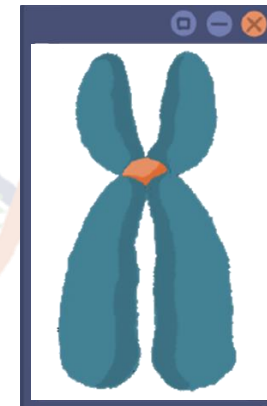


Caracterización de las secuencias repetidas y de las regiones centroméricas del lenguado senegalés *Solea senegalensis* (Kaup, 1858)



Autor: Aaron Gálvez Salido

Tutor: Rafael Navajas Pérez

Máster en Genética y Evolución

Generalidades iniciales

Taxonomía



http://www.ictiorterm.es/nombre_cientifico.php?nc=183

Filo: Chordata

Subfilo: Craniata

Superclase: Actinopterygii

Infraclase: Teleostei

Orden: Pleuronectiformes

Suborden: Pleuronectoidei

Familia: Soleidae

Género: *Solea*

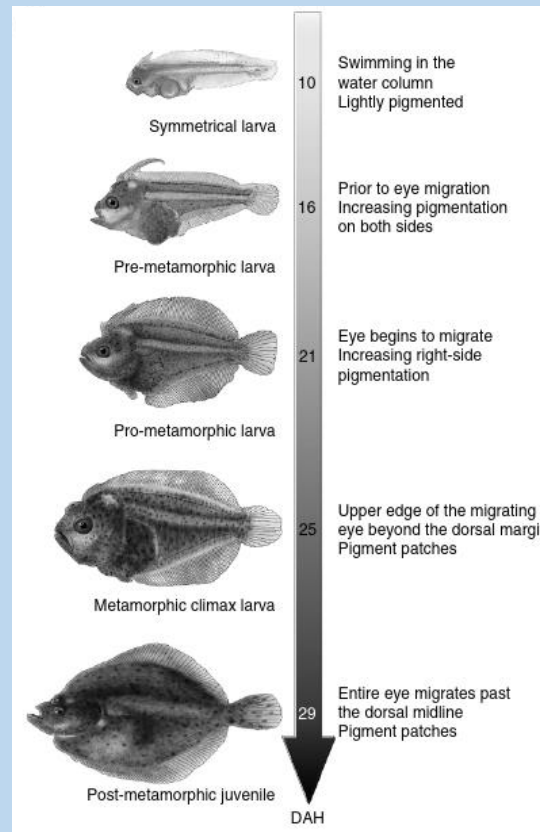
Especie: *S. senegalensis*

Distribución



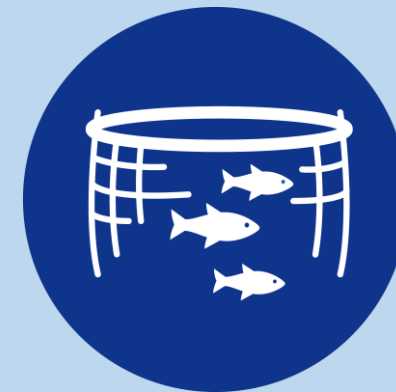
Su distribución natural
abarca desde el mar
Mediterráneo hasta las
costas africanas del
Atlántico oriental

Interés biológico



Shao *et al.*, 2017

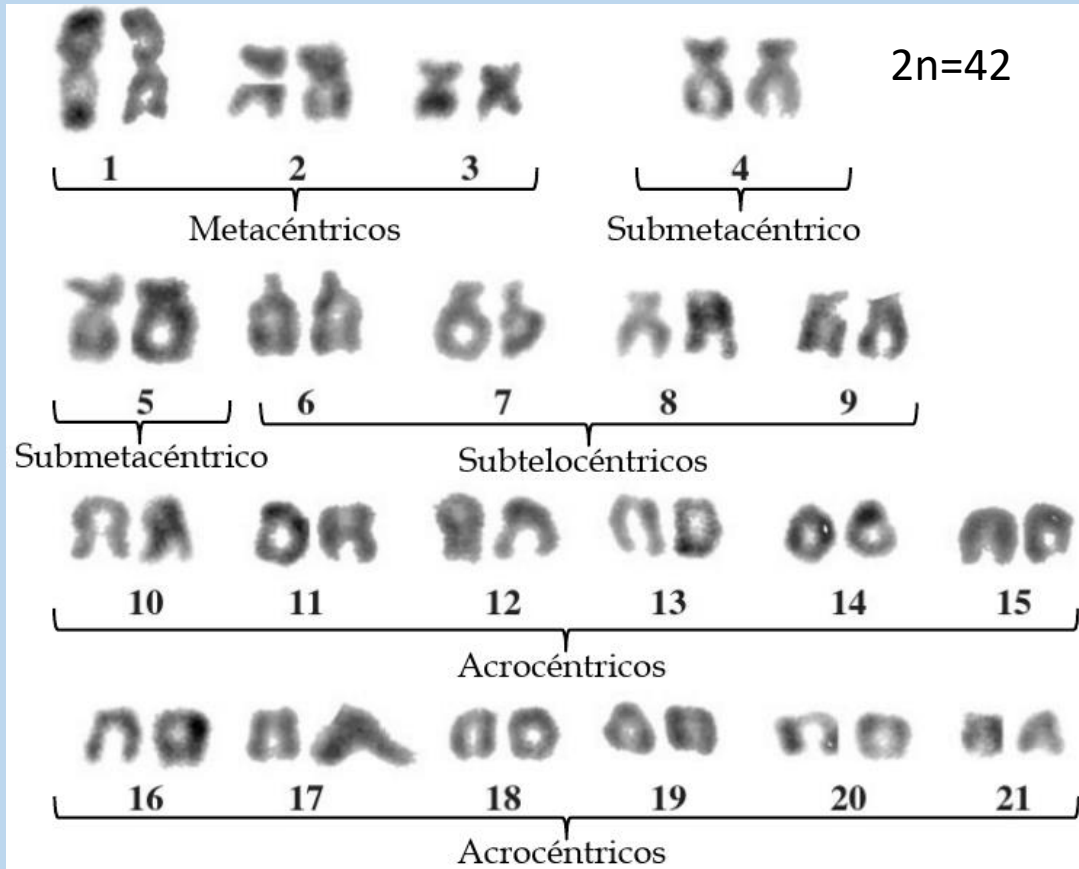
Interés económico



Producción en piscifactoría

Machos F1 infértiles

Composición cromosómica y determinación sexual



Sistema de determinación del sexo XX/XY

- Se propuso el par 1 como par sexual por la presencia única de un elemento móvil de la subfamilia Rex/Babar
- Se identificó el gen *fshr* como determinante del sexo, presente en el cromosoma 12

Estudios previos sobre la región centromérica del lengüado senegalés

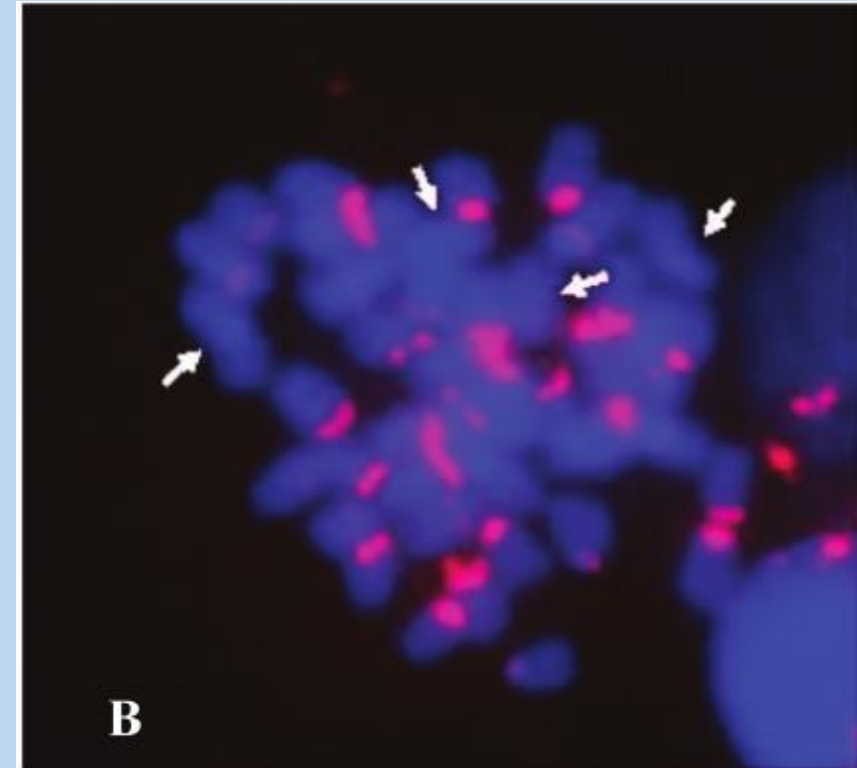
Estudios previos han encontrado familias de ADN satélite específicos de las regiones centroméricas y especie específicos en los peces planos

Nuevo ensamblaje del
genoma de *S. senegalensis*
(Tamaño 606,88 Mb)



Oportunidad para estudiar
las regiones centroméricas

Robles et al., 2017



Localización cromosómica de la familia de
ADN satélite *Pvull* mediante FISH en
metafases mitóticas de *S. senegalensis*

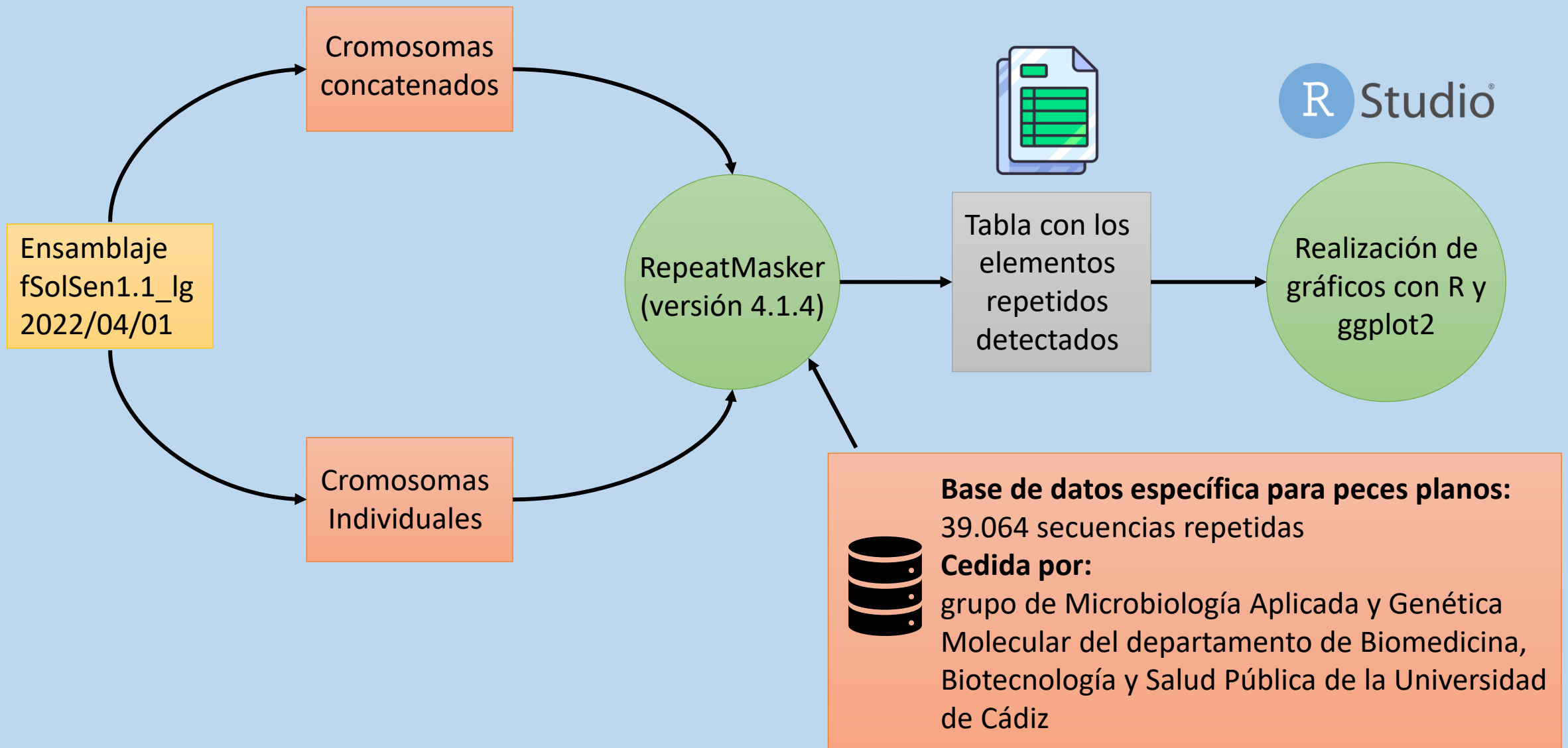
Objetivo principal:

- Caracterizar la fracción de ADN repetido en el lenguado senegalés

Objetivos específicos:

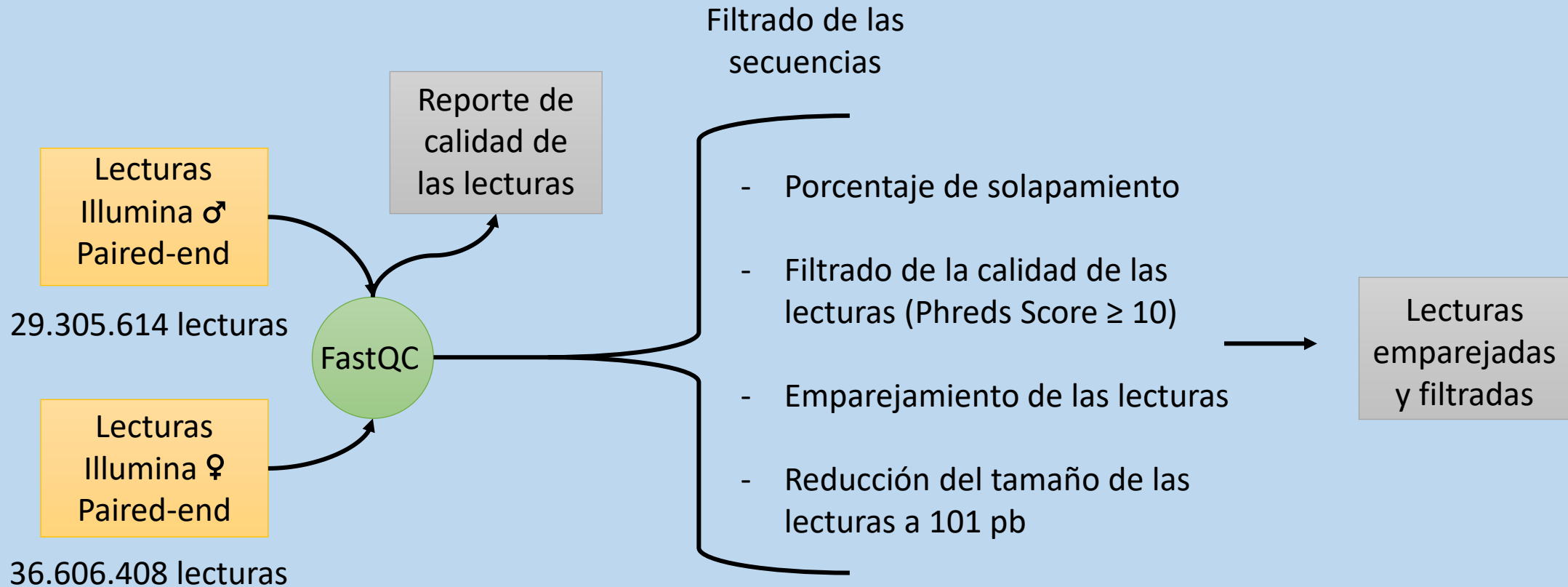
- Llevar a cabo una caracterización global de secuencias de ADN repetido sobre el ensamblaje del genoma y en lecturas de secuenciación de nueva generación
- Anotar y localizar *in silico* las secuencias repetidas
- Analizar comparativamente la acumulación de secuencias repetidas en machos y hembras y estudiar su posible implicación en la determinación sexual de esta especie
- Caracterizar las regiones centroméricas de esta especie y su contenido en elementos repetidos

Identificación de secuencias repetidas en el ensamblaje del genoma



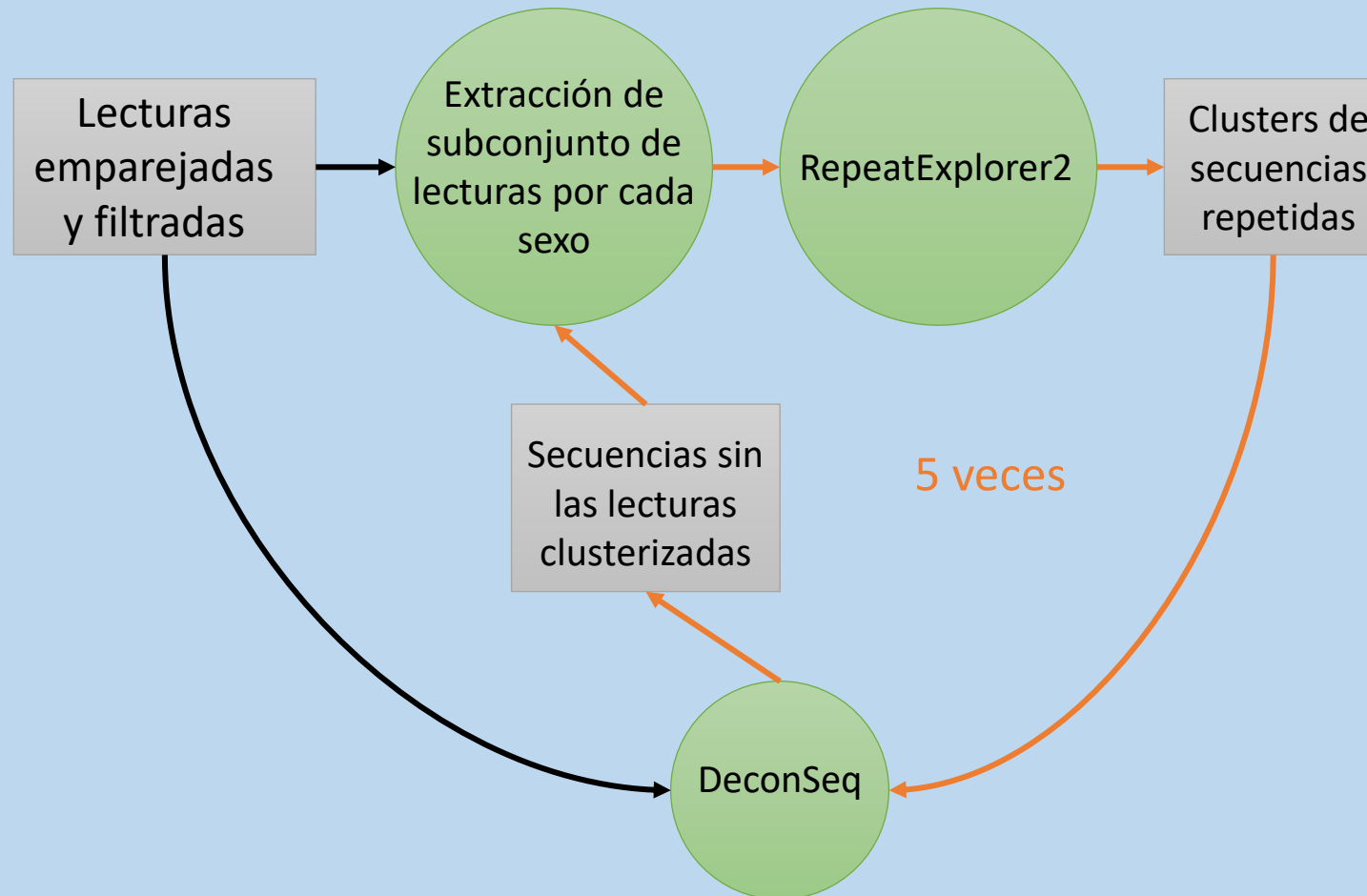
Identificación de secuencias repetidas y análisis comparativo entre sexos

Paso 1: Filtrado de las secuencias



Identificación de secuencias repetidas y análisis comparativo entre sexos

Paso 2: Rondas de clusterización



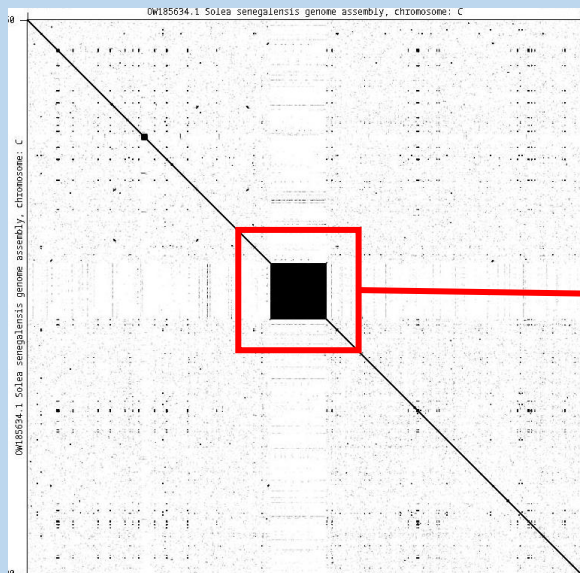
Paso 3: Anotación

La anotación manual de los clúster se realizó de dos formas:

1. Mediante BLASTn frente a la base de datos de peces planos considerando e-values de corte de 10^{-5}
2. Mediante la base de datos de Dfam ajustada al ensamblaje de *Danio rerio* (danRer10) se anotaron los resultados con menor e-values

Caracterización de las regiones centroméricas en el lenguado senegalés

Paso 1: Inspección visual de dotplots



Gepard
2.0



Paso 2: Extracción de secuencias potencialmente repetidas

Paso 3: Caracterización de las secuencias

Detección de repeticiones en tándem:
Tandem Repeat Finder

Estudio de la estructura secundaria:
RNAfold

Anotación mediante BLAST
con la base de datos de peces planos

Representación de las secuencias sobre el ensamblaje del genoma

BLASTn de las secuencias
frente a los cromosomas para
obtener las posiciones

Ideogramas

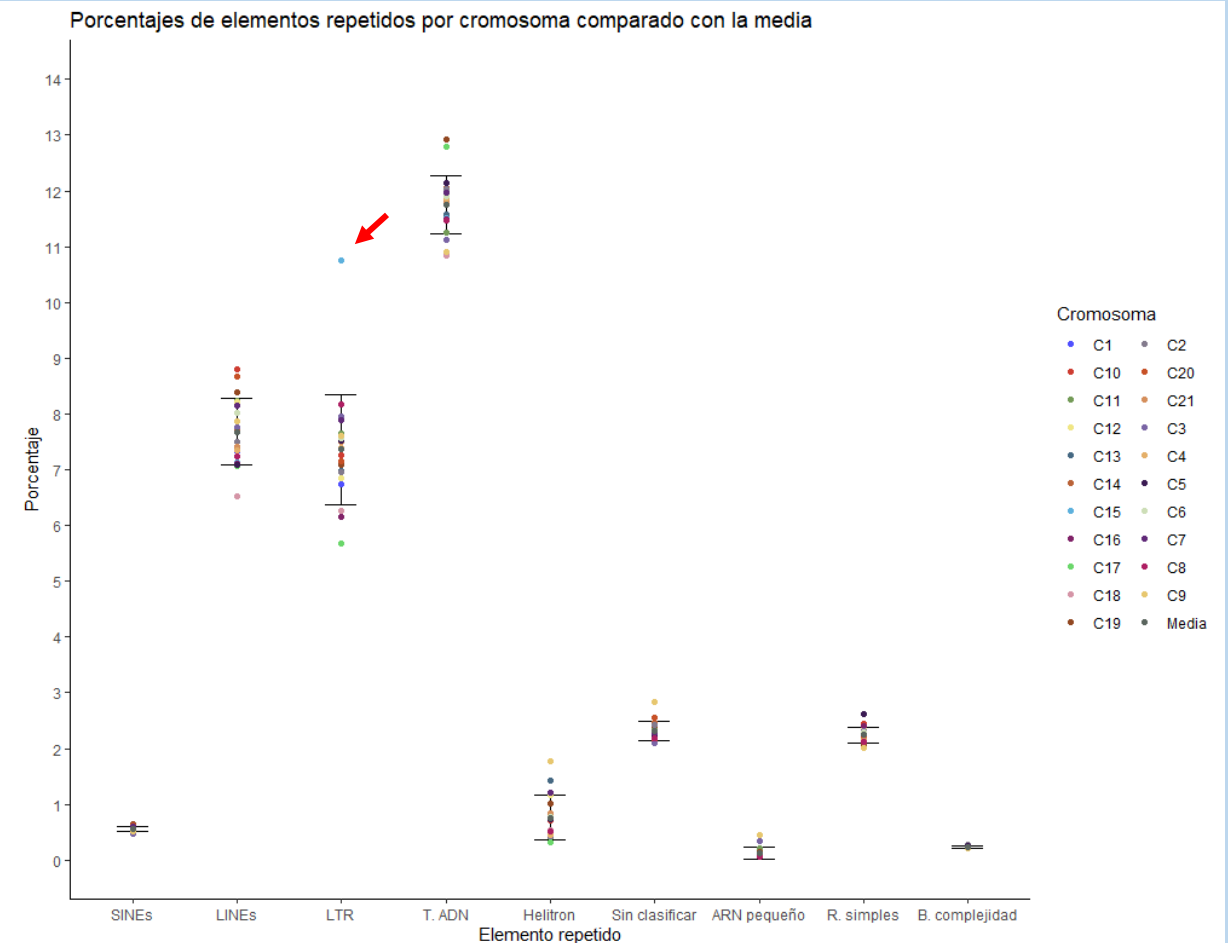
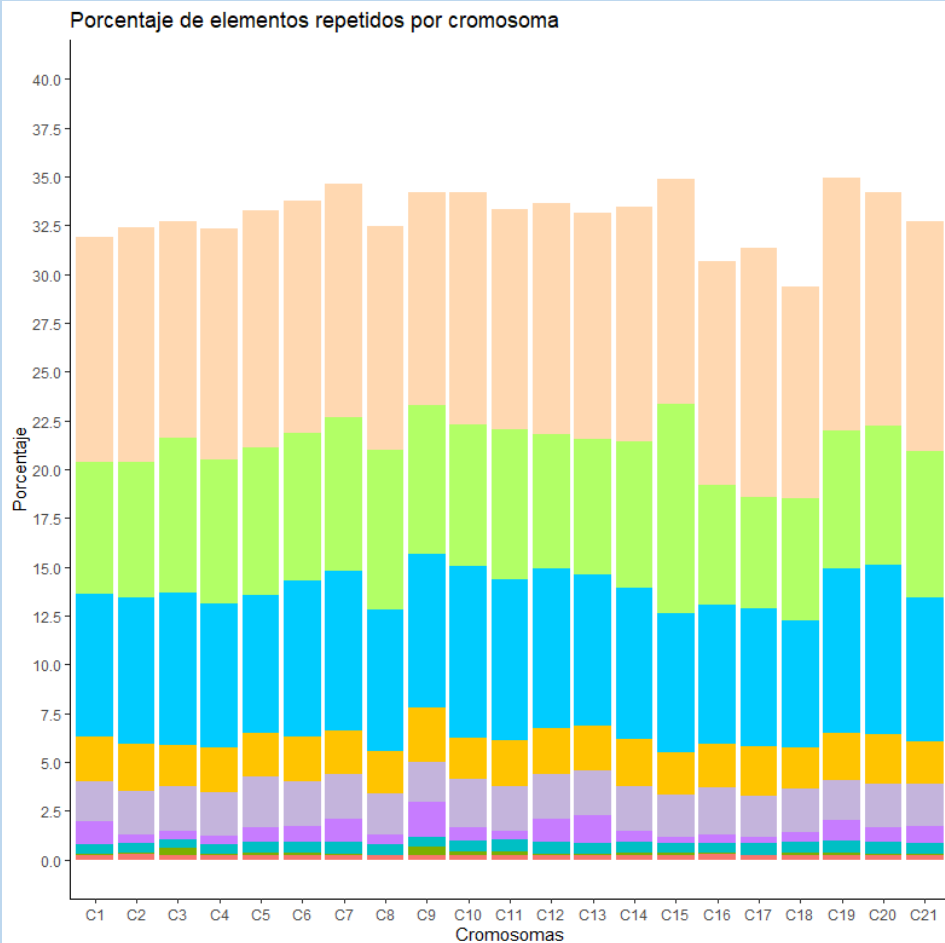
Scripts de Python

R junto a la librería KaryoploteR

Análisis del ADN repetido sobre el ensamblaje del genoma con RepeatMasker

40,70% de contenido en GC
32,98% de elementos repetidos

Transposones de ADN con un porcentaje de 11,76% > LINEs con un porcentaje de 7,75% > elementos LTRs con un porcentaje de 7,34%



Análisis del ADN repetido en lecturas de secuenciación de nueva generación

Clasificación	Nº Clusters
Transposones ADN	74
LINEs	46
Elementos LTR	26
Elementos repetidos desconocidos	15
ADN satélite	10
SINEs	7
Círculo rodante (Helitron)	3
ARN ribosómico	3
Repeticiones simples	3
Sin homología	26
Total	214

No se encontró acumulación de secuencias repetidas de forma diferencial entre sexos

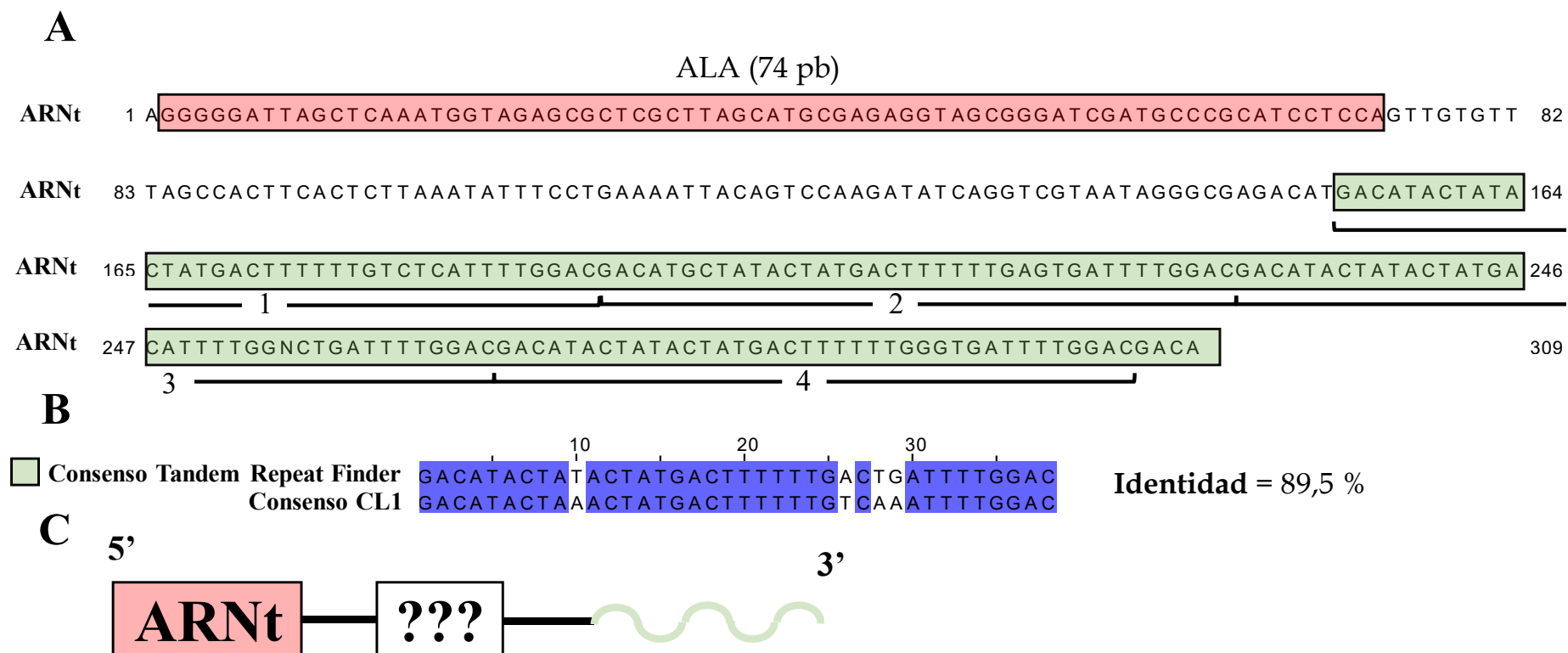


Incipiente estado de diferenciación sexual a nivel cromosómico en esta especie

Nº	Nº lecturas machos	Nº 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	tándem (baja confianza)	ARN _t
2	5910	5420	0,591	0,542	0	Asian_Seabass_LTR_161#LTR/Gypsy	tándem (baja confianza)	LTR
3	4700	4670	0,470	0,467	0	Sse_rnd-5_family-1529#LINE/L1	tándem (alta confianza)	Satélite

Análisis del ADN repetido en lecturas de secuenciación de nueva generación

Análisis de la secuencia repetida con ARNt y CL1

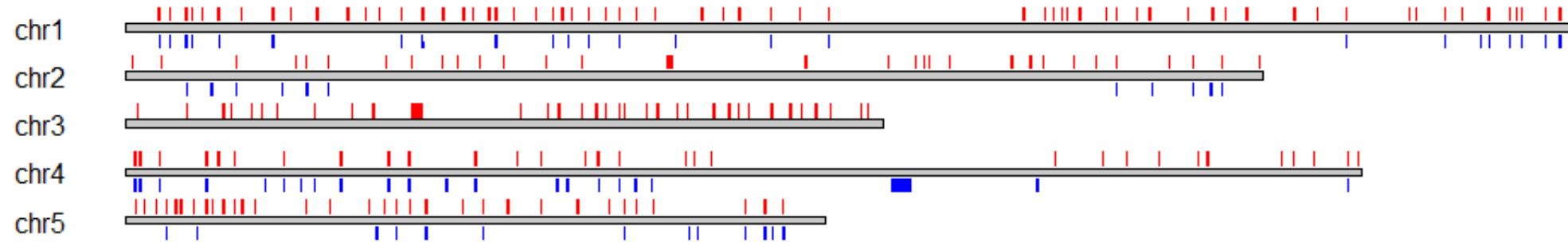


Análisis del ADN repetido en lecturas de secuenciación de nueva generación

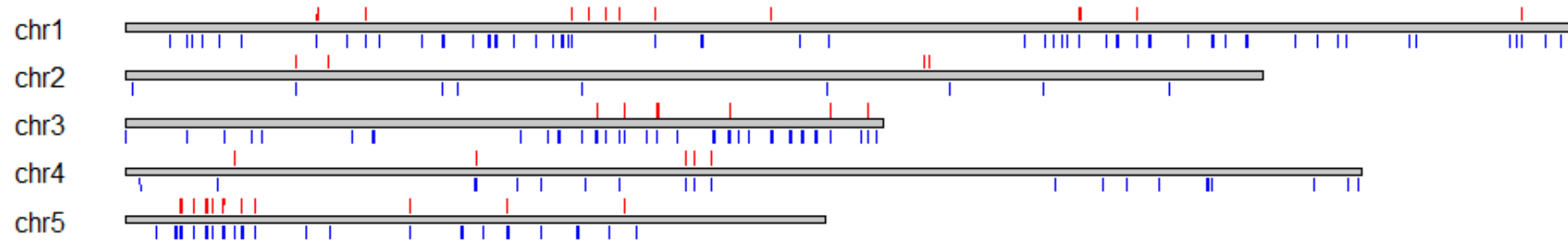
Identidad consenso CL1 y reverso complementaria del consenso CL2 = 89,5%

Contig más largo del CL2 → Homología con secuencia tipo Gypsy

CL1 ■ ARNt ■



CL2 ■ Gypsy ■



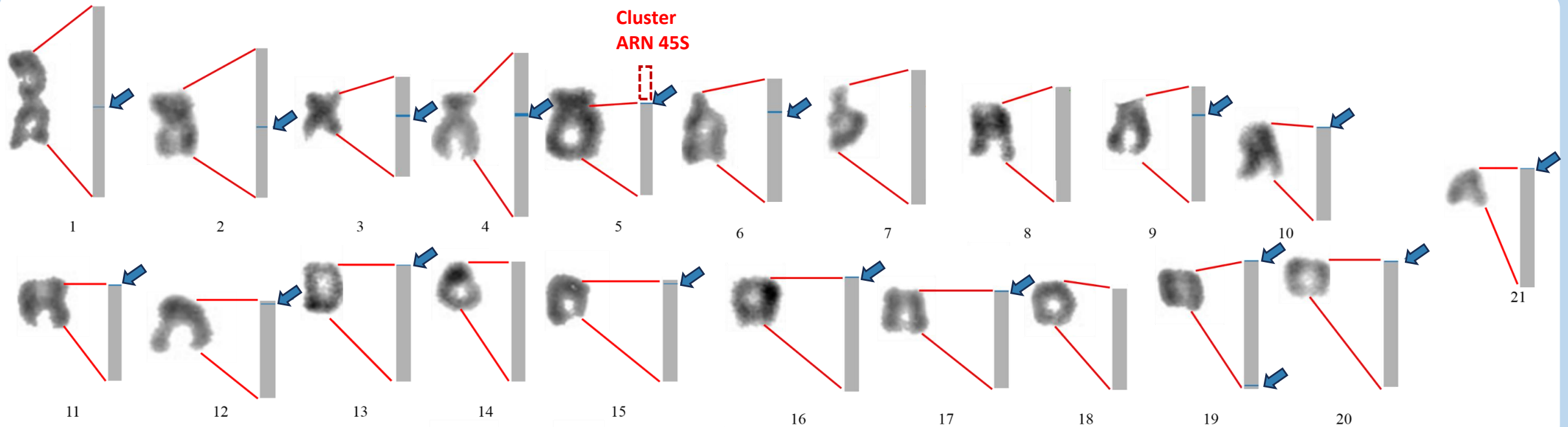
LINE/L1 → Origen del satélite centromérico *Pvull*

Caracterización de las regiones centroméricas

CL3 → *PvuII* → Satélite centromérico principal

Posición *PvuII* ■

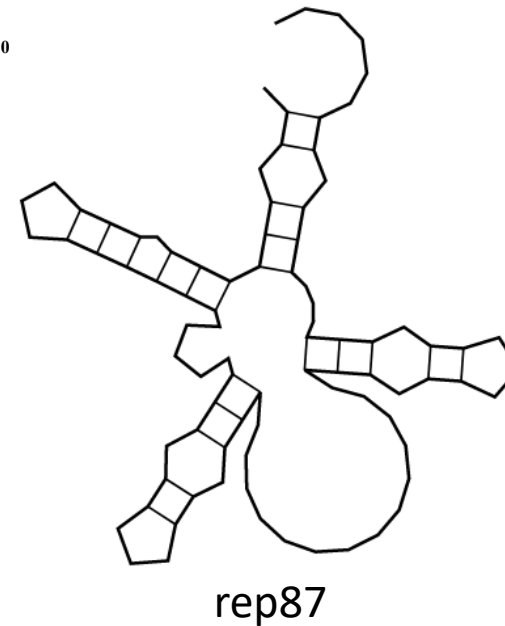
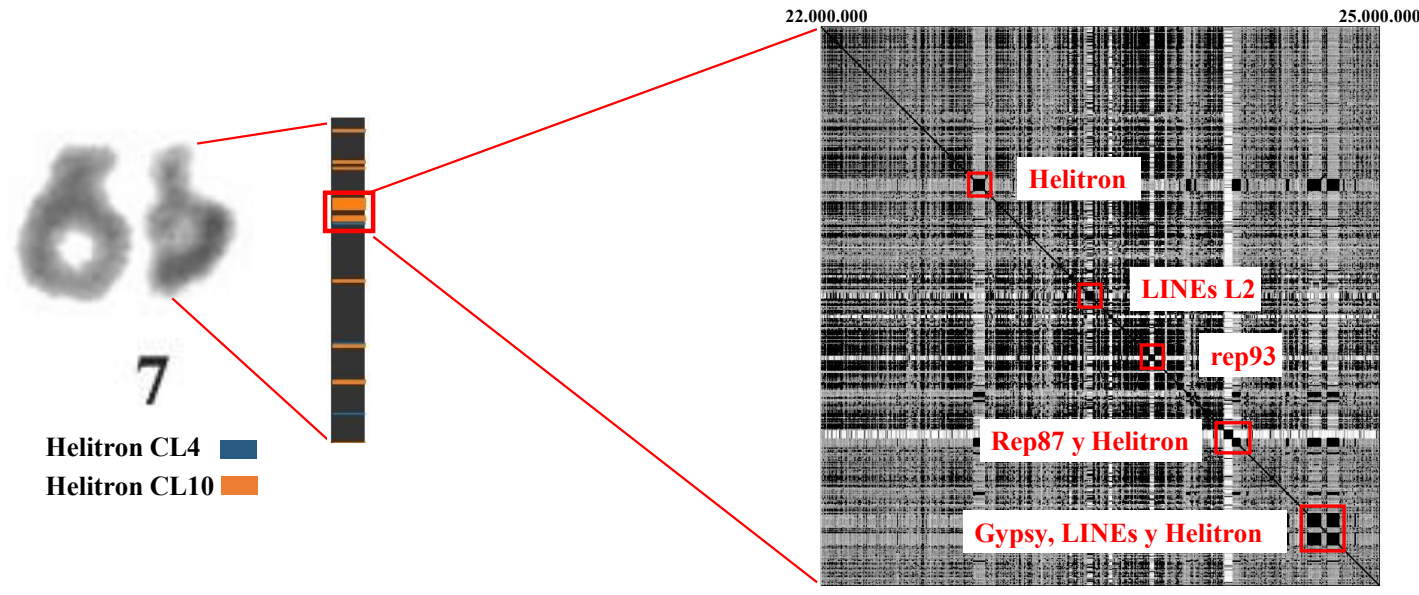
Helitron en regiones pericentroméricas en los cromosomas 3, 4, 5, 6, 9, 10, 12, 13, 16, 17 y 19



Caracterización de las regiones centroméricas

Región de búsqueda

Estudio de la estructura secundaria (RNAfold)

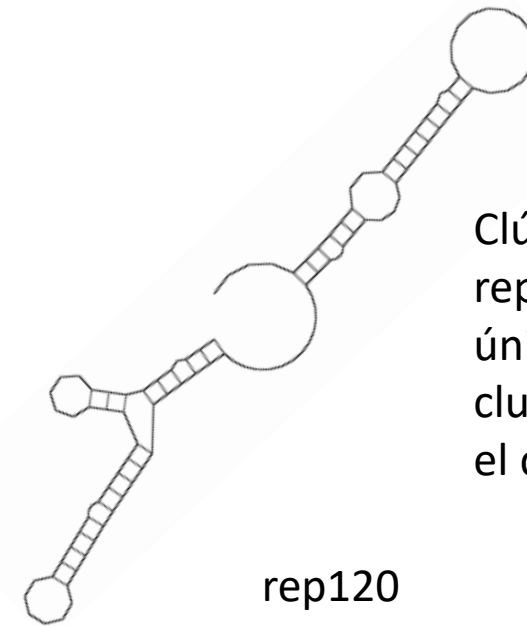
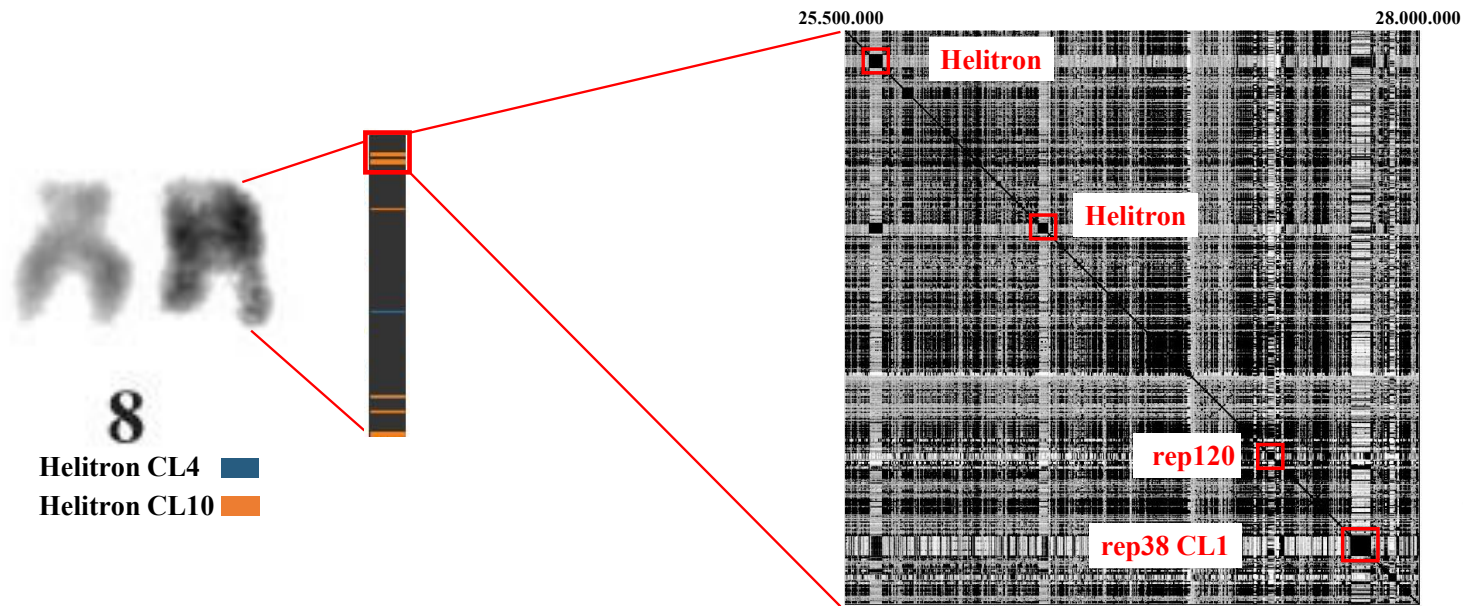


Clúster de 485 repeticiones y únicamente clusterizada en el cromosoma 7

Caracterización de las regiones centroméricas

Región de búsqueda

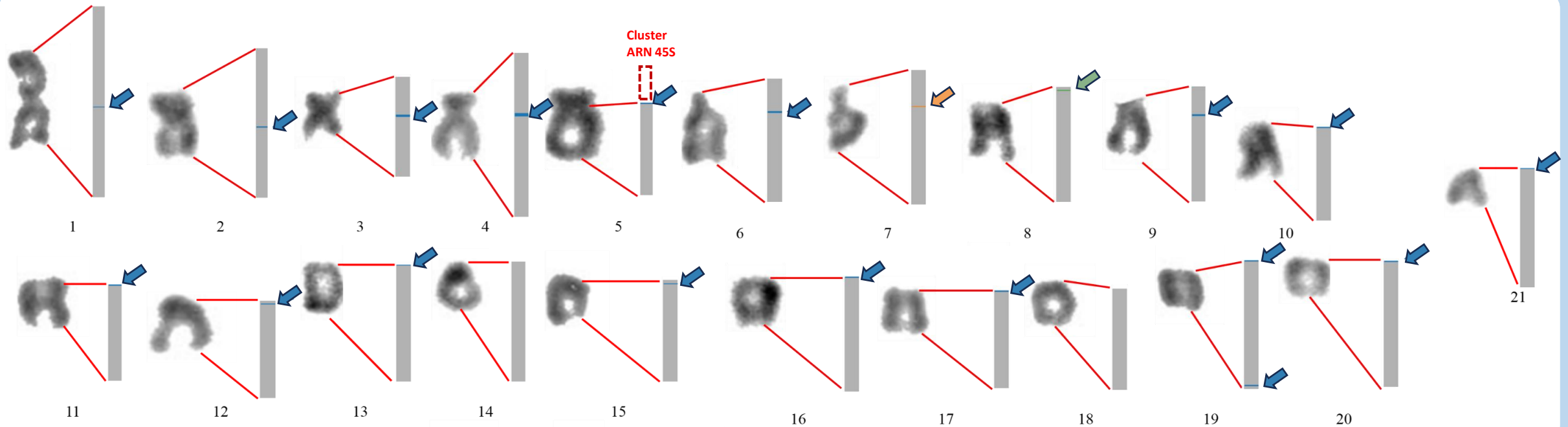
Estudio de la estructura secundaria (RNAfold)



Clúster de 220 repeticiones y únicamente clusterizada en el cromosoma 8

Caracterización de las regiones centroméricas

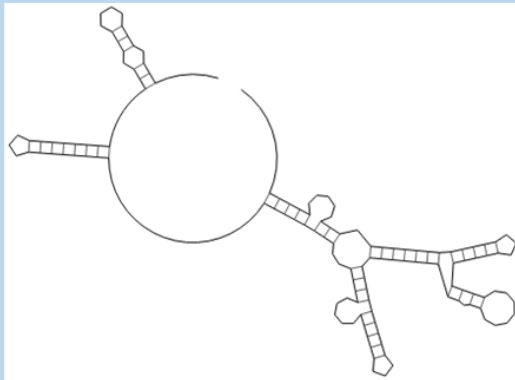
Posición *PvuII* ■ rep87 ■ rep120 ■



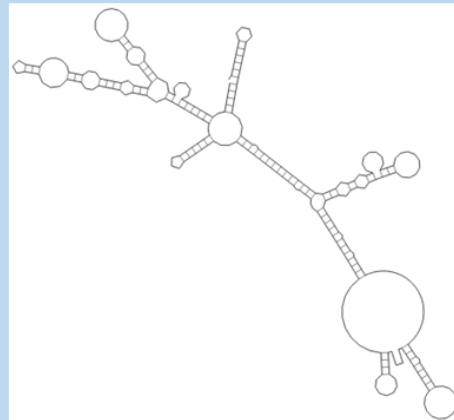
Conclusiones

- El ensamblaje del genoma de *S. senegalensis* está compuesto por un 32,98% de elementos repetidos y los más frecuentes son los transposones de ADN, las LINEs y los elementos LTRs
- No existen diferencias significativas en el contenido de secuencias repetidas entre los genomas de machos y hembras
- La mayoría de los cromosomas tienen los centrómeros ocupados por la familia de ADN satélite *PvuII* excepto en los cromosomas 7 y 8 que estarían ocupados por los satélites rep87 y rep120 respectivamente
- El clúster 1 mostró homología con un elemento repetido que contiene un ARNt, el clúster 2 con una secuencia de tipo Gypsy y el clúster 3 con la secuencia *PvuII* que presenta homología con una LINE/L1
- Los resultados de este estudio son coherentes con los datos existentes sobre morfología cromosómica de esta especie, excepto en el cromosoma 5, 14, 18 y 19 en los que no hemos podido identificar las regiones centroméricas

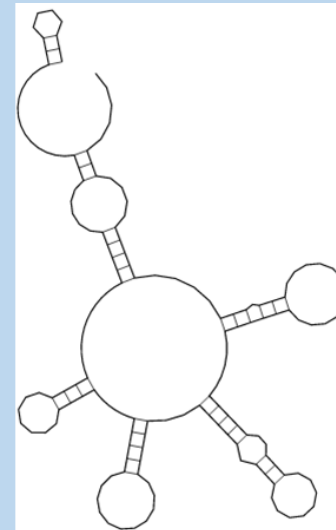
Satélites centroméricos



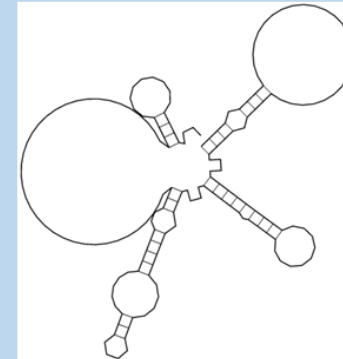
alpha-satellite (*Homo sapiens*)



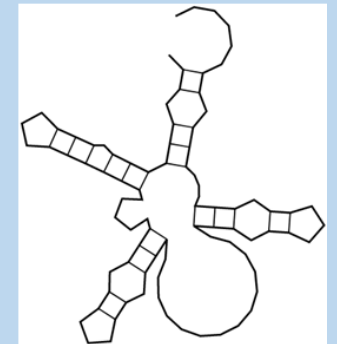
SacI
Hippoglossus hippoglossus



DraI
Dicologlossa cuneata



pvuII



rep87

Mecanismo de expansión de los elementos tipo LTR

