



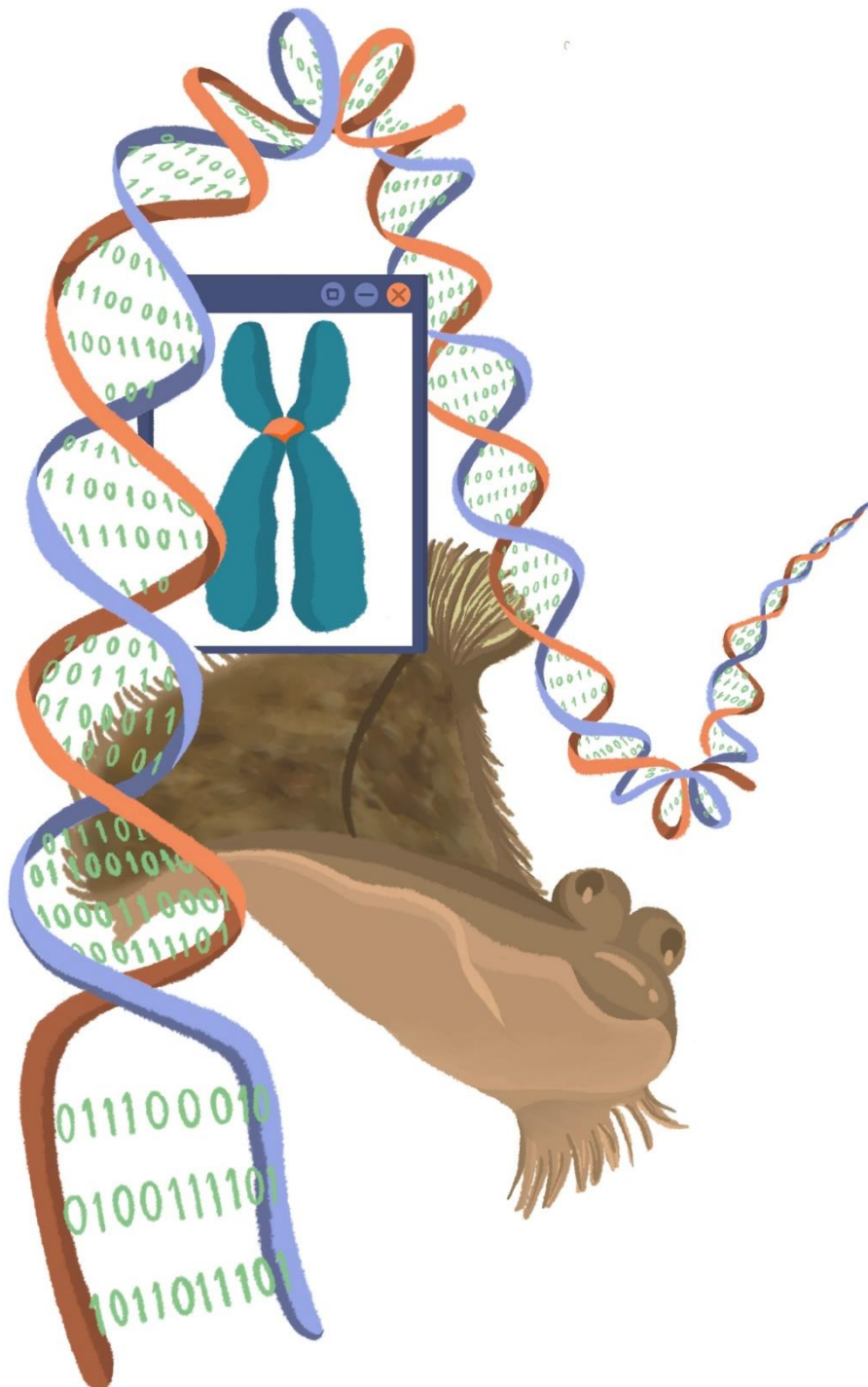
UNIVERSIDAD  
DE GRANADA

Facultad de Ciencias



Máster en Genética y Evolución

# Caracterización de las secuencias repetidas y de las regiones centroméricas del lengüado senegalés (*Solea senegalensis*, Kaup 1858)



## Resumen:

El lenguado senegalés, *Solea senegalensis*, es un pez plano perteneciente al orden Pleuronectiforme. Este grupo de peces presenta un creciente interés comercial, por su potencial en acuicultura, y han sido tradicionalmente foco de atención de múltiples investigaciones por la singular pérdida de simetría corporal que presenta. Recientemente, se ha secuenciado el genoma del lenguado senegalés. El estudio de este y otros genomas de especies emparentadas brinda una oportunidad única para caracterizar y conocer en profundidad la organización de sus genomas. En este estudio se han caracterizado los elementos repetidos del genoma de *S. senegalensis*. Nuestro análisis demuestra que este genoma contiene un 32,98% de elementos repetidos, siendo los más comunes los transposones de ADN (11,76%), las LINEs (7,75%), los elementos LTR (7,34%) y las secuencias de ADN satélite (0,91%). Utilizando estas secuencias repetidas hemos caracterizado las regiones centroméricas de 18 de las 21 parejas de esta especie. Nuestros datos demuestran que los centrómeros de esta especie están compuestos mayoritariamente por secuencias repetidas en tándem que, además, se encuentran asociadas a elementos móviles. En concreto: los centrómeros de los cromosomas 1 al 6, 9 al 13, 15 al 17, 20 y 21 están compuestos por la familia de ADN satélite *PvuII*, asociada a secuencias tipo LINE y Helitron, los centrómeros del par cromosómico 7 están ocupados por el ADN satélite rep87, asociado a una secuencia tipo Helitron y los centrómeros del par 8 están ocupados por el ADN satélite rep120, asociado a la secuencia repetida que incluye un ARN<sub>t</sub> descrita en este trabajo. Las posiciones de los centrómeros de los cromosomas 14, 18 y 19 permanecen inciertas. Los resultados obtenidos en este estudio aportan una base sólida para comprender la estructura y composición del genoma de *S. senegalensis*, lo cual puede ser de gran utilidad para mejorar la comprensión de la biología de esta especie y abordar los problemas pendientes en la producción en piscifactoría.

**Palabras clave:** *Solea senegalensis*, secuencias repetidas, ADN satélite, elementos móviles, centrómeros, bioinformática.

## Índice:

1.	Introducción .....	1
2.	Objetivos .....	8
3.	Material y Métodos.....	9
3.1.	Identificación de secuencias repetidas en el ensamblaje del genoma.....	9
3.2.	Identificación de secuencias repetidas en el lenguado senegalés y análisis comparativo entre sexos a partir de lecturas de Illumina .....	9
3.3.	Caracterización de las regiones centroméricas en el lenguado senegalés...	11
4.	Resultados .....	12
4.1.	Caracterización de elementos repetidos en <i>S. senegalensis</i> .....	12
4.1.1.	Resultados de la identificación de secuencias repetidas en el ensamblaje del genoma .....	12
4.1.2.	Resultados del análisis comparativo e identificación de secuencias repetidas a partir de lecturas de Illumina .....	15
4.2.	Resultados de la caracterización de las regiones centroméricas en el ensamblaje del genoma .....	22
5.	Discusión .....	28
6.	Conclusiones .....	34
7.	Perspectivas futuras.....	36
8.	Bibliografía .....	36
9.	Anexo.....	45

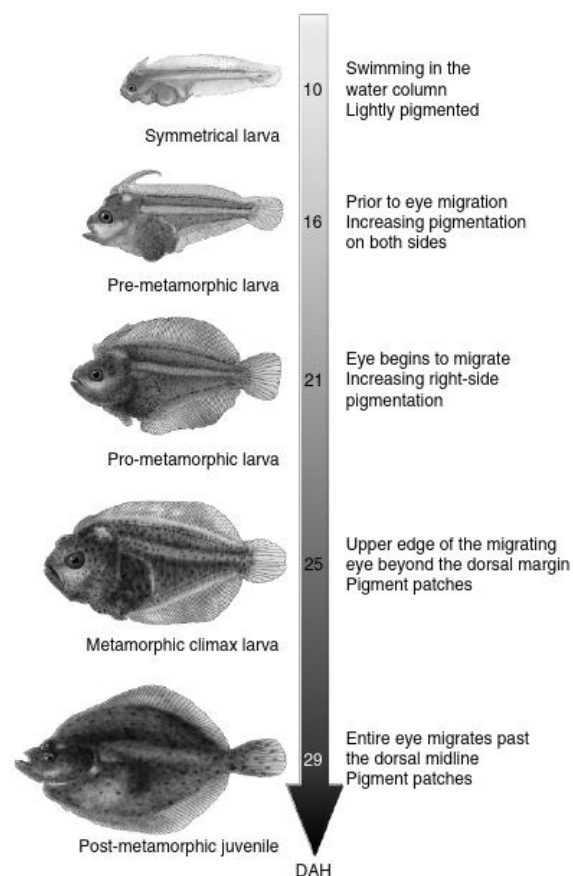
## 1. Introducción

### *Generalidades*

El lenguado senegalés (*Solea senegalensis*, Kaup 1858), es una especie perteneciente a la infraclase Teleostei y al orden Pleuronectiformes. Su distribución natural abarca desde el mar Mediterráneo hasta las costas africanas del Atlántico oriental. A nivel comercial, uno de los principales retos está en conseguir la producción intensiva de esta especie en piscifactorías (Morais et al., 2014). El cultivo de *S. senegalensis* podría ayudar a diversificar el mercado europeo de especies marinas, saturado por la producción de la dorada (*Sparus aurata*) y la lubina (*Dicentrarchus labrax*) (Cañavate, 2009; Chairi et al., 2010). Idealmente, para aumentar la producción y para no depender de las poblaciones naturales, el ciclo vital de *S. senegalensis* debería estar cerrado completamente en cautividad y evitar la dependencia de la disponibilidad de reproductores capturados del medio natural. Sin embargo, se han detectado problemas de fertilidad de los machos en la F<sub>1</sub> criados en cautividad. Este hecho se ha relacionado con determinados factores abióticos como el fotoperiodo y las estaciones (García-López et al., 2006). Además, se ha observado baja calidad de esperma en los machos de lenguado senegalés en cautividad, con una alta fragmentación del ADN y un alto porcentaje de células apoptóticas (Beirão et al., 2008). La baja producción de espermatozoides y la pobre capacidad de fertilización de los machos podría estar relacionada, también, con la regulación de proteínas implicadas en la actividad oxidoreductasa, el catabolismo proteico y el metabolismo lipídico, entre otros procesos (Forné et al., 2009). A pesar de estos avances y descubrimientos sobre la biología reproductiva de *S. senegalensis*, todavía las piscifactorías dependen de las capturas de individuos de las poblaciones naturales.

Además del valor económico de esta especie, también es destacable por el interés biológico que presenta el grupo al que pertenece. Los peces pleuronectiformes son uno de los grupos con mayor número de especies, se estima que existen unas 600 especies agrupadas en 11 familias. También se caracterizan por sufrir una metamorfosis única dentro del grupo de los vertebrados (**Figura 1**) en la que pierden

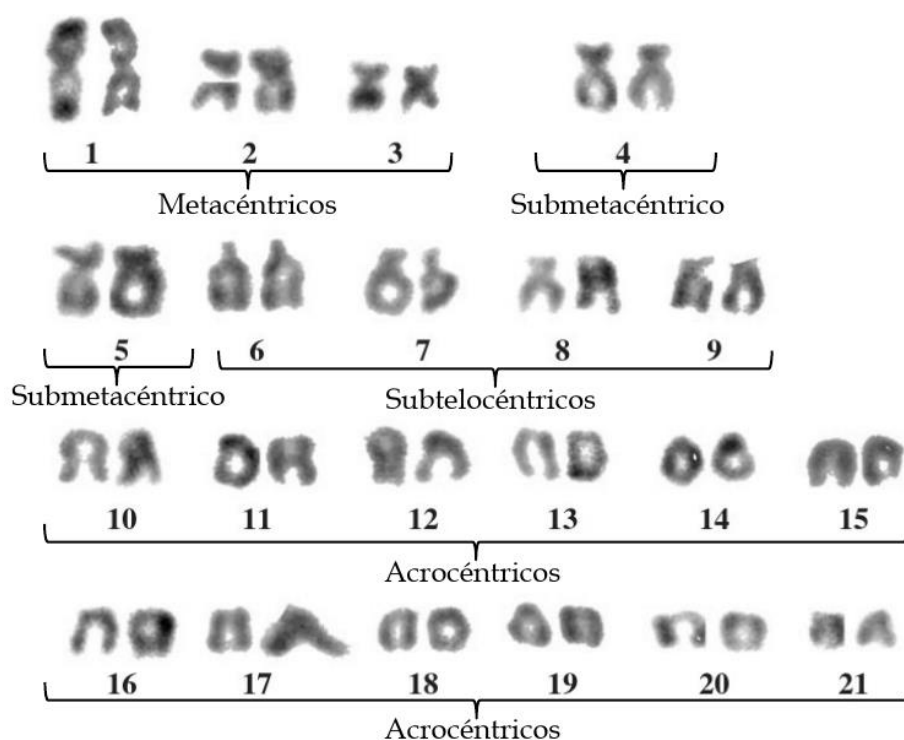
la simetría corporal durante el desarrollo de las larvas (Friedman, 2008). La migración de uno de los ojos, acompañada de cambios en la forma del cráneo, la pigmentación y la forma de las aletas, causa esta asimetría en los peces planos. Esto puede haber proporcionado ventajas adaptativas en términos de colonización y vida en entornos bentónicos. La capacidad de visión binocular mejorada que resulta de esta asimetría podría ser particularmente ventajosa para este grupo de especies, ya que son depredadores (Xing et al., 2020). Se alimentan principalmente de moluscos, poliquetos y pequeños invertebrados (Imshand et al., 2004). Aunque molecularmente no se conocen los mecanismos que producen la metamorfosis en los peces planos, se ha detectado mediante la comparación de genes de otros teleósteos que se ha producido una expansión de doce familias génicas relacionadas con la apoptosis celular, la regulación del tamaño del ojo, la neurotransmisión retiniana y la creación de un andamio de cartílago (Shao et al., 2017).



**Figura 1.** Infografía que muestra el proceso de metamorfosis en peces planos.  
Tomada del artículo Shao et al., 2017.

### Composición cromosómica del lenguado senegalés

Los teleósteos son un grupo de vertebrados que ha mostrado tener mucho éxito evolutivo y una gran diversidad (Ravi & Venkatesh, 2008). Además, se caracteriza por una duplicación propia del grupo producida hace unos 350 millones de años (Meyer & van de Peer, 2005) a parte de los dos eventos de duplicación del genoma completo durante la evolución de los vertebrados, el primero producido en la explosión cámbrica y el segundo en el Devónico temprano (Meyer & Schartl, 1999). Las duplicaciones de genomas completas producen grandes eventos de reorganización del material hereditario que, a su vez, tienen consecuencias como la pérdida de genes redundantes y la posibilidad de neofuncionalización de genes o incluso de rutas metabólicas completas (Jaillon et al., 2004). El genoma ancestral de los teleósteos está compuesto por  $2n=48$  cromosomas acrocéntricos (Ohno, 1970). *S. senegalensis* tiene un cariotipo de  $2n=42$  (Figura 2) formado por tres parejas de cromosomas metacéntricos, dos submetacéntricos, cuatro subtelocéntricos y doce acrocéntricos (Vega et al., 2002). Los estudios sobre la evolución y caracterización del cariotipo de *S. senegalensis* pueden ser muy interesantes por las peculiaridades expuestas anteriormente.



**Figura 2.** Cariotipo de *Solea senegalensis*. Tomada y modificada de Vega et al., 2002.

## *Determinación sexual en el lenguado senegalés*

El lenguado senegalés presenta un sistema de determinación del sexo XX/XY (Molina-Luzón et al., 2015). Aunque inicialmente se identificó el par metacéntrico más largo como candidato a ser el par sexual debido a la presencia única de un elemento móvil de la subfamilia Rex/Babar (García et al., 2019), estudios más recientes sobre el ensamblaje completo del genoma han arrojado nueva luz sobre el tema. En concreto, gracias a estos nuevos análisis, se ha propuesto al gen *fshr*, el cual codifica para la proteína del receptor de la hormona estimulante del folículo, como gen maestro en la determinación del sexo en esta especie (de la Herrán et al., 2023). Este gen está presente en el par cromosómico 12. Esto supone, además, un avance técnico del cultivo de esta especie en cautividad, ya que el gen *fshr* podría utilizarse como método de identificación temprana del sexo de los individuos.

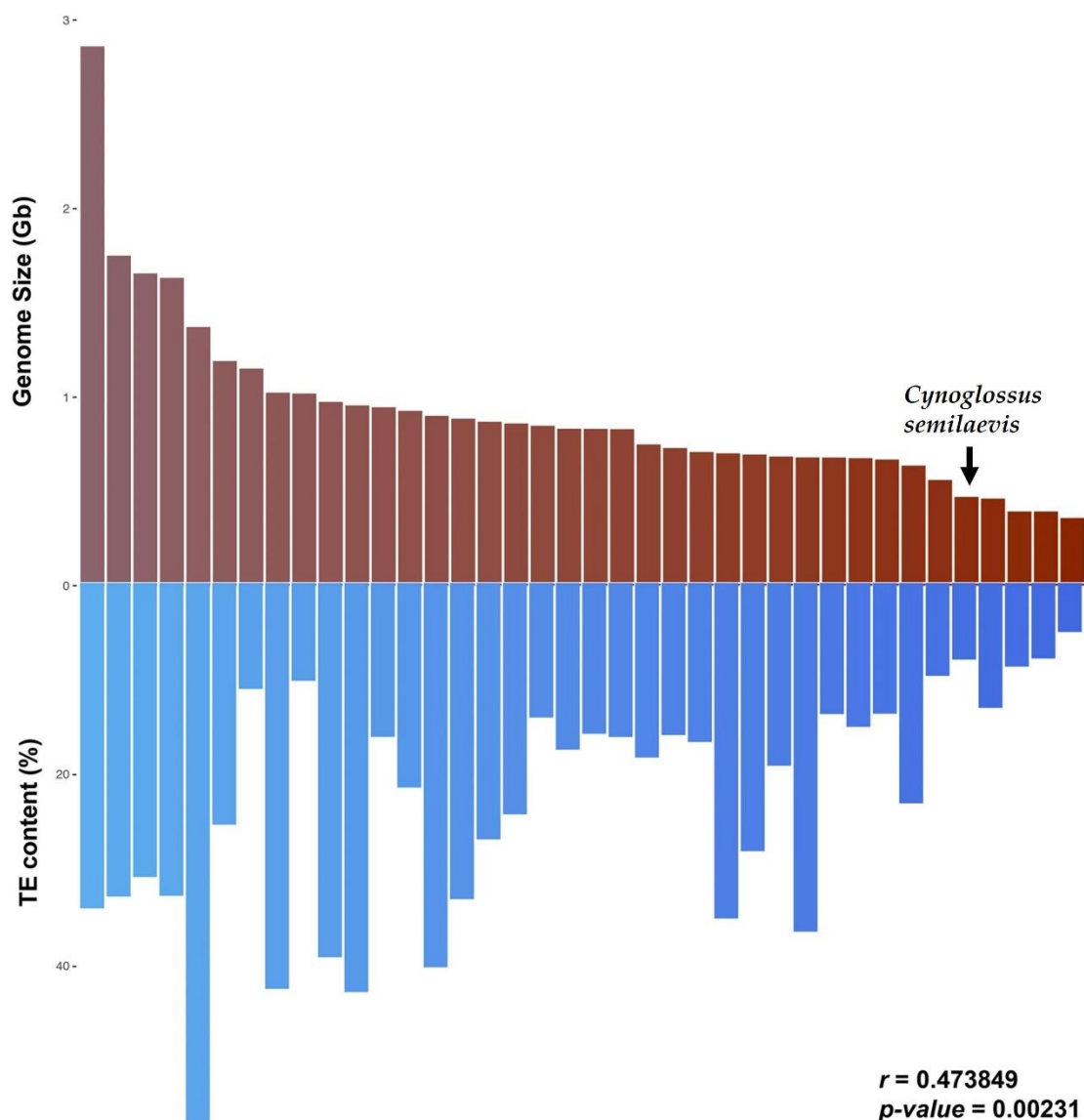
## *Análisis de las secuencias repetidas del genoma del lenguado senegalés*

La reciente secuenciación del genoma del lenguado senegalés con un tamaño de 606,88 Mb (de la Herrán et al., 2023), va a permitir profundizar en el conocimiento que actualmente tenemos de esta especie. Además, se prevé que, en un futuro próximo, dote a la comunidad científica de herramientas genómicas para el manejo y la mejora de su cultivo. Una fracción importante del genoma la constituyen las secuencias de ADN repetido. Estas tienen un papel fundamental en la evolución y diversificación de los genomas. En esta memoria nos centraremos, fundamentalmente, en los elementos móviles y en las secuencias organizadas en tándem.

Por un lado, los elementos móviles son una categoría muy diversa de secuencias repetidas en las que se pueden diferenciar: transposones de ADN, retrotransposones, elementos de tipo SINE (del inglés *Short Interspersed Nuclear Element*), elementos de tipo LINE (del inglés *Long Interspersed Nuclear Element*) y virus endógenos. Se los ha vinculado al incremento de variabilidad intrapoblacional entre individuos y a la modificación de la expresión génica, interrumpiendo o generando nuevas rutas metabólicas (Wessler, 2006). En algunos casos, estos elementos móviles pueden ser reclutados para cumplir funciones clave en los genomas; así ocurre con la función

telomérica en *Drosophila melanogaster* (Mason & Biessmann, 1995). Incluso se ha asociado el origen del gen de la telomerasa a elementos móviles (Pardue et al., 1997). También, los tamaños genómicos y diversidad de especies de peces se correlacionan positivamente con el contenido en elementos móviles y su diversificación (**Figura 3**) (Shao & Peng, 2019). Recientemente, en *S. senegalensis* se han detectado reorganizaciones estructurales en los cromosomas 2 y 4 mediante técnicas citogenéticas. A través de análisis bioinformáticos se asociaron estas reorganizaciones cromosómicas a LINEs tipo L1, algunas SINEs de tipo *Alu* y los elementos MITEs (del inglés *Miniature Inverted-repeat Transposable Elements*) (Rodríguez et al., 2021). Una caracterización profunda del contenido de elementos móviles en *S. senegalensis* en el nuevo ensamblaje del genoma podría proporcionar nueva información sobre la evolución de esta especie.



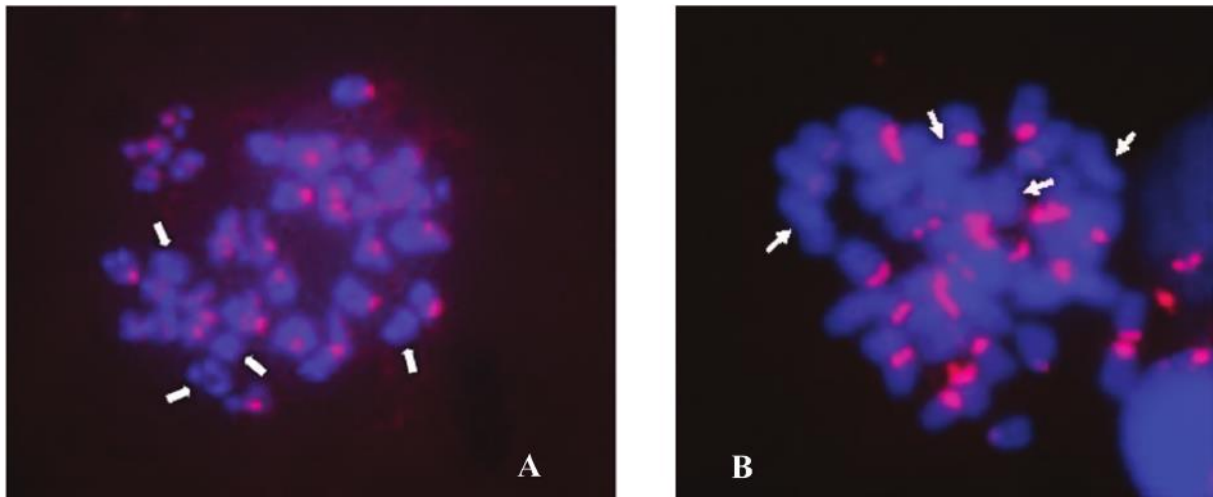


**Figura 3.** Gráfica de barras que muestra la correlación entre el tamaño del genoma de peces con el contenido de elementos móviles. La flecha indica el representante de pez plano que se utilizó en el estudio. Tomada y modificada de Shao & Peng, 2019.

Por otro lado, las secuencias repetidas en tándem son las componentes principales de las regiones heterocromáticas, fundamentalmente regiones asociadas a los centrómeros y telómeros. Por tanto, su función estructural es crucial para el correcto comportamiento de los cromosomas. Los centrómeros juegan un papel fundamental en el ciclo celular e intervienen en el proceso de unión entre cromátidas hermanas o cromosomas homólogos y su posterior segregación durante la mitosis y la meiosis, respectivamente. Sin embargo, no existe una secuencia conservada en esta región que sea común a todos los teleósteos. Las repeticiones en tándem centroméricas

suelen presentar un tamaño de 100-400 pares de bases (pb) (Talbert & Henikoff, 2020) y no se conservan características entre especies como el tamaño del monómero, el contenido en GC y los motivos de la secuencia (Melters et al., 2013). En este sentido, un estudio reciente apunta a que la función centromérica en eucariotas podría estar estrechamente relacionada con su estructura tridimensional y que, además, la unión a proteínas no es estrictamente necesaria (Kasinathan & Henikoff, 2018).

En relación a los peces planos, a pesar del alto número de especies que componen el grupo, únicamente se han caracterizado las regiones centroméricas de un reducido número de especies. En el caso del lenguado chino, *Cynoglossus semilaevis*, se han localizado marcadores de microsátélites asociados a la región centromérica (Ji et al., 2009). En el resto de especies, usando digestiones con enzimas de restricción, se han aislado familias de ADN satélite en esta posición. En concreto, en el lenguado *Achirus lineatus* se detectó la secuencia de ADN satélite *HindIII* (Carvalho de Azevedo et al., 2005), en la acedía, *Dicologlossa cuneata*, la familia *DraI* (de la Herrán et al., 2008), y en el halibut, *Hippoglossus hippoglossus*, se asoció el centrómero con la familia de ADN satélite *SacI*. Esta última familia se ha localizado mediante hibridación *in situ* y se sabe que se encuentra acumulada en la región centromérica de todos los cromosomas del complemento, excepto en dos parejas (**Figura 4A**) (Robles et al., 2017). En *S. senegalensis* se ha caracterizado una secuencia de ADN satélite, la familia *PvuII*, que se acumula específicamente en las regiones centroméricas de 19 parejas de cromosomas (Robles et al., 2017). De forma análoga a lo que ocurre en *H. hippoglossus*, dos parejas cromosómicas carecen de esta secuencia en sus regiones centroméricas (**Figura 4B**).



**Figura 4.** Fotografías de preparaciones cromosómicas mitóticas mostrando **A)** localización cromosómica de la familia de ADN satélite *SacI* mediante FISH en metafases mitóticas de *Hippoglossus hippoglossus*, y **B)** localización cromosómica de la familia de ADN satélite *PvuII* mediante FISH en metafases mitóticas de *Solea senegalensis*. Las flechas blancas indican los cromosomas que carecen de señal de hibridación. Tomada de Robles et al., 2017.

## 2. Objetivos

El objetivo principal de este trabajo es el de caracterizar la fracción de ADN repetido del genoma del lenguado senegalés. Los objetivos específicos pueden concretarse en:

- Llevar a cabo una caracterización global de secuencias de ADN repetido, fundamentalmente elementos móviles y ADN repetido en tándem en el lenguado senegalés, usando secuencias Illumina y el ensamblaje completo disponible.
- Anotar y localizar *in silico* las secuencias repetidas más importantes en términos cuantitativos usando el ensamblaje genómico disponible.
- Caracterizar las regiones centroméricas de esta especie y su contenido en elementos repetidos.

- Analizar comparativamente la acumulación de secuencias repetidas en machos y hembras y estudiar su posible implicación en la determinación sexual de esta especie.

En conjunto, estos objetivos permitirán una mejor comprensión de la organización genómica de *S. senegalensis* y podrían tener implicaciones importantes en la comprensión de eventos evolutivos relacionados con elementos repetidos, tanto clústerizados como dispersos, en los genomas de los peces planos.

### **3. Material y Métodos**

#### **3.1. Identificación de secuencias repetidas en el ensamblaje del genoma**

Se llevó a cabo la identificación de secuencias repetidas en los cromosomas de *S. senegalensis* utilizando una base de datos construida *ad hoc* para repetido de peces planos. Esta base de datos contenía un total de 39.064 secuencias y fue proporcionada por el grupo de Microbiología Aplicada y Genética Molecular del departamento de Biomedicina, Biotecnología y Salud Pública de la Universidad de Cádiz. La secuencia de los cromosomas se obtuvo del ensamblaje fSolSen1.1\_lg disponible en la base de datos "Assembly" del *National Center for Biotechnology Information*. Para la identificación de las secuencias repetidas, se utilizó el programa RepeatMasker versión 4.1.4 con los ajustes por defecto de forma local, excepto por la base de datos específica de peces planos. Este programa se aplicó tanto a las secuencias concatenadas de los cromosomas como a cada uno de ellos de forma individual. Para representar los resultados obtenidos en los análisis para cromosomas individuales se utilizó R v4.2.3 junto a la librería ggplot2 v3.4.1 (Wickham H., 2016).

#### **3.2. Identificación de secuencias repetidas en el lenguado senegalés y análisis comparativo entre sexos a partir de lecturas de Illumina**

La identificación de secuencias repetidas y el análisis comparativo de la composición de los genomas de macho y hembra de *S. senegalensis* se llevó a cabo usando RepeatExplorer2 (Novák et al., 2013) de acuerdo con el protocolo propuesto por Novák et al. (2020), con modificaciones menores. Este programa está incluido en

el servidor Galaxy (<https://repeatexplorer-elixir.cerit-sc.cz/galaxy/>), desarrollado y mantenido por el Laboratorio de Citogenética Molecular, el Instituto de Biología Molecular de Plantas y el Centro de Biología CAS de Ceske Budejovice en República Checa. Basándose en la similitud de las lecturas, identifica regiones repetidas y las agrupa en clústers.

Para el análisis, mediante la herramienta “Faster Download and Extract Reads in FASTQ”, se utilizaron lecturas Illumina *paired-end* previamente depositadas en la base de datos SRA del GenBank (códigos de acceso SRR18498474 -29.305.614 lecturas- y SRR18498475 -36.606.408- para macho y hembra, respectivamente). A continuación, se llevó a cabo un análisis previo de la calidad de las secuencias con “FastQC” y “Scan paired-end reads for overlap”. Posteriormente, las lecturas se emparejaron usando “Preprocessing of FASTQ paired-end reads” con los parámetros por defecto, un valor de “Quality cutoff” de 10 y un valor de “Trim reads” de 101. Para realizar el siguiente paso del protocolo, se utilizó medio millón de lecturas por sexo, aleatoriamente extraídas con “Read Sampling”. Esta ronda de análisis se repitió en cinco ocasiones. Para ello, las secuencias del análisis anterior, que no habían sido identificadas por el programa como repetidas, se aislaron con el script DeconSeq (deconseq\_run.py, <https://github.com/fjruizruano/satminer>) eliminando las lecturas clústerizadas del conjunto de lecturas. En cada ronda, el número de lecturas por sexo se fue duplicando, excepto en las dos últimas donde se utilizaron 900.000 y 200.000 respectivamente, debido a que no quedaban lecturas suficientes. El filtrado y duplicación de las lecturas en cada ronda permite la detección de secuencias repetidas menos representadas.

Para la anotación automática de los clústers detectados por RepeatExplorer2 se utilizó la base de datos REXdb (Metazoa 3.0). Los clústers no anotados por este sistema se anotaron manualmente mediante BLAST local (Altschup et al., 1990), usando la base de datos de secuencias repetidas de peces planos que se describió en el apartado anterior y considerando e-values de corte de  $10^{-5}$  o inferiores. En los casos en los que tampoco se pudo anotar mediante este método, se utilizó Dfam (Wheeler et al., 2013) con la opción “Organism” ajustada al ensamblaje de *Danio rerio* (danRer10) por ser la especie más cercana a *S. senegalensis* disponible y se anotaron los resultados con menor

e-values. Además, para los clúster 1, 2 y 3 de la primera vuelta de RepeatExplorer2 se realizaron alineamientos frente a las secuencias anotadas manualmente con el programa Clustal Omega (Goujon et al., 2010). Para saber qué tipo de ARN<sub>t</sub> era una de las secuencias en la base de datos de peces planos que daba homología con uno de los clústers se utilizó el programa tRNAscan-SE 2.0 (Lowe & Eddy, 1997).

### **3.3. Caracterización de las regiones centroméricas en el lengüado senegalés**

La búsqueda de regiones repetidas acumuladas específicamente en regiones centroméricas fue llevada a cabo utilizando gráficos de tipo dotplots y tomando como referencia las coordenadas de las secuencias centroméricas ya conocidas (Robles et al., 2017) y la morfología cromosómica (**Figura 2**). Los gráficos dotplots se construyeron con Genome Pair Rapid Dotter (gepard) (Krumdiek et al., 2007). Se identificaron secuencias potencialmente repetidas en tándem mediante inspección visual de los dotplots, se extrajeron las coordenadas y su disposición en tándem se confirmó mediante el programa Tandem Repeats Finder (Benson, 1999). También, se anotaron mediante BLAST usando la base de datos de secuencias repetidas de peces planos como referencia.

Los idiogramas que muestran la localización en el ensamblaje del genoma de dichas secuencias candidatas se construyeron usando scripts de Python v3.10.9 y R v4.2.3 junto a la librería karyoploteR v1.25.0 (Gel & Serra, 2017). Finalmente, para predecir la estructura secundaria de las secuencias candidatas a formar parte del ADN centromérico se utilizó el paquete ViennaRNA v2.5.1 (Lorenz et al., 2011). El comando utilizado en todos los casos fue: `RNAfold --noGU --noconv --paramFile dna_mathews2004.par -p -g < secuencia.fasta`.

Para automatizar gran parte de los procesos descritos en esta memoria, se han desarrollado scripts de Python que se encuentran disponibles en el siguiente repositorio Github: <https://github.com/AaronGS1999/TFM>.

## 4. Resultados

### 4.1. Caracterización de elementos repetidos en *S. senegalensis*

En este trabajo se ha llevado a cabo una caracterización de los elementos repetidos presentes en el genoma del lenguado senegalés. Para ello, se han seguido dos aproximaciones: por un lado, usando el ensamblaje del genoma, se estudió su composición en cuanto a elementos móviles. Por otro lado, usando lecturas Illumina, se realizó un estudio de clústerización para detectar, fundamentalmente, secuencias repetidas en tándem.

#### 4.1.1. Resultados de la identificación de secuencias repetidas en el ensamblaje del genoma

Los análisis con el programa RepeatMasker frente a los 21 cromosomas concatenados arrojaron como resultado que el genoma de *S. senegalensis* presenta un 40,70% de contenido en GC y que está compuesto por un 32,98% de elementos repetidos. Se detectó una gran diversidad de elementos repetidos, los cuales se pueden ver clasificados por categorías en la **Tabla 1**. Los que mayor representación tienen de forma global en el genoma son los transposones de ADN con un porcentaje de 11,76%, las LINEs con un porcentaje de 7,75% y los elementos LTR con un porcentaje de 7,34% respecto al genoma.

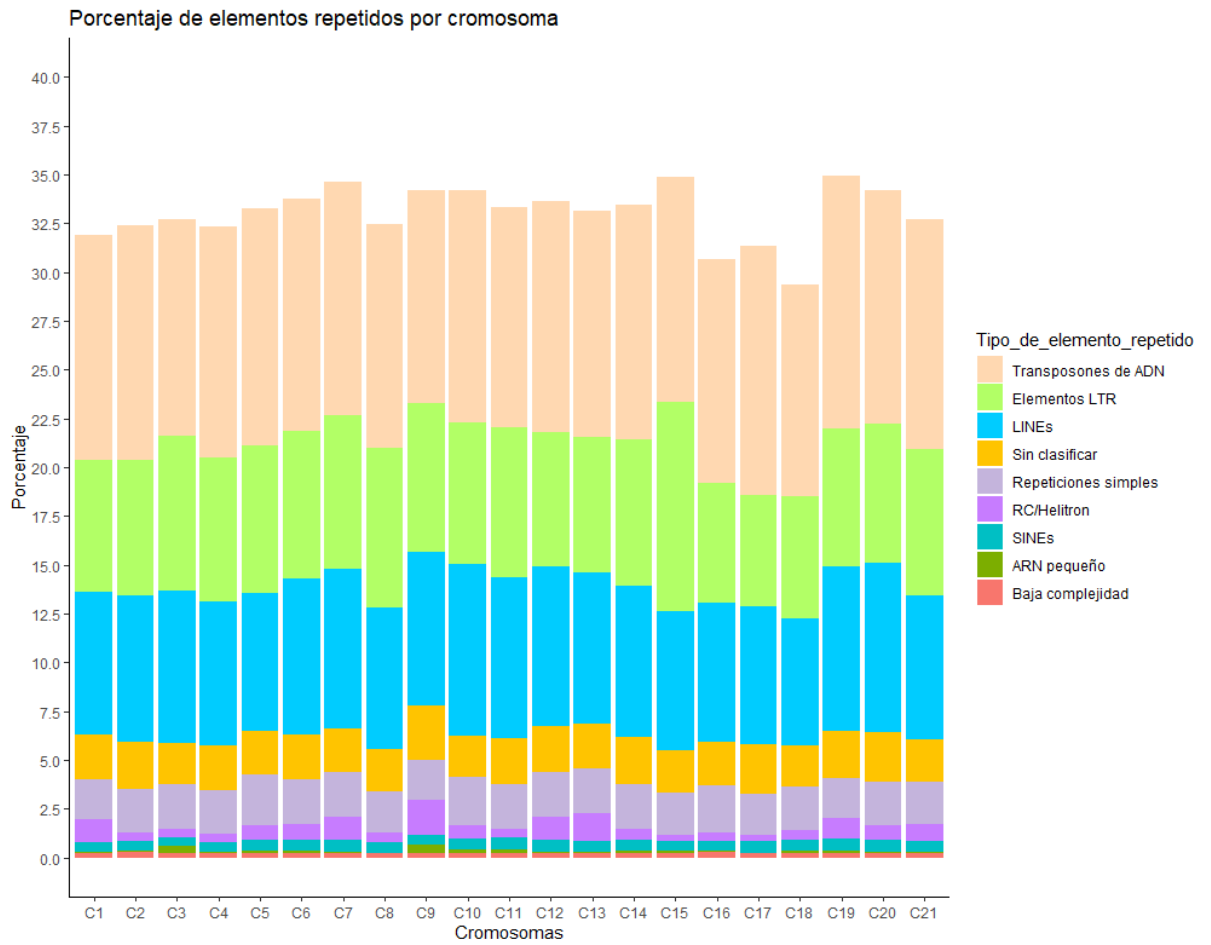
**Tabla 1.** Clasificación de elementos repetidos detectados en *S. senegalensis*. Resultado del análisis de RepeatMasker del genoma completo.

Tipo de elemento repetido	Nº de elementos	Longitud ocupada	Porcentaje de secuencia
<b>Retroelementos</b>	665153	94463879 bp	15,5%
SINEs	40202	3327559 bp	0,55%
LINEs:	407382	46581736 bp	7,75%
L2/CR1/Rex	138266	15850644 bp	2,61%
R1/LOA/Jockey	7186	721052 bp	0,12%
R2/R4/NeSL	37960	2626156 bp	0,43%
RTE/Bov-B	64817	6789351 bp	1,12%
L1/CIN4	53711	6143052 bp	1,01%
Penelope	3018	403561 bp	0,07%

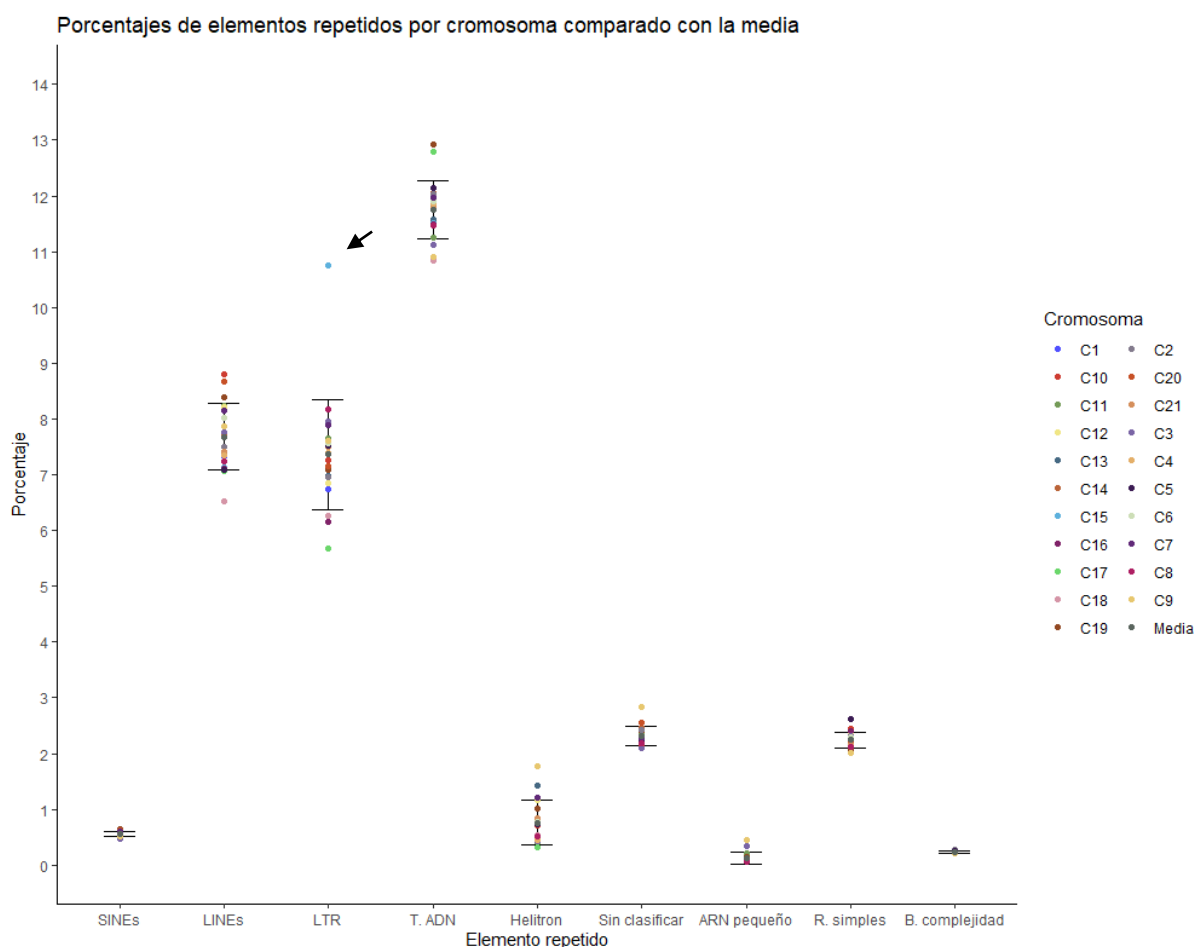
LTR:	217569	<u>44554584 bp</u>	7,34%
BEL/Pao	1545	817750 bp	0,13%
Ty1/Copia	427	94429 bp	0,02%
Gypsy/DIRS1	43535	18084060 bp	2,98%
Retroviral	4680	643123 bp	0,11%
<b>Transposones de DNA</b>	560060	71381281 bp	11,76%
hobo-Activator	187422	28848299 bp	4,75%
Tc1-IS630-Pogo	35998	5685578 bp	0,94%
MULE-MuDR	2647	363443 bp	0,06%
PiggyBac	2362	314498 bp	0,05%
Tourist/Harbinger	17274	2244656 bp	0,37%
<b>Otros (Mirage, P-element, Transib)</b>	440	23191 bp	0,004%
<b>RC/Helitron</b>	8096	4686611 bp	0,77%
<b>Sin clasificar</b>	132460	14071851 bp	2,32%
<b>Total de repeticiones intercaladas</b>	-	179917011 bp	29,65%
<b>ARN pequeño</b>	6332	706563 bp	0,12%
<b>Satélites</b>	96	19938 bp	0,003%
<b>Repeticiones simples</b>	306655	13563989 bp	2,24%
<b>Baja complejidad</b>	28968	1422150 bp	0,23%

Los distintos tipos de elementos repetidos se distribuyen de forma uniforme entre cromosomas (ver **Figura 5**), excepto en el caso del cromosoma 15, que representa una excepción en relación a los elementos LTR (ver **Figura 6**). En él se encontró un contenido de elementos de LTR del 10,75% siendo la media entre cromosomas de 7,34%. De estos elementos LTR cabe destacar que un 5,76% eran del tipo Gypsy y la media de este tipo de secuencia entre cromosomas es de 2,98%.





**Figura 5.** Representación gráfica del contenido total en porcentaje de elementos repetidos por cromosoma usando los datos obtenidos con RepeatMasker.



**Figura 6.** Representación gráfica mostrando el análisis de RepeatMasker de los cromosomas de forma individual. Comparación del contenido de elementos repetidos en porcentaje por clase y cromosoma. Las barras de error indican desviación estándar respecto a la media. La flecha negra destaca el valor de contenido en LTR para el cromosoma 15, el cual presenta una desviación destacable respecto a la media.

#### 4.1.2. Resultados del análisis comparativo e identificación de secuencias repetidas a partir de lecturas de Illumina

Tras los análisis de lecturas Illumina usando el programa RepeatExplorer2, se obtuvieron un total de 214 clústers, de los cuales se anotaron manualmente 188. Se asociaron 74 clústers con transposones de ADN; 46 con LINEs; 26 con elementos LTR; 15 con elementos repetidos desconocidos, sin clasificar en la base de datos de peces planos; 10 con ADN satélite (representando un total de 0,91% respecto al genoma); 7 con SINEs; 3 con círculo rodante, todos Helitron; 3 con ARN; 3 con repeticiones simples

y, finalmente, 26 clústers no mostraron ninguna homología (ver **Tabla Sup. 1**). El análisis comparativo entre machos y hembras de *S. senegalensis* no mostró diferencias importantes en el contenido de elementos repetidos por lo que no se logró identificar ningún clúster diferencial para alguno de los sexos.

Considerando los clústers de forma individual, los tres elementos repetidos más representados en el genoma del lenguado senegalés se corresponderían con elementos repetidos en tándem (**Tabla 2**) y se analizaron en mayor profundidad tal y como se muestra a continuación.

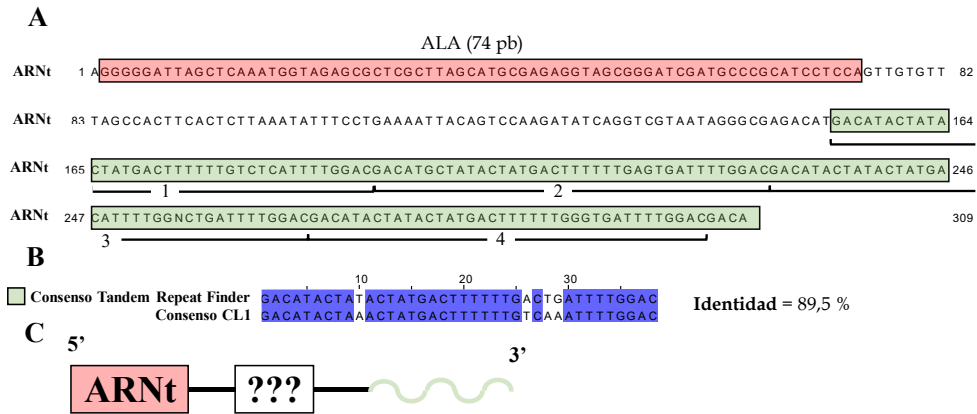
**Tabla 2.** Elementos repetidos más abundantes correspondiente a los primeros 15 clústers de RepeatExplorer2. La proporción para ambos sexos está corregida respecto al número de lecturas que se utilizó para realizar el RepeatExplorer2, en este caso 1.000.000.

N°	N° lecturas machos	N° lecturas hembras	Proporción macho (%)	Proporción hembra (%)	E-Value	Anotación	Repetido en tándem / Disperso	Clase superior
1	6350	6130	0,635	0,613	8,66E-52	Sse_rnd-1_family-0#tRNA	Repetido en tándem (baja confianza)	ARN <sub>t</sub>
2	5910	5420	0,591	0,542	0	Asian_Seabass_LTR_161#LTR/Gypsy	Repetido en tándem (baja confianza)	LTR
3	4700	4670	0,47	0,467	0	Sse_rnd-5_family-1529#LINE/L1	Repetido en tándem (alta confianza)	Satélite
4	4830	4370	0,483	0,437	0	Sse_rnd-1_family-8#RC/Helitron	Repetido en tándem (baja confianza)	RC
5	2230	2240	0,223	0,224	0,0017	Dfam: DF0004077.1 [Helitron-1N4_DR]	Disperso	RC
6	2200	1790	0,22	0,179	2,70E-10	Dfam: DF0004166.2 [SINE_TE]	Repetido en tándem (alta confianza)	Satélite
7	1840	2090	0,184	0,209	1,01E-162	Sse_rnd-1_family-71#DNA/Kolobok-T2	Disperso	Transposón de ADN
8	2080	1510	0,208	0,151	0	Sse_rnd-5_family-6723#rRNA	Repetido en tándem (alta confianza)	Satélite
9	1530	1470	0,153	0,147	0	Sse_rnd-1_family-49#LINE/L2	Disperso	LINE
10	1550	1400	0,155	0,14	0	Sse_rnd-1_family-7#RC/Helitron	Disperso	RC

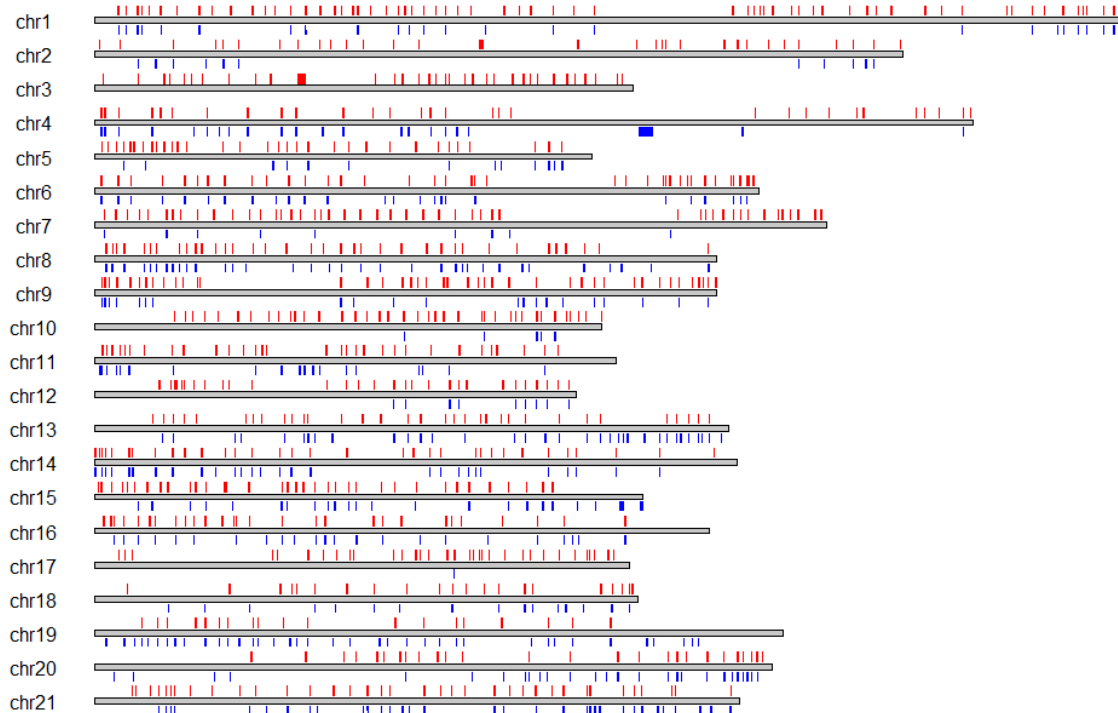
11	1350	1420	0,135	0,142	0	Sse_rnd-5_family-611#LINE	Disperso	LINE
12	1360	1320	0,136	0,132	0	Sse_rnd-1_family-113#LINE/L2	Disperso	LINE
13	1350	1270	0,135	0,127	6,90E-93	Sse_rnd-5_family-3283#DNA	Disperso	Transposón de ADN
14	974	1030	0,0974	0,103	7,38E-121	Sse_rnd-1_family-4#DNA/hAT-Ac	Disperso	Transposón de ADN
15	731	761	0,0731	0,0761	1,10E-99	Sse_rnd-4_family-608#DNA/hAT-Ac	Disperso	Transposón de ADN

El clúster 1, con un motivo de repetición de 38 pb, presenta homología con la parte final de una secuencia repetida de 309 pb anotada en la base de datos de peces planos como ARN<sub>t</sub>. Estudiando en detalle la estructura de esta última secuencia, se observa que en su extremo 5' está compuesta por una región de 74 pb codificante para un ARN<sub>t</sub> con un anticodón para la alanina (**Figura 7A** región sombreada en rojo) (Puntuación: 92,8 y anticodón detectado: AGC en tRNAscan-SE), seguida por una región donde no se encontró homología con ninguna secuencia y con un extremo 3' formado por una secuencia repetida cuatro veces en tándem (ver **Figura 7A** región sombreada en verde). En concreto, la secuencia consenso del clúster 1 presenta una gran homología con la secuencia consenso de esta repetición en tándem terminal (ver **Figura 7B**). En conjunto, el elemento repetido presenta una estructura que consiste en un ARN<sub>t</sub> en un extremo seguido de una región que no se ha podido caracterizar y en el otro extremo presenta una secuencia repetida en tándem (ver **Figura 7C**).

La localización *in silico* de la secuencia consenso de este clúster muestra que su distribución es dispersa y uniforme por todo el genoma de *S. senegalensis*. Al comparar esta localización con la del ARN<sub>t</sub> se observa como ambas secuencias colocalizan en gran cantidad de loci (ver **Figura 8**).



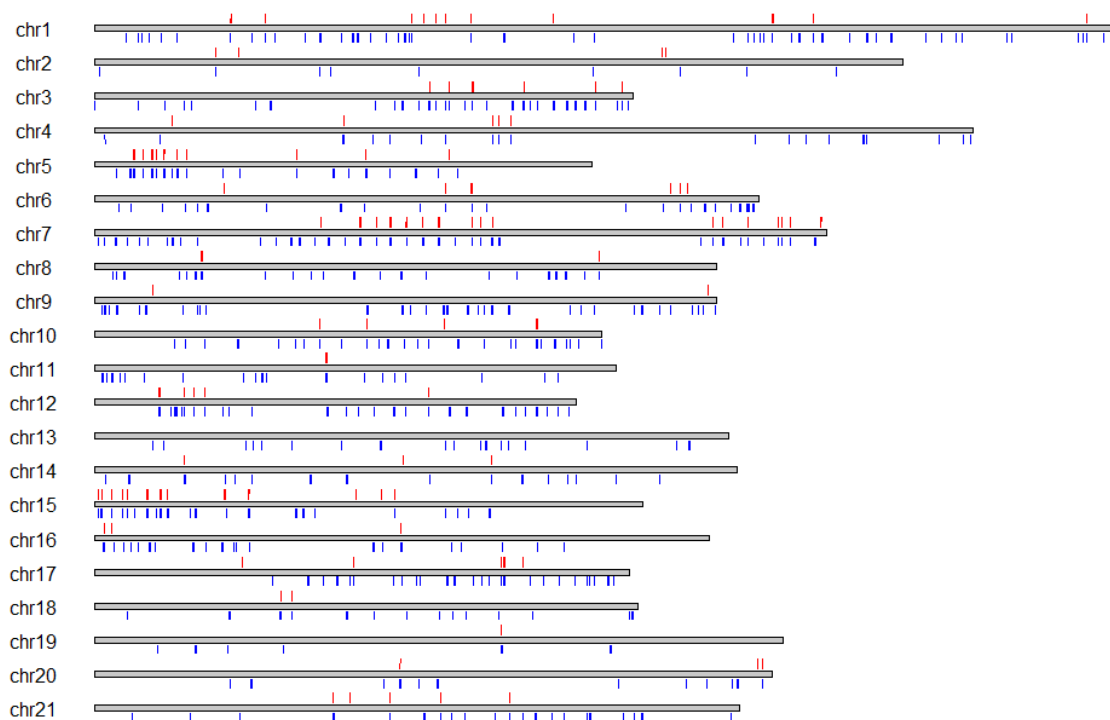
**Figura 7.** Infografía mostrando la estructura del clúster 1. **A:** Caracterización de la secuencia anotada como Sse\_rnd-1\_family-0#tRNA en la base de datos de peces planos. En rojo la secuencia identificada como ARN<sub>t</sub> Ala y en verde la región donde se encuentra las cuatro repeticiones en tándem detectadas por Tandem Repeats Finder. **B:** Alineamiento de la secuencia consenso de la repetición en tándem con la secuencia consenso del clúster 1. **C:** Esquema de la estructura del elemento repetido Sse\_rnd-1\_family-0#tRNA.



**Figura 8.** Representación de las posiciones sobre los cromosomas del ARN transferente anotado como Sse\_rnd-1\_family-0#tRNA en la base de datos de los peces planos.

peces planos (segmentos rojos) y la secuencia consenso del clúster 1 (segmentos azules). Se representaron las posiciones con e-values menores o iguales a  $10^{-40}$ .

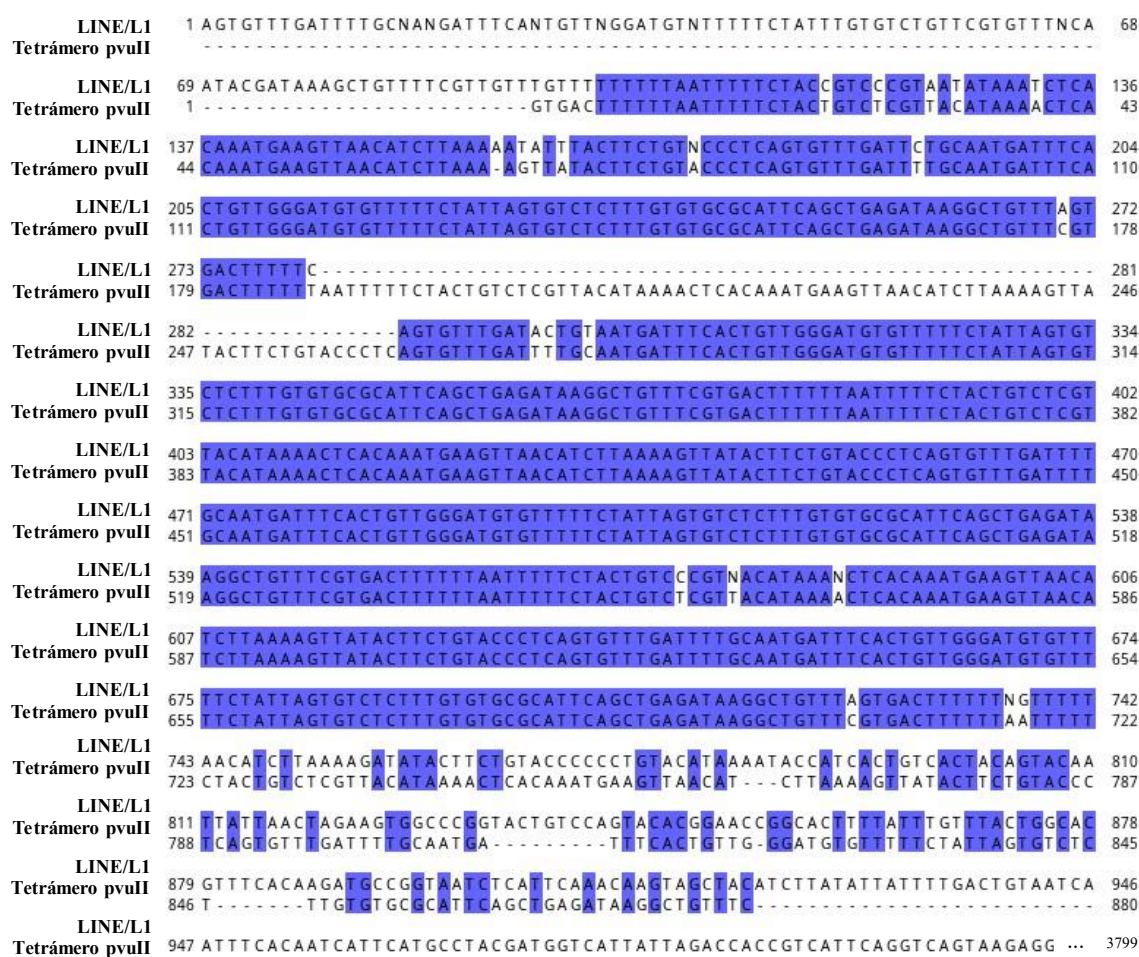
El segundo clúster en términos de abundancia, el clúster 2, presenta también un motivo de repetición de 38 pb. Significativamente, la secuencia reverso complementaria de este se corresponde con la secuencia consenso del clúster 1 con una identidad del 89,5%. Sin embargo, comparando el contig más largo de este clúster con las secuencias de la base de datos de peces planos, se detectó una mayor homología con la secuencia anotada como Asian\_Seabass\_LTR\_161#LTR/Gypsy en la base de datos (longitud del alineamiento = 2182 pb y e-value = 0,0) que con la secuencia anotada como Sse\_rnd-1\_family-0#tRNA (longitud del alineamiento = 156 pb y e-value =  $3,29 \times 10^{-63}$ ). Posiblemente esta sea una de las razones por la que, pese a ser secuencias similares y compartir un motivo, hayan aparecido en clústers diferentes. El análisis de la posición de este clúster y los elementos repetidos con los que presenta homología demostraron que se trata de secuencias distribuidas por todo el genoma de *S. senegalensis* y que existen algunas posiciones compartidas por dichos elementos (ver **Figura 9**). La presencia exclusiva del clúster 2 en el cromosoma 3 y su mayor abundancia frente al clúster 1 en otros cromosomas como en el 17, podría explicar también que hayan aparecido estas secuencias de forma separada en el análisis mediante RepeatExplorer2.



**Figura 9.** Representación de las posiciones sobre los cromosomas de la secuencia LTR/Gypsy anotada como Asian\_Seabass\_LTR\_161#LTR/Gypsy en la base de datos de los peces planos (segmentos rojos) y la secuencia consenso del clúster 2 (segmentos azules). Se representaron las posiciones con e-values menores o iguales a  $10^{-40}$ .

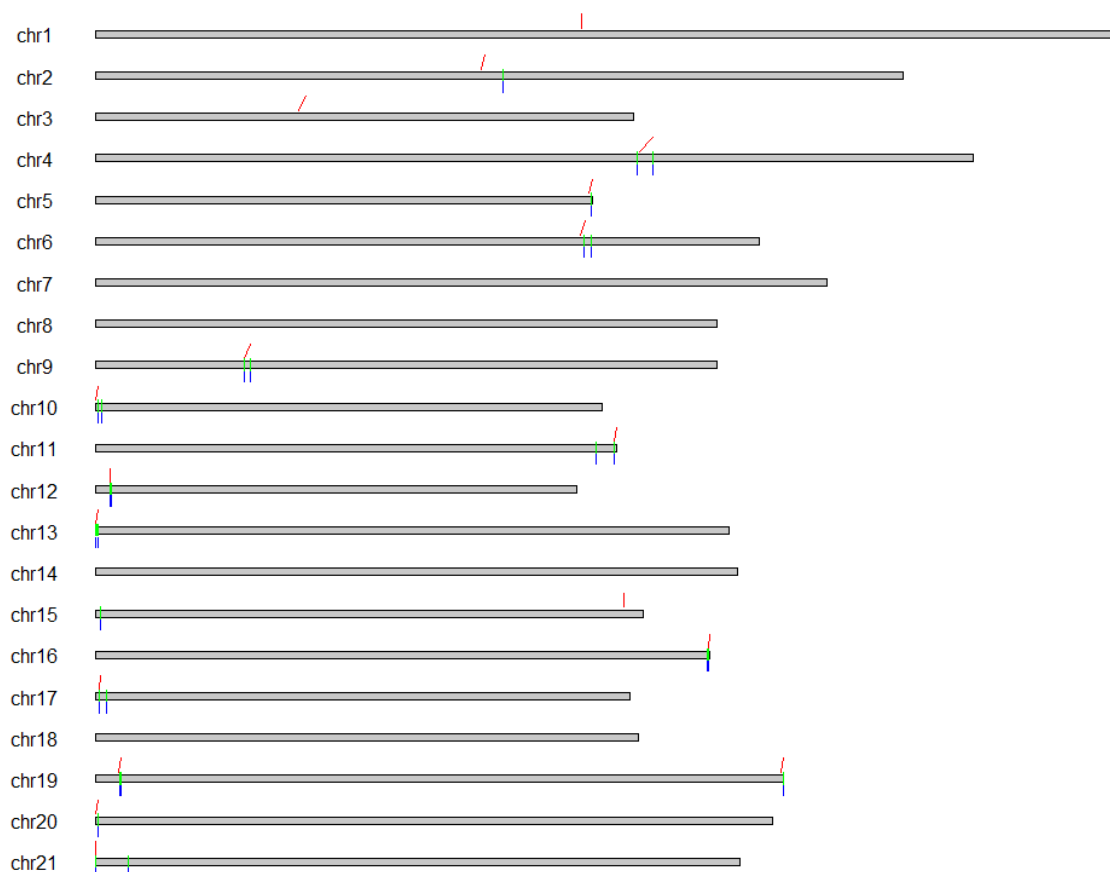
Por último, el clúster 3, con una secuencia consenso de 176 pb, se corresponde con una identidad del 91,5% a la familia de ADN satélite *PvuII*, previamente caracterizada en esta especie como una secuencia localizada en la región centromérica de todos los cromosomas, excepto en dos parejas (Robles et al., 2017). Al realizar la anotación manual, se descubrió que esta secuencia satélite se asocia con un elemento repetido encontrado en la base de datos de peces planos, clasificado como LINE de tipo L1. Posteriormente, se realizó un alineamiento de la secuencia consenso del clúster 3 frente a la secuencia de la LINE L1, para verificar su homología. El alineamiento mostró que dentro de la LINE L1 se encontraron cuatro repeticiones en tándem de la secuencia consenso del clúster 3 identificado como ADN centromérico *PvuII* (ver **Figura 10**).

Se estudiaron las posiciones de la LINE L1 y la secuencia del clúster 3, tanto de forma conjunta como individualmente, sobre el genoma de *S. senegalensis* y se observó cómo existía una asociación clara entre las posiciones. Cabe destacar que en los cromosomas 7, 8, 14 y 18 no se encontró la presencia de la secuencia LINE L1 ni de la secuencia *PvuII* y que en el cromosoma 19 se encontraron dos señales a ambos extremos tanto de la LINE L1 como de la secuencia del clúster 3 (ver **Figura 11**). Dado que se trata de una secuencia centromérica y uno de los objetivos de esta memoria era la caracterización de esta región, estudiamos estas y otras secuencias en profundidad tal y como se describe en el apartado 4.2.



**Figura 10.** Alineamiento de la secuencia LINE/L1 anotada como Sse\_rnd-5\_family-1529#LINE/L1 en la base de datos de los peces planos con un tetramero del satélite centromérico *PvuII* mediante Clustal Omega.





**Figura 11.** Representación de las posiciones sobre los cromosomas de la secuencia del satélite centromérico *PvuII* (segmentos rojos), la secuencia correspondiente a la LINE/L1 anotada como Sse\_rnd-5\_family-1529#LINE/L1 en la base de datos de los peces planos (segmentos verdes) y la misma secuencia extrayendo la región donde alinea la secuencia *PvuII* (segmentos azules).

#### 4.2. Resultados de la caracterización de las regiones centroméricas en el ensamblaje del genoma

Uno de los objetivos principales de esta memoria fue caracterizar las regiones centroméricas y pericentroméricas del genoma del lenguado senegalés. Para ello, se ha estudiado la localización y estructura de las secuencias repetidas candidatas (tanto en tándem como dispersas) caracterizadas en este trabajo (**Tabla 3**). Como criterios fundamentales se consideraron aquellas secuencias repetidas con alta representación en el genoma, aquellas previamente descritas como centroméricas en esta especie o

aquellas relacionadas con regiones pericentroméricas en otras especies, como las secuencias Helitron o LINEs (Cvitanić et al., 2020 y de Sotero-Caio et al., 2017).

En una primera aproximación, las coordenadas de las distintas secuencias se obtuvieron mediante BLAST contra los cromosomas y la localización se representó utilizando ideogramas. Una vez acotadas las regiones de interés, el análisis fino de las mismas se llevó a cabo mediante representaciones de tipo dotplot donde se puede observar regiones con secuencias repetidas. Por último, la composición y estructura de dichas regiones se exploró usando Tandem Repeat Finder y BLAST frente a la base de datos de peces planos.

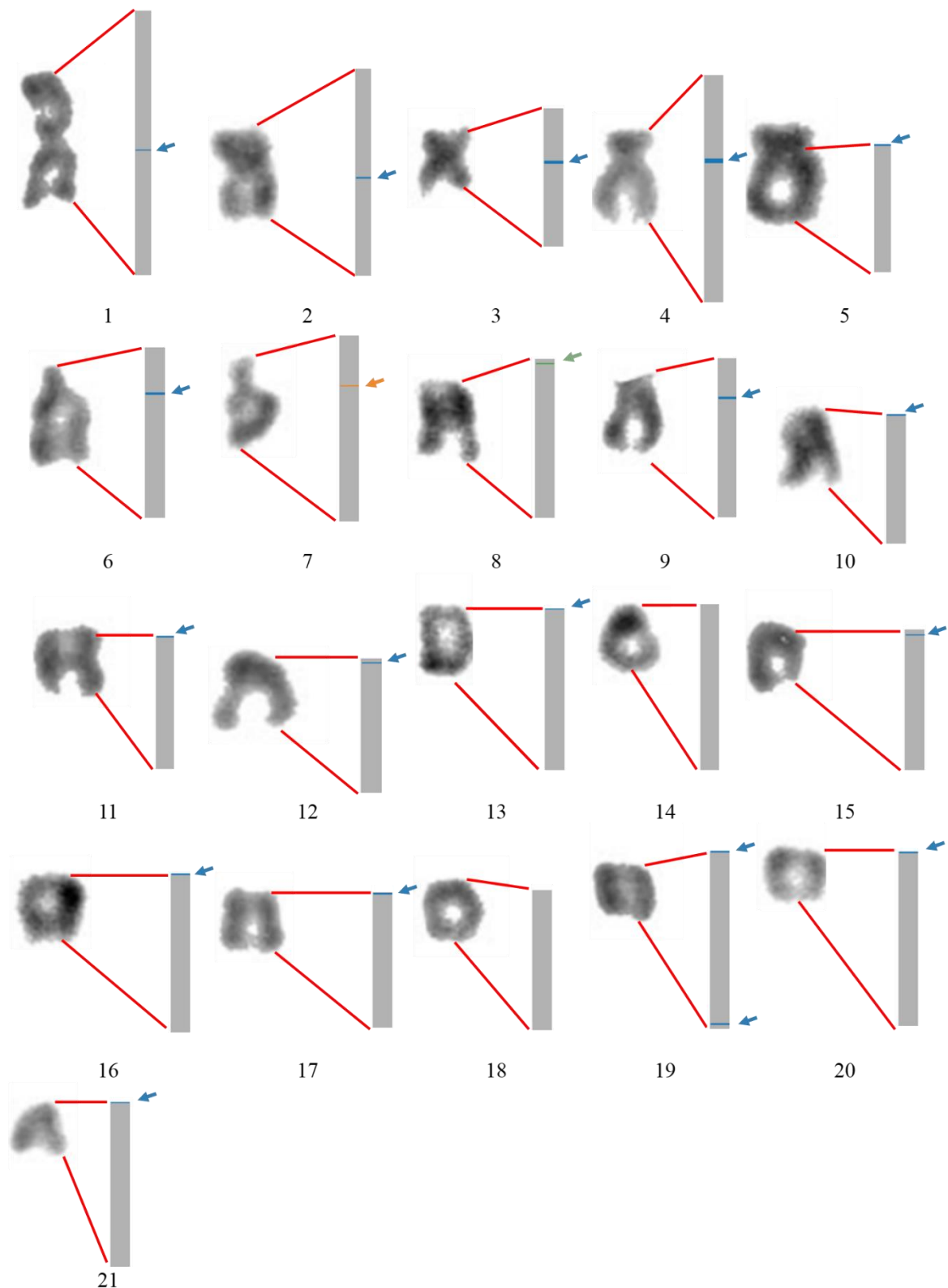
En un trabajo anterior, Robles y colaboradores (2017) determinaron mediante hibridación *in situ* que la familia de ADN satélite *PvuII* está presente en 19 parejas cromosómicas de las 21 de las que consta el genoma del lenguado senegalés. Como se ha comentado en el apartado 4.1.2, en este trabajo, nosotros hemos encontrado que el clúster 3 se corresponde con la secuencia de ADN satélite *PvuII*. Cuando mapeamos esta secuencia en el ensamblaje, se detectó su presencia en 17 parejas cromosómicas (los pares 1 al 6, 9 al 13, 15 al 17, y 19 al 21). Las secuencias de *PvuII* se agrupaban en clústers discretos (oscilando su tamaño entre el menor de 189 pb en el cromosoma 12 y hasta un máximo de 596.258 pb en el cromosoma 4). La localización *in silico* de las secuencias de *PvuII* nos permitió construir idiogramas coherentes con la morfología cromosómica de esta especie (tres parejas de cromosomas metacéntricos, dos submetacéntricos, cuatro subtelocéntricos y doce acrocéntricos, Vega et al., 2002) (ver **Figura 12**). Así las cosas, las coordenadas de la localización de este satélite fueron extraídas y consideradas las regiones centroméricas de los cromosomas implicados. Una excepción a esto la representa el cromosoma 19, para el que encontramos dos señales de *PvuII* (cada una en un extremo cromosómico), y el cromosoma 5 que, siendo submetacéntrico, aparece como telocéntrico de acuerdo a nuestros datos.

Las cuatro parejas cromosómicas restantes (los pares 7, 8, 14 y 18) no presentaron evidencias de la presencia del ADN satélite *PvuII*. Se abordó la búsqueda de sus regiones centroméricas tal y como se describe a continuación, se estudiaron los contigs más largos de los clústers de repetido más representados. De esta forma, encontramos

que el clúster 10 (ver **Tabla 2**) presentaba homología con varias familias de Helitron (fundamentalmente con la 7) y que se localizaba en las regiones pericentroméricas de los cromosomas 3, 4, 5, 6, 9, 10, 12, 13, 16, 17 y 19 (ver **Figura 12**). Esta relación con la región pericentromérica fue especialmente destacable en los cromosomas 4, 5, 6, 9, y 17 en la que los dotplots mostraban las regiones centroméricas de forma contigua a estas zonas enriquecidas en Helitron. Por todo ello, la consideramos secuencia candidata a formar parte de las regiones centroméricas de los cromosomas para los que todavía no estaba caracterizada y decidimos estudiarla en mayor profundidad. De forma significativa, las parejas 7 y 8 mostraron regiones enriquecidas en elementos Helitron, que fueron consideradas candidatas a contener la región centromérica. Además, se situaban en un extremo cromosómico, hecho coherente con la morfología de dichos pares, ambos acrocéntricos.

**Tabla 1.** Lista de elementos repetidos presentes en los centrómeros de *S. senegalensis*. Secuencias consenso del RepeatExplorer2 detectadas en los centrómeros o próximas a la región centromérica.

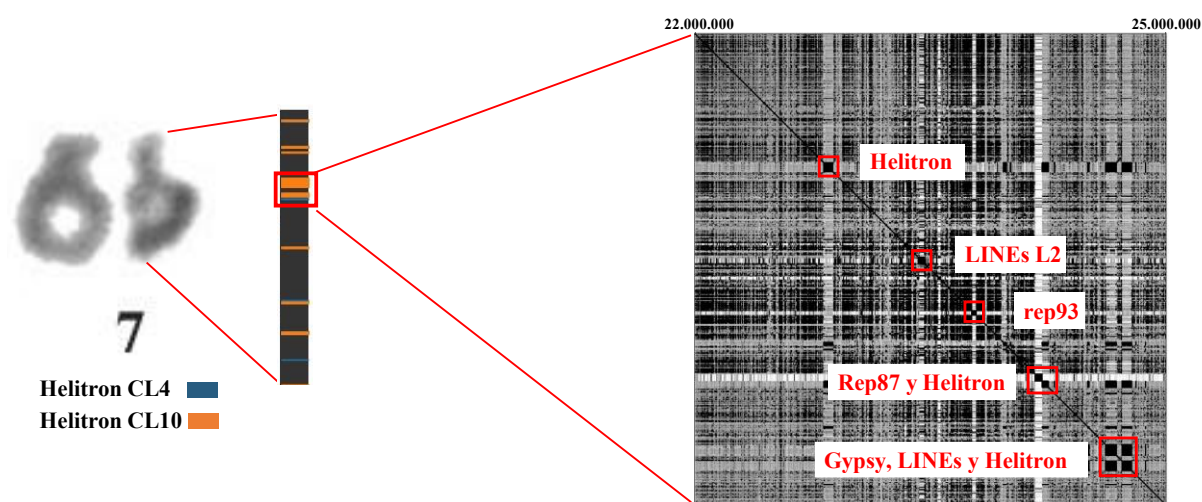
Nº clúster	Anotación	Presencia en los centrómeros (Nº cromosoma)	Próximo a los centrómeros (Nº cromosoma)
1	ARN <sub>t</sub>	4	-
3	Satélite ( <i>PvuII</i> )	1, 2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 15, 16, 17, 19, 20 y 21	-
4	RC (Helitron)	17	20
75	LTR	2, 3, 6, 9, 13, 19 y 21	-
89	Satélite	-	2
109	Satélite	3, 4, y 9	
202	Satélite	-	2
209	Satélite	-	3
214	No anotado	-	7 y 10



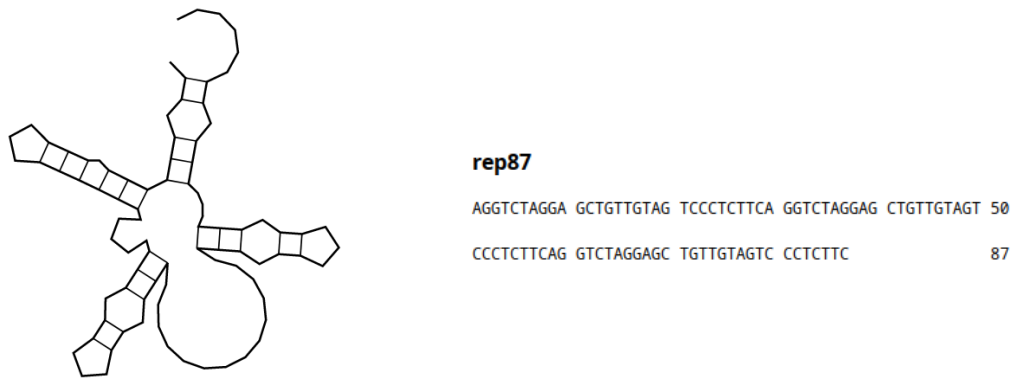
**Figura 12.** Cariotipo de *S. senegalensis* y su comparación con el ideograma construido utilizando la posición putativa de los centrómeros. Los cromosomas del cariotipo están a escala entre sí y los ideogramas de cada uno también. Las flechas azules

indican las posiciones donde se ha detectado *PvuII*, la flecha naranja indica la posición de la secuencia rep87 y la flecha verde indica la posición de la secuencia rep120. Cromosomas modificados del cariotipo de Vega et al., 2002.

Para el cromosoma 7, se examinó la región desde la posición nucleotídica 22.000.000 hasta la posición 25.000.000 del cromosoma. Durante este proceso, se identificaron elementos repetidos como Helitron, LINEs y Gypsy, además de dos repeticiones en tándem (ver **Figura 13**). La primera repetición, rep93, consta de una secuencia de 93 nucleótidos y forma un clúster con un total de 292 copias. No se encontró homología con otras secuencias ni se encontró en otros cromosomas, y tampoco presentaba estructura secundaria. Por otro lado, la repetición que denominamos rep87 y que está compuesta por una secuencia de 87 nucleótidos, forma un clúster de 485 repeticiones en tándem seguido de una región enriquecida en Helitron. Significativamente, esta última secuencia, sólo se encuentra clústerizada en el cromosoma 7, pero se detectó su presencia de forma dispersa en los cromosomas 8, 11, 14, 17 y 19. Aunque no se encontró homología entre rep87 y otras secuencias, sí se determinó que podía formar una estructura secundaria (ver **Figura 14**).



**Figura 13.** Caracterización de la región centromérica del cromosoma 7. Exploración de la región centromérica candidata mediante dotplots, Tandem Repeats Finder y BLASTn. Cromosoma tomado y modificado del cariotipo de *S. senegalensis* de Vega et al., 2002.



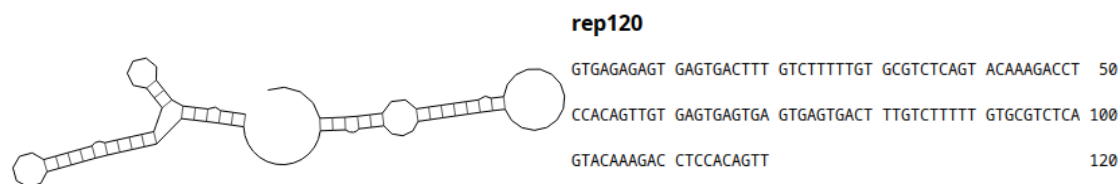
**Figura 14.** Estructura secundaria de rep87, propuesto como satélite centromérico del cromosoma 7, predicha con RNAfold y monómero de la repetición.

En la región donde potencialmente podría encontrarse la secuencia centromérica del cromosoma 8, comprendida entre la posición 25.500.000 hasta 28.000.000, donde se detectaron dos puntos en los que se acumulaban secuencias Helitron. Se encontró una secuencia repetida en tándem, que denominamos rep120 y un clúster que mostraba homología con la secuencia consenso del clúster 1 de RepeatExplorer2 (ver **Figura 15**). La secuencia rep120 está compuesta por 120 nucleótidos y forma un clúster de 220 repeticiones en tándem. Se encuentra únicamente en el cromosoma 8 y clústerizada. También determinó que puede formar una estructura secundaria (ver **Figura 16**).



**Figura 15.** Caracterización de la región centromérica del cromosoma 8. Exploración de la región centromérica candidata mediante dotplots, Tandem Repeats Finder y

BLASTn. Cromosoma tomado y modificado del cariotipo de *S. senegalensis* de Vega et al., 2002.



**Figura 16.** Estructura secundaria de rep120, propuesto como satélite centromérico del cromosoma 8, predicha con RNAfold y monómero de la repetición.

## 5. Discusión

### *Análisis del ADN repetido*

Los genomas de peces presentan un amplio repertorio de secuencias de ADN repetido (Cioffi & Bertollo, 2012), estando la proporción de los mismos correlacionada positivamente con el tamaño del genoma (Shao & Peng, 2019). En este trabajo, demostramos que el ensamblaje más reciente del genoma del lenguado senegalés está compuesto por un 32,98% de elementos repetidos, donde la fracción más representada corresponde a los transposones de ADN con un 11,76% (ver **Tabla 1**), seguida de las LINEs con un porcentaje de 7,75% y los elementos LTR con un porcentaje de 7,34%. La cantidad de elementos repetidos del lenguado senegalés es muy superior al de otras especies de peces planos como el rodaballo, *S. maximus*, el cual tiene un 8,5% de elementos repetidos en su genoma (Figueras et al., 2016) y el lenguado chino, *C. semilaevis*, con una representación de elementos transponibles del 5,8% (Chen et al., 2014). Estas discrepancias en contenidos de elementos repetidos podrían deberse a que, en este estudio, se ha utilizado una base de datos de elementos repetidos específica de peces planos con 39.064 secuencias y se han podido anotar un mayor número de secuencias que con bases de datos estándar de RepBase, utilizadas en los casos del rodaballo y el lenguado chino. Sin embargo, la cantidad de elementos repetidos que se ha detectado no es una cantidad alta comparada con el 52,2% que se ha encontrado en la especie modelo de pez *Danio rerio* (Howe et al., 2013). Un estudio

reciente, en el que se han comparado 52 genomas de peces, demuestra que los elementos transponibles de clase II son más abundantes en peces óseos de agua dulce, mientras que las secuencias repetidas en tándem son más comunes en los peces óseos marinos. Además, en este mismo estudio se observó que los transposones de clase I son más abundantes en especies primitivas, como los peces cartilaginosos y las lampreas, que en los peces óseos (Yuan et al., 2018). En contraste con estos datos, en el lenguado senegalés encontramos que existe una gran diversidad de clases diferentes de elementos repetidos que suponen una porción considerable del genoma, donde existe una predominancia de los elementos transponibles de clase I, representando un 15,5% del genoma frente a los de clase II, representando un 12,53% del genoma. Las cuestiones evolutivas de la predominancia de un tipo de elementos frente a otros en *S. senegalensis* aún están por esclarecer.

En cuanto a las secuencias repetidas en tándem presentes en genomas eucariotas, estas suelen tratarse de: (1) familias de ADN satélite, como los que ocupan las regiones heterocromáticas, en general, o las regiones diferenciales de cromosomas sexuales, o regiones centroméricas y teloméricas, en particular, (2) de secuencias asociadas a familias multigénicas, como los diferentes ARN ribosómicos o (3) de secuencias distribuidas de forma dispersa a lo largo del genoma y clústerizadas en localizaciones concretas. En este sentido, los elementos con mayor representación en el genoma del lenguado senegalés detectados mediante análisis de clústerización se corresponden con secuencias repetidas en tándem (ver **Tabla 2**).

Los cromosomas sexuales representan otra de las regiones genómicas en las que es frecuente encontrar acumulación de secuencias de ADN satélite. Los sistemas de determinación sexual en peces son complejos y dinámicos, seguramente debido a condiciones ambientales poco estables (Brykov, 2014). Las secuencias repetidas en tándem desempeñan un papel importante en el establecimiento y evolución de los sistemas de cromosomas sexuales en peces. Por ejemplo, en el pez guppy (*Poecilia reticulata*) se han realizado estudios que revelan la presencia de secuencias repetidas en tándem específicas del cromosoma Y (Nanda et al., 1990). En el caso de la especie *Megaleporinus elongatus*, que posee un sistema de cromosomas sexuales múltiples



$Z_1Z_1Z_2Z_2/Z_1W_1Z_2W_2$ , se observaron diferencias en la composición de ADN satélite entre los sexos. Estas secuencias satélites eran aproximadamente un 14% más abundantes en el genoma de las hembras y se encontró que su enriquecimiento estaba asociado con el cromosoma  $W_1$  (Crepaldi et al., 2021). El análisis comparativo de la acumulación de secuencias repetidas en machos y hembras de *S. senegalensis* no mostró diferencias apreciables en el contenido de elementos repetidos entre ambos sexos. Pese a presentar un sistema de determinación sexual XX/XY (Molina-Luzón et al., 2015), no hemos encontrado la acumulación selectiva de ninguna familia de ADN satélite en un sexo concreto en el cromosoma 1, par sexual propuesto en estudios anteriores (García et al., 2019), ni en el cromosoma 12 donde se encuentran el gen *fshr* determinante del sexo en esta especie (de la Herrán et al., 2023) ni en alguno de los otros pares autosómicos. De hecho, se ha encontrado aquí que la proporción de repetidos entre los cromosomas de *S. senegalensis* es homogénea (**Figura 5**). De acuerdo con la morfología cromosómica (**Figura 2**), no existiría una clara heteromorfía del par sexual, lo que podría indicar un escaso grado de diferenciación entre los cromosomas X e Y a nivel molecular. Esto queda corroborado en el descubrimiento por de la Herrán et al 2023 de que el gen determinante del sexo en esta especie es el receptor de la hormona estimulante del folículo (*fshr*), cuyas diferencias entre ambos sexos corresponden solo a varios SNPs, indicando un incipiente mecanismo molecular en la diferenciación sexual.

Respecto a las secuencias repetidas en tándem compuestas por genes ribosómicos, es común encontrar en los genomas de peces varios clústers de ARNr 45S, en algunos casos hasta 10 - 15 loci (Gromicho et al., 2006; Cioffi et al., 2009). Estos genes ribosómicos se suelen encontrar en los extremos de los cromosomas muy próximos a los telómeros. En el lenguado senegalés, Cross y colaboradores (2006) determinaron que se encuentran acumulados en el brazo corto del par cromosómico 5. En nuestro trabajo, no sólo no hemos podido identificar dentro del ensamblaje los genes ribosómicos 45S sino que, además, atendiendo a las coordenadas centroméricas, la morfología del par cromosómico 5 (telocéntrico), no sería coherente con lo descrito a nivel citogenético (submetacéntrico). En el ensamblaje utilizado en nuestros análisis (de la Herrán et al., 2023) esta secuencia no habría sido ensamblada. De hecho, de los

31 contigs sin anclar al ensamblaje, tres están compuestos mayoritariamente por ARNr 45S (ver **Tabla Sup. 2** con el análisis composicional de los contigs sin emplazar). Nuestra hipótesis es que en el ensamblaje del cromosoma 5 se ha perdido casi la totalidad del brazo corto, que incluye este locus de ARNr, alterando de esta forma la morfología cromosómica. Por otra parte, los genes ribosómicos 5S, se suelen localizar de forma intersticial en los cromosomas (Gromicho et al., 2006; Cioffi et al., 2009). En este trabajo hemos encontrado la presencia de genes ribosómicos 5S en el cromosoma 6, tal y como está descrito en la bibliografía (Cross et al., 2006). Adicionalmente, hemos encontrado señales de genes ribosómicos 5S en el cromosoma 11 y en el 20. Estos loci podrían haber pasado desapercibidos en la hibridación *in situ* o tratarse de reorganizaciones cromosómicas o de secuencias ensambladas incorrectamente.

En cuanto a otro tipo de secuencias en tándem, el clúster 1 se corresponde con un elemento que contiene un ARN<sub>t</sub>, cuya estructura ha sido descrito por primera vez en esta memoria (ver **Figura 7**). Esta secuencia aparece repartida por todo el genoma de *S. senegalensis* sin un patrón concreto, pero se ha localizado de forma intercalada con las secuencias de la familia *PvuII* en el cromosoma 4 (ver **Tabla 3**). El clúster 2, también con distribución dispersa, mostró similitud con un elemento de tipo Gypsy. Las secuencias consenso del clúster 1 y 2, a pesar de presentar 89,5% de identidad entre ellas, exhiben localizaciones distintas y se encuentran clústerizadas de forma independiente. Esto podría deberse a su asociación con elementos móviles diferentes. Otra explicación posible es que la secuencia que presenta homología con la secuencia consenso del clúster 1, contenga un ARN<sub>t</sub> implicado en la expansión de un LTR, el elemento Gypsy relacionado con las secuencias del clúster 2. Es conocido que los ARN<sub>t</sub> pueden actuar como moléculas cebadoras para guiar la transcripción inversa de los retrotransposones y, además, también se han encontrado que moléculas de ARN derivadas de los ARN<sub>t</sub> regulan el proceso de retrotransposición inhibiéndolo o fomentándolo (Martinez, 2017). Esto explicaría el patrón de localización de ambas repeticiones de 38 pb y el hecho de que una sea la reversa complementaria de la otra.

### *Regiones centroméricas*

Se ha demostrado que una de las localizaciones más frecuentes del ADN satélite de peces se corresponde con las regiones centroméricas. Esto sugeriría que desempeñan un papel importante en la estructura y función de los centrómeros (Martins, 2007). El clúster 3 resultó ser el satélite *PvuII*, presente en la mayoría de los centrómeros (Robles et al., 2017). Se han identificado familias de ADN satélite centroméricas en esturión (familia de ADN satélite *HindIII*) y en el pez *Hoplias malabaricus* (ADN satélite *5SHindIII-DNA*), perteneciente a la familia Erythrinidae, que presentan varias familias de ADN satélite de manera exclusiva en la región centromérica (Lanfredi et al., 2001; Vicari et al., 2010). En el caso de los peces planos, se han documentado varias familias de ADN satélite específicos de las regiones centroméricas. En el lenguado *A. lineatus* se describió la secuencia de ADN satélite *HindIII* (Carvalho de Azevedo et al., 2005), en la acedía, *D. cuneata*, la familia *DraI* (de la Herrán et al., 2008), y en el halibut, *H. hippoglossus*, la familia de ADN satélite *SacI* (Robles et al., 2017). En nuestra especie de estudio, hemos encontrado que las secuencias de ADN satélite también se distribuyen preferentemente en las regiones pericentroméricas y centroméricas por lo que deducimos que, como se ha observado en los casos anteriores, este tipo de secuencias podrían ser relevantes para el correcto funcionamiento de las regiones centroméricas.

En relación con el lenguado senegalés, a partir de la digestión de ADN genómico con enzimas de restricción e hibridación *in situ*, Robles y colaboradores (2017) describieron la familia de ADN satélite *PvuII* localizada en los centrómeros de 19 parejas de las 21 de *S. senegalensis* y dos especies estrechamente emparentadas, *S. solea* y *S. lascari*. En este trabajo, nosotros hemos confirmado *in silico*, mediante BLAST, la presencia de *PvuII* en las regiones centroméricas de 17 de los 21 cromosomas del genoma haploide de *S. senegalensis*. Atendiendo a nuestros datos, solo los pares 7, 8, 14 y 18 carecerían de esta secuencia. Mientras que en todos los casos en los que se ha detectado *PvuII* este se encuentra en una única región, el caso del cromosoma 19 es distinto y presentaría dos regiones de acumulación de este satélite.

Para caracterizar las regiones centroméricas de los cromosomas 7, 8, 14 y 18 hemos explorado en más detalle secuencias repetidas muy representadas y secuencias Helitron y LINEs, propuestas en otras especies como candidatas a ocupar las regiones

centroméricas y pericentroméricas (Cvitanić et al., 2020; de Sotero-Caio et al., 2017). Así, en el par cromosómico 7 hemos encontrado la acumulación específica de una familia de ADN satélite descrita aquí por primera vez, que hemos denominado rep87. Este ADN satélite se encontró de forma contigua a una región de Helitron y zonas próximas de LINEs y Gypsy, como en otros cromosomas y el satélite *PvuII*. Por su parte, en el par cromosómico 8 encontramos la acumulación específica de la familia de ADN satélite rep120 no muy alejada de regiones de Helitron y el elemento repetido descrito en esta memoria el cual incluye un gen de ARN<sub>t</sub> (secuencia del clúster 1). No se encontró la presencia de la secuencia LINE/L1 relacionada con el *PvuII* en el ensamblaje de los cromosomas 7 y 8. Tanto el satélite *PvuII* como rep87 y rep120, muestran repeticiones internas que permiten la formación de estructuras secundarias las cuales podrían tener un papel fundamental en la función centromérica como se ha descrito recientemente (Kasinathan & Henikoff, 2018).

Las coordenadas de estas secuencias en tándem (*PvuII*, rep87 y rep120) fueron consideradas como referencia para acotar las regiones centroméricas. Así las cosas, en cuanto a localización se refiere, todos los resultados obtenidos en esta memoria son coherentes con la morfología cromosómica descrita para esta especie (**Figura 12**), excepto en el caso de los cromosomas 5, por los motivos comentados anteriormente, y 19.

En cuanto al cromosoma 19, cuyos extremos están ocupados por secuencias de *PvuII*, observamos cómo en un extremo se encuentran, además, los telómeros en posición interna con secuencias *PvuII* aguas abajo (ver cromosoma 19 en **Figura 12**). En el otro extremo se encuentran también ambas secuencias, pero en posición invertida, esto es secuencias *PvuII* en posición interna y teloméricas en posición terminal. Con los datos disponibles, nos es imposible determinar cuál de las dos regiones se corresponde con la región centromérica.

Cuando analizamos en profundidad las secuencias de los cromosomas 14 y 18, no detectamos la presencia de ninguna secuencia centromérica candidata. Además, en el par 18 no detectamos secuencias teloméricas. En el 14 no se encontraron secuencias teloméricas terminales, aunque sí de forma intercalar en posiciones más internas del

cromosoma. Si atendemos a la caracterización llevada a cabo aquí y a los datos de Robles y colaboradores (2017), estos cromosomas son candidatos a ser portadores de secuencias *PvuII*. Sin embargo, estas no se detectarían a consecuencia de problemas del ensamblaje de esta región, como los datos mencionados apuntarían.

A la luz de los datos aquí expuestos, las regiones centroméricas de *S. senegalensis* están ocupadas mayoritariamente por elementos repetidos en tándem. Las secuencias de ADN satélite centromérico específicas de especies, o como en el caso de los esturiones, de un grupo de especies, podrían explicarse por un proceso de homogeneización del ADN satélite, necesario para el funcionamiento y la estabilidad de los centrómeros. De esta forma, y a medida que las secuencias de ADN satélite divergen considerablemente entre las diferentes especies, se pueden formar familias distintas de ADN satélite centromérico en cada una de ellas (Plohl et al., 2012).

Hemos detectado la asociación de rep87 y rep120 (satélites exclusivos de los centrómeros de los cromosomas 7 y 8 respectivamente) con elementos móviles LTR, tipo Helitron y el elemento repetido que incluye el ARNt que está relacionado con el clúster 1 (Tabla 3). Además, hemos encontrado que el ADN satélite *PvuII* tiene una alta identidad con un elemento móvil tipo LINE/L1. Comparando la distribución del satélite *PvuII* con la de la LINE, se observa una clara asociación que, además, está restringida a los centrómeros. Esto no es un caso aislado ya que está documentada la presencia de elementos transponibles Rex1, Rex3 y Rex6 exclusivos de las regiones heterocromáticas centroméricas en algunas especies de la familia de peces Cichlidae (Valente et al., 2011) y se ha hipotetizado en ostras con el origen de una secuencia centromérica repetida en tándem (familia *HindIII*) a partir de un elemento móvil denominado Perl de la clase TRIP (López-Flores et al., 2004).

## 6. Conclusiones

En el presente trabajo se ha realizado un análisis bioinformático del contenido de elementos repetidos del lengüado senegalés que, además, sirvió para caracterizar sus regiones centroméricas. Las principales conclusiones de este trabajo son:

- Los análisis con RepeatMasker evidenciaron que el genoma de *S. senegalensis* está compuesto por un 32,98% de elementos repetidos y los más frecuentes son transposones de ADN con un porcentaje de 11,76%, las LINEs con un porcentaje de 7,75% y los elementos LTR con un porcentaje de 7,34%.

- No existen diferencias significativas en el contenido de secuencias repetidas entre los genomas de machos y hembras, pudiendo estar relacionado con el incipiente estado de diferenciación sexual a nivel molecular en esta especie.

- Los centrómeros de *S. senegalensis* están ocupados principalmente por secuencias de ADN satélite: las regiones centroméricas de los cromosomas del 1 al 6, del 9 al 13, del 15 al 17, 20 y 21 estarían ocupadas por la familia de ADN satélite *PvuII*, las del cromosoma 7 por la familia rep87 y la del cromosoma 8 por la familia rep120.

- Las familias de ADN satélite *PvuII*, rep87 y rep120 se encuentran asociadas o tienen un origen común con elementos transponibles. En concreto, la secuencia *PvuII* presenta homología con una LINE/L1, y las secuencias rep87 y rep120 aparecieron próximas a elementos móviles de tipo Helitron y el elemento repetido con el gen de ARN<sub>t</sub> descrito en este trabajo, respectivamente.

- La localización *in silico* de las secuencias centroméricas aquí caracterizadas es coherente con los datos existentes sobre morfología cromosómica de esta especie, excepto en el cromosoma 5 y en aquellos cromosomas para los que no hemos podido identificar de forma inequívoca los centrómeros (14, 18 y 19).

- El par cromosómico 5 del ensamblaje disponible en la actualidad y que hemos usado en nuestros análisis, carece del brazo corto, portador del locus de ARNr 45S, siendo esta la explicación de la discrepancia detectada a nivel morfológico.

- Las regiones centroméricas de los cromosomas 14 y 18 no han sido caracterizadas a consecuencia de los problemas con el ensamblaje de dichas regiones en estos cromosomas. Hipotetizamos que son regiones en las que estaría presente el ADN satélite *PvuII*.

- La posición de la región centromérica del cromosoma 19 permanece incierta, ya que existen dos acumulaciones de secuencias *PvuII* en ambos extremos cromosómicos.

## 7. Perspectivas futuras

Con la información recopilada en este trabajo, se propone completar el estudio bioinformático con técnicas de genética molecular y citogenética. Podría ser interesante realizar hibridaciones *in situ* de las secuencias relevantes como los satélites específicos de los cromosomas 7 y 8 (rep87 y rep120 respectivamente) con el objetivo de alcanzar una mejor comprensión de las regiones centroméricas de los cromosomas del lenguado senegalés. También, podría ser interesante realizar una resecuenciación del genoma para conseguir un mayor número de lecturas para el ensamblaje. Esto podría solucionar los problemas aquí detectados, como la pérdida del brazo corto del cromosoma 5 y la ausencia del satélite centromérico *PvuII* en los cromosomas 14 y 18. De forma adicional, la estructura y composición genómica descrita en este trabajo podrían suponer una base o un punto de partida para una mejor comprensión de la biología de esta especie y su aprovechamiento en piscifactorías.

## 8. Bibliografía

- Altschup, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215(3).  
[https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Beirão, J., Cabrita, E., Soares, F., Herráez, M. P., & Dinis, M. T. (2008). Cellular damage in spermatozoa from wild-captured *Solea senegalensis* as detected by two different assays: Comet analysis and Annexin V-Fluorescein staining. *Journal of Applied Ichthyology*, 24(4), 508–513.  
<https://doi.org/10.1111/j.1439-0426.2008.01144.x>
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, 27(2), 573–580.  
<https://academic.oup.com/nar/article/27/2/573/1061099>

- Bernat Gel & Eduard Serra. (2017). karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*, 33(19), 3088–3090. <https://doi.org/10.1093/bioinformatics/btx346>
- Brykov, V. A. (2014). Mechanisms of sex determination in fish: Evolutionary and practical aspects. *Russian Journal of Marine Biology*, 40(6), 407–417. <https://doi.org/10.1134/S1063074014060145>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10, 421. <https://doi.org/10.1186/1471-2105-10-421>
- Cañavate, J.P., 2009. Relevancia de la reproducción en la diversificación de la piscicultura marina en España, La reproducción de los peces: aspectos básicos y sus aplicaciones en acuicultura. *Publicaciones Científicas y Tecnológicas de la Fundación Observatorio Español de Acuicultura*, 597-642. <https://www.observatorio-acuicultura.es/recursos/publicaciones/la-reproduccion-de-los-peces-aspectos-basicos-y-sus-aplicaciones-en>
- Carvalho De Azevedo, M. F., Oliveira, C., Martins, C., Wasko, A. P., & Foresti, F. (2005). Isolation and characterization of a satellite DNA family in *Achirus lineatus* (Teleostei: Pleuronectiformes: Achiridae). *Genetica*, 125(23), 205–210. <https://doi.org/10.1007/s10709-005-8419-0>
- Chairi, H., Fernández-Díaz, C., Navas, J. I., Manchado, M., Rebordinos, L. & Blasco, J., 2010. In vivo genotoxicity and stress defences in three flatfish species exposed to CuSO<sub>4</sub>. *Ecotox Environ Safe*, 73, 1279-1285. <https://doi.org/10.1016/j.ecoenv.2010.07.028>
- Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., Song, W., An, N., Chalopin, D., Volff, J. N., Hong, Y., Li, Q., Sha, Z., Zhou, H., Xie, M., Yu, Q., Liu, Y., Xiang, H., Wang, N., ... Wang, J. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nature Genetics*, 46(3), 253–260. <https://doi.org/10.1038/ng.2890>



- Cioffi, M. B., & Bertollo, L. A. C. (2012). Chromosomal Distribution and Evolution of Repetitive DNAs in Fish the Repetitive Fraction of the Genome. *Genome dynamics*, 7, 197–221. <https://doi.org/10.1159/000337950>
- Cioffi, M. B., Martins, C., & Bertollo, L. A. C. (2009). Comparative chromosome mapping of repetitive sequences. Implications for genomic evolution in the fish, *Hoplias malabaricus*. *BMC Genetics*, 10, 34. <https://doi.org/10.1186/1471-2156-10-34>
- Crepaldi, C., Martí, E., Gonçalves, É. M., Martí, D. A., & Parise-Maltempi, P. P. (2021). Genomic differences between the sexes in a fish species seen through satellite DNAs. *Frontiers in Genetics*, 12, 1-11. <https://doi.org/10.3389/fgene.2021.728670>
- Cross, I., Merlo, A., Manchado, M., Infante, C., Cañavate, J. P., & Rebordinos, L. (2006). Cytogenetic characterization of the sole *Solea senegalensis* (Teleostei: Pleuronectiformes: Soleidae): Ag-NOR, (GATA)<sub>n</sub>, (TTAGGG)<sub>n</sub> and ribosomal genes by one-color and two-color FISH. *Genetica*, 128(1-3), 253–259. <https://doi.org/10.1007/s10709-005-5928-9>
- Cvitanić, M. T., Zeljko, T. V., Pasantes, J. J., García-Souto, D., Gržan, T., Despot-Slade, E., Plohl, M., & Šatović, E. (2020). Sequence composition underlying centromeric and heterochromatic genome compartments of the pacific oyster *Crassostrea gigas*. *Genes*, 11(6), 1–20. <https://doi.org/10.3390/genes11060695>
- de la Herrán, R., Hermida, M., Rubiolo, J. A., Gómez-Garrido, J., Cruz, F., Robles, F., Navajas-Pérez, R., Blanco, A., Villamayor, P. R., Torres, D., Sánchez-Quinteiro, P., Ramirez, D., Rodríguez, M. E., Arias-Pérez, A., Cross, I., Duncan, N., Martínez-Peña, T., Ríaza, A., Millán, A., ... Martínez, P. (2023). A chromosome-level genome assembly enables the identification of the follicle stimulating hormone receptor as the master sex-determining gene in the flatfish *Solea senegalensis*. *Molecular Ecology Resources*, 23(4), 886–904. <https://doi.org/10.1111/1755-0998.13750>
- de la Herrán, R., Robles, F., Navas, J. I., López-Flores, I., Herrera, M., Hachero, I., Garrido-Ramos, M. A., Ruiz Rejón, C., & Ruiz Rejón, M. (2008). The centromeric satellite of the wedge sole (*Dicologlossa cuneata*,

- Pleuronectiformes) is composed mainly of a sequence motif conserved in other vertebrate centromeric DNAs. *Cytogenetic and Genome Research*, 121(3–4), 271–276. <https://doi.org/10.1159/000138896>
- de Sotero-Caio, C. G., Cabral-de-Mello, D. C., Calixto, M. da S., Valente, G. T., Martins, C., Loreto, V., de Souza, M. J., & Santos, N. (2017). Centromeric enrichment of LINE-1 retrotransposons and its significance for the chromosome evolution of Phyllostomid bats. *Chromosome Research*, 25(3–4), 313–325. <https://doi.org/10.1007/s10577-017-9565-9>
- Ferreira, D. C., Porto-Foresti, F., Oliveira, C., & Foresti, F. (2011). Transposable elements as a potential source for understanding the fish genome. *Mobile Genetic Elements*, 1(2), 112–117. <https://doi.org/10.4161/mge.1.2.16731>
- Figueras, A., Robledo, D., Corvelo, A., Hermida, M., Pereiro, P., Rubiolo, J. A., Gómez-Garrido, J., Carreté, L., Bello, X., Gut, M., Gut, I. G., Marcet-Houben, M., Forn-Cuní, G., Galán, B., García, J. L., Abal-Fabeiro, J. L., Pardo, B. G., Taboada, X., Fernández, C., ... Martínez, P. (2016). Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): A fish adapted to demersal life. *DNA Research*, 23(3), 181–192. <https://doi.org/10.1093/dnares/dsw007>
- Forné, I., María, J. A., Esther Asensio, Joaquín Abián, & Cerdà, J. (2009). 2-D DIGE analysis of Senegalese sole (*Solea senegalensis*) testis proteome in wild-caught and hormone-treated F1 fish. *Proteomics*, 9(8), 2171–2181. <https://doi.org/10.1002/pmic.200800696>
- Friedman, M. (2008). The evolutionary origin of flatfish asymmetry. *Nature*, 454(7201), 209–212. <https://doi.org/10.1038/nature07108>
- García, E., Cross, I., Portela-Bens, S., Rodríguez, M. E., García-Angulo, A., Molina, B., Cuadrado, A., Liehr, T., & Rebordinos, L. (2019). Integrative genetic map of repetitive DNA in the sole *Solea senegalensis* genome shows a Rex transposon located in a proto-sex chromosome. *Scientific Reports*, 9(17146). <https://doi.org/10.1038/s41598-019-53673-6>
- García-López, A., Anguis, V., Couto, E., Canario, A. V. M., Cañavate, J. P., Sarasquete, C., & Martínez-Rodríguez, G. (2006). Non-invasive assessment of

- reproductive status and cycle of sex steroid levels in a captive wild broodstock of Senegalese sole *Solea senegalensis* (Kaup). *Aquaculture*, 254(1–4), 583–593.  
<https://doi.org/10.1016/j.aquaculture.2005.10.007>
- Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., & Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*, 38. <https://doi.org/10.1093/nar/gkq313>
- Gromicho, M., Coelho, M. M., Alves, M. J., & Collares-Pereira, M. J. (2006). Cytogenetic analysis of *Anaocypris hispanica* and its relationship with the paternal ancestor of the diploid-polyploid *Squalius alburnoides* complex. *Genome*, 49(12), 1621–1627. <https://doi.org/10.1139/G06-121>
- Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J. C., Koch, R., Rauch, G. J., White, S., ... Stemple, D. L. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496(7446), 498–503.  
<https://doi.org/10.1038/nature12111>
- Imsland, A.K., Foss, A., Conceição, L.E.C., Dinis, M.T., Delbare, D., Schram, E., Kamstra, A., Rema, P., & White, P., 2004. A review of the culture potential of *Solea solea* and *S. senegalensis*. *Reviews in Fish Biology and Fisheries*, 13, 379–407. <https://doi.org/10.1007/s11160-004-1632-6>
- Jaillon, O., Aury, J.-M., de Ric Brunet, F., Petit, J.-L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Cille Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Atrice Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., ... Roest Crollius, H. (2004). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*, 431(7011), 946–957. <https://doi.org/10.1038/nature03025>
- Ji, X. S., Chen, S. L., Liao, X. L., Yang, J. F., Xu, T. J., Ma, H. Y., Tian, Y. S., Jiang, Y. L., & Wu, P. F. (2009). Microsatellite-centromere mapping in *Cynoglossus semilaevis* using gynogenetic diploid families produced by the use of homologous and non-homologous sperm. *Journal of Fish Biology*, 75(2), 422–434. <https://doi.org/10.1111/j.1095-8649.2009.02257.x>

- Kasinathan, S., & Henikoff, S. (2018). Non-B-form DNA is enriched at centromeres. *Molecular Biology and Evolution*, 35(4), 949–962.  
<https://doi.org/10.1093/molbev/msy010>
- Krumsiek, J., Arnold, R., & Rattei, T. (2007). Gepard: A rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics*, 23(8), 1026–1028.  
<https://doi.org/10.1093/bioinformatics/btm039>
- Lanfredi, M., Congiu, L., Garrido-Ramos, M. A., de la Herrán, R., Leis, M., Chicca, M., Rossi, R., Tagliavini, J., Rejón, C. R., Rejón, M. R., & Fontana, F. (2001). Chromosomal location and evolution of a satellite DNA family in seven sturgeon species. *Chromosome Research*, 9(1), 47–52.  
<https://doi.org/10.1023/A:1026739616749>
- López-Flores, I., de La Herrán, R., Garrido-Ramos, M. A., Boudry, P., Ruiz-Rejón, C., & Ruiz-Rejón, M. (2004). The molecular phylogeny of oysters based on a satellite DNA related to transposons. *Gene*, 339(1–2), 181–188.  
<https://doi.org/10.1016/J.GENE.2004.06.049>
- Lorenz, R., Bernhart, S. H., Höner Zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6, 26. <http://www.tbi.univie.ac.at/RNA>.
- Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25(5), 955–964. <https://doi.org/10.1093/nar/25.5.955>
- Martinez, G. (2017). tRNAs as primers and inhibitors of retrotransposons. *Mobile Genetics Elements*, 7(5), 1–6. <https://doi.org/10.1080/2159256X.2017.1393490>
- Martins, Cesar. (2007). Chromosomes and Repetitive DNAs: A Contribution to the Knowledge of the Fish Genome. CRC Press. *Fish Cytogenetics* (pp.421–453)  
<http://dx.doi.org/10.1201/b10746-16>
- Mason, J. M., & Biessmann, H. (1995). The unusual telomeres of *Drosophila*. *Trends in Genetics*, 11(2), 58–62. [https://doi.org/10.1016/S0168-9525\(00\)88998-2](https://doi.org/10.1016/S0168-9525(00)88998-2)
- Melters, D. P., Bradnam, K. R., Young, H. A., Telis, N., May, M. R., Ruby, J. G., Sebra, R., Peluso, P., Eid, J., Rank, D., Garcia, J. F., DeRisi, J. L., Smith, T., Tobias, C., Ross-Ibarra, J., Korf, I., & Chan, S. W. L. (2013). Comparative analysis of

- tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biology*, 14(1). <https://doi.org/10.1186/gb-2013-14-1-r10>
- Meyer, A., & Scharl, M. (1999). Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Current Opinion in Cell Biology*, 11(6), 699-704. [https://doi.org/10.1016/s0955-0674\(99\)00039-3](https://doi.org/10.1016/s0955-0674(99)00039-3)
- Meyer, A., & van de Peer, Y. (2005). From 2R to 3R: Evidence for a fish-specific genome duplication (FSGD). *BioEssays*, (27(9), 937-945. <https://doi.org/10.1002/bies.20293>
- Molina-Luzón, M. J., López, J. R., Robles, F., Navajas-Pérez, R., Ruiz-Rejón, C., de la Herrán, R., & Navas, J. I. (2015). Chromosomal manipulation in Senegalese sole (*Solea senegalensis* Kaup, 1858): induction of triploidy and gynogenesis. *Journal of Applied Genetics*, 56(1), 77-84. <https://doi.org/10.1007/s13353-014-0233-x>
- Morais, S., Aragão, C., Cabrita, E., Conceição, L.E.C., Constenla, M., Costas, B., Dias, J., Duncan, N., Engrola, S., Estevez, A., Gisbert, E., Mañanós, E., Valente, L.M.P., Yúfera, M., & Dinis, M.T. (2014). New developments and biological insights into the farming of *Solea senegalensis* reinforcing its aquaculture potential. *Reviews in Aquaculture*, 6, 1-37. <https://doi.org/10.1111/raq.12091>
- Nanda, I., Feichtinger, W., Schmid, M., Schröder, J. H., Zischler, H., & Epplen, J. T. (1990). Simple repetitive sequences are associated with differentiation of the sex chromosomes in the guppy fish. *Journal of Molecular Evolution*, 30, 456-462. <https://doi.org/10.1007/BF02101117>
- Novák, P., Neumann, P., & Macas, J. (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nature Protocols*, 15(11), 3745-3776. <https://doi.org/10.1038/s41596-020-0400-y>
- Novák, P., Neumann, P., Pech, J., Steinhaisl, J., & Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic

- repetitive elements from next generation sequence reads. *Bioinformatics*, 29(6), 792-793. <https://doi.org/10.1093/bioinformatics/btt054>
- Ohno, S. (1970). The Enormous Diversity in Genome Sizes of Fish as a Reflection of Nature's Extensive Experiments with Gene Duplication. *Transactions of the American Fisheries Society*, 99(1), 120–130. [https://doi.org/10.1577/1548-8659\(1970\)99%3C120:TEDIGS%3E2.0.CO;2](https://doi.org/10.1577/1548-8659(1970)99%3C120:TEDIGS%3E2.0.CO;2)
- Oliveira, C., Chew, J. S., Porto-Foresti, F., Dobson, M. J., & Wright, J. M. (1999). A LINE2 repetitive DNA sequence from the cichlid fish, *Oreochromis niloticus*: Sequence analysis and chromosomal distribution. *Chromosoma*, 108(7), 457-468. <https://doi.org/10.1007/s004120050397>
- Pardue M. L., Danilevskaya O. N., Traverse K. L., & Lowenhaupt K. (1997) Evolutionary links between telomeres and transposable elements. *Genetics*, 100(1-3), 73-84. [https://doi.org/10.1007/978-94-011-4898-6\\_7](https://doi.org/10.1007/978-94-011-4898-6_7)
- Plohl, M., Meštrović, N., & Mravinac, B. (2012). Satellite DNA Evolution. *Genome Dynamics*, 7, 126-152 <https://doi.org/10.1159/000337122>
- Ravi, V., & Venkatesh, B. (2008). Rapidly evolving fish genomes and teleost diversity. *Current Opinion in Genetics and Development*, 18(6), 544–550. <https://doi.org/10.1016/j.gde.2008.11.001>
- Robles, F. R., de La Herrán, R., Navajas-Pérez, R., Cano-Roldán, B., Sola-Campoy, P. J., García-Zea, J. A., & Rejón, C. R. (2017). Centromeric satellite DNA in flatfish (Order Pleuronectiformes) and its relation to speciation processes. *Journal of Heredity*, 108(2), 217–222. <https://doi.org/10.1093/jhered/esw076>
- Rodríguez, M. E., Cross, I., Arias-Pérez, A., Portela-Bens, S., Merlo, M. A., Liehr, T., & Rebordinos, L. (2021). Cytogenomics unveil possible transposable elements driving rearrangements in chromosomes 2 and 4 of *Solea senegalensis*. *International Journal of Molecular Sciences*, 22(4), 1–17. <https://doi.org/10.3390/ijms22041614>
- Shao, C., Bao, B., Xie, Z., Chen, X., Li, B., Jia, X., Yao, Q., Ortí, G., Li, W., Li, X., Hamre, K., Xu, J., Wang, L., Chen, F., Tian, Y., Schreiber, A. M., Wang, N., Wei, F., Zhang, J., ... Chen, S. (2017). The genome and transcriptome of Japanese

- flounder provide insights into flatfish asymmetry. *Nature Genetics*, 49(1), 119–124. <https://doi.org/10.1038/ng.3732>
- Shao, F., Han, M., & Peng, Z. (2019). Evolution and diversity of transposable elements in fish genomes. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-51888-1>
- Symonová, R., & Suh, A. (2019). Nucleotide composition of transposable elements likely contributes to AT/GC compositional homogeneity of teleost fish genomes. *Mobile DNA*, 10(1). <https://doi.org/10.1186/s13100-019-0195-y>
- Talbert, P. B., & Henikoff, S. (2020). What makes a centromere?. *Experimental Cell Research*, 389(2), 111895. <https://doi.org/10.1016/j.yexcr.2020.111895>
- Valente, G. T., Mazzuchelli, J., Ferreira, I. A., Poletto, A. B., Fantinatti, B. E. A., & Martins, C. (2011). Cytogenetic mapping of the retroelements Rex1, Rex3 and Rex6 among cichlid Fish: new insights on the chromosomal distribution of transposable elements. *Cytogenetic and Genome Research*, 133(1), 34–42. <https://doi.org/10.1159/000322888>
- Vega, L., Díaz, E., Cross, I., & Rebordinos, L. (2002). Caracterizaciones citogenética e isoenzimática del lenguado *Solea senegalensis* Kaup, 1858. *Boletín del Instituto Español de Oceanografía*, 18(4), 245–250. <http://www.repositorio.ieo.es/e-ieo/handle/10508/1213>
- Vicari, M. R., Nogaroto, V., Noleto, R. B., Cestari, M. M., Cioffi, M. B., Almeida, M. C., Moreira-Filho, O., Bertollo, L. A. C., & Artoni, R. F. (2010). Satellite DNA and chromosomes in Neotropical fishes: Methods, applications and perspectives. *Journal of Fish Biology*, 76(5), 1094–1116. <https://doi.org/10.1111/j.1095-8649.2010.02564.x>
- Wessler, S. R. (2006). Transposable elements and the evolution of eukaryotic genomes. *PNAS*, 103(47), 17600–17601. <https://doi.org/10.1073/pnas.0607612103>
- Wheeler, T. J., Clements, J., Eddy, S. R., Hubley, R., Jones, T. A., Jurka, J., Smit, A. F. A., & Finn, R. D. (2013). Dfam: A database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Research*, 41(1). <https://doi.org/10.1093/nar/gks1265>



- Wickham H (2016). Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Xing, J., Ke, Z., Liu, L., Li, C., Gong, X., & Bao, B. (2020). Eye location, cranial asymmetry, and swimming behavior of different variants of *Solea senegalensis*. Aquaculture and Fisheries, 5(4), 182–186.  
<https://doi.org/10.1016/j.aaf.2019.11.003>
- Yuan, Z., Liu, S., Zhou, T., Tian, C., Bao, L., Dunham, R., & Liu, Z. (2018). Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. BMC Genomics, 19(1). <https://doi.org/10.1186/s12864-018-4516-1>

## 9. Anexo

- **Tabla suplementaria 1:**  
[https://github.com/AaronGS1999/TFM/blob/1025ed8fc90b36923befe6edb57b817012a9da30/Anexo/anexo\\_1\\_anotacionRE\\_230623.xlsx](https://github.com/AaronGS1999/TFM/blob/1025ed8fc90b36923befe6edb57b817012a9da30/Anexo/anexo_1_anotacionRE_230623.xlsx)
- **Tabla suplementaria 2:**  
[https://github.com/AaronGS1999/TFM/blob/984405eea48c490308875ff4ec0f6a6b364fabdb/Anexo/unplaced\\_scaffolds.xlsx](https://github.com/AaronGS1999/TFM/blob/984405eea48c490308875ff4ec0f6a6b364fabdb/Anexo/unplaced_scaffolds.xlsx)
- **Scripts de Python:**
  - [https://github.com/AaronGS1999/TFM/blob/8cbbe9e9a3addfb88fc94a71a3cfdc46852e4b87/Python\\_Scripts/BlastSeqFreqAnalyzer.py](https://github.com/AaronGS1999/TFM/blob/8cbbe9e9a3addfb88fc94a71a3cfdc46852e4b87/Python_Scripts/BlastSeqFreqAnalyzer.py)
  - [https://github.com/AaronGS1999/TFM/blob/8cbbe9e9a3addfb88fc94a71a3cfdc46852e4b87/Python\\_Scripts/GenoMapper.py](https://github.com/AaronGS1999/TFM/blob/8cbbe9e9a3addfb88fc94a71a3cfdc46852e4b87/Python_Scripts/GenoMapper.py)