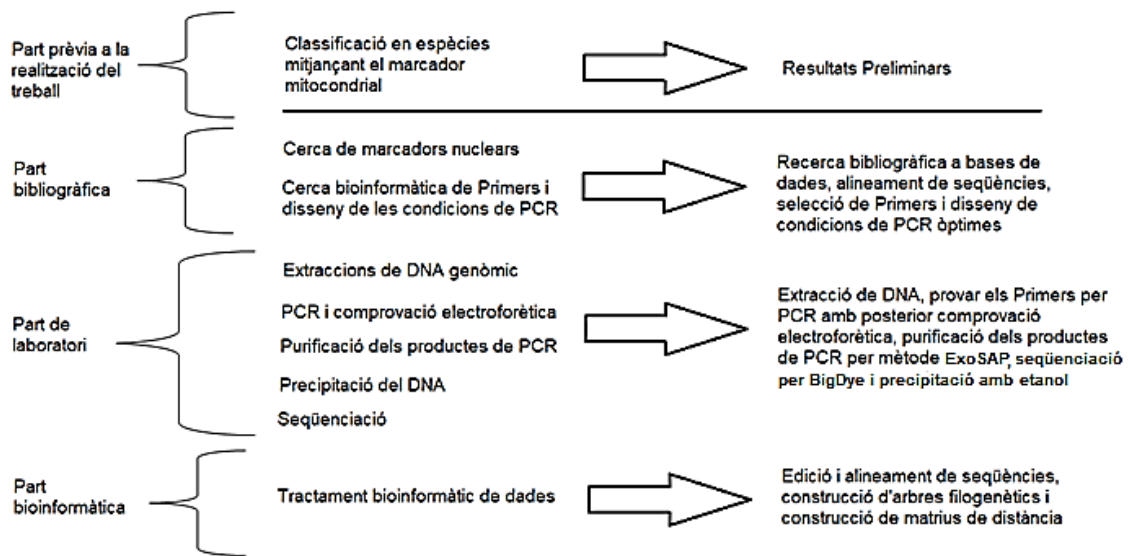


Figura 8: Esquema que actua com a resum de la totalitat del protocol que s'ha seguit per a l'obtenció dels resultats.

3.8. Comentari ètic

Durant la realització d'aquest estudi no s'han dut a terme pràctiques que involucrin l'experimentació amb animals, sinó que s'han adquirit exemplars de les dues espècies de sonso que ja havien estat capturades prèviament per a la posterior conservació. Es considera que el compliment d'aquest estudi permetrà determinar les poblacions existents de cadascuna de les dues espècies estudiades, per tal de posar en marxa nous criteris ètics per al manteniment i sostenibilitat dels recursos a llarg termini evitant la pèrdua de poblacions. Alhora, es creu que la metodologia desenvolupada en aquest estudi pot ser extrapolada a futures investigacions relacionades amb la genètica de poblacions que involucrin altres espècies.

4. Resultats i discussió

A partir de la metodologia que s'ha realitzat s'han anat obtenint diferents resultats, que corresponen principalment als marcadors nuclears que s'han escollit, el resultat de les PCR que s'han dut a terme amb les seves corresponents electroforesi i els resultats de les anàlisis bioinformàtiques. Conjuntament són els que permeten determinar quins dels marcadors nuclears han funcionat i quina classificació taxonòmica final se'ls pot assignar a aquestes dues espècies hipotèticament diferents.

4.1. Selecció dels marcadors nuclears

Ràpidament, s'ha pogut observar que no hi ha cap estudi previ que hagi intentat comparar les dues espècies problema mitjançant marcadors moleculars, és a dir, aquest estudi tracta un nou tema del qual no hi ha cap referència prèvia.

Degut a aquest fet, no s'han pogut trobar marcadors moleculars que hagin estat utilitzats conjuntament en les dues espècies. Per tant, s'ha procedit a cercar investigacions anteriors que hagin realitzat estudis genètics mitjançant marcadors moleculars amb espècies del mateix gènere (*Gymnammodytes*) o família (*Ammodytidae*). Això és així simplement pel fet que s'espera que compartint certa relació genètica el marcador molecular presenti el suficient nivell de conservació com per resultar eficient.

Després d'una àmplia recerca, cercant *Gymnammodytes semisquamatus* a l'apartat *Nucleotide* de NCBI s'han trobat una sèrie de seqüències que corresponen a marcadors nuclears utilitzats en estudis que involucren aquesta espècie. En aquestes, s'hi ha trobat referenciat l'article de Laoutredou et al., (2013), on pretenien estudiar les relacions genètiques entre els diferents *Serraniformes* (família de peixos inclosa dins de l'ordre dels *Perciformes*) utilitzant un conjunt de marcadors nuclears que ja s'havien utilitzat prèviament per tal de definir la filogènia dels *Teleostis*, que són una de les tres infraclases dels *Actinopterygii* (peixos ossis que inclouen les dues espècies d'estudi).

L'objectiu de l'estudi era ambiciós, és a dir, pretenien estudiar al complet les diferents classificacions taxonòmiques dins dels *Serraniformes* i com es relacionen entre elles. Per dur-ho a terme van utilitzar una gran quantitat de mostres i diversos marcadors nuclears.

Les mostres que van recol·lectar pertanyen a diferents espècies dins de la família dels *Serraniformes* així com a espècies incloses dins de l'ordre dels *Perciformes* però distribuïdes en famílies diferents. Entre aquestes últimes s'hi troben tres espècies que pertanyen a la família *Ammodytidae*: *Ammodytes marinus*, *Hyperoplus lanceolatus* i *Gymnammodytes semisquamatus*. D'aquesta manera, s'han seleccionat com a possibles candidats els marcadors nuclears que es van utilitzar en aquestes tres espècies. Considerant que una de les tres espècies és *Gymnammodytes semisquamatus* la possibilitat que aquests marcadors funcionessin era força elevada.

Seguint aquest criteri, els marcadors nuclears que s'han seleccionat són aquells que els investigadors d'aquest estudi van utilitzar conjuntament en les tres espècies, descartant aquells que no s'utilitzessin en alguna d'elles. Els marcadors nuclears

escollits es mostren a continuació juntament amb els *accession number* de *G. semisquamatus*, *H. lanceolatus* i *A. marinus* respectivament:

- IRBP (AN: JX628250, JX628251, JX628252): És una proteïna coneguda com a *interphotoreceptor retinoid-binding protein* i es troba codificada pel gen RBP3. Correspon a una llarga glicoproteïna que s'uneix als retinoides i es troba situada principalment a la matriu interfotoreceptora de la retina, entre l'epiteli del pigment de la retina i les cèl·lules fotoreceptores. Actualment es creu que la seva funció és la de transportar els retinoides entre aquestes dues parts, un procés crític per al bon funcionament visual (NCBI, 2018).
- RNF213 (AN: JX628500, JX628501, JX628502): És una proteïna coneguda com a *ring finger protein 213* i es troba codificada pel gen RNF213. És una proteïna formada per dits de zinc que intervé com a mediadora de les interaccions entre certes proteïnes per tal de degradar-les. Intervé en funcions com ara el creixement i multiplicació cel·lular, la transducció del senyal i la inducció a l'apoptosi. Per altra banda, presenta dominis amb funció ATPasa que intervenen en el desempaquetament de proteïnes i de DNA (NCBI, 2018).
- MLL2 (AN: JX628118, JX628120, JX628119): És una proteïna coneguda com a *mixed-lineage leukemia-like protein 2*. Actualment no s'ha descobert la funció exacta que desenvolupa aquesta proteïna.
- MLL4 (AN: X628250, JX628251, JX628252): És una proteïna coneguda com a *mixed-lineage leukemia-like protein 4*. Actualment no s'ha descobert la funció exacta que desenvolupa aquesta proteïna.
- Pkdt (AN: JX628380, JX628381, JX628382): És un gen anomenat *polycystic kidney disease* que codifica per a una glicoproteïna. Actualment no s'acaba de conèixer la funció que desenvolupa, tot i que se sap que es troba involucrada en les adhesions cèl·lula-cèl·lula i cèl·lula-matriu (Laotredou et al., 2013).

4.2. Cerca bioinformàtica de primers i disseny de les condicions de PCR

Tal i com s'ha mencionat a la metodologia, s'ha realitzat un alineament múltiple de les seqüències de les tres espècies candidates per a cada un dels cinc marcadors (Figura 9). Al realitzar-los s'ha pogut observar, en tots els marcadors, que les seqüències alineades presenten longituds diferents, tot i així, la longitud de la secció alineada ha estat suficient com per poder generar una seqüència consens de qualitat en la majoria de marcadors.

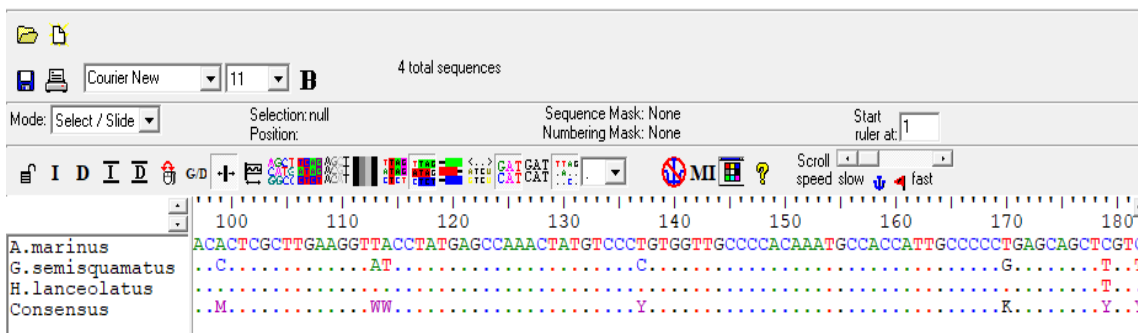


Figura 9: Alineament realitzat pel marcador IRBP. Permet veure el nivell de conservació entre les tres seqüències alineades. En aquest cas es pot observar que presenten zones amb un gran nivell de conservació, fent-les bones candidates per a l'ús de primers.

Pel que fa a la cerca bioinformàtica de primers s'han pogut obtenir dos parells de primers de qualitat pels marcadors IRPB, RNF213, MLL2 i Pkdt (Taula 2), en canvi, pel marcador MLL4 no s'ha trobat ni un sol parell de primers de qualitat. El motiu d'això és que l'alineament obtingut per a aquest marcador presentava molts gaps entremig de les seqüències, de manera que el programa Primer3 únicament era capaç de generar parells de primers que no complien amb els requisits de qualitat que es demanen. En aquest cas concret, la mida del fragment que s'amplificaria presentava una longitud inferior a 350 parells de bases amb un percentatge de C i G inferior als acceptats, donant lloc així a unes condicions de PCR deficientes. Per tant, el marcador nuclear MLL4 ha estat descartat de l'estudi i s'han demanat els parells de primers, recomanats pel programa Primer3 (Figura 10), de millor qualitat dels altres quatre.

Taula 2: Primers seleccionats juntament amb les seves característiques.

Marcador Nuclear	Primer	Primer (Pb)	Amplicó (Pb)	GC (%)	MT (°C)	Seqüència
IRBP	Forward	19	500	57,9	59	AGTCAGTGACCCTCGCTTG
	Reverse	20		55	58	TCAGTTCTGGAGGCAGTAG
MLL2	Forward	20	508	55	59,7	ACTGCAGGCTCTTTCTCAG
	Reverse	20		50	61,4	CATGGGGCCTTGAATAAAC
Pkdt	Forward	20	474	50	60	GTTTGTGACAGCCGAAAGCTC
	Reverse	20		45	60,1	GGCCTTGATGGCTTTGTTTA
RNF213	Forward	18	499	55,6	62,9	TTTGACGGCCTTCCGAGA
	Reverse	20		50	61	TGTGCTTCGGTAAGGGTTG

Pair 1:					
<input checked="" type="checkbox"/> Left Primer 1:	Primer_F				
Sequence:	AGTCAGTGACCCTCGCTTG				
Start: 89	Length: 19 bp	Tm: 59.0 °C	GC: 57.9 %	ANY: 7.0	SELF: 1.0
<input checked="" type="checkbox"/> Right Primer 1:	Primer_R				
Sequence:	TCAGGTTCTGGAGGCAGTAG				
Start: 588	Length: 20 bp	Tm: 58.0 °C	GC: 55.0 %	ANY: 3.0	SELF: 0.0
Product Size: 500 bp	Pair Any: 4.0	Pair End: 0.0			
<input type="button" value="Send to Primer3Manager"/>	<input type="button" value="Reset Form"/>				
1	TCGCAGCAGC	TGGCAGCAAC	ACAGAGATTC	TCAACATCCC	AGACGGCGAA
51	GCCTTGGCTA	ATGTTCTCAG	TGCTGGAGTG	CAGTCCACAG	TCAGTGACCC
101	TCGCTTGAAG	GTATCCTATG	AGCCAAACTA	TGTCCCCTGT	GTTGCCCCAC
151	AAATGCCACC	ATTGCCCCCG	GAGCAGCTTG	TTGCCATCCT	CCAGACCTCA
201	ATCAAGCTCG	ACATCCTTGA	GGGTAACATT	GGCTACCTGC	GAATTGATCA
251	CATCCTTGGG	GAGGATGTAG	CTGAAAAGAT	TGGCCCTATG	CTCCTGGACC
301	TTGTCTGGAA	TAAGATCCTG	CCAACCTCAG	CTCTTATATT	TGATCTACGC
351	TACACCAGCA	CGGGGACAT	CTCAGGAATC	CCCTATATTG	TGTCTTACTT
401	TACCGAAGCT	GAGCCTGCAA	TTCACATTGA	CAGTGTATAT	GACCGTCCCT
451	CAAAACACCAC	AACTAAGATG	TTCTCTATGA	CGACACTACT	GGGGGAAAGA
501	TATGGCACTA	CCAAAACCCCT	TATTATGCTC	ACCAAGTAAA	ACACAATTGG
551	CATTGCTGAG	GATCTCGCCT	ACTGCCTCCA	GAACCTGAAG	AGGGCAACAA
601	TTGTGGGTGA	GAAGACGGCA	GGGGGCTCTG	TGAACATCGA	TAAATCAGA
651	GTGGCTGACA	CTGACTTCTA	TGTAACCGTG	CCAACCTGCAA	AGTCCATCAA
701	CCCCATCACC	GGCTCCACCT	GGGAAATTAG	AGGAGTGACC	CCTGAT

Figura 10: Resultats del programa Primer3 pel marcador IRBP. S'observen els dos primers seleccionats pel programa així com les seves característiques i la mida del producte final.

4.3. Resultat de les PCR i comprovacions electroforètiques

Tal com s'ha explicat a l'apartat de metodologia, prèviament a testar els primers dels corresponents marcadors nuclears s'ha fet la prova amb els primers del marcador mitocondrial utilitzats pel LIG-UdG per tal de verificar la qualitat de les extraccions de DNA i la capacitat per realitzar les corresponents PCR. S'han obtingut bons resultats, procedint així a l'ús dels marcadors nuclears.

A nivell de resultats, es pot apreciar que la PCR dels marcadors MLL2 i Pkdt (Figura 11) no ha funcionat correctament, és a dir, els primers utilitzats no han pogut hibridar amb el DNA i per tant no s'ha amplificat el fragment desitjat. Tot i així, sembla que s'observa una lleugera banda en algun dels individus que segurament no serà més que soroll de fons o amplificació residual. Per altra banda, pels marcadors IRBP (Figura 12) i RNF213 (Figura 13) s'observa que han amplificat tots els individus de l'espècie 2 (*G. semisquamatus*) però no els de l'espècie 1 (*G. cicereus*). S'han considerat diferents hipòtesis per tal d'explicar què ha pogut passar. La primera d'elles és que al haver seleccionat els primers a partir d'uns alineaments on no hi era present l'espècie 1 possiblement el resultat final no és compatible amb aquesta espècie.

Per altra banda s'ha considerat el fet que la PCR podria no haver funcionat correctament o que les extraccions de DNA no s'haguessin realitzat del tot bé.

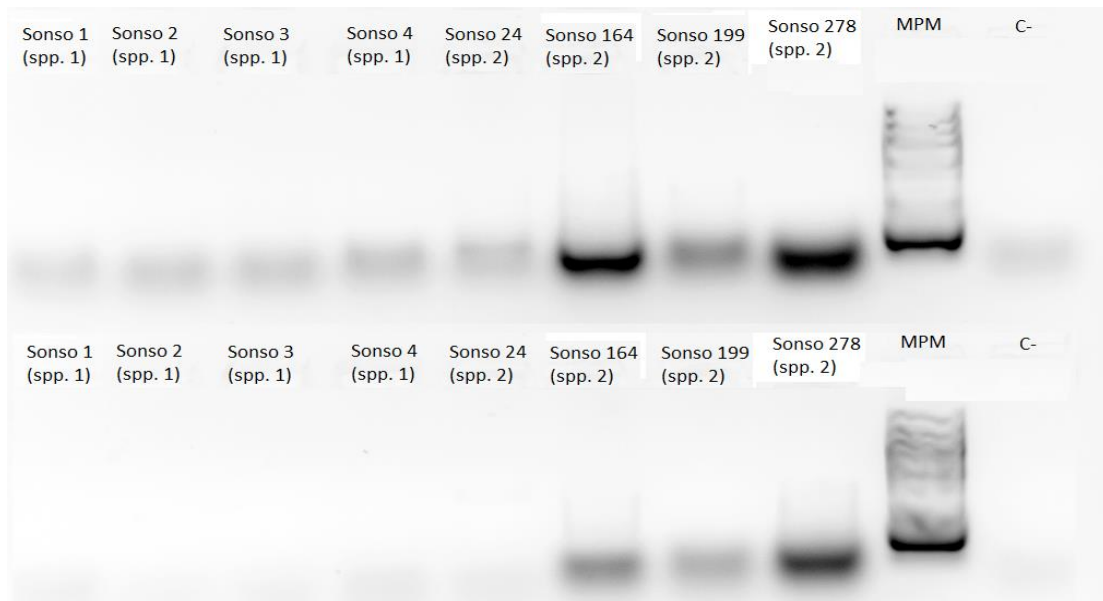


Figura 11: Resultats d'electroforesi corresponents a una PCR on s'han utilitzat els marcadors nuclears MLL2 (carril superior) i Pkdt (carril inferior). No s'observa amplificació a cap individu independentment del marcador nuclear.

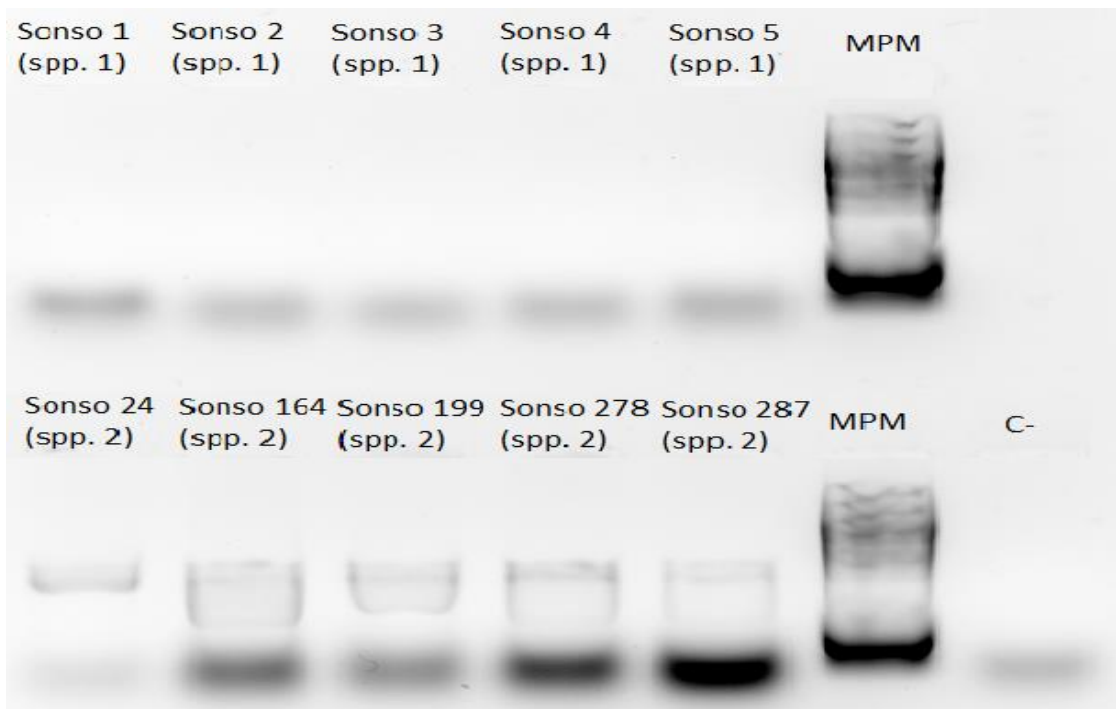


Figura 12: Resultats d'electroforesi corresponents a una PCR on s'ha utilitzat el marcador nuclear IRBP. S'observa amplificació a tots els individus de l'espècie 2 (carril inferior) però a cap de l'espècie 1 (carril superior). No s'observa amplificació en el control negatiu.

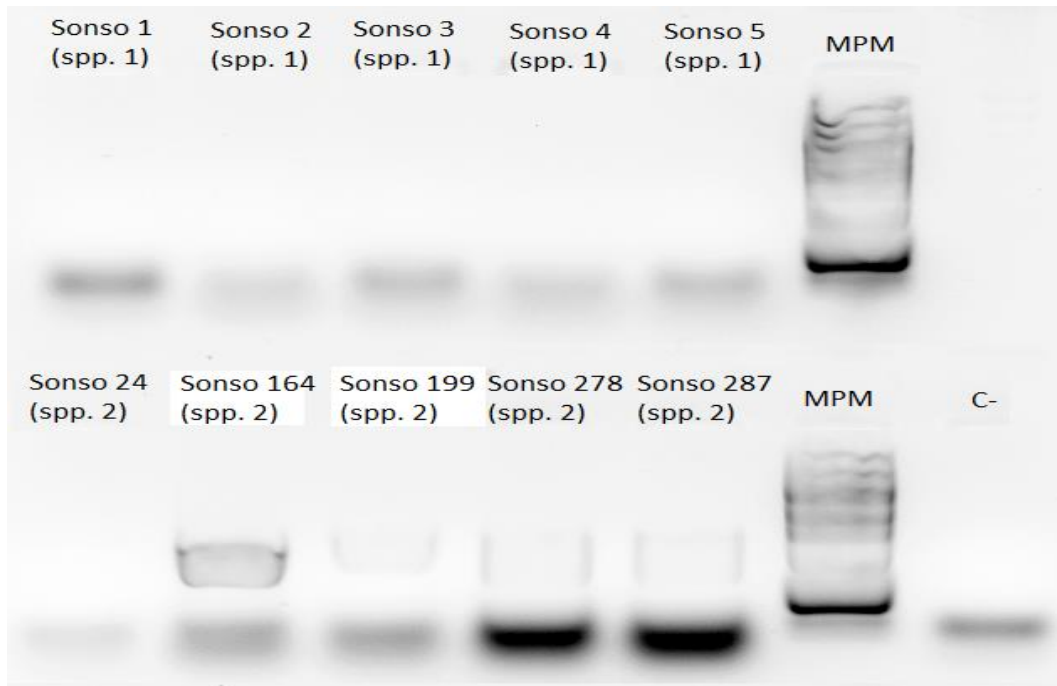


Figura 13: Resultats d'electroforesi corresponents a una PCR on s'ha utilitzat el marcador nuclear RNF213. S'observa amplificació a tots els individus de l'espècie 2 (carril inferior) però a cap de l'espècie 1 (carril superior). No s'observa amplificació en el control negatiu.

Considerant aquests resultat, s'han repetit les extraccions de DNA i s'han repetit tots els passos realitzats per tal de tornar a provar els marcadors IRBP i RNF213. Al fer-ho s'han obtingut els resultats mostrats a la Figura 14. En aquest cas s'observa que, efectivament, el problema es trobava en les extraccions de DNA que no s'havien realitzat del tot bé i segurament no contenien la quantitat suficient de DNA i/o hi havia restes de proteïnes. Com es pot apreciar, han amplificat tots els individus d'ambdues espècies a tots dos marcadors nuclears, exceptuant els controls negatius. Gràcies a aquests resultats es pot concloure que efectivament els *primers* seleccionats funcionen per a ambdues espècies.

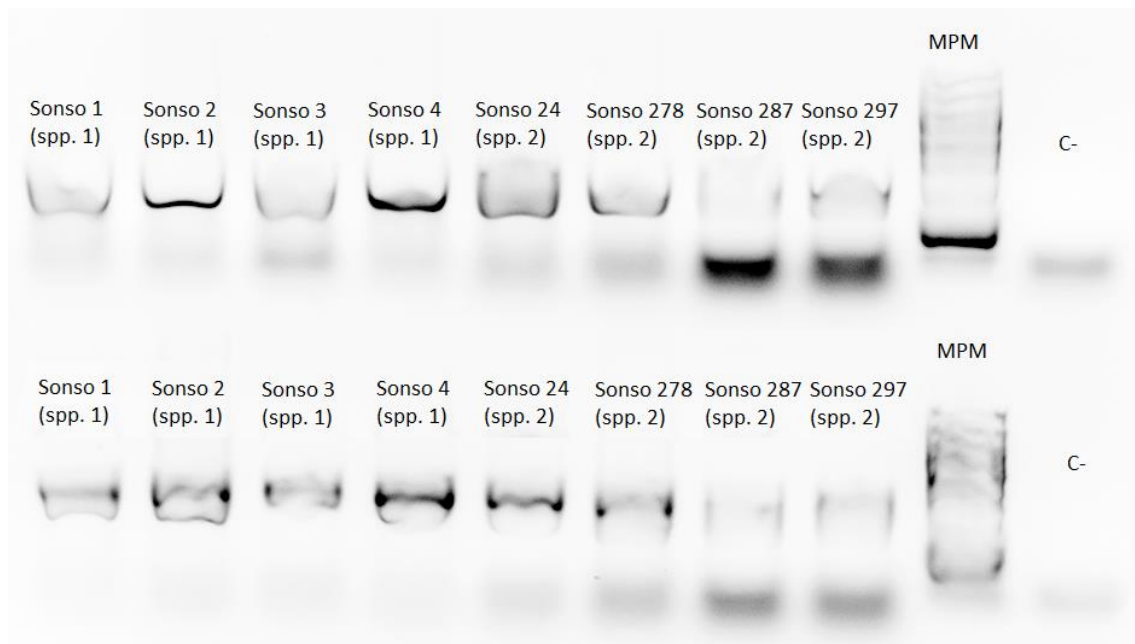


Figura 14: Resultats d'electroforesi corresponents a una PCR on s'han utilitzat els marcadors nuclears IRBP (carril superior) i RNF213 (carril inferior) en noves extraccions de DNA. S'observa amplificació a tots els individus independentment del marcador utilitzat.

4.4. Anàlisi bioinformàtiques

Considerant els resultats anteriors, s'ha procedit a la seqüenciació de tots els individus utilitzats pels marcadors IRBP i RNF213. Per algun motiu desconegut les seqüenciacions amb els *primers forward* no han funcionat correctament, per tant s'han realitzat utilitzant els *primers reverse*. Tal i com s'ha explicat a la metodologia s'han editat les seqüències obtingudes, s'han alineat amb les de referència (Figura 15) i s'ha construït un arbre filogenètic per a cada un dels dos marcadors nuclears (Figures 16 i 17).

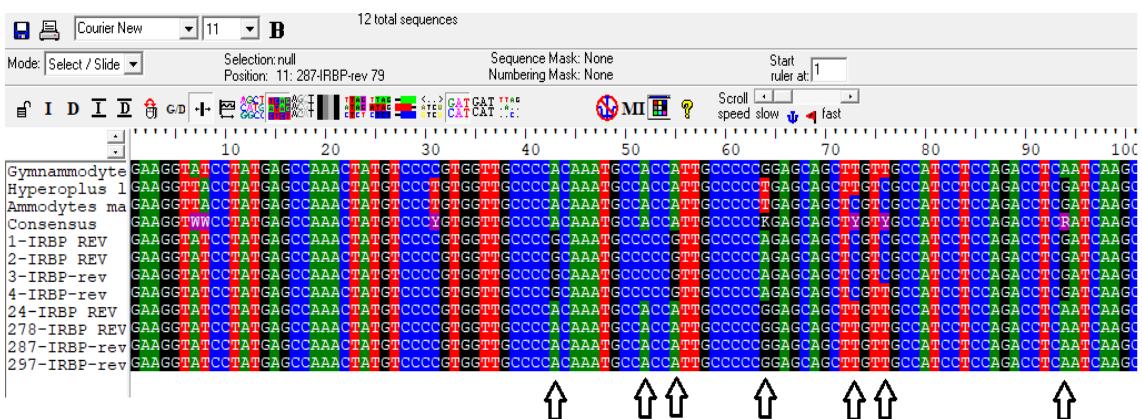


Figura 15: Alineament múltiple obtingut pel marcador IRBP, on els individus 1, 2, 3 i 4 corresponen a l'espècie 1 (*G. cicereus*) i els individus 24, 278, 287 i 297 a l'espècie 2 (*G. semisquamatus*). Es poden veure certs patrons en les seqüències que varien entre una espècie i l'altre.

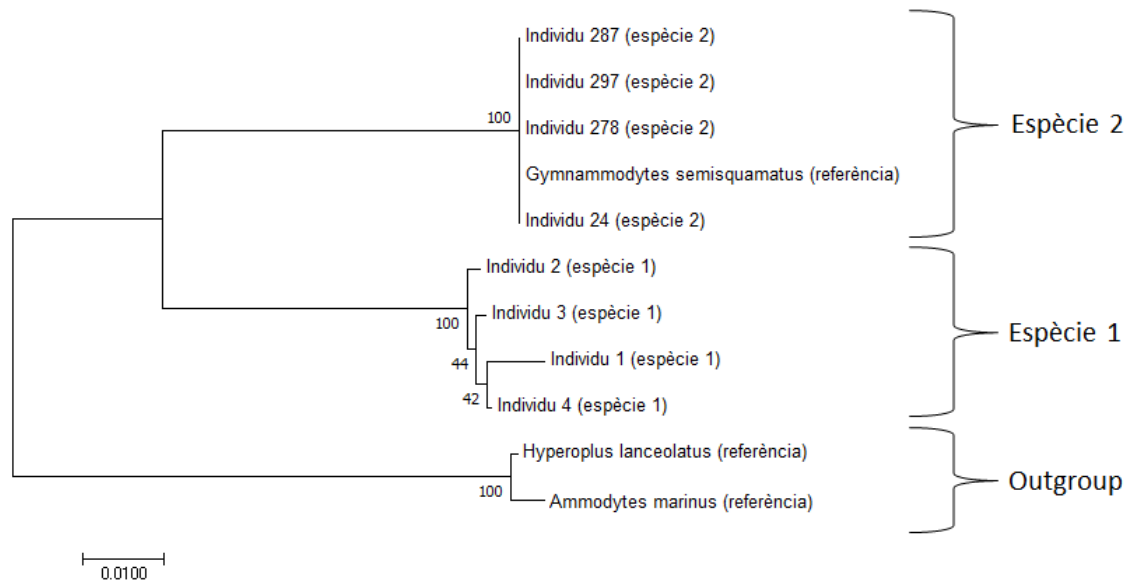


Figura 16: Arbre filogenètic construït pel mètode *neighbor joining* amb les seqüències del marcador nuclear IRBP. L'espècie 1 correspon a *G. cicereus* mentre que l'espècie 2 correspon a *G. semisquamatus*.

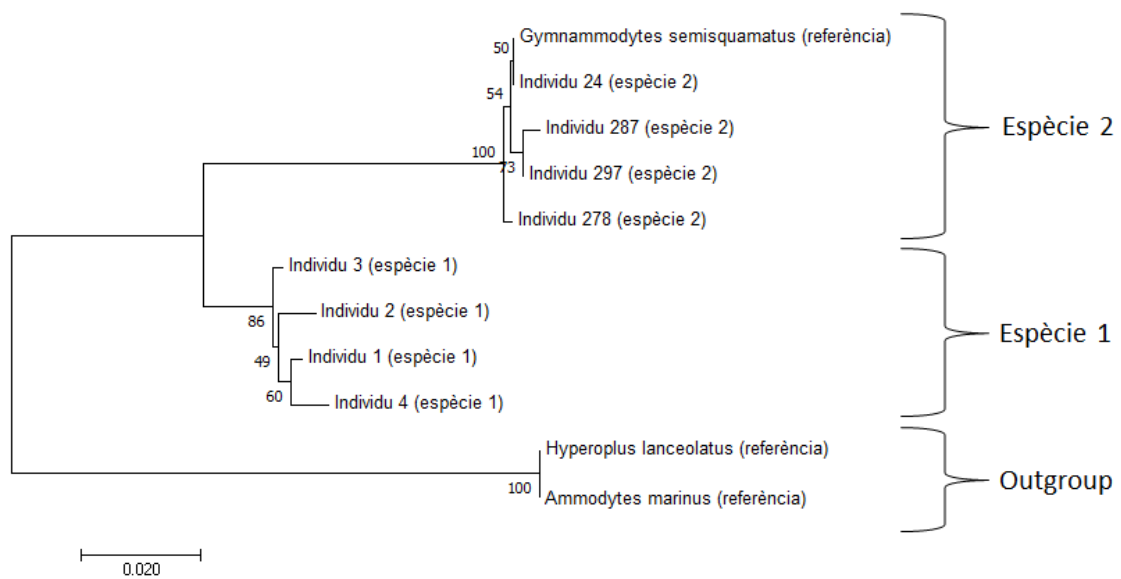


Figura 17: Arbre filogenètic construït pel mètode *neighbor joining* amb les seqüències del marcador nuclear RNF213. L'espècie 1 correspon a *G. cicereus* mentre que l'espècie 2 correspon a *G. semisquamatus*.

A grans trets s'observa que els dos marcadors nuclears donen lloc a un arbre que agrupa les diferents mostres d'una manera molt similar. Es pot apreciar que tots els individus de l'espècie 1 s'agrupen en una branca i els de l'espècie 2 s'agrupen en una altra, cosa que permet determinar preliminarment que efectivament si que es tracta de dues espècies diferents. Pel que fa a les espècies de referència es pot veure que *G. semisquamatus* queda agrupat juntament amb els individus de l'espècie 2, que són de la mateixa espècie, cosa que augmenta en gran mesura la fiabilitat d'aquests resultats. Això és així bàsicament pel fet que si individus de la hipotètica espècie

G. semisquamatus quedessin agrupats separadament d'un individu d'aquesta mateixa espècie, que s'utilitza com a referència, seria senyal que s'ha produït algun error metodològic. A més, es pot observar que els altres individus de referència (*H. lanceolatus* i *A. marinus*) queden agrupats separadament de la resta actuant com a *outgroup*. Aquest fet dóna encara més robustesa als resultats ja que és normal que les branques de *G. cicereus* i *G. semisquamatus* s'uneixin entre elles en un únic node abans de fer el mateix amb les espècies que conformen l'*outgroup*, ja que aquestes últimes tot i formar part de la mateixa família que les altres no pertanyen al mateix gènere, cosa que sí que passa amb les dues espècies d'estudi.

Finalment, es poden observar les distàncies entre els diferents grups que s'han establert (Taules 3 i 4), en aquest cas són el grup d'individus de l'espècie 1, el grup d'individus de l'espècie 2 i el grup d'individus que conformen l'*outgroup*. Ràpidament es pot observar que els resultats són similars per als dos marcadors utilitzats. A més, són molt coherents si es comparen amb els obtinguts en l'arbre filogenètic, on la mínima distància s'observa entre el grup representatiu de l'espècie 1 i el representatiu de l'espècie 2 i la màxima distància s'observa entre el grup representatiu de l'espècie 1 i l'*outgroup*. És normal, doncs, que la mínima distància es trobi entre l'espècie 1 i la 2 ja que són espècies diferents dins d'un mateix gènere, mentre que amb les espècies que conformen l'*outgroup* només comparteixen família.

Taula 3: Matriu de distàncies entre els grups formats pels individus de l'espècie 1 (*G. cicereus*), els de l'espècie 2 (*G. semisquamatus*) i els que conformen l'*outgroup*. A la diagonal inferior s'hi mostren les distàncies i a la diagonal superior les desviacions estàndard. Matriu realitzada a partir de les seqüències obtingudes pel marcador IRBP.

	Espècie 2	Outgroup	Espècie 1
Espècie 2		0,019	0,014
Outgroup	0,123		0,018
Espècie 1	0,082	0,118	

Taula 4: Matriu de distàncies entre els grups formats pels individus de l'espècie 1 (*G. cicereus*), els de l'espècie 2 (*G. semisquamatus*) i els que conformen l'*outgroup*. A la diagonal inferior s'hi mostren les distàncies i a la diagonal superior les desviacions estàndard. Matriu realitzada a partir de les seqüències obtingudes pel marcador RNF213.

	Espècie 2	Outgroup	Espècie 1
Espècie 2		0,024	0,013
Outgroup	0,17		0,019
Espècie 1	0,063	0,13	

5. Conclusions

From all results obtained several conclusions can be drawn, as well as improvement proposals that could have allowed to get a greater robustness and reliability of the results obtained.

First, it has already been mentioned that two of the five used molecular markers (MLL2 and Pkdt) haven't worked. This could have been caused by the fact that the conservation level of these markers, in the three reference aligned species, was smaller than the present in the nuclear markers that have worked correctly, therefore, it's possible that at the moment to realize the alignment the obtained consensus sequence wasn't of good quality, giving rise to a deficient primers selection. Despite this, it has been proved that the first DNA extractions lacked of good quality and the primers of these two markers has never been tested again with the new extractions. Considering this, it could be that if the PCR had been repeated with the new extractions the amplification of the desired fragment could have been achieved for future sequencing.

However, two of the five markers initially considered have yielded good results which, at the same time, are quite similar. The trees generated from the sequences of both markers separate clearly *G. cicereilus* from *G. semisquamatus*. In addition, the branch of each species ends up joining each other before doing the same with the outgroup. As mentioned above this is coherent by the fact that both *G. cicereilus* and *G. semisquamatus* are two species who share the same genus, while *H. lanceolatus* and *A. marinus* are species of the same family but of separate genus. It's important to mention that the trees have been repeated using a higher bootstrap and still has given the same results.

Finally, the distance matrices generated have allowed the confirmation of the phylogenetic trees results, where it can be checked that effectively the minor distance is situated between the group composed by the individuals of *G. cicereilus* and the group composed by the individuals of *G. semisquamatus*, while the higher distances are the represented by this two groups with de outgroup.

Based on all of this it can be concluded that, effectively, *Gymnamodytes cicereilus* and *Gymnamodytes semisquamatus* are two different species inside the same genus. Despite this, future researches using more nuclear markers could be necessary in order to confirm these results.

6. **Bibliografia**

Avise, J.C. and R.M. Ball, Jr. (1990). *Principles of genealogical concordance in species concepts and biological taxonomy*. Oxford Surveys in Evolutionary Biology 7:45-67.

Azuma, N., Hangui, J., Wakahara, M., & Michimae, H. (2013). *Population Structure of the Salamander *Hynobius retardatus* Inferred from a Partial Sequence of the Mitochondrial DNA Control Region*. Zoological Science, 30(1), 7–14.
<https://doi.org/10.2108/zsj.30.7>

Canyelles Ferrà, X., & Mas Ferrà, X. (2000). *Peixos de les Illes Balears*. Editorial Moll. Recuperat de http://cabib.uib.es/search*cat/i?SEARCH=8427360134

Froese, R. and D. Pauly. Editors. 2018. *FishBase*. World Wide Web electronic publication. www.fishbase.org, version (02/2018).

Hall, T.A., *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT*. Nucleic Acids Symposium Series, 1999. 41: p. 95-98.

Hauser, L., G. J. Adcock, P. J. Smith, J. H. B. Ramírez & G. R. Carvalho. (2002). *Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*)*. Proceedings of the National Academy of Sciences, USA 99:11742-11747.

Larkin, M.A., et al., *Clustal W and Clustal X version 2.0*. Bioinformatics, 2007. 23(21): p. 2947-8.

Lautredou, A.-C., Motomura, H., Gallut, C., Ozouf-Costaz, C., Cruaud, C., Lecointre, G., & Dettai, A. (2013). *New nuclear markers and exploration of the relationships among Serraniformes (*Acanthomorpha*, *Teleostei*): The importance of working at multiple scales*. Molecular Phylogenetics and Evolution, 67(1), 140–155.
<https://doi.org/10.1016/j.ympev.2012.12.020>

Lleonart, J., M. Demestre, P. Martín, J. Rodón, S. Sainz Trápaga, P. Sánchez, I. Segarra & S. Tudela. (2014). *The co-management of the sand eel fishery of Catalonia (NW Mediterranean): the story of a process Scientia Marina*. vol 78, 87-93.

Mitchell, A., McCarthy, E., & Verspoor, E. (1998). *Discrimination of the North Atlantic lesser sandeels *Ammodytes marinus*, *A. tobianus*, *A. dubius* and *Gymnammodytes semisquamatus* by mitochondrial DNA restriction fragment patterns*. Fisheries Research, 36(1), 61–65. [https://doi.org/10.1016/S0165-7836\(98\)00081-2](https://doi.org/10.1016/S0165-7836(98)00081-2)

NCBI. (2018). *RBP3 retinol binding protein 3 [*Homo sapiens* (human)]*. Recuperat de <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd>ShowDetailView&TermToSearch=5949>

NCBI. (2018). *RNF213 ring finger protein 213 [*Homo sapiens* (human)]*. Recuperat de <https://www.ncbi.nlm.nih.gov/gene/57674>

Punt, A. E. (2006). *The FAO Precautionary Approach after almost 10 years: have we progressed towards implementing simulation-tested feedback-control management systems for fisheries management?* Natural Resource Modeling 19:441-464.

Rozen, S., Skaletski, H. (2012). *Primer3*. Recuperat de <http://primer3.sourceforge.net/history.php>

Saitou, N., & M, Nei. (1987). *The neighbor-joining method: a new method for reconstructing phylogenetic trees*. Molecular Biology and Evolution, 4(4), 406–425.
<https://doi.org/10.1093/oxfordjournals.molbev.a040454>

Sambrook, J., E.J. Fritsch, and T. Maniatis., *Molecular Cloning. A Laboratory Manual*. 1989, New York: Cold spring Harbor, Laboratory Press.

Sambrook, J., Russel, D. (2001). *Molecular Cloning: A Laboratory Manual (3rd ed. edición)*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. ISBN 0-87969-576-5.

Stock, S. P. (2009). *Insect pathogens : molecular approaches and techniques*. CABI. Retrieved from https://books.google.es/books?id=yFqTeKMS_T0C&pg=PA22&redir_esc=y#v=onepage&q&f=false

Tamura, K., et al., MEGA6: *Molecular Evolutionary Genetics Analysis Version 6.0*. *Molecular Biology and Evolution*, 2013. 30(12): p. 2725-2729.

Technelysium. (1996). *Chromas*. Recuperat de <https://technelysium.com.au/wp/>

The IUCN Red List of Threatened Species. (2015). *Gymnamodytes cicereus*. Recuperat de <http://www.iucnredlist.org/details/198630/0>

Waples, R. S., A. E. Punt & J. M. Cope. (2008). *Integrating genetic data into management of marine resources: how can we do it better?* *Fish and Fisheries* 9:423-449.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). *Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction*. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-13-134>

Yu-Huei, C. (2015). *Estimation of Teaching-Learning-Based Optimization Primer Design Using Regression Analysis for Different Melting Temperature Calculations*. *IEEE Transactions on NanoBioscience*, 14(1), 3–12. <https://doi.org/10.1109/TNB.2014.2352351>