

Títol del treball:

Estudi bioinformàtic de la variabilitat genètica de la tonyina vermella al Mediterrani

Estudiant: Anna Sanchez-Reseco Villaró

Grau en Biotecnologia

Correu electrònic: asanchez.reseco@gmail.com

Tutor: Jordi Viñas De Puig

Empresa / institució: UNIVERSITAT DE GIRONA

Vistiplau tutor:



Nom del tutor: Jordi Viñas De Puig

Empresa / institució: UNIVERSITAT DE GIRONA

Correu electrònic: jordi.vinas@udg.edu

Data de dipòsit de la memòria a secretaria de coordinació: 22 de juliol de 2016

RESUM

Amb l'objectiu d'estudiar la variabilitat genètica de la tonyina vermella (*Thunnus thynnus*) s'han analitzat 689 seqüències de la regió control de DNA mitocondrial d'individus d'aquesta espècie mitjançant eines bioinformàtiques. S'han dut a terme quatre anàlisis basades en l'estudi del nombre d'haplotipus i amb diferents tractaments de les dades en cada cas. Per detectar a quin haplotipus correspon cada seqüència s'ha fet un arbre filogenètic i s'ha determinat que dos o més seqüències corresponen al mateix haplotipus quan no hi ha distància genètica entre elles. Analitzant l'arbre filogenètic s'ha vist la gran variabilitat que presenten les poblacions de *T. thynnus*, on un alt nombre d'haplotipus estan formats només per una seqüència.

En els quatre anàlisis s'ha estudiat com apareixen els haplotipus a mesura que es va augmentant el volum de seqüències analitzades per veure en quin punt deixen d'aparèixer haplotipus nous, la qual cosa significaria que ja està representada tota la variabilitat genètica de la població. Dels quatre anàlisis realitzats s'han seleccionat els que millor responen davant els objectius plantejats, degut principalment al tractament de les dades, i s'ha pogut extreure que cal analitzar més de 657 seqüències per obtenir la representació de tota la població de *T. thynnus*.

També s'ha fet l'estudi de la diversitat haplotípica (*h*). S'ha vist que els valors d'aquest estadístic són generalment alts, tot i que van disminuint com més augmenta el volum de seqüències analitzades. Uns valors de diversitat haplotípica elevats signifiquen que en la població de *T. thynnus* hi ha una gran diversitat genètica i això comporta que calgui analitzar un elevat nombre d'individus per poder representar tota aquesta variabilitat. Aquest fet és essencial quan es fan estudis de genètica de poblacions, filogènia i filogeografia. A més, és de vital importància realitzar correctament els estudis poblacionals en una espècie com *T. thynnus*, ja que és una espècie molt valorada i amenaçada. Per això és necessari saber quan un d'aquests estudis representa o no a tota la població de *Thunnus thynnus*.

Paraules clau: Tonyina vermella, *Thunnus Thynnus*, regió control de DNA mitocondrial, haplotipus, diversitat haplotípica, variabilitat genètica

RESUMEN

Con el objetivo de estudiar la variabilidad genética del atún rojo (*Thunnus thynnus*) se han analizado 689 secuencias de la región control de DNA mitocondrial de individuos de esta especie mediante herramientas bioinformáticas. Se han realizado cuatro análisis basados en el estudio del número de haplotipos y con diferentes tratamientos de los datos en cada caso. Para detectar a que haplotipo corresponde cada secuencia se ha elaborado un árbol filogenético y se ha considerado que dos o más secuencias corresponden al mismo haplotipo cuando no hay distancia genética entre ellas. Analizando el árbol filogenético se ha visto la gran variabilidad que presentan las poblaciones de *T. thynnus*, donde un elevado número de haplotipos están formados sólo por una secuencia.

En los cuatro análisis se ha estudiado como aparecen los haplotipos a medida que se va aumentando el volumen de secuencias analizadas para ver en qué punto dejan de aparecer haplotipos nuevos, lo que significaría que ya queda representada toda la variabilidad genética de la población. De los cuatro análisis realizados se han seleccionado los que mejor responden ante los objetivos planteados, debido principalmente al tratamiento de los datos, y se ha podido inferir que es necesario analizar más de 657 secuencias para obtener la representación de toda la población de *T. thynnus*.

También se ha estudiado la diversidad haplotípica (h). Se ha visto que los valores de este estadístico son generalmente altos, a pesar de que van disminuyendo cuanto más aumenta el volumen de secuencias analizadas. Unos valores de diversidad haplotípica elevados significan que en la población de *T. thynnus* hay una gran diversidad genética y esto conlleva que sea necesario analizar un elevado número de individuos para poder representar toda esta variabilidad. Esto es esencial en los estudios de genética de poblaciones, filogenia y filogeografía. Además, es de vital importancia realizar correctamente los estudios poblacionales en una especie como *T. thynnus*, ya que es una especie muy valorada y amenazada. Por eso es necesario saber cuándo uno de estos estudios representa o no toda la población de *Thunnus thynnus*.

Palabras clave: Atún rojo, *Thunnus thynnus*, región control de DNA mitocondrial, haplotipo, diversidad haplotípica, variabilidad genética

ABSTRACT

With the aim of studying genetic variability of Bluefin Tuna (*Thunnus thynnus*) 689 sequences of the mitochondrial DNA control region of this species have been analyzed using bioinformatics tools. Four analysis have been performed based on the study of the number of haplotypes and using a different data treatment in each case. In order to detect which haplotype belongs to each sequence a phylogenetic tree has been created and it has been established that two or more sequences belong to the same haplotype when there isn't genetic distance between them. Analyzing the phylogenetic tree the huge variability of *T. thynnus* populations has been detected, where a large number of haplotypes are made up of only one sequence.

In the four analysis it has been studied how haplotypes appear while the volume of analyzed sequences increases in order to see at what point new haplotypes don't appear anymore, which means that complete genetic variability of the population has already been represented. The analysis that best represent the objectives suggested have been selected between the four analysis, principally due to data treatment, and it has been concluded that more than 657 sequences have to be analyzed in order to achieve the total representation of *T. thynnus* population.

Haplotype diversity (h) has also been studied. It has been observed that these statistic values are generally high, despite the fact that they decrease while the volume of analyzed sequences is increasing. High haplotype diversity values indicate that *T. thynnus* populations present a huge genetic diversity and this implies that an extremely large number of individuals has to be analyzed to represent all this variability. This is essential when population, phylogenetic and phylogeography studies are performed. Moreover, it is crucial to perform population studies correctly when they center around species such as *T. thynnus*, because it is a highly valued and threatened species. For this reason it is necessary to know when one of these studies do or do not represent the entire *Thunnus thynnus* population.

Key words: Bluefin Tuna, *Thunnus thynnus*, mitochondrial DNA control region, haplotype, haplotype diversity, genetic variability

ÍNDEX

| | |
|----------------------------------------------------------------------------------------|----|
| INTRODUCCIÓ | 5 |
| Ètica i sostenibilitat | 8 |
| OBJECTIU | 9 |
| MATERIALS I MÈTODES | 10 |
| Anàlisi 1: Estudi d'haplotipus independent de les seqüències anteriors..... | 11 |
| Anàlisi 2: Estudi d'haplotipus dependent de les seqüències anteriors..... | 12 |
| Anàlisi 3: Estudi d'haplotipus diferents independent de les seqüències anteriors | 12 |
| Anàlisi 4: Estudi d'haplotipus diferents dependent de les seqüències anteriors | 13 |
| RESULTATS | 14 |
| Arbre filogenètic: detecció d'haplotipus..... | 14 |
| Anàlisi 1: Estudi d'haplotipus independent de les seqüències anteriors..... | 15 |
| Anàlisi 2: Estudi d'haplotipus dependent de les seqüències anteriors..... | 15 |
| Anàlisi 3: Estudi d'haplotipus diferents independent de les seqüències anteriors | 16 |
| Anàlisi 4: Estudi d'haplotipus diferents dependent de les seqüències anteriors | 17 |
| DISCUSSIÓ | 19 |
| CONCLUSIONS | 22 |
| AGRAÏMENTS..... | 23 |
| BIBLIOGRAFIA..... | 24 |

INTRODUCCIÓ

La tonyina vermella (*Thunnus thynnus*), representada en la Figura 1, és l'espècie més gran d'entre els túnids i presenta un important comportament migratori (Rooker et al., 2007). El seu cos és fusiforme, pot arribar a superar els 4 metres de longitud i el seu pes màxim oficial està registrat en 726kg, tot i que s'han arribat a trobar individus de fins a 900kg (Fromentin, 2006). És una espècie molt valorada, culinàriament i de gran valor comercial, que en els últims anys s'ha vist afectada per la sobrepesca (Cloutier et al., 2013) fins al punt que va ser proposada com a espècie en perill per l'ICCAT (International Commission for the Conservation of Atlantic Tuna) (Collette et al., 2011).



Figura 1. Imatge d'un individu de *Thunnus Thynnus* (European Commission, 2016).

Segons l'ICCAT, l'organisme internacional responsable de la gestió d'aquesta espècie, actualment es consideren dos *stocks* o poblacions de *T. thynnus*, una situada a l'oest Atlàntic i una altra a l'est Atlàntic –la qual inclou el Mediterrani– (Figura 2). Aquesta separació es basa principalment en la diferència dels llocs de posta, per la població oest situat al Golf de Mèxic, i per la població oriental al Mar Mediterrani (Di Natale, 2015). Les dues poblacions també són diferents en mida, la població de l'est Atlàntic és aproximadament deu vegades més gran que la població occidental. Actualment ambdues poblacions són susceptibles a la sobrepesca, fins al punt que es va arribar a considerar que s'estaven esgotant (Giulia Riccioni et al., 2010). Actualment hi ha plans de recuperació per a *T. thynnus*, l'objectiu dels quals és restaurar les poblacions (ICCAT, 2015).

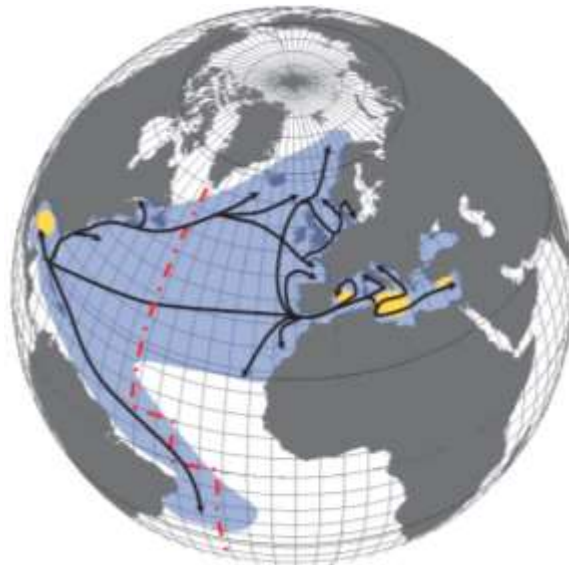


Figura 2. Mapa de la distribució espacial de *T. thynnus* (color blau). Les fletxes indiquen les principals rutes migratòries i les zones de color groc són els punts destacats on es dona la posta dels ous. La línia vermella discontinua representa la delimitació entre les poblacions de l'oest i de l'est Atlàntic (Fromentin, 2006).

A nivell biològic la separació entre les dues poblacions de *T. thynnus* és més indeterminada. Tot i que hi ha estudis que indiquen que hi pot haver heterogeneïtat genètica entre les poblacions (Carlsson et al., 2004), quan s'estudia el DNA mitocondrial (mtDNA) i s'amplia el nombre d'individus de les mostres s'observa com les diferències desapareixen (Alvarado Bremer et al., 2005). Això es deu a l'alt comportament migratori d'aquests peixos així com la manca de barreres geogràfiques que delimitin les poblacions, permetent el flux de gens (Boustany et al., 2008).

El DNA mitocondrial (mtDNA) és un marcador clàssic que s'ha utilitzat en genètica de poblacions, filogènia i filogeografia (Alvarado Bremer et al., 2005; Avise et al., 1987; Boustany et al., 2008; Carlsson et al., 2004). En una gran varietat d'espècies el mtDNA s'utilitza per l'estudi de les relacions que s'estableixen entre individus, poblacions i espècies gràcies a les seves característiques: té una mida petita, és abundant a la cèl·lula, s'hereta per via materna i té una elevada taxa de variació molecular (Manchado et al., 2004).

El mtDNA és un material genètic circular tancat, de doble cadena i que es localitza a l'interior dels mitocondris cel·lulars (Domínguez, 2000). La llargada total del mtDNA en *T. thynnus* és de 16526 pb, similar a la llargada del mtDNA d'altres animals. L'estructura del mtDNA de *T. thynnus* és semblant a la que presenten altres genomes mitocondrials. En aquest cas el mtDNA conté 13 gens que codifiquen per proteïnes, els gens que codifiquen per les subunitats petita (12S) i gran (16S) de RNA ribosomal (rRNA) i 22 gens de RNA de transferència (tRNA). També conté una regió no codificant D-loop o regió control (mtDNA CR) de 865 pb considerada una regió conservada, a més aquesta regió no conté repeticions en tàndem (Manchado et al., 2004).

Quan s'han fet estudis a partir de la seqüenciació de mtDNA sempre s'ha examinat la variabilitat de seqüències de la regió control. Per exemple en el cas de *T. thynnus* s'ha vist que quan s'analitza el mtDNA CR de mostres recollides en el Golf de Mèxic, l'est i l'oest Mediterrani no hi ha diferenciació significativa entre les tres poblacions (Boustany et al., 2008). A més, la variació genètica d'aquesta regió és molt elevada, Ely et al. (2002) van afirmar que gairebé tots els individus analitzats tenien haplotipus únics, entenent per haplotipus com un conjunt d'al·lels provinents de diferents loci d'un cromosoma i que es transmeten junts. En el cas de DNA mitocondrial, es consideren concretament com un al·lel o haplotipus mitocondrial tots aquells loci que tinguin una combinació de mutacions, insercions, delecions i transversions diferent. Quan es treballa amb haplotipus únics implica que també es treballi amb diversitats haplotípiques molt elevades.

Cal tenir en compte que s'han donat casos en què la regió control de DNA mitocondrial de *T. thynnus* és més semblant a *Thunnus alalunga* o *Thunnus orientalis* que a la regió control d'altres *T. thynnus*. Es creu que això és degut a introgressions entre espècies, i si es tinguessin en compte en els estudis podrien donar estimacions de diversitat errònies (Alvarado Bremer et al., 2005; Boustany et al., 2008).

Un dels paràmetres principal per determinar la variabilitat genètica d'una població és la diversitat haplotípica (h). Aquesta es defineix com la probabilitat de que dos haplotipus seleccionats aleatòriament siguin diferents en una mostra (Excoffier & Lischer, 2015). La diversitat haplotípica es calcula amb la següent fórmula:

$$h = \frac{n}{n-1} \left(1 - \sum_{i=1}^k p_i^2 \right)$$

On n és el nombre d'individus que s'han analitzat en aquella mostra, k és el nombre d'haplotipus i p_i és la freqüència haplotípica de l'haplotipus i (Excoffier & Lischer, 2015).

En els estudis previs que s'han fet sobre *T. thynnus* es treballa amb valors de diversitats haplotípiques molt elevats. Algunes d'aquestes dades, seleccionades de publicacions amb què s'ha treballat amb mtDNA CR, queden recollides a la Taula 1:

Taula 1. Recopilació de dades referents a diferents estudis de la regió control de DNA mitocondrial de de *Thunnus Thynnus*.

| Estudi | Població | n | Hpl | h |
|--------------------------------------------------|------------------|-----|-------|--------------|
| Alvarado Bremer et al. (2005) | Mediterrani | 436 | 253 | 0.992 |
| | Nord Atlàntic | 141 | 101 | 0.988 |
| | TOTAL | 577 | 314 | 0.992 |
| Carlsson, McDowell, Carlsson, & Graves (2007) | Golf de Mèxic | 40 | 38 | 0.997 |
| | Oest Mediterrani | 77 | 62 | 0.992 |
| | Est Mediterrani | 23 | 21 | 0.949 |
| | TOTAL | 140 | 121 | 0.979 |
| Boustany et al. (2008) | Golf de Mèxic | 61 | 57 | 0.997 |
| | Carib | 8 | 7 | 0.964 |
| | Oest Mediterrani | 47 | 42 | 0.986 |
| | Est Mediterrani | 62 | 45 | 0.973 |
| TOTAL | 178 | 151 | 0.987 | |
| Carlsson et al. (2004) | Balear | 24 | 19 | 0.982 |
| | Tirrena | 23 | 21 | 0.984 |
| | Jònic | 22 | 21 | 0.996 |
| | TOTAL | 69 | 54 | 0.991 |
| Ely et al. (2002) | Mediterrani | 138 | 120 | 0.989 |
| | Atlàntic | 72 | 58 | 0.987 |
| | TOTAL | 210 | 178 | 0.988 |
| Viñas et al. (2011) | Oest Mediterrani | 40 | 36 | 0.994 |
| | Est Mediterrani | 40 | 37 | 0.996 |
| | TOTAL | 80 | 72 | 0.997 |
| Mitjana de les diversitats haplotípiques: | | | | 0.985 |

L'apartat *Estudi* fa referència a quina publicació pertanyen les dades, *Població* dona informació del lloc d'on es van pescar els individus per recollir les mostres, n és el nombre d'individus amb què es treballa, *Hpl* fa referència al nombre d'haplotipus de cada mostra i h és la diversitat haplotípica de cada mostra.

Segons la definició de h , que la mitjana de la diversitat haplotípica sigui 0.985 significa que només hi ha una probabilitat de l'1.5% d'obtenir dos individus iguals, de cada 100 individus 98.5 són diferents. Per tant quan es treballa amb aquests valors de diversitat haplotípica no es pot considerar que la mostra sigui representativa de la població de *Thunnus thynnus*. Aquest fet indica que hi ha una problemàtica a l'hora de treballar amb *T. thynnus*: quan s'utilitza la regió control

destaca una elevada diversitat on pràcticament cada individu és una seqüència diferent. Amb aquests nivells de diversitat tant elevats es pot arribar a preguntar quina quantitat d'individus cal mostrejar per tenir una representació real de la variabilitat genètica. Aquest fet és molt important per saber quin nombre d'individus cal mostrejar en cada punt en estudis de genètica de poblacions, on és essencial tenir una representació total de la variabilitat de la població.

Ètica i sostenibilitat

Moltes espècies de túnids formen una valuosa i amenaçada base en algunes pesqueries. *T. thynnus* és molt apreciada en la indústria del sushi i sashimi a Japó (Manchado et al., 2004) i també és una espècie susceptible a la pesca esportiva especialment a països com les Estats Units i Canadà (Collette et al., 2011). Per tant comprendre l'estructura de les poblacions d'aquests peixos és summament important per aquesta indústria (Boustany et al., 2008).

Degut a la sobrepesca, s'estima que la població oriental de *T. thynnus* ha decaigut un 80% des de principis dels anys 1970s (European Commission, 2016), i a nivell global es calcula un descens de l'espècie del 51% en referència a tres generacions enrere (39 anys), fent que *T. thynnus* es considerés en perill d'extinció (Collette et al., 2011). Per exemple, el 2011 la mortalitat deguda a la pesca estava per sobre del rang màxim per conservar la sostenibilitat, i la biomassa de la població de *T. thynnus* capaç de pondre ous era molt inferior a aquest rang (Collette et al., 2011).

Per aquest motiu s'han ideat plans de gestió i recuperació de *T. thynnus*. A la reunió anual de l'any 2006 de la Comissió Internacional per la Conservació de la Tonyina Atlàntica (ICCAT) es va acordar un pla de 15 anys per la recuperació de *T. thynnus*, basat en la reducció del nombre total de captures permeses, en l'establiment d'una mida mínima de pesca més alta i en mesures de control i execució més estrictes (European Commission, 2016).

Així doncs, per tal d'identificar i definir correctament les poblacions d'aquesta espècie i procurar mesures de gestió correctes, és de vital importància entendre l'estructura genètica de *Thunnus Thynnus* (Manchado et al., 2004), i saber quants individus s'han d'analitzar és un primer pas per desenvolupar correctament estudis de genètica de poblacions.

OBJECTIUS

Fins el moment s'ha vist la importància de conèixer l'estructura de les poblacions de *T. thynnus*, i per això s'han fet diversos estudis que ho investiguen (Boustany et al., 2008; Carlsson et al., 2004, 2007; G. Riccioni et al., 2010; Giulia Riccioni et al., 2010). Però, com es pot saber quan una mostra és representativa de tota la població? On és el límit d'individus que cal analitzar per què ho sigui?

Respondre aquestes preguntes és el principal objectiu que es planteja: determinar si hi ha un nombre límit d'individus que s'ha analitzar per tal de representar tots els haplotipus d'una població de *T. thynnus* i, en cas afirmatiu, quin és aquest volum d'individus que cal incorporar en una mostra quan es vol fer un estudi amb *T. thynnus*. Així es podrà establir un llinar mínim d'individus que cal analitzar per fer que les mostres siguin representatives d'una població.

També es pretén determinar com van apareixent els diferents haplotipus a mesura que s'afegeixen més seqüències a les anàlisis, així com estudiar la diversitat haplotípica per poder comparar els resultats amb altres estudis de *Thunnus thynnus* o fins i tot altres espècies.

OBJECTIVES

Up until now, the importance of knowing the structure of *T. thynnus* populations has been seen, and for this reason several studies investigating this issue have been performed (Boustany et al., 2008; Carlsson et al., 2004, 2007; G. Riccioni et al., 2010; Giulia Riccioni et al., 2010). Nevertheless, how can we know when a sample is representative of overall population? Which is the limit of individuals that have to be analysed in order to represent all the population?

Answering these questions is the main objective raised: determining if a limited number of individuals that have to be analysed in order to represent all the *T. thynnus* population exists, and, if that is the case, set which is the amount of individuals needed in a sample when studying *T. thynnus*. Thus a minimum threshold number of individuals which have to be analysed in order to obtain representative samples of a population can be established.

Furthermore, this study aims to determine how different haplotypes appear as more sequences are added to the analysis, as well as to study haplotype diversity in order to compare the obtained results with other *Thunnus Thynnus* analysis or even with other species.

MATERIALS I MÈTODES

Per assolir els objectius plantejats, s'ha realitzat un treball estrictament bioinformàtic sense generar noves dades. Part de les seqüències analitzades ja estaven disponibles al LIG (Laboratori d'Ictiologia Genètica de la UdG) obtingudes a partir d'estudis preliminars (Alvarado Bremer et al., 2005; Viñas et al., 2011) i també seqüències no publicades. A més també s'han analitzat noves seqüències disponibles a les bases de dades.

Per fer els estudis s'han utilitzat 689 seqüències de la regió control de mtDNA, 607 publicades per Viñas et al. (2005) i 82 de no publicades. Degut a les introgressions que es donen entre espècies de *T. thynnus*, principalment per presència de seqüències similars a *Thunnus orientalis* i *Thunnus alalunga* en genomes de *T. thynnus* s'han eliminat de l'anàlisi aquestes seqüències i per tant s'ha acabat treballant amb un total de 657 seqüències. La procedència dels individus és diversa: est i oest de l'Atlàntic (141 seqüències) i Mediterrani (516 seqüències), però s'ha considerat com una única població degut a que els estudis basats en aquest marcador no mostren que hi hagi diferències (Alvarado Bremer et al., 2005). Això significa que els resultats que s'obtenen són independents de la localitat.

Cal tenir en compte que en un primer moment no es va treballar amb aquest volum de seqüències, sinó que es van fer les anàlisis amb un submostratge per tal d'obtenir una idea inicial de les anàlisis i poder estimar quin és el procediment més adequat per a l'anàlisi.

També es va avaluar la utilització de seqüències disponibles en bases de dades, principalment Genbank mitjançant la cerca booleana per paraules clau "*mitochondrial control region*" i amb filtres per a organisme *Thunnus thynnus* i autor *NOT Viñas*. També es va considerar utilitzar BLAST, concretament blastn (Altschul 1990), utilitzant com a referència les seqüències proporcionades pel LIG. Els requisits necessaris per tal d'afegir les noves seqüències a les anàlisis serien que siguin seqüències pertanyents a individus, que no fossin haplotipus i que estiguessin citades en articles científics.

Per determinar els haplotipus s'ha realitzat una anàlisi bioinformàtica que es desglossa en 4 passos:

1. *Alineament*: s'ha fet l'alineament de totes les seqüències disponibles amb el programa Bioedit (Hall 1999) i l'aplicació que integra ClustalW (Larkin MA 2007). L'alineament resultat ha estat millorat manualment.
2. *Filogènia*: la reconstrucció filogenètica s'ha dut a terme utilitzant el programa MEGA versió 6 (Tamura et al., 2013). Es basa en la construcció d'un arbre filogenètic Neighbor-Joining amb les següents preferències: nombre de Bootstrap (Felsenstein 1986) de 1000 replicacions, distància de les branques basada en el nombre de diferències entre les seqüències, tenint en compte tant les transicions i transversions, i tractant la informació amb *Pairwise deletion*.
3. *Detecció d'haplotipus*: gràcies a l'arbre filogenètic resultant (Figura 3) s'ha determinat quines seqüències són iguals i, per tant, pertanyen al mateix haplotip. S'ha considerat que aquelles seqüències que mostren una distància genètica nul·la –es situen a la mateixa branca de l'arbre– són iguals i formen un haplotip.

4. *Comprovació final*: quan s'han identificat les seqüències de la mateixa branca s'ha comprovat la seva identitat utilitzant el programa Bioedit (Hall 1999), a través d'un alienament de les seqüències. Finalment s'ha obtingut el nombre d'haplotipus diferents de les seqüències analitzades.

Amb les dades obtingudes s'han realitzat quatre anàlisis per tal d'assolir els objectius proposats:

1. Anàlisi de les seqüències per determinar en quin moment es col·lapsa el nombre d'haplotipus estudiant les seqüències aleatòriament amb mostratges en grups de $n+25$ seqüències –on n és el volum de seqüències del punt anterior– i independentment en cada punt.
2. Anàlisi de les seqüències per determinar en quin moment es col·lapsa el nombre d'haplotipus estudiant les seqüències aleatòriament en grups múltiples de 25 i afegint dependència d'un punt respecte l'anterior.
3. Anàlisi de com apareixen els haplotipus a mesura que es van afegint individus a l'anàlisi estudiant les seqüències aleatòriament en grups de $n+25$ seqüències i independentment en cada punt.
4. Anàlisi de com apareixen els haplotipus a mesura que es van afegint individus a l'anàlisi estudiant les seqüències aleatòriament en grups múltiples de 25 i amb dependència d'un punt respecte l'anterior.

Aquestes quatre opcions s'han desenvolupat per determinar quina és la que millor pot respondre al problema que es planteja. És per això que són anàlisi aparentment semblants però en realitat tenen diferències significatives. A més, per cada anàlisi s'ha fet el càlcul de la diversitat haplotípica utilitzant la fórmula descrita en la introducció.

Anàlisi 1: Estudi d'haplotipus independent de les seqüències anteriors

L'objectiu d'aquesta anàlisi és determinar si hi ha un nombre límit d'individus que calgui estudiar per tal de representar tots els haplotipus d'una població.

Es basa en l'exploració de les 657 seqüències en blocs múltiples de 25 i independentment dels blocs anteriors –s'escullen els nous blocs sense tenir en compte les seqüències que ja s'han analitzat–. Inicialment s'han escollit 25 seqüències no repetides entre elles i aleatòries, s'ha mirat a quins haplotipus corresponen i s'ha recomptat quants haplotipus diferents representen. Després s'han escollit 50 seqüències, també aleatòriament, sense repeticions i sense tenir en compte si s'han afegit a l'anàlisi les 25 anteriors, i s'ha determinat a quin haplotipus correspon cadascuna i fet l'anàlisi dels haplotipus per determinar quants de diferents n'apareixen en total. Seguidament s'ha repetit el procés successivament, augmentant en blocs de 25 seqüències, fins a analitzar la totalitat de les seqüències.

Es pot resumir que en aquesta anàlisi l'estratègia es basa en l'estudi de les seqüències i els seus haplotipus a mesura que s'afegeixen més individus a la investigació, independentment entre ells i de forma aleatòria.

En total s'han fet 5 rèpliques per cada nou grup de seqüències aleatòries analitzades, per tant el procés anteriorment descrit s'ha repetit 5 vegades independents. D'aquesta manera el que es representa en la Figura 4 són les mitjanes de les 5 rèpliques a cada punt.

Anàlisi 2: Estudi d'haplotipus dependent de les seqüències anteriors

La principal diferència respecte l'anàlisi anterior és que en aquest cas en el nou bloc de seqüències s'inclouen de forma acumulativa les seqüències anteriors. Aquesta dependència també permet l'observació de com variaria un mateix estudi referent a *T. thynnus* si s'agafessin més o menys individus d'una mateixa població. A més també comparteix objectiu amb la primera anàlisi: determinar si hi ha un nombre límit d'individus, i per tant de seqüències, a partir del qual no apareguin més haplotipus diferents, de manera que aquest nombre correspondria al mínim d'individus que cal estudiar per representar la totalitat d'una població.

Per fer l'anàlisi 2 s'han estudiat les 657 seqüències en blocs múltiples de 25 i s'han anat afegint 25 noves seqüències de forma acumulativa en cada punt de l'anàlisi. En primer lloc s'han generat 25 seqüències aleatòries i no repetides, s'ha determinat a quins haplotipus corresponen i s'ha recomptat el total d'haplotipus que representen. Seguidament s'han generat 25 seqüències aleatòries, no repetides i diferents a les anteriors, les quals s'han afegit a la llista anterior. Així s'obté l'anàlisi de 50 seqüències, on s'ha determinat a quin haplotipus correspon cadascuna i quants haplotipus diferents hi ha en total. A continuació s'ha repetit el procés: s'han generat 25 seqüències aleatòries, no repetides i diferents a les 50 anteriors, que s'han afegit a la llista per obtenir un total de 75 seqüències i s'ha fet l'anàlisi dels haplotipus. Aquest procés s'ha repetit fins que l'estudi avarca les 657 seqüències totals.

En aquest cas l'estratègia que segueix l'anàlisi és l'estudi de les seqüències i els haplotipus corresponents a mesura que s'afegeixen més individus a la investigació, però sempre tenint en compte les seqüències ja analitzades i de manera aleatòria.

Igualment que en el cas anterior, l'anàlisi s'ha repetit en 5 rèpliques independents per cada punt seguint la metodologia ja explicada. Així en la Figura 5 es representa la mitjana de les rèpliques per cada punt

Anàlisi 3: Estudi d'haplotipus diferents independent de les seqüències anteriors

L'objectiu de l'anàlisi 3 és determinar com van apareixent haplotipus diferents a mesura que s'estudien més individus d'una població. En aquest cas no es representen els haplotipus totals, sinó que només són representats els que no han aparegut de nou en l'anàlisi a mesura que aquest va avançant. Aquesta anàlisi permet observar en quin punt deixen d'aparèixer haplotipus nous.

Per a realitzar aquesta anàlisi s'han estudiat les 657 seqüències en blocs múltiples de 25 seqüències i tenint en compte que els nous blocs que s'agafen són diferents als anteriors –s'escullen els nous blocs sense tenir en compte les seqüències que ja s'han analitzat– en un procediment igual a l'anàlisi 1. Primerament s'han seleccionat 25 seqüències no repetides i aleatòries, de les quals se n'ha determinat l'haplotipus. Els haplotipus diferents que van apareixent a l'anàlisi són els que s'han recomptat i representat al gràfic. A continuació s'han generat 50 seqüències, també aleatòriament, sense repeticions i independentment de les anteriors, i s'ha mirat a quin haplotipus correspon cadascuna. Els diferents haplotipus apareguts s'han comparat amb els que ja s'han trobat anteriorment, i el recompte d'aquells que no han aparegut s'ha representat. Seguidament s'ha repetit aquest procés per 75 seqüències i s'ha continuat l'anàlisi augmentant en blocs de 25 seqüències fins a arribar a l'estudi de la totalitat de la mostra.

L'estratègia que segueix aquesta anàlisi es basa en l'exploració dels haplotipus diferents que van apareixent a mesura que s'afegeixen individus aleatòria i independentment dels anteriors a l'estudi.

També s'han fet 5 rèpliques independents per cada punt, la mitjana de les quals es representa a la Figura 6. S'ha de considerar que en aquest cas el procediment utilitzat és el mateix que el ja descrit.

Anàlisi 4: Estudi d'haplotipus diferents dependent de les seqüències anteriors

La principal diferència respecte l'anàlisi 3 és que en aquesta les seqüències amb què es treballa inclouen les seqüències anteriors a cada punt d'aquesta anàlisi, similar al procediment de l'anàlisi 2. Això permet, respecte l'anàlisi anterior, veure com van apareixent els haplotipus a mesura que s'analitzen més individus d'una mateixa població de *T. thynnus* en un mateix estudi. A més, com en el cas anterior, aquesta anàlisi també permet fer un seguiment de com apareixen haplotipus diferents a mesura que s'analitzen més individus d'una població, ja que es representen els haplotipus diferents que van sorgint a mesura que s'avança en l'anàlisi.

L'anàlisi 4 es basa en l'estudi de les 657 seqüències en blocs múltiples de 25 i de forma acumulativa. Primer s'han escollit 25 seqüències no repetides i aleatòries i se n'ha determinat l'haplotipus. El recompte d'haplotipus diferents que ha aparegut s'han representat. Seguidament s'han seleccionat 25 seqüències aleatòries, no repetides i diferents a les anteriors que s'han afegit a les 25 amb què ja s'ha treballat. Amb això s'han aconseguit 50 seqüències de les quals se n'ha determinat l'haplotipus. Els haplotipus als que corresponen s'han comparat amb els anteriors, fent la diferència entre els haplotipus diferents ara apareguts i els anteriors. Així s'ha obtingut el nombre d'haplotipus diferents que sorgeixen respecte els haplotipus corresponents a les seqüències anteriors ja analitzades. Després s'ha repetit el procés: s'han generat 25 seqüències aleatòries, no repetides i diferents a les 50 anteriors, i juntament amb aquestes 50 seqüències corresponen a l'anàlisi de 75 seqüències. Amb aquest bloc de dades s'ha repetit el procés ara descrit, i successivament fins a analitzar la totalitat de les seqüències.

En aquesta anàlisi l'estratègia que s'ha seguit es basa en l'estudi de com apareixen els diferents haplotipus a mesura que s'afegeixen individus a l'anàlisi. Per tant es tenen en compte les seqüències anteriors ja analitzades.

Com en tots els casos, també s'han fet 5 rèpliques independents per cada punt, i s'ha representat la mitjana de cada punt en la Figura 7.

Aquesta anàlisi es pot resumir com una representació diferent de les dades que s'obtenen en fer l'anàlisi 2. En la segona anàlisi es representa el total d'haplotipus que van apareixent a mesura que s'afegeixen seqüències per estudiar i en aquest cas s'ha representat la diferència entre els nous haplotipus que apareixen i els anteriors.

RESULTATS

S'ha treballat amb un volum de 657 seqüències. Aquestes han estat escollides dins del volum total de seqüències amb què es disposava inicialment –format per les 689 disponibles al LIG i també per una selecció de seqüències de Genbank– degut a que s'ha fet un filtratge per eliminar les seqüències que provenen de la introgressió mitocondrial entre espècies de *T. thynnus* (Alvarado Bremer et al., 2005). El nombre de seqüències final amb què s'ha treballat, però, no és un valor tancat. S'ha intentat integrar seqüències disponibles a la base de dades Genbank a les anàlisis mitjançant la cerca booleana per paraules clau. Tot i això en intentar fer-ho s'ha vist que l'alineament no era del tot correcte degut a les diferències entre les seqüències, que s'han generat en laboratoris diferents. Aquest fet provoca problemes a l'hora d'inferir els haplotipus. Per això l'opció d'afegir seqüències de Genbank ha quedat descartada. Donada aquesta dificultat, també s'ha valorat utilitzar BLAST (Altschul 1990) fent servir com a referència les seqüències disponibles al LIG. Utilitzant aquesta eina s'eliminarien els problemes d'alineament, ja que les seqüències trobades serien més semblants. Tot i això per manca de temps no s'ha pogut arribar a incorporar noves seqüències a les anàlisis.

Arbre filogenètic: detecció d'haplotipus

Amb l'alineament de les seqüències s'ha construït l'arbre filogenètic (Figura 3). Gràcies a l'arbre resultant s'ha determinat quines seqüències comparteixen haplotipus a partir de l'observació de com es situen en l'arbre: es consideren iguals aquelles seqüències amb una distància genètica –nombre de posicions variables– de zero, i per tant formen un sol haplotipus. Les seqüències que possiblement conformen un haplotipus s'han comprovat mitjançant una alineament dos a dos. En canvi les que es troben en branques separades representen haplotipus diferents. Per exemple observant la Figura 3 es veu clarament com les seqüències anomenades 258 i 262 són un únic haplotipus, mentre que tota la resta constitueixen haplotipus diferents.

S'ha determinat que les 657 seqüències amb què es treballa constitueixen un total de 354 haplotipus diferents, dels quals 262 (74%) són únics per una seqüència.

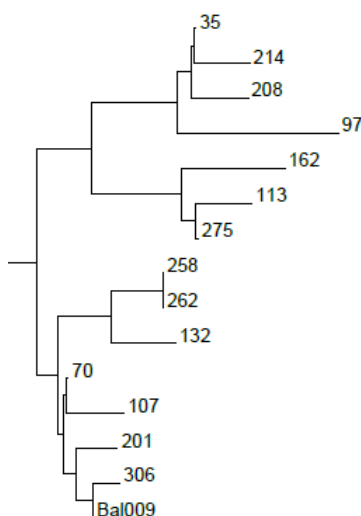


Figura 3. Fragment de l'arbre filogenètic obtingut a partir de l'alineament de les seqüències. Cada punta de l'arbre correspon a una seqüència. En aquest cas es pot veure com les seqüències anomenades 258 i 262 comparteixen el mateix haplotipus.

Anàlisi 1: Estudi d'haplotipus independent de les seqüències anteriors

En els resultats obtinguts en l'anàlisi 1 (Figura 4) s'observa com en la sèrie de dades corresponent als haplotipus com més gran és el nombre de seqüències que s'estudien el recompte d'haplotipus va augmentant gradualment, sense col·lapsar en cap moment. És a dir, com més elevat és el volum de seqüències que s'analitzen més augmenta el nombre d'haplotipus que representen i fins que no s'arriba a l'anàlisi complet de les seqüències no s'observen els haplotipus totals.

Els valors entre els quals oscil·la la diversitat haplotípica es troben entre 0.969 i 0.983. Tot i que es poden observar petites irregularitats entre la tendència que marquen els punts, es veu que aquest estadístic tendeix a disminuir a mesura que augmenta el nombre de seqüències estudiades.

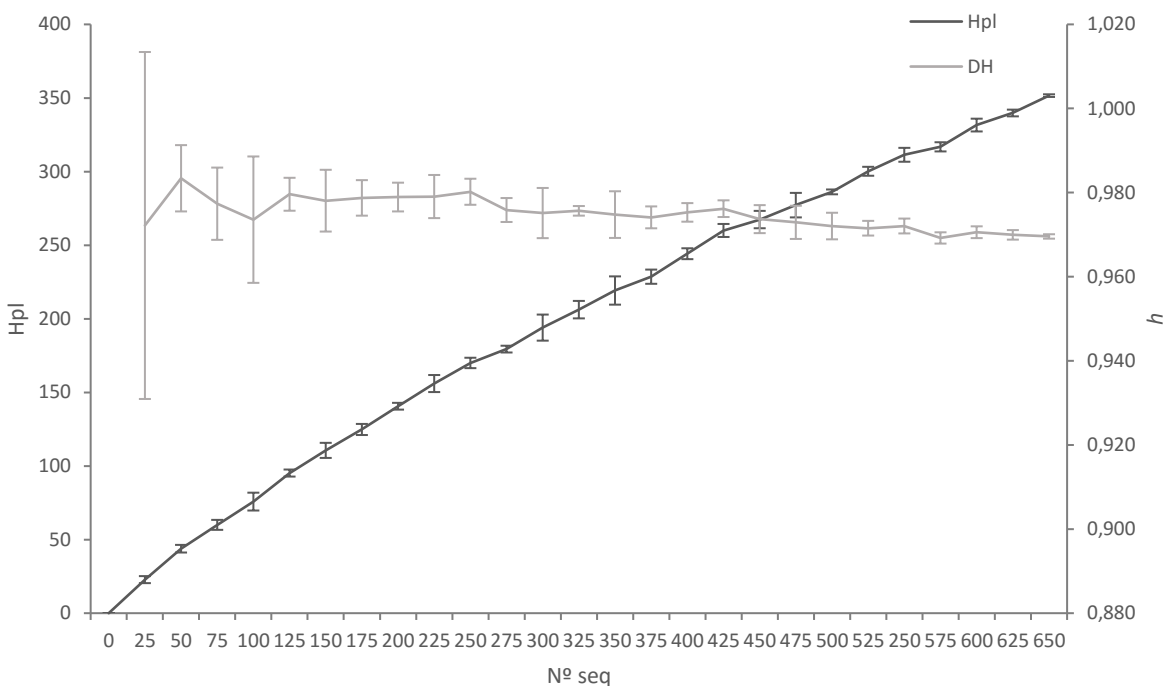


Figura 4. Anàlisi 1: Estudi d'haplotipus independent de les seqüències anteriors. Es representa amb un color més fosc el nombre total d'haplotipus (Hpl) que apareixen en funció del del nombre de seqüències (Nº seq) que s'analitzen. De color més clar hi ha resrepresentades les diversitats haplotípiques (h) per cada punt.

Anàlisi 2: Estudi d'haplotipus dependent de les seqüències anteriors

En l'anàlisi 2 (Figura 5) també s'observa com el recompte d'haplotipus totals augmenta a mesura que va incrementant el nombre de seqüències analitzades. Cal tenir en compte que, respecte l'anàlisi 1, en aquest cas l'augment és més progressiu i regular, sense alts i baixos. També que, de la mateixa manera que en la Figura 4, fins que no s'analitzen la totalitat de les seqüències no apareixen tots els haplotipus amb què es treballa.

En referència a la diversitat haplotípica en aquest cas igualment disminueix. Ho fa gradualment i sense alts i baixos entre els punts de la gràfica. Els valors entre els quals oscil·la aquest estadístic van de 0.954 a 0.980.

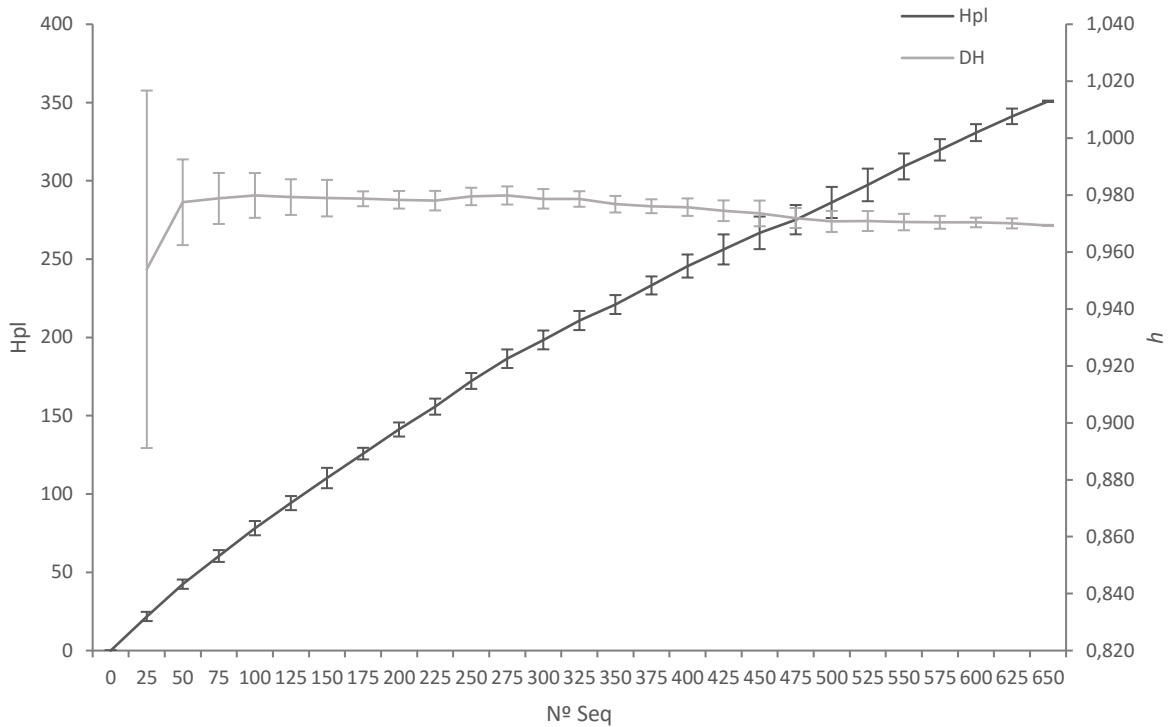


Figura 5. Anàlisi 2: Estudi d'haplotipus dependent de les seqüències anteriors. Es representa amb un color més fosc el nombre total d'haplotipus (Hpl) que apareixen en funció del del nombre de seqüències (Nº seq) que s'analitzen. De color més clar hi ha respresentades les diversitats haplotípiques (h) per cada punt.

Anàlisi 3: Estudi d'haplotipus diferents independent de les seqüències anteriors

En els resultats de l'anàlisi 3 (Figura 6) s'observa com inicialment hi ha un augment del nombre d'haplotipus diferents que apareixen respecte els haplotipus ja representats. A partir de les 100 seqüències el nombre d'haplotipus diferents que encara no han aparegut en l'anàlisi va disminuint, fins al punt en què quan es treballa amb més de 450 seqüències el gràfic col·lapsa, ja ha aparegut com a mínim una vegada la totalitat dels haplotipus. Per tant a partir d'aquest punt no apareixen més haplotipus diferents als anteriors i així el recompte d'haplotipus diferents es manté constant a 0.

Per altra banda en la Figura 6 també s'observa la tendència que té la diversitat haplotípica. En aquest cas l'estadístic és molt irregular, la qual cosa s'observa en les variacions que es manifesten entre els punts consecutius del gràfic. Tot i això es pot afirmar que la tendència que mostra la diversitat haplotípica és disminuir a mesura que es treballa amb un nombre de seqüències més elevat. Els valors entre els quals fluctua es troben entre 0.969 i 0.989.

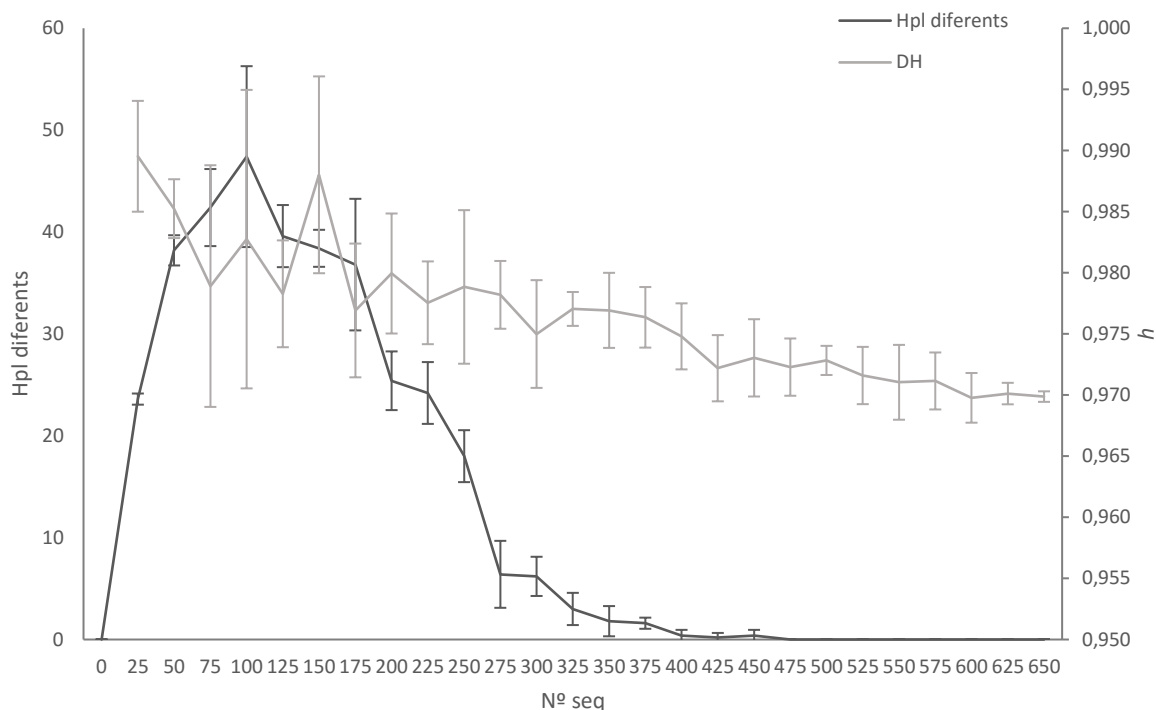


Figura 6. Anàlisi 3: Estudi d'haplotipus diferents independent de les seqüències anteriors. Es representa amb un color més fosc el nombre total d'haplotipus diferents als anteriors ja representats (Hpl diferents) que apareixen en funció del nombre de seqüències (Nº seq) que s'analitzen. De color més clar hi ha representades les diversitats haplotípiques (h) per cada punt.

Anàlisi 4: Estudi d'haplotipus diferents dependent de les seqüències anteriors

En l'anàlisi 4 (Figura 7), al contrari que en l'anàlisi anterior, es pot observar com el nombre d'haplotipus que encara no han aparegut en l'anàlisi no col·lapsa. És a dir, fins que no s'estudien les 657 seqüències no apareixen tots els 354 haplotipus diferents. Es pot observar que la tendència general amb què apareixen haplotipus nous va disminuint a mesura que s'afegeixen més seqüències a l'anàlisi, tot i que els punts consecutius mostren variacions entre ells fent que els haplotipus nous apareguin de forma irregular.

Pel que fa referència a la diversitat haplotípica, en aquesta anàlisi oscil·la entre 0.969 i 0.987. Com en els casos anteriors, la diversitat igualment tendeix a disminuir a mesura que s'afegeixen més seqüències a l'anàlisi.

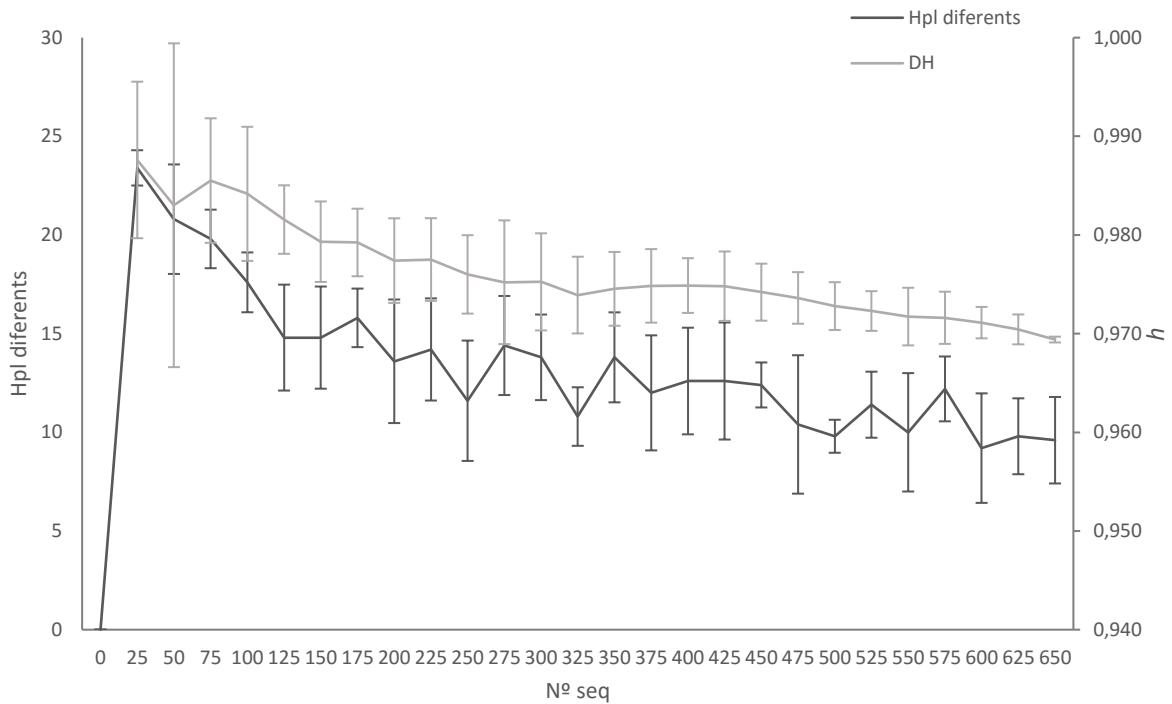


Figura 7. Anàlisi 4: Estudi d'haplotipus diferents dependent de les seqüències anteriors. Es representa amb un color més fosc el nombre total d'haplotipus diferents als anteriors ja representats (Hpl diferents) que apareixen en funció del del nombre de seqüències (Nº seq) que s'analitzen. De color més clar hi ha resrepresentades les diversitats haplotípiques (h) per cada punt.

DISCUSSIÓ

En el cas del primer anàlisi s'evidencia quants haplotipus s'espera trobar si s'analitzen aleatòriament un nombre concret de mostres de *T. thynnus*. En canvi l'anàlisi 2 no només mostra això, sinó que també permet veure quants haplotipus apareixerien en un mateix mostratge depenent dels individus que aquest contingui, és a dir, en el segon anàlisi s'ensenyava l'evolució de com augmenta el nombre d'haplotipus depenent del nombre d'individus que s'analitzin a mesura que es van afegint seqüències dins d'un mateix mostratge. Aquest fet es dona gràcies a la dependència que hi ha entre les seqüències anteriors i les noves que es van afegint a l'anàlisi.

Analitzant les Figures 4 i 5, corresponents a la primera i segona anàlisi, es pot extreure que si es treballa amb una mostra que contingui un nombre inferior a 657 individus aquesta mostra no és representativa de tota la població de *T. thynnus* que es vulgui estudiar, ja que no inclourà tots els haplotipus presents en la població. Aquest fet es pot afirmar degut a que el nombre d'haplotipus que apareixen en cap cas es col·lapsa, sinó que sempre augmenta. A més encara que s'utilitzin 657 individus per fer les anàlisis és probable que encara no s'incloguin en els estudis la totalitat dels haplotipus presents a la població, degut a que en el submostratge inicial s'ha vist que com més seqüències s'afegeixen el nombre d'haplotipus diferents total augmenta, fent que el nombre d'haplotipus no sigui un paràmetre tancat de moment.

Mentre que l'anàlisi 1 només mostra el nombre d'haplotipus que hi ha en un nombre de mostres determinat de *T. thynnus*, en l'anàlisi 3 es fa una aproximació de com apareixen els diferents haplotipus en base a un sistema de mostratge de les seqüències igual al mètode 1. Gràcies a això s'observa com en l'anàlisi 1 el nombre d'haplotipus no deixa d'augmentar a mesura que el volum de seqüències aleatòries també augmenta (Figura 4), però l'anàlisi 3 demostra que no apareixerien haplotipus nous a partir de les 450 seqüències analitzades aproximadament (Figura 6). És a dir, en el recorregut de les dues anàlisis deixen d'aparèixer haplotipus nous a partir de les 450 seqüències estudiades, però tot i això en un sol punt de mostratge no es trobarien la totalitat dels haplotipus fins que no s'analitzessin totes les seqüències.

Si l'observació dels resultats es centrés només en l'anàlisi 3 es podria extreure una conclusió errònia, ja que es deduiria que a partir de les 450 seqüències analitzades en un estudi referent a *T. thynnus* no apareixen més haplotipus diferents. Això és un error degut a com s'han tractat les dades perquè s'ha vist que el nombre total d'haplotipus no col·lapsa en cap moment. Per aquest motiu observar només els resultats de l'anàlisi 3 no és correcte, sinó que calen utilitzar els resultats que proporciona per complementar la informació obtinguda a partir de l'anàlisi 1. Per altra banda si s'observa només l'anàlisi 1 es podria arribar a la conclusió de que en aquest anàlisi no apareixen tots els haplotipus diferents fins que no s'analitzen la totalitat de les seqüències. Això també és un error degut al tractament de les dades, el qual queda demostrat mirant l'anàlisi 3.

Per tal d'eliminar la possibilitat d'extreure conclusions errònies observant només un gràfic s'ha fet posteriorment l'anàlisi 4, on es representa el mateix que en l'anàlisi 3 però amb un tractament de les dades diferent. En la Figura 7 s'observa com apareixen haplotipus diferents durant tot un mateix estudi, la qual cosa indica com apareixen els haplotipus a mesura que es van afegint seqüències noves a les que ja s'ha analitzat. D'aquesta manera s'evidencia l'error de l'anàlisi 3 i s'elimina –comparant les Figures 6 i 7 es veu que són molt diferents–, ja que en aquest cas es té en

compte tot el recorregut del gràfic, és a dir els punts anteriors, a l'hora de representar un nou valor del gràfic.

Referent a l'anàlisi 4 es pot dir que és complementari a l'anàlisi 2, ja que utilitza el mateix sistema de mostratge. En el primer cas es representa la diferència entre els haplotipus d'un determinat nombre de seqüències respecte les anteriors –tenint en compte el correcte tractament de les dades diferent a l'anàlisi 3– (Figura 7), i en canvi en el segon es representa directament el nombre d'haplotipus totals –seguint el mateix tractament de les dades– (Figura 5). Per aquest motiu es pot dir que l'anàlisi 4 és addicional al segon anàlisi, ja que representen el mateix resultat de formes diferents.

Tenint en compte tots aquest arguments es pot concloure que el millor anàlisi que s'ajusta als objectius és l'anàlisi 2 (Figura 5) complementat amb els resultats de l'anàlisi 4 (Figura 7).

Un cop atesa aquesta conclusió, es pot afirmar que quan es treballa amb 657 seqüències encara no s'ha assolit una representació de la variabilitat de la població de *T. thynnus*, ja que van apareixent nous haplotipus. Aquest resultat es pot veure en la Figura 5, on no col·lapsa el nombre d'haplotipus en cap moment i s'ha vist que quan s'afegeixen més seqüències al gràfic el nombre d'haplotipus continua augmentant. Per tant es pot dir que encara que en un estudi referent a *T. thynnus* s'analitzessin 657 mostres aquest estudi encara no contemplaria tota la diversitat de la població.

Per altra banda cal observar la informació que proporciona la diversitat haplotípica (h). Com més s'aproxima a 1 aquest estadístic més difícil serà que el nombre d'haplotipus arribi a col·lapsar. Això es deu a que la probabilitat de que dos haplotipus siguin diferents és més elevada com més gran és aquest estadístic. Com s'ha vist en totes les anàlisis en augmentar el volum de seqüències analitzades el valor de h disminueix, indicant que com més individus formin un mostreig el nombre d'haplotipus diferents que inclogui serà menor. Això suggereix que si s'ampliessin les anàlisis la diversitat haplotípica cada vegada seria menor, fins arribar al punt en què no apareixerien més haplotipus diferents encara que s'analitzin més seqüències de *T. thynnus*. Tot i això no es coneix fins a quin punt s'haurien d'ampliar les anàlisis.

A més, tenir una diversitat haplotípica molt elevada té una clara relació amb la grandària efectiva poblacional. El que segurament indica l'alt nivell de variabilitat que s'observa és que la grandària efectiva poblacional és molt alta i, per tant, no sembla una espècie en declivi. Tot i això es necessitaria dur a terme noves anàlisis amb marcadors més adequats –microsatèl·lits, SNPs– per determinar aquest fet.

Així doncs una futura direcció seria ampliar el nombre de seqüències amb què es treballa, amb això s'aconseguiria poder determinar el nombre d'individus totals que cal analitzar en estudis referents a *T. thynnus* perquè les mostres siguin representatives. Com s'ha dit inicialment es va treballar amb un submostratge de seqüències, els resultats del qual indicaven que les anàlisis podien respondre als objectius plantejats però de moment eren insuficients, ja que tampoc col·lapsaven. En veure aquests resultats es van repetir les anàlisis amb la totalitat de seqüències disponibles. És important començar les anàlisis de nou quan s'afegeixen seqüències noves, ja que si només s'ampliessin les obtingudes anteriorment no hi hauria un correcte tractament de les dades, es perdria l'aleatorietat. Per tant seria necessari repetir com a mínim l'anàlisi 2, ja que s'ha

determinat que és el millor, amb les seqüències ja estudiades i afegint-ne de noves, comparant els haplotipus amb les seqüències ja estudiades.

Per afegir seqüències noves s'haurien de cercar utilitzant BLAST (Altschul 1990). També hi ha l'opció de fer una cerca booleana a Genbank, però en aquest cas s'ha vist que els alineaments obtinguts no són del tot correctes degut a les diferències entre les seqüències i, per tant, caldria invertir molt més temps buscant seqüències que després podria ser que no fossin útils per aquest anàlisi. En canvi si l'eina escollida és BLAST (Altschul 1990) i s'utilitzen com a referència les seqüències ja disponibles es pot assegurar que les noves seqüències que s'obtidran seran semblants a les model i, per tant, és probable que s'obtingui un bon alineament i que el temps invertit en la cerca sigui menor que si s'emprés Genbank. Cal tenir en compte que les noves seqüències que s'afegissin a l'anàlisi haurien de complir uns requisits: que vinguin d'una font fiable, que estiguin citades en articles científics, que pertanyin a individus, que no representin haplotipus i sobretot que no siguin les seqüències que ja s'han analitzat amb anterioritat provinents del LIG i publicades per Viñas et al. (2005). També caldria comprovar el bon alineament de les seqüències.

El fet que el nombre d'haplotipus nous no col·lapsi és important a l'hora de realitzar mostres per futurs estudis, ja que ni amb 657 seqüències es veu representada la diversitat de la població i, per tant, s'han de mostrejar més individus –encara no se sap quants– per tenir la representació total de la variabilitat. Així en els futurs estudis poblacionals idealment es necessitarien més de 657 individus per cada punt, sense conèixer el límit, per poder tenir representada la totalitat de la variabilitat, la qual cosa fa que les anàlisis i la realització dels propers estudis de *T. thynnus* sigui molt complicada.

També convé considerar que s'ha vist que la majoria d'haplotipus (74%) corresponen únicament a una seqüència, per tant si s'ampliés l'anàlisi també es podria determinar si la freqüència d'aparició d'haplotipus nous continua augmentant –el que ha passat fins el moment– o bé arriba un punt on disminueix i fins i tot és nul·la. Donat aquest punt s'esperaria veure com mica en mica l'anàlisi amb què es treballa col·lapsa, i per tant es podria arribar a trobar el punt en què per més individus que s'estudiessin el nombre d'haplotipus no augmentaria. També caldria observar l'evolució de la diversitat haplotípica, ja que si el nombre d'haplotipus diferents que apareixen descendeix també ho farà aquest estadístic.

Per altra banda, cal tenir en compte que els anàlisis que s'han realitzat i els resultats que s'han obtingut es poden extrapolar a altres espècies. Per fer-ho caldria comparar les diversitats haplotípiques i, en el cas que aquestes fossin semblants, s'esperaria trobar uns resultats similars per a les mateixes anàlisis. Per exemple Hammer et al. (1999) van calcular la diversitat haplotípica del cromosoma Y en humans. Els resultats mostraven que en la població africana $h=0.841\pm 0.001$ quan s'estudien 360 individus (n), en la població europea $h=0.852\pm 0.000$ quan $n=507$ i en la població asiàtica $h=0.904\pm 0.000$ quan $n=1415$. En aquest cas les diversitats haplotípiques per al cromosoma Y són més baixes que les que s'han obtingut en l'estudi del mtDNA CR de *T. thynnus*, tot i que el nombre d'individus que formen les mostres és més petit que amb el nombre d'individus amb què s'ha treballat en aquest cas (657). Això indica que probablement quan es facin estudis referents al cromosoma Y en humans siguin necessaris menys individus per mostra respecte a aquelles estudis referents al mtDNA CR de *T. thynnus*.

CONCLUSIONS

1. L'estudi de la regió control del DNA mitocondrial en *Thunnus Thynnus* és una bona eina per dur a terme estudis sobre les poblacions d'aquesta espècie i així procurar una bona gestió i conservació de les poblacions.
2. Perquè una mostra sigui representativa de la població de *T. thynnus* aquesta ha de comptar amb un nombre superior a 657 individus, ja que fins que no s'analitzen tots els individus disponibles no apareixen tots els haplotipus diferents.
3. Seria necessari ampliar les anàlisis, o com a mínim l'anàlisi 2 i opcionalment el 4, per determinar quin és el nombre d'individus que cal mostrejar per representar tota la població, començant des de zero els anàlisis per comptar amb un correcte tractament de les dades. Fent això possiblement s'aconseguiria trobar el nombre mínim d'individus que cal analitzar per representar tota una població de *T. thynnus*.
4. En els futurs estudis referents a *T. thynnus* caldria mostrejar més de 657 individus per a cada punt per poder tenir una representació de tota la variabilitat.
5. Mitjançant els valors de la diversitat haplotípica es pot deduir que la grandària efectiva poblacional de *T. thynnus* és molt alta, la qual cosa sembla indicar que no es tracta d'una espècie en declivi.
6. Els resultats obtinguts per *T. thynnus* es poden extrapolar a altres anàlisis de *T. thynnus* o a altres espècies comparant els valors de la diversitat haplotípica.

CONCLUSIONS

1. Studying the *Thunnus Thynnus* mitochondrial DNA control region is a positive tool to perform population analysis of this species as well as to achieve a favourable management and conservation of the populations.
2. A sample has to contain more than 657 individuals to represent a population of *T. thynnus*, due to the fact that until all available individuals are analysed the overall different haplotypes doesn't appear.
3. It would be required to expand the analysis, or at least analysis 2 and optionally 4, in order to establish which number of individuals is needed in a sample to represent all the population. Analysis should be started from zero to achieve good data processing. By doing this, the minimum number of individuals that has to be analysed to represent overall *T. thynnus* population will probably be found.
4. In future studies related to *T. thynnus* sampling more than 657 individuals for each point would be needed in order to obtain an overall variability representation.
5. By using haplotype diversity values it can be deduced that the effective *T. thynnus* population size is very high, so this could indicate that this species doesn't decline.
6. Results obtained for *T. thynnus* can be extrapolated to other *T. thynnus* analysis or other species by comparing the haplotype diversity values.

AGRAÏMENTS

Agraeixo al Dr. Jordi Viñas De Puig el seu suport i per guiar-me durant la realització de les anàlisis dutes a terme, així com per les indicacions i coneixements transmesos a l'hora de la redacció d'aquesta memòria.

BIBLIOGRAFIA

- Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
- Alvarado Bremer, J. R., Viñas, J., Mejuto, J., Ely, B., & Pla, C. (2005). Comparative phylogeography of Atlantic bluefin tuna and swordfish: The combined effects of vicariance, secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. *Molecular Phylogenetics and Evolution*, 36(1), 169–187. <http://doi.org/10.1016/j.ympev.2004.12.011>
- Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., ... Saunders, N. C. (1987). Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annual Review of Ecology and Systematics*, 18(1), 489–522. <http://doi.org/10.1146/annurev.es.18.110187.002421>
- Boustany, A. M., Reeb, C. A., & Block, B. A. (2008). Mitochondrial DNA and electronic tracking reveal population structure of Atlantic bluefin tuna (*Thunnus thynnus*). *Marine Biology*, 156(1), 13–24. <http://doi.org/10.1007/s00227-008-1058-0>
- Carlsson, J., McDowell, J. R., Carlsson, J. E. L., & Graves, J. E. (2007). Genetic identity of YOY bluefin tuna from the eastern and western Atlantic spawning areas. *Journal of Heredity*, 98(1), 23–28. <http://doi.org/10.1093/jhered/esl046>
- Carlsson, J., McDowell, J. R., Díaz-Jaimes, P., Carlsson, J. E. L., Boles, S. B., Gold, J. R., & Graves, J. E. (2004). Microsatellite and mitochondrial DNA analyses of Atlantic bluefin tuna (*Thunnus thynnus thynnus*) population structure in the Mediterranean Sea. *Molecular Ecology*, 13(11), 3345–3356. <http://doi.org/10.1111/j.1365-294X.2004.02336.x>
- Cloutier, O., Cloutier, R., Libert, E., Ortega, A., & de la Gándara, F. (2013). Condrogénesis y malformaciones craneales de larvas y juveniles de atun rojo (*Thunnus thynnus*) criados en cautividad: identificación de una fuente importante de mortalidad de las larvas. *Proceedings of the XIV Spanish Aquaculture Congress (Congreso Nacional de Acuicultura) Gijón (Spain), 24-25 September 2013*, (June 2016), 166–167.
- Collette, B., Amorim, A. F., Boustany, A., Carpenter, K.E., D., Oliveira Leite Jr., N., Di Natale, A., ... Uozumi, Y. (2011). *Thunnus thynnus*. The IUCN Red List of Threatened Species, 8235. Retrieved from http://bioweb.uwlax.edu/bio203/s2009/schwobe_dust/Geography.htm
- Di Natale, A. (2015). REVIEW OF THE HISTORICAL AND BIOLOGICAL EVIDENCES ABOUT A POPULATION OF BLUEFIN TUNA (*THUNNUS THYNNUS*) IN THE EASTERN MEDITERRANEAN AND THE BLACK SEA, 71(3), 1098–1124.
- Domínguez, E. F. (2000). *Introducción: El ADN mitocondrial*. Universitat de Barcelona.
- Ely, B., Stoner, D. S., Alvarado Bremer, J. R., Dean, J. M., Addis, P., Cau, A., ... Quattro, J. M. (2002). Analyses of nuclear *IdhA* gene and mtDNA control region sequences of Atlantic northern bluefin tuna populations. *Marine Biotechnology*, 4(6), 583–588. <http://doi.org/10.1007/s10126-002-0040-y>

- European Commission - Fisheries (2016). *Bluefin tuna (Thunnus thynnus)* [en línia]. <http://ec.europa.eu/fisheries/marine_species/wild_species/bluefin_tuna/index_en.htm> [Consulta: 17 juliol 2016]
- Excoffier, L. and, & Lischer, H. E. L. (2015). Arlequin Ver 3.5.2 User Manual, 176.
- Felsenstein, Joseph. Discussion: Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *Ann. Statist.* 14 (1986), no. 4, 1304--1305. doi:10.1214/aos/1176350146. <http://projecteuclid.org/euclid.aos/1176350146>.
- Fromentin, J. M. (2006). Manual de ICCAT. Capítulo 2.1.5. El atún rojo. *Manual de ICCAT*, 99–118.
- Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41:95-98.
- Hammer, M. F., Karafet, T. M., Redd, A. J., Jarjanazi, H., Santachiara-benerecetti, S., Soodyall, H., & Zegura, S. L. (1999). Hierarchical Patterns of Global Human Y-Chromosome Diversity, (Stoneking 1998), 1189–1203. <http://doi.org/10.1093/oxfordjournals.molbev.a003906>
- ICCAT. (2015). Report of the 2014 Atlantic bluefin tuna stock assessment session. *Collective Volume of Scientific Papers ICCAT*, 71(2), 692–945.
- Koichiro Tamura, Glen Stecher, Daniel Peterson, Alan Filipski, and Sudhir Kumar (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*:30 2725-2729.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG *Bioinformatics* (2007) ClustalW and ClustalX version. 2 2007 23(21): 2947-2948. doi:10.1093/bioinformatics/btm404
- Manchado, M., Catanese, G., & Infante, C. (2004). Complete mitochondrial DNA sequence of the Atlantic bluefin tuna *Thunnus thynnus*. *Fisheries Science*, 70(1), 68–73. <http://doi.org/10.1111/j.1444-2906.2003.00772.x>
- Riccioni, G., Landi, M., Ferrara, G., Milano, I., Cariani, a., Zane, L., ... Tinti, F. (2010). Spatio-Temporal Genetic Patterns in Mediterranean Bluefin Tuna: Population Structuring and Retention of Genetic Diversity. *Collect. Vol. Sci. Pap. ICCAT*, 65(3), 996–1003.
- Riccioni, G., Landi, M., Ferrara, G., Milano, I., Cariani, A., Zane, L., ... Tinti, F. (2010). Spatio-temporal population structuring and genetic diversity retention in depleted Atlantic bluefin tuna of the Mediterranean Sea. *Proceedings of the National Academy of Sciences of the United States of America*, 107(5), 2102–2107. <http://doi.org/10.1073/pnas.0908281107>
- Rooker, J. R., Alvarado Bremer, J. R., Block, B. A., Dewar, H., De Metrio, G., Corriero, A., ... Secor, D. H. (2007). *Life History and Stock Structure of Atlantic Bluefin Tuna (Thunnus thynnus)*. *Reviews in Fisheries Science* (Vol. 15). <http://doi.org/10.1080/10641260701484135>
- Viñas, J., Gordo, A., Fernández-Cebrián, R., Pla, C., Vahdet, Ünal, & Araguas, R. M. (2011). Facts and uncertainties about the genetic population structure of Atlantic bluefin tuna (*Thunnus thynnus*) in the Mediterranean. Implications for fishery management. *Reviews in Fish Biology and Fisheries*, 21(3), 527–541. <http://doi.org/10.1007/s11160-010-9174-6>