# Identification of genetic markers associated with growth and morphology quality in Senegalese sole (*Solea senegalensis*) to boost aquaculture production

TESIS DOCTORAL

Israel Guerrero Cózar

Cádiz, 2021

PROGRAMA DE DOCTORADO EN CIENCIAS Y TECNOLOGÍAS MARINAS

UCA Universidad de Cádiz

Junta de Andalucía
Consejería de Agricultura, Ganadería, Pesca y Desarrollo Sostenible
INSTITUTO DE INVESTIGACIÓN Y FORMACIÓN AGRARIA Y PESQUERA

# Identification of genetic markers associated with growth and morphology quality in Senegalese sole (*Solea senegalensis*) to boost aquaculture production


Memoria presentada por

**Israel Guerrero Cózar**

para obtar al grado de

**Doctor con mención internacional**

Tesis realizada bajo la dirección de

Dr. **Manuel Manchado Campaña**, Investigador Principal del Instituto de Investigación y Formación Agraria y Pesquera (IFAPA)

**Dr. Manuel Manchado**

**D. Israel Guerrero**

UCA Universidad de Cádiz

Junta de Andalucía
Consejería de Agricultura, Ganadería, Pesca y Desarrollo Sostenible
INSTITUTO DE INVESTIGACIÓN Y FORMACIÓN AGRARIA Y PESQUERA

The studies carried out in this PhD Thesis have been financed by the following projects and scholarships:

- Proyecto RTA2013-00023-C02-01, titulado "Implementación de tecnología innovadoras de mejora genética en lenguado senegalés (*Solea senegalensis*) y dorada (*Sparus aurata*) para la optimización de su producción industrial (INNOTECSS)", financiado por INIA y EU a través del programa FEDER2014-2020 "Programa Operativo de Crecimiento Inteligente.

- Proyecto RTA2017-00054-C03-01, titulado "Estudio de los factores genéticos que regulan la morfología y sus alteraciones en dorada y lenguado senegalés (MORFOGEN)", financiado por MCIU/AEI/FEDER, UE.

- Proyecto PR.AVA.AVA201601.9, titulado "Evaluación transcriptómica de compuestos funcionales basados en microalgas y desarrollo de nuevas herramientas de análisis genómico que impulsen la acuicultura del lenguado (SOLEALGAE)", financiado por FEDER, UE.

- Beca predoctoral del Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria. FPI-INIA 2015.

The results derived from this PhD Thesis have been partially published in:

## PEER-REVIEWED JOURNALS (see ANEXO I)

**Guerrero-Cozar, I.**; Perez-Garcia, C.; Benzekri, H., Sánchez, J.; Seoane, P.; Cruz, F.; Gut, M.; Zamorano, M.J.; Claros, M.G.; Manchado, M. Development of whole-genome multiplex assays and construction of an integrated genetic map using SSR markers in Senegalese sole. *Sci. Reports* **2020**, *10*, 21905 DOI: 10.1038/s41598-020-78397-w

**Guerrero-Cozar, I.**; Jimenez-Fernandez, E.; Berbel, C.; Cordoba-Caballero, J.; Claros, M.G.; Zerolo, R.; Manchado, M. Genetic parameter estimates and identification of SNPs associated with growth traits in Senegalese sole. *Aquaculture* **2021**, *539*, 736665 DOI: 10.1016/j.aquaculture.2021.736665

**Guerrero-Cozar, I.**; Jimenez-Fernandez, E.; Berbel, C.; Espinosa, E.; Claros, M.G.; Zerolo, R.; Manchado ,M. Genetic Estimates for Growth and Shape-Related Traits in the Flatfish Senegalese Sole. *Animals* **2021**, 11, 1206 DOI: 10.3390/ani11051206

**Guerrero-Cozar, I.**; Gomez-Garrido, J.; Berbel, C.; Martinez-Blanch, J.F.; Alioto, T.; Claros, M.G.; Gagnaire, P.A.; Manchado, M. Chromosome anchoring in Senegalese sole (Solea senegalensis) reveals sex-associated markers and genome rearrangements in flatfish. *Sci. Reports* **2021,** 11**,** 13460 DOI: 10.1038/s41598-021-92601-5

## INTERNATIONAL CONFERENCES

**Guerrero-Cózar, I**.; Gomez-Garrido J.; Martinez-Blanch, J.F.; Alioto T; Claros, M.G.; Gagnaire P.A.; Manchado, M. Construcción y anclaje de un mapa genético de ligamiento de alta densidad en el lenguado senegalés (*Solea senegalensis*). Oral presentation. Conyba 2020. Nayarit (Mexico). November 2020.

**Guerrero-Cozar, I**.; Cordoba, J.; Seoane, P.; Claros, M.G.; Zerolo, R.; Manchado, M. Identification of SNP markers associated to somatic growth in the flatfish Senegalese sole (*Solea senegalensis*). Poster presentation. Aquaculture Europe 2019. Berlín (Alemania). October 2019.

## NATIONAL CONFERENCES

**Guerrero-Cózar, I**.; Córdoba, J.; Seoane, P.; Claros, M.G.; Zerolo, R.; Manchado, M. Identificación de marcadores SNP ligados al sexo en el lenguado senegalés (*Solea senegalensis*). Poster presentation. Congreso Nacional Acuicultura 2019. Cartagena, Murcia (España). May 2019.

**Guerrero-Cózar, I**.; Aparicio, M.; Berbel, C.; Jimenez-Fernandez, E.; Manchado, M. Estudio genético de las malformaciones esqueléticas en lenguado senegalés (*Solea senegalensis*). Poster presentation. IX Jornadas de Acuicultura en el Litoral Suratlántico. Cartaya, Huelva (España). May 2018.

Aparicio, M.; **Guerrero-Cózar, I**.; Martín, N.; Crespo, A.M.; Manchado, M.; Berbel, C. Identificación de patrones morfológicos asociados a las anomalías esqueléticas en lenguado senegalés (*Solea senegalensis*). Poster presentation. IX Jornadas de Acuicultura en el Litoral Suratlántico. Cartaya, Huelva (España). May 2018.

**Guerrero-Cózar, I**.; Aparicio, M.; Berbel, C.; Benzekri, H.; Claros, M.G.; Manchado, M. Optimización de nuevos ensayos de genotipado mediante PCR múltiple en el lenguado senegalés (*Solea senegalensis*). Poster presentation. XVI Congreso nacional de acuicultura 2018. Zaragoza (España). October 2017.

Román-Padilla, J.; **Guerrero-Cózar, I**.; Hachero-Cruzado, I.; Manchado, M. Apolipoproteína B: Expresión tisular, funciones e importancia en el transporte lipídico de tres parálogos en *Solea senegalensis*. Poster presentation. XVI Congreso nacional de acuicultura 2018. Zaragoza (España). October 2017.

# AGRADECIMIENTOS

Ya se acerca el final de esta importante y bonita etapa de mi vida en la que tengo muchas cosas que agradecer a todas las personas que me han ayudado a llegar hasta aquí, que han compartido mi camino y que me han hecho ser quien soy, porque sin ellas jamás lo hubiera conseguido. Muchas de estas personas siguen ahí en mi día a día compartiendo mis buenos y malos momentos, otras han pasado de manera transitoria por mi vida pero han dejado su huella y algunas ya no están pero es como si nunca se hubieran ido porque siempre van a viajar a mi lado.

Para empezar me gustaría agradecer a todos mis amigos y profesores del colegio SAFA de Linares, donde comenzó todo y en especial a mi "profe" de biología Susana Muñoz por transmitirme de esa manera tan bonita su pasión por la ciencia que hizo que me decidiese por estudiar biología.

De Granada, donde realicé la carrera, me llevo una infinidad de recuerdos y personas, en especial a María a la que debo el poder estar aquí y le agradezco de corazón todo lo que hizo por mí durante esos años y ser mi guía cuando más perdido estaba.

Me gustaría dar las gracias también a Juan Alberto Marchal, mi director del TFM y a María Arroyo, mi compañera de laboratorio de la Universidad de Jaén, por todo lo que me enseñaron y todos los consejos que me dieron, siempre con una sonrisa y que me hicieron crecer mucho en mi etapa en Jaén.

Sinceramente la oportunidad de hacer el doctorado me llegó cuando menos lo esperaba, ni siquiera conocía el IFAPA, de hecho la primera vez me costó encontrarlo pero ahora más de 5 años después le he acabado cogiendo cariño y es que desde el primer momento me he sentido como en casa. Muchas gracias a Javi, Paula, Marian, Eugenia, Silvia, Pablo, Ismael, María del Mar, Raúl, Moi, Javier, Joaquín, Teresa, Lina, Pedro, César y en especial a mi equipo, el "Elite Team" por todos los momentos que hemos compartido y seguimos compartiendo todavía. Gracias a Vitorio por alegrar los muestreos con sus historias de marinero, a Rafael que siempre me ha ayudado cuando lo he necesitado y al que aprecio mucho, a Concha que aunque se empeñe en asustarme cada vez que entra al despacho siempre me saca una sonrisa, a Manolo Aparicio que fue el primero con quien compartí poyata y me enseñó la diferencia entre hacer las cosas en modo biólogo y modo químico, a Carlos, el "McGyver" del grupo que vale para todo por enseñarme tantas cosas y guiarme siempre que he tenido dudas, a Aniela, la organizadora de eventos, que desde el primer momento en que llegué me integró en el grupo y con la que he vivido muchos buenos momentos de risas y cotilleos infinitos, a Nuria, la pequeña de gran corazón, uno de los motores de este equipo

siempre dispuesta a ayudar a los demás, a Patri, por toda la alegría que me transmite cada mañana con esa intensidad tan característica que la hacen única en su especie, a la *gamer* Cris, la última en llegar y que se apunta hasta a un bombardeo, y al líder del equipo y director de esta tesis Manuel Manchado por todo su esfuerzo y dedicación por su trabajo, por soportarme todos estos año aún siendo un desastre y aunque me ha metido de lleno en la secta de la manzana siempre le estaré agradecido.

A Laureana Rebordinos, mi tutora en la Universidad de Cádiz por toda su colaboración y amabilidad a lo largo de todo este proceso en el que siempre me ha ayudado mucho.

Agradecer también Kostas, Sophia, Dimitra, Eleni, Vana, Sophia Lesiou, Eva, Elisavet, Stamos y Maria de Apivita y Joao, el infiltrado, por hacer tan especial mi estancia en Atenas.

Dar las gracias a Pierre-Alexandre y todo su equipo del Institut des Sciences de l'Évolution Montpellier (ISEM), donde a pesar de haber una pandemia por medio disfruté y aprendí muchísimo.

A Mamen, mi madre adoptiva en el Puerto y a toda su familia que me abrieron las puertas de su casa y me acogieron como uno más desde el principio haciendo que mi adaptación aquí fuera mucho más fácil.

Al artista argentino y gran amigo Ernesto Merino por el diseño de la portada y a su preciosa familia Cristina, Paloma, Mateo, Bauti y Ernestito, porque esté donde esté siempre serán mis vecinos preferidos.

A Pilar, a la que quiero mucho y que ha marcado sin duda este tramo de mi vida con bonitos momentos compartidos y a la que pronto le llegará su turno de escribir estos agradecimientos.

No me iba a olvidar de dar las gracias todos mis "critters", Balti, Jorgito, Ángel, Danieles, Ñete, Jonche, Fosky, Dela, Efren, Juanillo, Ransi, Relovader, Javi, Jose, David y Jabu, no he podido tener más suerte al encontraros todos en mi vida.

Y por supuesto a mi familia, a todos mis tíos y primos, en especial a Esteban, Alberto, Chiqui, Mari Carmen, Alejo y Amparo, por apoyarme siempre incondicionalmente y por todo lo que nos ayudáis día a día. Por último a mis pilares más importantes, a mi inseparable Tara por su amor incondicional. A mis hermanos Rubén y Noemí, por quererme tanto y hacer posible que vaya cumpliendo metas y sueños en mi vida. A mi padre Pedro, por inculcarme tantos valores, enseñarme a ser una mejor persona y ser una referencia para mí. Y a mi madre Emi a quien le debo la vida y todo lo que soy, por su esfuerzo, por su sacrificio, por ser una luchadora nata, ojalá algún día pudiera llegar a ser como ella.

# INDEX

# ABBREVIATIONS

**A**: area

**AD**: amoebic disease

**AFLP**: amplified fragment length polymorphism

**BAC**: bacterial artificial chromosome

**BHP**: body height at the insertion of the pectoral fin

**BLUP**: best linear unbiased prediction

**BMH**: body maximum height

**Chr**: chromosome

**cM**: centimorgan

**CPH**: caudal peduncle height

**DI**: Digital Image Analysis

**EB**: evaluation batches

**EBV**: estimated breeding values (EBV)

**FG**: fast-growing

**GBS**: genotyping by sequencing

**GLM**: general linear model

**GWAS**: genome-wide association analysis

**HAT**: histone acetyltransferase activity

**He**: expected heterozygosity

**Ho**: observed heterozygosity

**HWE**: Hardy–Weinberg equilibrium

**LG**: linkage groups

**MAS**: marker-assisted selection

**mFISH**: multiple fluorescent in situ hybridisation

**MLM**: mixed linear model

**NE-PP**: Non-exclusion probabilities for pair parent

**NGS**: Next generation sequencing

**PCA**: Principal component analysis

**PIC**: Polymorphic information content

**QTL**: Quantitative trait loci

**RAD**: restriction site associated DNA markers

**RAS**: recirculating aquaculture system

**REML**: restricted maximum likelihood

**RFm**: recombination frequency estimates

**RR**: recombination rate

**SCF**: Scaffold

**SD**: sex determination

**SG**: slow-growing

**SL**: Standard length

**SLI**: digital image analysis to estimate standard length

**SM**: supermultiplex

**SNP**: single nucleotide polymorphisms

**SSR**: simple sequence repeats

**W**: body weight

**Wi**: Standard width

**WiI**: Standard width estimated by digital image analysis

# SUMMARY

## Identification of genetic markers associated with growth and morphology quality in Senegalese sole (*Solea senegalensis*) to boost aquaculture production.

The Senegalese sole (*Solea senegalensis*) is one of the most valuable flatfish in aquaculture in Southern Europe and although the production has grown exponentially in the last decade, reproduction success, health status, and the improvement of growth rates and morphology quality still remain as important limitations for industrial exploitation. The development of breeding programs is a fundamental tool to solve these problems but requires estimate genetic components of economically valuable traits such as those growth and morphology related under industrial conditions. By other side, the advances in genomics provide new highly powerful analysis tools to determine more accurately genetic components productive traits.

In this thesis, new genomic tools and molecular markers as well as the genetic components of growth and morphology quality traits have been developed in sole. For this purpose, firstly, a high-density SNP genetic map and a *de novo* sole genome assembly were generated. Later, genetic and physical maps were anchored and integrated into 21 linkage groups (SseLGs) corresponding to the expected number of chromosomes of this species. Genetic map was bigger in female than male (1.49) observing also a different recombination rate landscape between sexes. The integrated physical map obtained was used for an association study to identify sex-linked markers. Seven families were analyzed using ddRAD and 30 significant sex-associated SNP markers located onto SseLG18 were identified. Searching for candidate genes for sex determination identified the follicle stimulating hormone receptor (*fshr*) that it was located within a hot recombination region although with an incomplete penetrance.

In addition to SNP markers, genome information was used for searching and identifying SSR markers. Hence, 108 new SSR markers distributed throughout the genome were identified. They were structured in 13 PCR Multiplex assays (with up to 10-plex) and the amplification conditions were optimized and validated with a high-quality score. A subset of 40 highly polymorphic markers were selected to optimize four supermultiplex PCR Multiplex assays (8-11 SSRs per assay) were designed for use in pedigree analysis. Moreover, a new integrated genetic map with 229 SSRs distributed in 21 SseLGs was created by *in silico* genomic analysis. Both maps generated in this thesis were used to carry out evolutive genome studies in flatfish to identify lineage-specific Robertsonian fusions and several other rearrangements that explain changes in chromosome number in the karyotype of Pleuronectiformes.

To investigate the genetic components of growth and morphology-related traits, estimates for different variables were determine before on-growing (400 days) and at harvest (800 days). Growth-related traits such as body weight (W), standard length (SL), width (W) and body area (A) showed high heritabilities (ranging from 0.568 to 0.609 at 400 d and from 0.424 to 0.500 at 800 d) with very high genetic correlations (>0.94) at both ages. With respect to morphology quality traits, six quality predictors including ellipticity (E), body height at the pectoral fin base (BHP), body maximum height (BMH) and caudal peduncle height (CPH) and two ratios (BMH/BHP and BMH/CPH) were evaluated. Results showed high heritabilities (0.463-0.774) for E, BHP, BMH and CPH which were higher at 400 d than 800 d. In contrast, the BMH/BHP and BMH/CPH ratios showed low-moderate heritabilities (0.144-0.306). High positive correlations (>0.95) were found between growth traits and the three heights, which decreased with age. In contrast, ellipticity showed negative and medium-high genetic correlations with growth traits and heights, indicating that fish selected for larger size will also be less elliptical. Finally, an association study to find genetic markers linked to growth traits was carried out. A low-density DNA chip was designed and validated for 49 SNPs distributed in 17 SseLGs. The analysis of fast and slow-growing families identified two significant markers within the general transcription factor 3C polypeptide 4 and the mitochondrial fission process protein 1.

All these results provide powerful tools for genomic analysis as well as genetic highly valuable information to design genetic breeding programs in Senegalese sole to optimize to boost the industrial production in aquaculture.

# RESUMEN

## Identificación de marcadores genéticos asociados al crecimiento y calidad morfológica en lenguado (*Solea senegalensis*) para el impulso de su producción en acuicultura.

El lenguado senegalés (*Solea senegalensis*) es uno de los peces planos de mayor valor económico en acuicultura del sur de Europa y aunque su producción ha crecido exponencialmente en la última década, aún existen limitaciones importantes en el éxito reproductivo, control sanitario, así como de las tasas de crecimiento y calidad morfológica de cara a su explotación industrial. El desarrollo de programas de mejora genética es una herramienta fundamental para solventar estos problemas, pero requiere estimar los componentes genéticos de caracteres económicamente valiosos como los relacionados con el crecimiento y la morfología bajo condiciones industriales. Por otro lado, los avances en genómica proporcionan nuevas capacidades analíticas de alto rendimiento para determinar de forma más precisa la evaluación genética de los caracteres productivos.

En esta tesis se han desarrollado nuevas herramientas genómicas y marcadores moleculares así lo componentes genéticos de caracteres ligados al crecimiento y calidad morfológica en lenguado. Para ello, en primer lugar, se realizó un mapa genético de alta densidad basado en SNPs así como un ensamblaje *de novo* del genoma del lenguado. Posteriormente, los mapas genético y físico se anclaron e integraron en 21 grupos de ligamiento (SseLGs) correspondientes al número de cromosomas en esta especie. El mapa genético fue mayor en la hembra que en el macho (1.49) obteniendo diferentes perfiles de recombinación entre ambos sexos. El mapa físico obtenido se utilizó para un estudio de asociación para encontrar marcadores ligados al sexo. Para ello se analizaron siete familias mediante ddRAD que permitió identificar 30 marcadores SNPs asociados significativamente al sexo en el SseLG18. La búsqueda de genes candidatos para la determinación del sexo identificó el gen del receptor de la hormona folículo estimulante (*fshr*) ubicado en una región caliente de recombinación, aunque con una penetrancia incompleta.

Además de marcadores SNPs, la información del genoma se usó en la búsqueda e identificación de marcadores SSRs. Así se identificaron 108 nuevos marcadores SSR polimórficos distribuidos por todo el genoma. Estos se estructuraron en 13 ensayos PCR multiplex (con hasta 10 *loci*) cuyas condiciones de amplificación se optimizaron y validaron con una alta calidad de análisis. Además, se seleccionó un subgrupo de 40 SSR muy polimórficos con los que se diseñaron 4 ensayos PCR supermultiplex (8-11 SSRs por ensayo) para su uso en el análisis del pedigrí. Además, se creó un nuevo mapa genético integrado con 229 SSRs distribuidos en 21 SseLGs mediante un análisis

genómico *in silico*. Ambos mapas generados en esta tesis se usaron para realizar estudios evolutivos del genoma de peces planos para identificar las fusiones robertsonianas específicas de linaje y otros reordenamientos que explican los cambios en el número cromosómico del cariotipo entre Pleuronectiformes.

Para estudiar los componentes genéticos de caracteres ligados al crecimiento y calidad morfológica, se obtuvieron estimas genéticas para distintas variables antes de entrar en el preengorde (400d) y al sacrificio (800d). Los caracteres relacionados con el crecimiento tal como peso corporal (W), la longitud estándar (SL), la anchura (Wi) y el área corporal (A) presentaron heredabilidades altas (0,568 y 0,609 a los 400 d y entre 0,424 y 0,500 a los 800 d) con correlaciones genéticas muy altas (>0,94) a ambas edades. Respecto a la calidad morfológica, se usaron 6 predictores incluyendo la elipsidad (E), la anchura del cuerpo en la base de la aleta pectoral (BHP), la anchura corporal máxima (BMH), la anchura del pedúnculo caudal (CPH) y dos ratios (BMH/BHP y BMH/CPH)). Los resultados indicaron altas heredabilidades para E, BHP, BMH y CPH (0,463-0,774) que fueron mayores a 400 d que a 800 d. Por el contrario, las ratios BMH/BHP y BMH/CPH presentaron una heredabilidad baja-moderada (0,144-0,306). Las correlaciones genéticas fueron positivas y positivas (>0,95) entre los caracteres de crecimiento y las tres anchuras, que disminuyeron con la edad. Por el contrario, la elipsidad presentó correlaciones genéticas negativas y medianamente altas con los caracteres de crecimiento y las anchuras, lo que indica que los peces que se seleccionen por su mayor tamaño serán también menos elípticos. Finalmente se realizó un estudio de asociación para encontrar marcadores genéticos ligados a los caracteres de crecimiento. Para ello, se diseñó y validó un chip ADN de baja densidad con 49 SNPs distribuidos en 17 SseLGs. El análisis de familias de alto y bajo crecimiento identificó dos marcadores significativos ligados al factor de transcripción general 3C polipéptido 4 y la proteína del proceso de fisión mitocondrial 1.

Los resultados obtenidos proporcionan herramientas muy potentes de análisis genómico, así como información genética muy valiosa para diseñar programas de mejora genética en lenguado con el fin de su impulsar su producción en acuicultura.

# CHAPTER 1

# GENERAL INTRODUCTION

## 1.1- Current status and importance of aquaculture

World population is increasing year by year, which leads to a higher demand for food. According to the Food and Agricultural Organization (FAO), it is expected that by 2030 there will be more than eight billion people in the world, so food production needs to increase at least as fast as the population (annual increase of 1.6%) (FAO 2020). Due to the saturation of land-based products (agriculture and livestock) and their limited growth, global demand for aquatic products from fisheries and aquaculture has increased considerably over the last three decades, during this time, production of these products has doubled, showing a growth of vital importance to cover the world's nutritional needs (average annual increase of 2.5% in the last 30 years). Global aquatic production (aquaculture and fisheries) in 2018 was 211.9 million tonnes (t), 2.6% more than the previous year. According to the APROMAR 2020 report, global aquaculture production reached 114.5 million t in 2018, 2 % more than the previous year, exceeding fisheries production by 17.1 million t. (APROMAR 2020) The great impact of aquaculture over last 30-40 years also shows a capacity for innovation in activity, a sustainable use of the resources at its disposal and environmentally friendly, this is why it has become a very important economic activity for many developed and developing countries (APROMAR 2020; FAO 2020).

In terms of global aquaculture production, Asia leads the ranking with 91.8% of production, followed by America (3.3%), Europe (2.7%) and Africa (2%). The leading country in 2018 was China with a production of 66.1 million t, far ahead of Indonesia in second place with a production of 14.7 million t. Concerning species, Japanese laminaria or kombu algae (*Saccharina japonica*) with 11.4 millions t, the eucheuma algae (genera *Eucheuma* and *Kappaphycus*) with 9.2 million t and the Japanese oyster (*Crassostrea gigas*) with 5.8 million t (APROMAR 2020; FAO 2020).

Focusing on aquaculture in Spain, in 2018, production was 347,825 t, occupying 20th place in the world ranking with an increase of 11.8% over the previous year and the first place in the European Union in aquaculture production. The main species produced were mussels (261,513 t), sea bass (27,335 t), rainbow trout (18,955 t) and sea bream (13,521 t). Most productive regions were Valencia, Galicia, Canarias, Murcia y Andalucía (APROMAR 2020).

## 1.2- Aquaculture of Senegalese sole (*Solea senegalensis*)

World production of Senegalese sole during 2019 increased compared to the previous year by 2.2% reaching 1,651 t, of which 818 t (49.5%) were harvested in Spain, showing itself as the main producer country ahead of Iceland, France and Portugal.

### 1.2.1 Relevant aspect of biology

Senegalese Sole (*Solea senegalensis* Kaup, 1858), is a demersal marine flatfish of the family Soleidae that inhabits mainly the Southern Atlantic and Western Mediterranean. One of the most important and characteristic aspects of Senegalese sole and other flatfishes is the metamorphosis from pelagic larva to benthic juvenile involving a series of regulated processes at the tissue, biochemical, physiological and molecular levels, in which hormones play a very important role, especially thyroid hormones. (Klaren et al. 2008; Manchado et al. 2008; Isorna et al. 2009). During this process a series of changes occur such as migration of one eye to the opposite side, remodeling of the head, drastic reorganization of the abdominal cavity, pigmentation patterns of the skin and development of sensory structures. These modifications provide them with a flattened shape specific for swimming and camouflage mechanisms that favor their adaptation to benthic life (Akkaynak et al. 2017). Senegalese sole is recognizable by its tall, elliptical bodies, short jaws, long fins and a great plasticity of skeletal components, such as the number of vertebrae ranging from 44 to 48 (mode = 45) with 8-9 in the abdominal region, 34-35 in the caudal region and 3-4 in the caudal complex (de Azevedo et al. 2017; Fernandez et al. 2017). This fish is one of the most valuable flatfish in southern European aquaculture due in part to the commercial value and quality of its flesh, which has led to an exponential increase in production in recent years. Moreover, a high attention was paid to this new species due to rapid increase in the production volume of sea bream (*Sparus aurata*) and sea bass (*Dicentrarchus labrax*), imports from third countries and the fall in prices in recent years (Dinis et al. 1999; Imsland et al. 2003). Significant advances in sole aquaculture such as larval rearing, optimization of dietary requirements and the use of recirculation technologies (RAS) for on growing have been done (Manchado et al. 2016; Morais et al. 2016; Manchado et al. 2019). However, despite the remarkable progress achieved on their biology and some technical aspects of cultivation, there are still several aspects that require optimization, especially with a view to their industrial exploitation such as improving production yields of juvenile sole fry and juvenile sole by having greater control over the genetic factors that determine

the growth and quality of the fry, but the two most important problems to solve are the skeletal malformations and the reproductive problems suffered by F1 males.

### 1.2.2- Skeletal abnormalities and morphology quality

Skeletal malformations are not an exclusive problem in the aquaculture production of Senegalese sole, since high incidence are found in many species, such as sea bass, sea bream, sea bream and yellowtail (Gavaia et al. 2002) among others, which leads to significant economic losses as these deformations affect the morphology of the fish, which has a direct impact on both the consumer and the animals themselves, as it affects their correct development and welfare (Lee-Montero et al. 2015).

In the case of Senegalese sole, skeletal malformations can affect more than 70% of individuals in a culture, which is a real problem for producers, as it is a species prone to suffer from this type of anomaly (de Azevedo *et al.* 2017). The origin of these malformations may have a genetic and/or environmental component, although the genetic factors are not well characterized as they have not yet been evaluated. At the environmental level, it has been seen that there are two important modulators that influence the malformations and morphological traits of this species: growing conditions and nutrition. Temperature above 18°C during larval stage has been shown to lead to increased vertebral anomalies in the caudal region and complex, although direct effects on external morphology were not assessed (Dionisio et al. 2012). Stocking density also influences the incidence of malformations when crop densities are high (29.8 kg $m^{-2}$), as it shifts relative body proportions towards a wider head and a shorter caudal region with a larger peduncle (Ambrosio et al. 2008). High dietary vitamin A levels lead to an increase in the average number of vertebrae and in the rate of caudal fin and vertebral malformations (Fernández et al. 2009).

The most common malformations in sole are vertebral fusions in the caudal region and deformities in the caudal complex, which often have a low morphological impact or go unnoticed (approximately 46 % of the animals with vertebral deformities were classified as normal) (de Azevedo *et al.* 2017). Moreover, it should be highlighted the great skeletal plasticity of this species coupled with a high incidence of malformations that can have an impact on the fish product at the time of marketing due to modifications of body ellipticity directly influencing the shape of the fish. Hence, it is very important to identify the phenotypic and genetic determination of the main morphological traits and the association with other productive parameters. For example, in the

closely related species *Solea solea*, body ellipticity measured by image analysis was proposed as an optimal trait to assess the quality of the external shape of sole (Blonk et al. 2010c).

### 1.2.3 Reproductive dysfunctions

Reproductive dysfunction of captive-bred animals (F1 generation) is another problem faced by companies, the major bottleneck for the production of viable larvae and for genetic breeding programs. Sole is a gonochoric species that shows a high population hierarchy and performs a characteristic courtship (stay quiet, rest head, guardian, following, coupled swim) when reproducing (Carazo et al. 2016).

Causes of this F1 reproductive dysfunction are not entirely clear: firstly, they do not display correct courtship behavior, which is necessary for successful spawning (Fatsini et al. 2016; Fatsini et al. 2017; Fatsini et al. 2020), on the other hand, although capable of producing viable gametes, clutches of captive-born soles are infrequent and unfertilized compared to those obtained from wild soles that lay without problem after acclimatization to captive conditions. F1 individuals have been shown to have lower fertilization capacity (Forne et al. 2011), low sperm concentration (<130 µl) and sperm quality (Cabrita et al. 2011; Chauvigne et al. 2016; Riesco et al. 2019), which makes it impossible to use *in vitro* techniques routinely in hatcheries (Chauvigne et al. 2017) that there are endocrine differences between F1 and wild males during spawning (Guzman et al. 2009; Riesco *et al.* 2019) and that hormonal therapies revealed as unsuccessful to release fertilized eggs (Agulleiro et al. 2006). Due to this limitation new strategies based on environmental control were developed based on thermocycles applied to males and females that partly mitigate this problem (Martin et al. 2019). This new approach can be used for the design of breeding programs based on mass spawning.

It should be indicated that sex ratios in sole are biased towards males indicating epigenetic regulation of sexual differentiation (Blanco-Vives et al. 2011). This is highly relevant for the industry since sole experience differential sex-linked growth, with females growing 50-100% larger than males (Sanchez et al. 2010). Identification of genetic sex-associated markers or development of procedures to reprogram larvae are a priority for the industry to increase female rates in the cultivated population to improve growth rates.

## 1.3- Genetic Breeding programs

The goal of a genetic breeding program is to select animals that are able to make the most efficient use of all the resources available in their environment by domesticating, so that they are adapted to living in captivity and do not suffer from high levels of stress. This leads to remarkable behavioral changes between wild and captive-bred animals to increase the productivity and quality and animal welfare (Gjedrem 2005b).

The first breeding programs were developed in livestock and have been in use for several decades. The increase of productivity has led to a reduction in the production costs and a considerable improvement in performance, such as the amount of milk produced by dairy cows (Berglund 2008), the number and size of eggs and weight in hens (Ducrocq et al. 2000), an increase in the weight of pigs for meat production (Davoli & Braglia 2008). Focusing on aquaculture species, situation is slightly different, although fish farming started many years ago, with the first described documents dating back to the 12th century BC in China, as well as the domestication of fish in the 5th century BC, applications of new methods and technologies in breeding programs have always been several steps behind the livestock and another farm-bred animal. In Europe, the common carp and rainbow trout were among the first species where captive breeding was possible, but currently the species at the forefront of aquaculture is the Atlantic salmon (*Salmo salar*) which began to be cultivated in the late 1960s (Gjedrem 2005a).

Although aquaculture is playing an increasing role in food production and is continuously growing, important factors such as the absence of genetically improved stocks in several species (preventing inbreeding), are slowing down this expansion, partly because most aquaculture species are bred without an advanced selection program. It should be indicated that aquatic organisms offer important advantages for the implementation of breeding because of their high fecundity rate linked to high levels of genetic variability and their phenotypic plasticity in response to hormones and environmental changes (Gjedrem 2005a). In *S senegalensis* spite of the notable advances achieved on their biology and some technical procedures of culture, there are still several aspects that require optimization, especially for their industrial exploitation. One of the key aspects are genetic breeding programs and genetic tools.

Under production conditions, breeding programs require both physical identification systems (Navarro et al. 2006; Rosyara et al. 2016) and genetic identification through molecular marker analysis (Navarro et al. 2008; Lee-Montero et al. 2013), in order to avoid costly genetic analysis at all points of commercial interest and sampling during the growth process (Toro & López-Fanjúl

2007). The combination of both technologies allows the evaluation of a wide range of production parameters, at the lowest possible cost (Lee-Montero *et al.* 2013), during the complete production cycle, i.e. offering companies the possibility to evaluate their animals from fingerlings to market size. A key point is the development of multiplex PCR systems for parental assignment under a mass-production system that enables the implementation of genetic improvement. This multiplex PCR system using simple sequence repeats (SSRs) or microsatellites has been used in other species such as sea bream (Lee-Montero *et al.* 2013), sea bass (Novel et al. 2010) or brown trout (Lerceteau-Köhler & Weiss 2006). A microsatellite multiplex PCR has also been described for Senegalese sole but with only 5 *loci* (Porta et al. 2006). The analysis based on best linear unbiased prediction (BLUP) allows us to obtain estimated breeding values (EBV) at family level from phenotypic traits and pedigree to be used to select the best candidates (Gjedrem 2012).

SSRs are short tandem repeat motifs that vary in the number of repeats (di-, tri-, tetra-or penta-nucleotide) and evolve on the basis of two opposing mutational forces: length mutations (increase the number of repeats) and point mutations (break these repeat motifs). Although single nucleotide polymorphisms have become the most widely used molecular markers (see below), SSRs still have features that make them very valid for multiplex PCR construction for parentage assignment and genetic variability analysis. Among these features are their genome-wide distribution, their high mutation rates compared to non-repeat sequences, their high polymorphism, their codominant inheritance and reproducibility. Another advantage of using this methodology is that it is cheap, fast and easy to analyze, which makes it affordable for use in laboratories that do not have large equipment (Lee-Montero *et al.* 2013; Zarouri et al. 2015), this makes it a highly accurate, accessible and easy to automate data collection tool. When designing a multiplex PCR, several essential aspects must be taken into account. In the multiplexing of loci, the primers must be designed very well as they must be combined with each other and must amplify simultaneously without unspecific or primer binding, the amplification conditions must be similar, the allelic range for markers with the same fluorophore must not overlap. All these conditions imply a prior knowledge of the genome information, with in silico study, experimental validation and refinement of the multiplex PCR.

For genetic breeding programs, it is essential to obtain genetic estimates of the productive traits. In aquaculture, growth, body shape quality or flesh characteristic traits are of particular relevance because of their economic impact. In Senegalese sole, heritability estimates and genetic correlations for growth and morphology traits have not yet been described until this thesis. In the

closely related flatfish common sole (*Solea solea*) the heritability for body weight and body length at slaughter is known to be between 0.21-0.28 and for ellipticity at 0.34 and these variables have a negative genetic correlation (-0.44) (Blonk et al. 2010b; Blonk *et al.* 2010c). These data indicate the possibility of using selection programs to improve production and economic performance for companies (Blonk *et al.* 2010b, a; Mas-Muñoz et al. 2013). This indicates that selection for size in genetic selection schemes should take morphological aspects into consideration to avoid excessive rounding of fish between generations. It has also been shown in common sole a good pedigree assessment increases the precision in the estimation of the breeding values (Blonk *et al.* 2010b). In addition, a genotype-environment interaction has been demonstrated in growth traits for soles cultivated in RAS vs open systems, demonstrating the importance of selected specific genetic lines for sole production (Mas-Muñoz *et al.* 2013). Therefore, more knowledge is needed to understand the genetic determination of productive traits in sole to be transferred to the companies.

## 1.4- New technologies for massive sequencing (NGS): Applications

New technologies for massive sequencing (NGS) have transformed the way in which the genome and transcriptome can be analyzed. These NGS technologies have become a tool with great potential for different genetic applications that include polymorphism analyses (such as GBS (Genotyping-by-sequencing) or RAD-Seq (Restriction-site associated DNA sequencing)) and quantitative (RNA-Seq) and qualitative transcriptome studies (Metzker 2010; Cerda & Manchado 2013; De Donato et al. 2013). NGS represents a significant progress for the use of markers for genetic selection and the implementation of novel genomic selection. Genomic selection is a marker-assisted selection method that covers the entire genome and is based on linkage disequilibrium of linkage of at least one marker with respect to the quantitative trait loci (QTL) (Zenger et al. 2019) and allows the estimation of interest values in a test population so that genomic estimated breeding values (GEBV) can then be predicted through the genomic ratio matrix (GRM) obtained from all the marker values used in the genome and applying genomic best linear unbiased prediction methods (GBLUP) (Goddard & Hayes 2007). This has an advantage compared to the breeding programs used in most aquaculture species as get 100% of the genetic variation and not only 50% corresponding to the interfamily variation used when measuring traits that require animal slaughtering and can be tested in the siblings of the candidates, e.g. meat quality or resistance to pathogens and diseases (Sonesson & Meuwissen 2009). This type of selection has been applied in mollusks like Marine shrimp and Pearl oysters (Goddard & Hayes 2007) and in fish such as

Atlantic salmon (Odegard et al. 2014), Large yellow croaker (Dong et al. 2016) or European sea bass (Vandeputte et al. 2019). Moreover, has facilitated the development of several other genetic studies such as genetic maps, association studies or the design of genetic tools as described below.

### 1.4.1- Genetics maps

Genetic linkage maps are considered an indispensable tool that allows us to go deeper into the genome organization of a species, revealing key aspects on the evolution and chromosomal architecture or divergence of species and at the same time opening the door to use of new complementary techniques that enhance the applicability of linkage mapping (Yue 2013). The basic principle is to place known molecular markers along the whole genome or a chromosome, quality and resolution of the mapping is closely related to the coverage and density of markers, as well as the type of marker used (Duran et al. 2009). Despite the rapid progress in the use of genetic maps in aquaculture, some problems such as low number of individuals per family or low density of markers or marker quality have limited their implementation in breeding programs (Robledo et al. 2018).

Molecular markers used have varied along with the genotyping techniques. Amplified fragment length polymorphism (AFLP) and simple sequence repeats (SSRs) were the most common some years ago (Vignal et al. 2002), but they present a series of disadvantages: while the firsts ones are dominants, which makes it difficult to transfer information between research teams, the latter, although co-dominant, present difficulties, as has been seen in genotyping of some crustacean and mollusk species where due to long repeating motifs and null alleles (Scarbrough et al. 2002), so AFLP and SSRs have been displaced by SNPs with the lowering of genotyping and sequencing costs. SNPs are bi-allelic and co-dominant markers widely distributed throughout the genome of an organism (Vignal *et al.* 2002). These characteristics have made them the current preferred marker for the construction of high-density genetic maps. They are widely used in non-model species such as most of the species used in aquaculture, including Senegalese sole, where only a genetic linkage map based on SSRs (Molina-Luzon et al. 2015a) and an integrated map using BAC clones and repetitive DNA families using a multiple fluorescent in situ hybridization (mFISH) technique (Garcia et al. 2019) have been reported.

### 1.4.2- Genome wide association studies (GWAS) and quantitative traits loci (QTL)

As high-density linkage maps are more easily available, association studies for searching chromosomal regions associated to valuable traits or QTLs have exploded (Laghari et al. 2015). QTLs are chromosomal regions that are associated with a specific phenotypic trait. These regions of DNA of interest can include a single gene or several, usually continuous variables are polygenic, highly influenced by the environment and genotype-environment interaction, knowing all the QTLs that determine a character indicates the genetic architecture of the same (Laghari et al. 2014). Many of valuable quantitative traits of aquaculture species fit these conditions (Massault et al. 2008).

Primary quantitative trait in most species and most evaluated is growth through many associated variables (weight, length, width, growth rates...), its medium-high heritability and the ease of evaluation as it does not require the slaughter of the animal, measurements are simple and many of them can even be carried out by image analysis (Liu & Cordes 2004) making it an ideal candidate that has already been evaluated in evaluated in numerous species of commercial interest such as bighead carp (Fu et al. 2016), Japanese flounder (Song et al. 2012b), Asian seabass (Wang et al. 2015a) or turbot (Wang et al. 2015b).

Another feature of economic importance is morphology, when the consumer is going to buy a fish, he chooses the one that best fits aesthetic canons of the chosen species, so that those with deformities or that are out of these canons will have no outlet and this leads to large losses for production companies. Studies to identify chromosomal regions associated with morphological variables have been carried out in species such as sea bass (Massault et al. 2010) and sea bream (Loukovitis et al. 2013). This type of analysis would be very useful in Senegalese sole where skeletal malformations are frequent and lead to economic damage.

In addition to growth and morphology, sex determination (SD) and identification of sex-linked genes is another trait to be considered in aquaculture (Martinez et al. 2014). These genes can be found on sex chromosomes and/or autosomes. SD is important as there are species in which one of the sexes (male or female) shows faster growth or matures earlier than the other. In halibut and tilapia, these techniques have allowed the identification of sex-linked genes and a major sex determination locus (Palaiokostas et al. 2013a; Palaiokostas et al. 2013b). In case of Senegalese sole, where females grow faster and more than males, decoding SD and understanding sex-linked chromosomal regions is important in this species, where so far, only cytogenetic studies have

shown absence of heteromorphic sex chromosomes, so such studies are necessary in this species to understand these key aspects of sex determination (Portela-Bens et al. 2017).

### 1.4.3- *De novo* genomes assemblies and re-scaffolding

The progressive reduction of DNA sequencing prices and the development of new sequencing technologies for long and short-reads and powerful analysis software capable of processing large amounts of information have driven the emergence of *de novo* assemblies for many species, including aquaculture species (Simpson & Durbin 2012). Genome assembly involves extracting and sequencing fragments of the genome, which are joined with other fragments to form contigs, which in turn are grouped with adjacent contigs to create scaffolds (Fierst 2015). With reduced costs, these sequencing and assembly techniques are affordable for small and medium budget laboratories. The problem of these draft genomes generated is that they contain a large number of short sequences without any information on how they assemble into larger linkage groups or chromosomes, making further molecular or evolutionary studies impossible (Pop 2009). Size of fragments generated is very important when it comes to ordering a genome, a high number of reads and short sequences increases flexibility when it comes to assembly, reducing chimeras, but the computation requires time and very complex algorithms, and it is also difficult to detect duplications or repeated sequences (Liao et al. 2019). Using longer fragments with fewer reads facilitates overlap assembly and requires less powerful software, but is difficult to correct sequencing errors within these long fragments and number of chimeric contigs is higher (Henson et al. 2012). An increasingly common alternative is hybrid assemblies using both short and long fragments to facilitate assembly and correct errors. This method has been used for example in hard-shelled mussel (Li et al. 2020) or pikeperch (Nguinkal et al. 2019).

Once a draft genome has been obtained and high density linkage maps are available, they can be anchored to build pseudochromosomes or linkage groups (LG) to validated bioinformatic algorithms using genetic recombination rates (Fierst 2015). One challenge for many of species of interest in aquaculture is that linkage groups or scaffolds fit to the expected number of chromosomes. In Senegalese sole, as mentioned above, a linkage map of SSRs clustered on 27 LGs (Molina-Luzon *et al.* 2015a) has been reported. However further efforts should be made to get to 21 chromosomes present in this species and anchoring of genetic and physical maps seems a feasible procedure. Moreover, this physical-genetic anchoring validate genome-wide sorting which makes possible to comparative genomics between species and to carry out in-depth synteny

studies to understand how chromosome structure has evolved within species lineages with respect to their ancestor (Palti et al. 2012).

### 1.4.4- DNA chips

As the number of SNPs increase thanks to NGS, they can be used in genetic tools such as DNA chips that offer a technology for genotyping in a rapid and cost-effective of a high number of genetic markers. Due to the vast amounts of information generated through the new massive sequencing technologies that allow the tracking of a complete genome, these chips can be used as a chromosomal tracking tool or specifically designed to study a specific trait once the chromosomal regions of interest related to that trait are known (Srivastava et al. 2013). The density of markers used can vary with high, medium and low-density chips. These chips have the advantage that they can be easily and routinely used in the laboratory and are a powerful tool to support genetic selection plans.

These strategies have proven successful in economically important aquaculture species, such as salmon (Houston et al. 2014), as they provide a wealth of information quickly, easily and inexpensively. These chips have been successfully applied to the identification of genes related to growth, immunity, sex determination or disease resistance, the identification of genes associated with response to environmental variations or the study of genes in other organisms (heterogeneous microarray hybridization) (Zhang et al. 2009; Eisbrenner et al. 2014).

# CHAPTER 2

# OBJECTIVES

The general objective of this PhD Thesis is the improvement of sole aquaculture production at industrial scale through the development and optimization of genetic technologies, as well as the development of innovative strategies in the selection of broodstock and their evaluation by means of the latest mass sequencing technologies (NGS).
To achieve this objective, molecular markers and genetic estimates of growth and morphology-related traits were studied in a collaborative framework linked to the development of the latest mass sequencing technologies. The information provide should serve as a basis for producing significant progress to increase the competitiveness and productivity of companies through innovation in their production procedures and new applications.

The specific objectives of this PhD Thesis are:

1- Generate a high-density map of SNP markers and a high-quality physical map to be later integrated in a reference genome assembly. Identify sex-linked chromosomal regions through GWAS analysis as well as chromosome rearrangements in flatfish through intra- and interspecific comparative mapping (Results available in chapter 3 of the present PhD Thesis).

2- Development, design and optimization of SSR multiplex assays for pedigree analysis in genetic breeding programs in sole. Integration of SSR markers in Senegalese sole in a genetic map and synteny analysis with the other flatfish species to understand chromosome evolution. (Results available in chapter 4 of the present PhD Thesis).

3- Study the genetic determination of growth and morphological quality traits during production cycle and its application to broodstock selection. Estimate of heritabilities and genetic correlations. Design of a low-density array and application in association analysis for growth traits. (Results available in chapter 5 and 6 of the present PhD Thesis).

# CHAPTER 3

## Chromosome anchoring in Senegalese sole (Solea senegalensis) reveals sex-associated markers and genome rearrangements in flatfish

The results of this chapter were published in: Guerrero-Cozar, I.; Gomez-Garrido, J.; Berbel, C.; Martinez-Blanch, J.F.; Alioto, T.; Claros, M.G.; Gagnaire, P.A.; Manchado, M. Chromosome anchoring in Senegalese sole (*Solea senegalensis*) reveals sex-associated markers and genome rearrangements in flatfish. Sci. Reports 2021, 11, 13460 DOI: 10.1038/s41598-021-92601-5

## 3.1- Introduction

Genetic maps represent essential tools for genomic research in aquaculture. Originally, linkage mapping studies were mainly based on microsatellite (SSR) and AFLP markers (Bouza et al. 2007; Reid et al. 2007); nevertheless, they recently reached a milestone with the development of genotyping methods based on cost-effective massive parallel sequencing. The genomic revolution has made single-nucleotide polymorphisms (SNPs) very popular, opening up access to a simple biallelic marker with a wide distribution and high abundance across the genome. As consequence, an increasing number of high-density genetic maps is nowadays reported in non-model organisms including aquaculture fish (Maroso et al. 2018; Nguyen et al. 2018). These maps have proven to be useful to provide new clues on genome evolution and speciation between closely related lineages, and to unravel the genetic architecture of both simple Mendelian and complex quantitative traits in many fish species, thus facilitating marker-assisted selection in aquaculture (Liu & Cordes 2004; Wang *et al.* 2015b). More recently, a new application of high-density linkage maps as backbones to anchor *de novo* genome assemblies into pseudo-chromosomes has become more widespread (Rastas 2017; Catchen et al. 2020). Although long-read sequences have significantly enhanced the average size of scaffolds in *de novo* assembled genomes (Goodwin et al. 2015), the total number of scaffolds are still far beyond the expected number of chromosomes. The large arrays of repeated sequences and the degree of conservation for some tandem repeats families widely distributed across the genome still remain a major obstacle for most *de novo* assembly algorithms, resulting in fragmented scaffolds or even misassembled sequences within chimeric contigs. Linkage maps thus provide highly valuable tools to anchor physical maps into pseudo-chromosomes, while enabling the identification of chimeric or misassembled contigs towards enhancing the quality of new genome assemblies (Rastas 2017).

Flatfish (Pleuronectiformes) is an attractive group of fish that have long been investigated due to the drastic morphological, physiological and behavioural remodelling changes that occur during metamorphosis from a pelagic larva to a benthic juvenile stage. Several flatfish species are worldwide exploited in fisheries and aquaculture, thus representing an important resource for human consumption. This taxonomic group diverged from carangimorphs in the early Paleocene, and underwent a major diversification in the middle Paleocene (Shi et al. 2018). Cytogenetic studies have suggested that the Pleuronectiformes ancestor should have 2n = 48 chromosomes in agreement with the most frequent number of chromosomes found in the sister clade Carangidae,

and in the most deep-branching flatfish families (Pleuronectidae and Paralichthyidae) (Azevedo et al. 2007). However, the number of chromosomes in flatfish encompasses a wide range varying from 2n = 26 to 2n = 50 (Azevedo *et al.* 2007; Garcia-Angulo et al. 2018). An intense cascade of Robertsonian rearrangements and pericentromeric inversions seems to have shaped flatfish genome evolution, especially reducing the chromosome number in most recently diverged families of Soleidae, Cynoglossidae and Achiridae (Azevedo *et al.* 2007). A recent comparison of the turbot genome with other fish assemblies clearly pointed out the high degree of conserved synteny across chromosomes in Pleuronectiformes, although with high rates of intrachromosomal reorganisations. Moreover, some chromosome fusions identified through comparative mapping are thought to have given arise to a new karyotype organization in turbot (Maroso *et al.* 2018). Hence, integrated genetic and physical maps are important genomic resources to understand chromosome evolution in flatfish.

The Senegalese sole is an important flatfish in aquaculture and fisheries. A genetic linkage map based on 129 SSRs grouped into 27 linkage groups (LG) was previously reported (Molina-Luzon et al. 2015b). Moreover, an integrated map using BAC clones and repetitive DNA families was also developed using a multiple fluorescence in situ hybridization (mFISH) technique with at least one BAC mapped to each chromosome arm (Garcia *et al.* 2019). This cytogenetic study evidenced a lack of heteromorphic sex chromosomes and identified the largest metacentric chromosome to result from a Robertsonian fusion of two acrocentric chromosomes during flatfish evolution (Portela-Bens *et al.* 2017; Rodriguez et al. 2019). Moreover, a preliminary draft genome sequence of a female Senegalese sole was reported (600.3 Mb, N50 of 85 kb), and then further improved with a hybrid assembly using Nanopore and Illumina reads (608 Mb long, N50 of 340 kb) (Manchado *et al.* 2016; Manchado *et al.* 2019). This genome information was used to design whole-genome multiplex PCR and create a new integrated SSR map with 234 markers. Nevertheless, further efforts are required to better assemble and anchor scaffolds onto the 21 expected chromosomes, and to better understand the genomic architecture of sex-determination.

The aim of this study was to: 1) generate an improved *de novo* assembly of a male Senegalese sole based on a combination of long and short read sequencing; 2) build a high-density genetic map using ddRAD markers; 3) anchor the physical to the genetic map in order to 4) improve the scaffolding of the reference genome assembly; 5) estimate genome-wide variation in recombination rates; and 6) carry out GWAS analysis to identify sex-associated markers and intra-

and interspecific comparative mapping to better understand the evolutionary history of chromosome rearrangements in flatfish.

## 3.2- Material and methods

### 3.2.1- Animals

Soles used for the preparation of ddRAD libraries and sequencing were selected from the genetic breeding program carried out by the IFAPA in collaboration with a commercial aquaculture company (CUPIMAR S.A.). Production of families used in this study, genotyping and parentage assignment were previously published (Guerrero-Cozar et al. 2020; Guerrero-Cozar et al. 2021). Five families (three full-sib and two maternal half-sib families) containing between 48 and 96 individuals per family (total n = 356) were selected to construct the genetic linkage map (Table 1). Moreover, seven families with sex ratios close to 1:1 were selected for genome-wide association analysis (GWAS). Average weight and length of each family are depicted in Table 1. As genotyping of parents was also required to build the genetic map, five fathers and three mothers involved in family production were sampled for blood by puncturing in the caudal vein using a heparinized syringe, adding heparin (100 mU) and keeping at -20°C until use. To obtain high-molecular weight genomic DNA for genome sequencing, a wild male from the broodstock (weight higher than 2 kg; code Sse05_10M) was sampled for blood as indicated above.

**Table 1**. Data about the families used to construct the genetic linkage map. Father (F) and Mother (M) of each family, the average weight and length and the number of specimens originally selected for analysis (n) are indicated. Moreover, the number of animals that passed that DNA quality analysis (nQ) and the final number of animals that passed after checking for Mendelian errors.

| Family name | Parents | Weight | Length | n | nQ | Final |
|---|---|---|---|---|---|---|
| Fam1 | F1/M1 | 161.6±94.3 | 20.6±4.0 | 76 | 76 | 73 |
| Fam2 | F2/M2 | 244.5± 157.8 | 22.7±4.4 | 95 | 95 | 90 |
| Fam3 | F3/M3 | 219.3± 95.9 | 22.4±3.5 | 68 | 67 | 65 |
| Fam4 | F4/M4 | 460.8± 195.4 | 27.8±4.1 | 99 | 79 | 77 |
| Fam5 | F5/M5 | 216.2±67.1 | 22.5±2.3 | 48 | 48 | 47 |
| Fam6 | F6/M5 | 345.5±136.2 | 25.6±3.4 | 71 | 65 | 63 |
| Fam7 | F7/M2 | 540.4±211.3 | 28.6±3.6 | 66 | 62 | 54 |
| Fam8 | F1/M8 | 129.8±72.7 | 19.5±3.9 | 76 | 73 | 73 |
| Total | | | | 599 | 565 | 543 |

All procedures were authorized by the Bioethics and Animal Welfare Committee of IFAPA and given the registration number 10/06/2016/101 by the National authorities for regulation of animal care and experimentation. The study was carried out in compliance with the ARRIVE guidelines and all procedures were performed in accordance with Spanish national (RD 53/2013) and European Union legislation for animal care and experimentation (Directive 86\609\EU).

### 3.2.2- Genome sequencing and assembly

High-molecular weight genomic DNA was prepared from heparinized whole blood using the MagAttract HMW DNA kit (Qiagen). Once confirmed quality, four libraries were prepared for sequencing using the Oxford nanopore Technology (ONT) MinION platform. Overall, 19.2 Gb of genome information was generated with an average read length of 4.3 kb. In parallel, the same sample was also sequenced in a NextSeq550 sequencer (Illumina, USA) that overall generated 43 Gb of sequence from 143 million reads (average length 147 nt). The raw read data were deposited to the NCBI Sequence Read Archive (SRA) under accession number SAMN16809702. The hybrid genome assembly was carried out using MaSuRCA*v*3.2.3 (Zimin et al. 2013; Zimin et al. 2017) with the Illumina libraries (57.3x coverage) and the error-corrected Nanopore reads (25.5x). The LR-hybrid assembly was characterized for completeness using Benchmarking Universal

Single-Copy Orthologs (BUSCO*v*3.0.2) (Simao et al. 2015; Waterhouse et al. 2018) containing 4,854 single-copy orthologs from actinopterygii_odb9.

### 3.2.3- ddRAD-seq library preparation and sequencing

Genomic DNA from the caudal fin (offspring) or whole blood (parents) were purified using the Isolate II Genomic DNA Kit (Bioline). DNA was sent to the company LifeSequencing S.L. (Valencia, Spain) and a total of 346 samples were selected for library construction (Table 1). Libraries were constructed based on the protocol described by Peterson et al. (2012) using the EcoRI/NcoI enzyme combination that generated as average 24,874 SNPs per sample. Pools of libraries were loaded on a Novaseq 6000 sequencer (Illumina), following the manufacturer's instructions and the specifications mentioned above.

### 3.2.4- Genetic linkage map and scaffold anchoring

Illumina reads were processed using Stacks v2.3e (Rochette & Catchen 2017). To construct the map, SNPs were filtered using Plink v1.9 (Purcell et al. 2007) to remove markers that segregated with Mendelian errors in more than 10% of individuals. Moreover, those individuals with more than 5% of markers with Mendelian errors were removed. The final SNP dataset contained 40,041 markers from 327 individuals (Table 1) and 8 parents that were imported in LepMap3 (Rastas 2017). The SNPs were assigned to 21 linkage groups (named as SseLGs) corresponding to the expected number of chromosomes (2n = 42) using the·"SeparateChromosomes" module. A LOD threshold of 11 and a size limit of 200 were selected as the most adequate parameters to keep an optimal number of markers grouped in the expected number of SseLGs (Figure 1A and 1B). Module JoinSingles2 was run to assign additional single SNPs to existing SseLG using decreasing LOD score iterations from 10 to 5 ( Figure 1B). Finally, the genetic distances between markers on each SseLG was calculated with the OrderMarkers2 module (male, female, sex average (SA)) using the Kosambi mapping function. The resulting genetic map was visualized using the software linkagemapview (Ouellette et al. 2018). Scaffolds anchoring was carried out using the Lep-Anchor program following the author's recommendation (Rastas 2020).

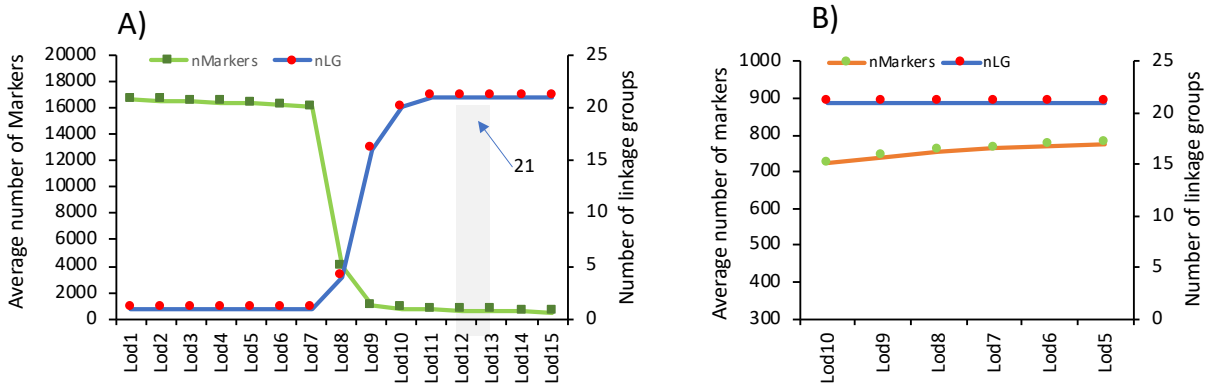


**Figure 1. Selection of LOD score limit (Lod) to construct genetic map in LepMap3.** A. The average of number of markers (nMarkers) positioned in linkage groups (left Y axis) and the number of linkage groups (nLG; right Y axis) for Lod values from 1 to 15 as implemented in the "SeparateChromosomes" module. Lod11 (shaded) indicates the value selected that grouped the markers in 21 LGs. B) Average number of markers recovered and added to the 21 LGs using decreasing LOD score iterations from 10 to 5 in the JoinSingles2 module.

### 3.2.5- Genome annotation

Genome annotation was performed by combining alignments of *Danio rerio*, *S. maximus* and *S. semilaevis* proteins, RNAseq from several tissues and developmental stages alignments and ab initio gene predictions. Functional annotation was performed on the male annotated proteins with Blast2GO (Conesa et al. 2005). After performing an alignment-based strategy to determine equivalences between female and male genomes, the female proteins inherited the functional annotation of their male equivalences. Next, functional annotation was performed in the female genes that remained unannotated after this step. Gene Ontology (GO) enrichment was carried out with topGO in those genes that were unique to one of the genomes.

### 3.2.6- Recombination rates, association analyses and cross-species comparisons

Recombination rate variation along the genome was evaluated by comparing the consensus linkage map for both sexes and SA and the physical map of each pseudo-chromosome using MareyMap (Rezvoy et al. 2007). The cumulative recombination frequency (RFm) along LGs was used to infer the chromosome type as previously described (Limborg et al. 2016). GWAS analysis were carried out with seven families (Table 1) using a logistic mixed model (multi-step) approach as implemented in the R package GENABEL (v1.8-0) (Aulchenko et al. 2007) for binary traits (Female= 0 and Male= 1). A highly detailed analysis of synteny across flatfish is beyond the scope

of this study, but a chromosome alignment analysis was carried out to identify chromosomal rearrangements in flatfish using D-Genies (Cabanettes & Klopp 2018). We then used the SatsumaSynteny to compute whole-genome synteny blocks (Grabherr et al. 2010) that were later represented using Shinycircos (Yu et al. 2018).

## 3-3 Results

### 3.3.1- Male genome assembly and annotation

A *de novo* hybrid genome for a male sole was assembled using a combination of Illumina and Nanopore long-reads. The hybrid assembly draft sequence was generated using MaSuRCA and later refined with Pilon to correct bases, mis-assemblies and filling gaps. The new assembly consists of 3,403 contigs with a total length of 609,359,514 bp, and a N50 of 513 kb. Overall, 49.4% of contigs had a size longer than 50 kb and the largest fragment was 4.5 Mb long. The estimated gene integrity, as determined by BUSCO analysis, revealed 97.0% completeness. For comparison purposes, the assembly statistics for a recent female genome draft of *S. senegalensis* are shown in (Claros et al. 2020; Guerrero-Cozar *et al.* 2020). Both genome assemblies had a similar size (608-610 Mb) although the newly assembled male genome had longer contigs with higher N50 values. A dot-plot alignment using the scaffolds of both genomes indicated that with 92.8% of genomic information highly similar (>75%) and only 5.3% had no similarity (average similarity 94%) (Figure 2).

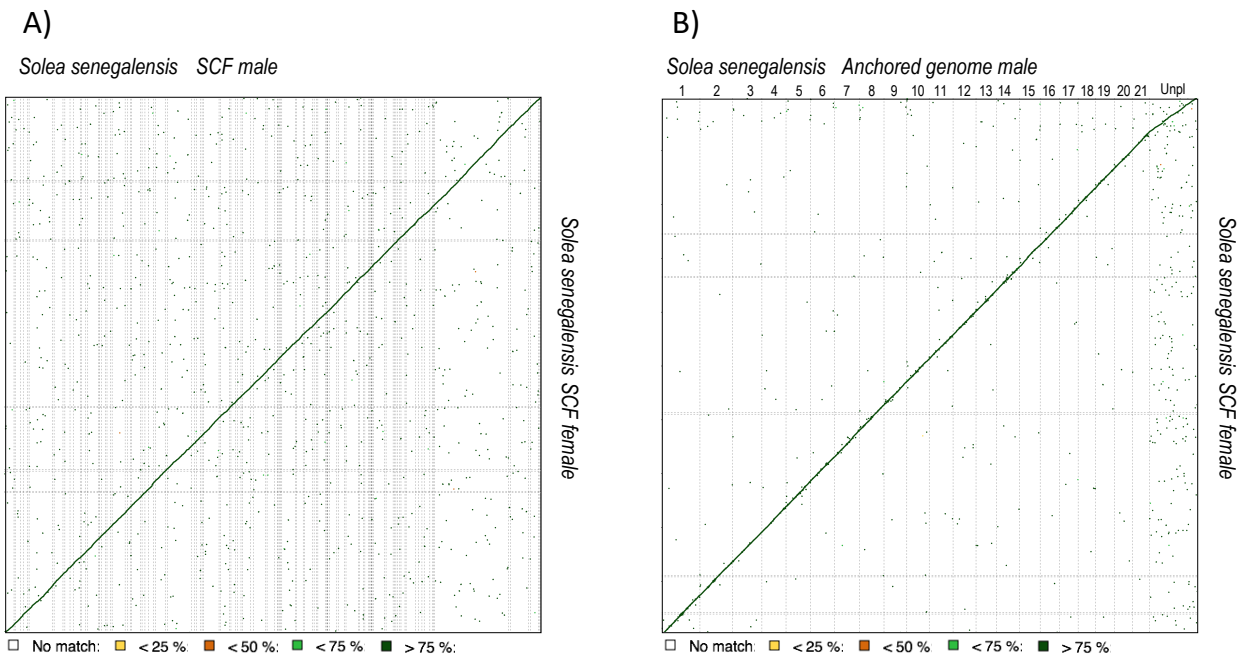


**Figure 2.** Dot plot comparison of scaffolds (SCF) assembled (A) or 21 pseudo-chromosomes (B) in the male with respect to SCF in the female. Scale is indicated below.

Assembly annotation statistics are depicted in Table 2. The number of protein-coding genes in the male assembly (27,175) was slightly lower than in the female (28,988) but with a longer mean length (7.4 vs 6.7 kb). The estimated percentages of annotated transcripts (69.4-72.1%) and gene density (45.03-47.68) were similar between both assemblies. Around 85% of the annotated genes in each assembly had an equivalent gene in the other assembly. However, a few genes were only present in one of the genomes (unique genes). Some of these might be due to genome heterozygosity and repeat content or even sex-specific genes. A GO enrichment analysis using these unique genes indicated that categories related to the cell-cycle regulation and regulation of transcription, involving canonical histones H3.2 and H4 and retinoid X receptor alpha (*rxra*), were highly significantly overrepresented in the female (p-value $<10^{-3}$). Mapping of these two histone genes on female assembly showed that they were co-localized in five scaffolds (Sosen1_s0284, Sosen1_s0324, Sosen1_s1454, Sosen1_s1522, Sosen1_s1726), four of which clustered in SseLG1 and one in SseLG16. In male, the most significant enriched categories for unique genes were skeletal system development and morphogenesis although with *P*-values >0.001. Some short, single-exonic unique genes might be the result of scaffold splitting or annotation processes. The

non-coding gene annotation resulted in 23,822 female and 21,123 male transcripts, respectively. From these, 6,549 and 6,007 female and male transcripts were long non-coding RNAs (lncRNAs) and the rest short non-coding RNAs.

**Table 2**. Summary annotation statistics for male and female assemblies.

| | Male | Female[#] |
|---|---|---|
| Repeat content | 23.55% | 23.41% |
| Number of protein-coding genes | 27,175 | 28,988 |
| Median gene length (bp) | 7,368 | 6,721 |
| Number of transcripts | 50,133 | 51,844 |
| Number of exons | 303,132 | 307,753 |
| Number of coding exons | 284,414 | 288,788 |
| Coding GC content | 52.67% | 52.57% |
| Median UTR length (bp) | 1,231 | 1,222 |
| Median intron length (bp) | 388 | 371 |
| Exons/transcript | 11.88 | 11,53 |
| Transcripts/gene | 1.84 | 1.79 |
| Multi-exonic transcripts | 0.956 | 0.941 |
| Gene density (gene/Mb) | 45.026 | 47.679 |
| Functionally annotated transcripts | 36,130 (72.1%) | 35,999 (69.4%) |
| Unique genes | 3,806 (14%) | 4,643 (16%) |
| non-conding RNAs | 21,123 | 23,822 |

[#] Sequence deposited in figshare https://doi.org/10.6084/m9.figshare.12472100.v1.

### 3.3.2- ddRAD sequencing and SNP detection for genetic linkage map

Three full-sib and two half-sib families consisting of 47 to 95 individuals were used for ddRAD analysis (Table 1). The total number of paired-end reads generated for each family ranged between 280,609,738 (F5) and 398,313,256 (F2) with an average length of 150 nt (Table 3). The average number of reads per individual in each family varied between 6,444,752 (F1) and 11,692,072 (F5) (Table 3). For parents, the average number of reads was 8,847,913.

The new assembled male genome was used as reference to map the ddRAD reads. The average fraction of primary alignments onto this reference genome ranged between 88.04 (F6) and 89.71% (F2). An average of 10.5% of reads had insufficient mapping qualities or excessively soft-clipped primary alignments while less than 0.34% were unmapped. A total of 199,188 ddRAD *loci* were

reconstructed with an average number of loci per sample ranging between 23,828 (F1) and 30,550 (F7) and a mean insert length of 330.7 bp. The effective coverage per sample was 193.3±110.4 (ranging from 146 to 242 between families) and the estimated mean number of sites per locus was 242.8 (Table 3).

**Table 3**. Main statistics of ddRAD libraries, mapping and SNP detection. The total number of individuals analyzed (n), the total reads per family, the average number of paired-end reads per Invidual, the average number reads used by stacks, the % of primary alignment and unmapped reads, number of loci, effective coverage, and number of genotypes (n_gts)

| | n | Total reads family | Av. raw reads | Av. reads stacks | PA (%) | Unmapped | loci | mean cov | n_gts |
|---|---|---|---|---|---|---|---|---|---|
| F1 | 76 | 244,900,564 | 6,444,752 | 6,215,911 | 88.23 | 0.34% | 23,828 | 146 | 22,040 |
| F2 | 95 | 398,313,256 | 8,385,542 | 8,090,267 | 89.71 | 0.33% | 24,978 | 190 | 22,823 |
| F5 | 48 | 280,609,738 | 11,692,072 | 11,384,985 | 88.13 | 0.33% | 30,005 | 237 | 27,011 |
| F6 | 65 | 363,499,961 | 11,184,614 | 10,899,007 | 88.04 | 0.31% | 27,742 | 242 | 24,883 |
| F7 | 62 | 337,573,225 | 10,889,459 | 10,627,007 | 88.93 | 0.34% | 30,550 | 226 | 26,773 |
| Parents | 8 | 39,815,609 | 8,847,913 | 8,323,338 | 86.08 | 0.36% | 17,632 | 242 | 15,898 |

### 3.3.3- Construction of a linkage genetic map and anchoring to physical map

To construct the genetic map, only those SNPs detectable in at least 80% of samples with a coverage of 10 reads per sample were considered. Moreover, SNPs with a significant deviation from Mendelian segregation were also removed (a total of 2,439 markers, 5.7% SNPs). By family, the number of markers with Mendelian errors ranged from 1.5 to 1.7%. Moreover, those animals with markers that had more than 5% of Mendelian errors (19 specimens) were also removed. Overall, the final dataset contained 40,041 SNPs segregating in eight parents and their 327 offspring.

For linkage analysis, the ParentCall2 module retained only 16,287 informative markers after checking for segregation distortion ($P < 0.05$). Markers grouped into 21 SseLGs (via the SeparateChromosomes2 module) with a LOD = 11 ( Figure 1), which is consistent with the number of chromosomes in *S. senegalensis*. Each SseLG contained between 530 and 1,337 markers with

an average number of 21.9 markers per Mb (Figure 3, Table 4 "Anchoring genetic map and physical map"). In total, the genetic map allowed the anchoring and positioning of 1,665 out of 3,403 total contigs, ranging between 50 to 129 contigs in each SseLG. The genome sequence positioned on the linkage map was larger (746.3 bp) than the assembly size, mainly due to the presence of chimeric contigs (n = 133) positioned in various chromosomes.

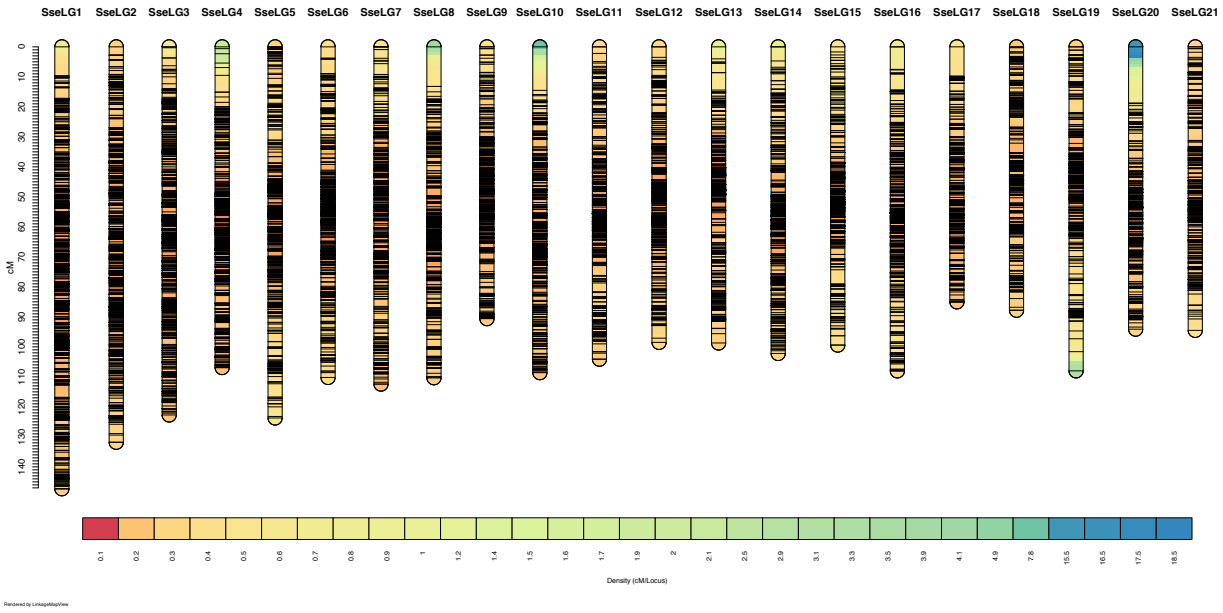


**Figure 3.** Genetic distance (cM) and SNP distribution across 21 linkage groups (SseLG) of the Senegalese sole.

### 3.3.4- Rescaffolding of reference genome with the genetic map

SNP marker information was further used for fine-scale correction of genome contigs to build 21 pseudo-chromosomes. After masking the repetitive sequences, the contigs were orientated and sorted within each SseLG (Table 4 "Genome re-scaffolding"). The total number of positioned contigs reduced from 1,665 to 1,563. Lep-anchor corrected the contig errors removing six contigs, splitting another 105 into two fragments, 20 in three fragments, and two in more than four fragments. After these corrections, the total number of markers assigned to the SseLGs decreased by 1.3% (16,075 SNPs) and 212 markers were moved to unplaced with an average density of 10.3 markers per contig. After these corrections, 548.6 Mb out of the 610.4 Mb total assembly length (89.9%) were assigned to the 21 SseLGs and only 61.9 Mb remained as unanchored (Table 4).

**Table 4.** Information for anchored physical map (LepMap3 step), after genome re-scaffolding (Lep-anchor3 step) and after removal of markers with discrepancies between genetic and physical maps (MareyMap step). The physical (bp) and genetic (cM) length of each linkage group, number of markers (nMar), number of contigs (nCon), average contig length (ACL), marker density density (markers per megabase; M/Mb) and the ratio physical to genetic length (Mb/cM) for sex-average genetic-physical map are indicated.

| | Anchoring genetic map and physical map | | | | | Genome re-scaffolding | | | | | | Marker refining | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Length (bp)** | **nMar** | **nCont** | **ACL** | **M/Mb** | **Length(bp)** | **NM** | **nCont** | **ACL** | **L(cM)** | **M/Mb** | **NMar** | **Mb/cM** | **M** |
| **1** | 59,220,137 | 1,337 | 129 | 459,071 | 22.6 | 42,924,012 | 1,323 | 124 | 343,392 | 147.3 | 30.8 | 1,296 | 0.29 | 30 |
| **2** | 42,658,310 | 1,054 | 91 | 468,773 | 24.7 | 36,396,255 | 1,046 | 88 | 413,594 | 131.8 | 28.7 | 1,032 | 0.28 | 28 |
| **3** | 47,587,809 | 1,015 | 85 | 559,857 | 21.3 | 33,319,822 | 1,006 | 80 | 416,498 | 136.6 | 30.2 | 978 | 0.24 | 29 |
| **4** | 42,630,187 | 920 | 83 | 513,617 | 21.6 | 27,129,084 | 899 | 73 | 366,609 | 106.9 | 33.1 | 885 | 0.25 | 32 |
| **5** | 32,366,427 | 891 | 86 | 376,354 | 27.5 | 27,692,037 | 872 | 78 | 350,532 | 142.5 | 31.5 | 811 | 0.19 | 29 |
| **6** | 34,539,569 | 864 | 80 | 431,745 | 25.0 | 26,866,643 | 860 | 77 | 348,917 | 114.0 | 32.0 | 832 | 0.24 | 31 |
| **7** | 36,891,773 | 849 | 87 | 424,043 | 23.0 | 28,334,760 | 836 | 77 | 367,984 | 133.8 | 29.5 | 795 | 0.21 | 28 |
| **8** | 36,615,909 | 784 | 86 | 425,766 | 21.4 | 27,361,452 | 769 | 82 | 333,676 | 119.3 | 28.1 | 756 | 0.23 | 27 |
| **9** | 32,328,246 | 804 | 65 | 497,358 | 24.9 | 25,679,769 | 802 | 63 | 407,615 | 105.1 | 31.2 | 765 | 0.24 | 29 |
| **10** | 35,518,751 | 768 | 88 | 403,622 | 21.6 | 25,170,845 | 762 | 84 | 299,653 | 113.7 | 30.3 | 748 | 0.22 | 29 |
| **11** | 37,595,336 | 780 | 99 | 379,751 | 20.7 | 26,846,769 | 769 | 93 | 288,675 | 126.2 | 28.6 | 732 | 0.21 | 27 |
| **12** | 37,197,923 | 763 | 80 | 464,974 | 20.5 | 25,840,656 | 752 | 77 | 335,593 | 98.5 | 29.1 | 731 | 0.26 | 28 |
| **13** | 34,656,556 | 665 | 50 | 693,131 | 19.2 | 23,154,965 | 658 | 48 | 482,395 | 98.7 | 28.4 | 637 | 0.24 | 27 |
| **14** | 33,597,656 | 668 | 76 | 442,074 | 19.9 | 26,091,242 | 665 | 74 | 352,584 | 109.5 | 25.5 | 637 | 0.24 | 24 |
| **15** | 36,416,189 | 644 | 66 | 551,760 | 17.7 | 22,903,974 | 632 | 59 | 388,203 | 113.1 | 27.6 | 601 | 0.20 | 26 |
| **16** | 26,721,177 | 630 | 58 | 460,710 | 23.6 | 21,637,702 | 618 | 52 | 416,110 | 108.0 | 28.6 | 602 | 0.20 | 27 |
| **17** | 30,251,165 | 616 | 79 | 382,926 | 20.4 | 21,095,432 | 610 | 75 | 277,572 | 103.3 | 28.9 | 563 | 0.20 | 26 |
| **18** | 24,300,965 | 587 | 62 | 391,951 | 24.2 | 19,718,726 | 577 | 57 | 345,943 | 87.8 | 29.3 | 561 | 0.23 | 28 |
| **19** | 36,478,108 | 584 | 75 | 486,375 | 16.0 | 21,051,312 | 575 | 70 | 296,497 | 108.0 | 27.3 | 562 | 0.20 | 26 |
| **20** | 24,034,263 | 534 | 62 | 387,649 | 22.2 | 20,166,255 | 530 | 62 | 325,262 | 105.7 | 26.3 | 497 | 0.19 | 24 |
| **21** | 24,720,343 | 530 | 78 | 316,928 | 21.4 | 19,202,697 | 514 | 70 | 270,461 | 98.3 | 26.8 | 490 | 0.20 | 25 |
| **ST** | 746,326,799 | 16,287 | 1,665 | 453,259 | 21.9 | 548,584,409 | 16,075 | 1,563 | 349,640 | 2,408.1 | 29.3 | 15,511 | 0.23 | 28 |
| Not-anchored | | | 1,738 | | | 61,859,804 | 212 | 1,840 | | | | 776 | | |
| **Total** | 746,326,799 | 16,287 | 3,403 | 453,259 | 21.9 | 610,444,213 | 16,287 | 3,403 | | | | 16,287 | | |

The total map length was 2,408.1 cM, SseLG1 was the largest group (42,924,012 bp and 147.3 cM) and SseLG4 showed the highest marker density per megabase (33.1). The average marker interval reached 0.155 cM. A further refining of anchored markers was carried out through the comparison of physical and genetic distance in MareyMap. The average genome-wide recombination rate (RR) was 4.35 cM/Mb (ranging between 3.45 and 5.26 cM/Mb among chromosomes) (Table 4 "Marker refining"). An alignment of the anchored and refined reference male genome with the scaffolds of the female assembly ( Figure 2B) slightly increased to 93.2% the regions with more than 75% similarity and provided a clear sequence alignment in the diagonal with only dispersion in unplaced scaffolds.

### 3.3.5- Analysis of recombination rates

Consensus genetic maps for female and male were 2,698.4 cM (15,022 markers) and 2,036.6 cM (15,390 markers), respectively. These differences in map size were observable for the 21 SseLGs (Figure 4A and Table 5). Overall, the female-to-male ratio (F:M) for genetic distances was 1.32, ranging from 1.08 (SseLG15) to 1.77 (SseLG5) (Table 5). The genetic map length of chromosomes was highly positively correlated with their physical length in both males (r = 0.43) and females (r = 0.60) ( Figure 4B).

**Table 5**. Refined genetic maps for male (M) and female (F). The genetic (cM) length of each linkage group, number of markers (nMar), the ratio physical to genetic length (Mb/cM), marker density (markers per megabase; M/Mb), the F:M ratio of genetic map lenght, the recombination rates (RR) in both sexes and the F:M ratio of RR are indicated.

| | Male genetic map | | | | Female genetic map | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | nMar | L(cM) | Mb/cM | M/Mb | nMar | Length (cM) | Mb/cM | M/Mb | F:M (cM) | RRM | RRF | F/M (RR) |
| **1** | 1,297 | 117.7 | 0.37 | 30.2 | 1,254 | 175.7 | 0.24 | 29.2 | 1.49 | 2.56 | 4.05 | 1.58 |
| **2** | 1,027 | 105.5 | 0.35 | 28.2 | 998 | 156.2 | 0.23 | 27.4 | 1.48 | 2.64 | 4.09 | 1.55 |
| **3** | 976 | 124.9 | 0.27 | 29.3 | 962 | 145.9 | 0.23 | 28.9 | 1.17 | 3.05 | 4.15 | 1.36 |
| **4** | 881 | 83.3 | 0.33 | 32.5 | 868 | 128.8 | 0.21 | 32.0 | 1.55 | 2.58 | 3.57 | 1.38 |
| **5** | 811 | 101.4 | 0.27 | 29.3 | 811 | 179.5 | 0.15 | 29.3 | 1.77 | 3.38 | 5.65 | 1.67 |
| **6** | 833 | 103.6 | 0.26 | 31 | 814 | 122.2 | 0.22 | 30.3 | 1.18 | 3.24 | 4.50 | 1.39 |
| **7** | 786 | 126.3 | 0.22 | 27.7 | 777 | 138.2 | 0.21 | 27.4 | 1.09 | 2.73 | 4.75 | 1.74 |
| **8** | 737 | 112.9 | 0.24 | 26.9 | 758 | 140 | 0.20 | 27.7 | 1.24 | 3.15 | 3.94 | 1.25 |
| **9** | 757 | 84.8 | 0.30 | 29.5 | 762 | 106.4 | 0.24 | 29.7 | 1.25 | 2.78 | 4.12 | 1.48 |
| **10** | 732 | 86.6 | 0.29 | 29.1 | 713 | 115 | 0.22 | 28.3 | 1.33 | 3.50 | 4.50 | 1.28 |
| **11** | 722 | 111.8 | 0.24 | 26.9 | 724 | 137.6 | 0.20 | 27.0 | 1.23 | 3.16 | 3.85 | 1.22 |
| **12** | 709 | 77.3 | 0.33 | 27.4 | 677 | 118.2 | 0.22 | 26.2 | 1.53 | 2.47 | 4.70 | 1.90 |
| **13** | 628 | 84.6 | 0.27 | 27.1 | 613 | 110.7 | 0.21 | 26.5 | 1.31 | 2.76 | 4.15 | 1.50 |
| **14** | 645 | 100.3 | 0.26 | 24.7 | 608 | 116.4 | 0.22 | 23.3 | 1.16 | 2.99 | 4.10 | 1.37 |
| **15** | 609 | 110.5 | 0.21 | 26.6 | 574 | 119.3 | 0.19 | 25.1 | 1.08 | 2.64 | 4.41 | 1.67 |
| **16** | 575 | 91.6 | 0.24 | 26.6 | 580 | 119.6 | 0.18 | 26.8 | 1.31 | 3.60 | 5.15 | 1.43 |
| **17** | 585 | 80.1 | 0.26 | 27.7 | 540 | 123.7 | 0.17 | 25.6 | 1.54 | 3.38 | 5.17 | 1.53 |
| **18** | 552 | 75.4 | 0.26 | 28 | 542 | 98.5 | 0.20 | 27.5 | 1.31 | 3.05 | 4.87 | 1.60 |
| **19** | 555 | 91.2 | 0.23 | 26.4 | 543 | 111.8 | 0.19 | 25.8 | 1.23 | 3.58 | 5.33 | 1.49 |
| **20** | 502 | 84.1 | 0.24 | 24.9 | 458 | 122.7 | 0.16 | 22.7 | 1.46 | 2.64 | 4.26 | 1.61 |
| **21** | 471 | 82.7 | 0.23 | 24.5 | 446 | 112.1 | 0.17 | 23.2 | 1.36 | 3.47 | 5.38 | 1.55 |
| **ST** | 15,390 | 2,036.6 | 0.27 | 28.1 | 15,022 | 2,698.4 | 0.20 | 27.4 | 1.32 | 3.02 | 4.51 | 1.49 |
| **NA** | 897 | | | | 1,265 | | | | | | | |
| **Total** | 16,287 | | | | 16,287 | | | | | | | |

The average genome-wide RR was estimated 3.02 ± 0.37 cM/Mb in males and 4.51 ± 0.57 cM/Mb in females (Table 5). The overall female-to-male ratio (F: M) for RR was 1.49, ranging from 1.43

to 1.90 across chromosomes. In the case of males, SseLG12 showed the lowest (2.47 cM/Mb) and SseLG16 the highest (3.60) mean RR values. In females, SseLG4 had the lowest (3.57 cM/Mb) and SseLG5 the highest (5.65 cM/Mb) mean RR values.

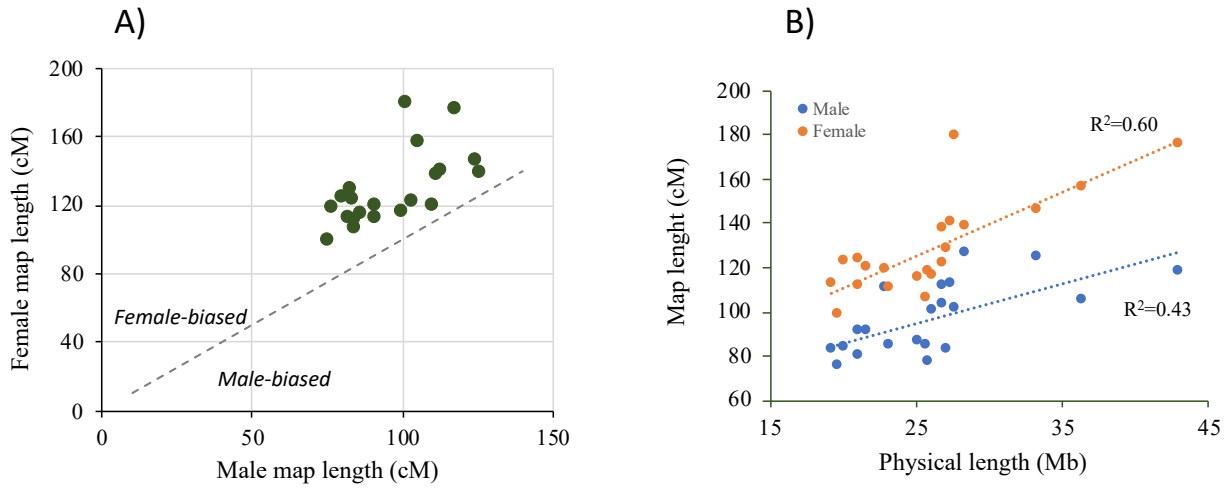


**Figure 4.** Comparison of male and female genetic maps. (A) Male *vs* female linkage groups lengths (cM) for the 21 Senegalese sole chromosomes. All chromosomes exhibit female-biased recombination. (B) Correlation between recombination map and physical map lengths in both males (blue) and females (orange). The determination coefficient $R^2$ is shown separately for each sex.

The local RR value as estimated by the relative distance to the nearest telomere was clearly different between males and females. High RR values were mainly concentrated close to the telomeres in males (Figure 5A), while they were more uniformly distributed in females with higher RR being found around 15% of the distance to the nearest telomere (Figure 5B). This was illustrated by contrasted chromosomal RR landscapes between males and females, as shown Figure 5 C&D for SseLG1. We detected some regions within SseLGs (i.e 5, 11, 13, 14, 15, 18) with very low RR. In the case of SsseLG18, partially restricted male or female RR was detected in the region 9.5-10.9 Mb. This region had very low RR in males (1.2) and females (0.6) compared with average SseLG18 (3.0 and 4.9 RR, respectively). Cumulative RR crossed between both sexes around chromosomal position 10 Mb with female RR closed to zero in 10.8-10.9 Mb ( Figure 6). Moreover, recombination frequencies were used to describe and classify chromosome morphologies. Figure 7 depicts the typical RFm plots for an acrocentric (SseLG20) and a metacentric (SseLG1) chromosome.

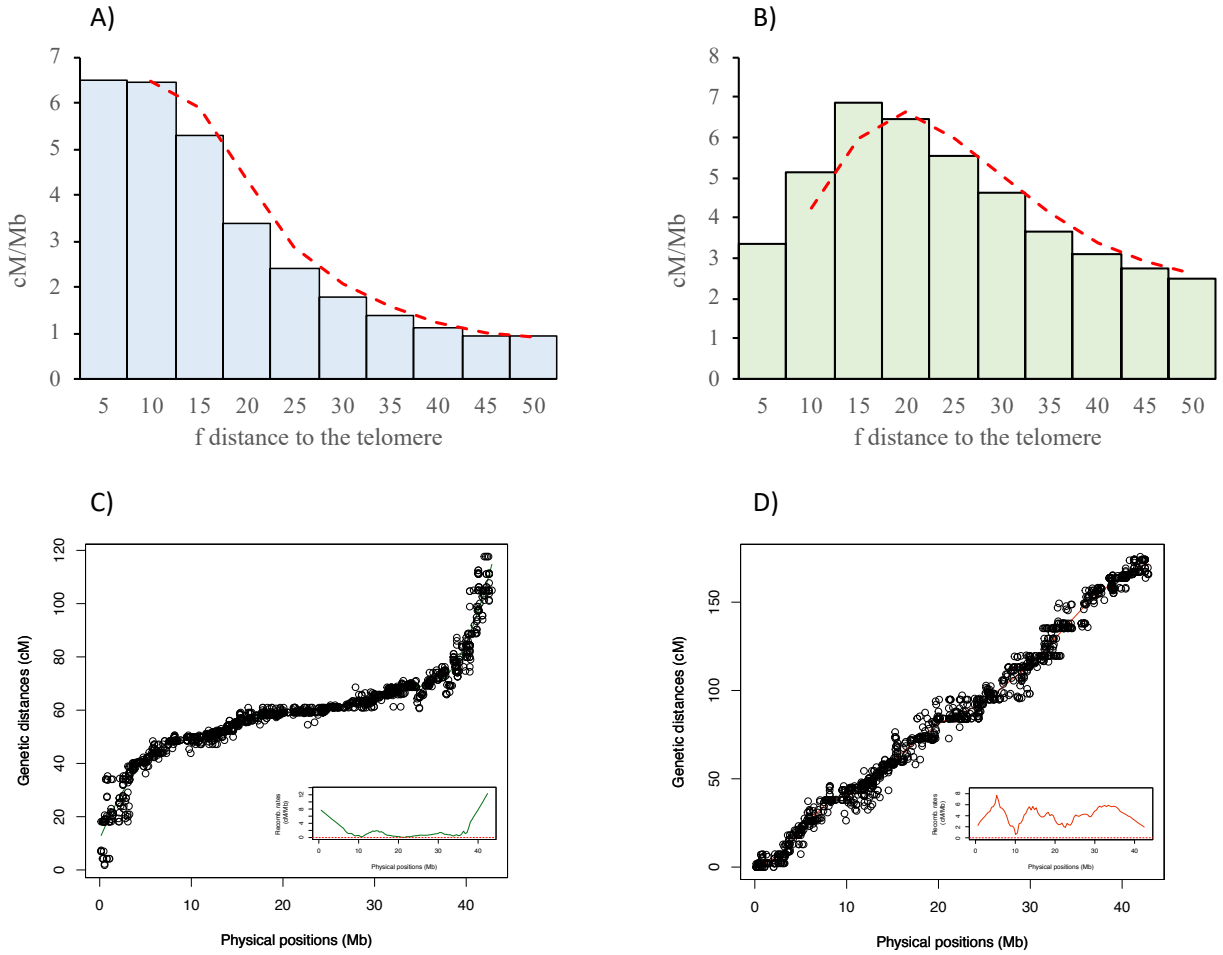


**Figure 5.** Recombination landscape averaged across linkage groups for (A) male and (B) female. The recombination rates (cM/Mb) and the relative distance from the nearest telomere scaled by the chromosome length (f) is represented. The red dashed line indicates the observed tendency. Panels (C) and (D) show the relationship between physical and genetic distances for SseLG1 in male and female, respectively. The square inside the panels C and D show the specific recombination landscape.

### 3.3.6- Association analyses for sex

To identify genome regions associated with sex, a GWAS analysis was carried using seven families (Table 1) and a total of 10 426 markers. Data for RAD-seq data and markers are indicated in Table 3. The results showed 30 markers significantly associated with sex after bonferroni correction using seven families ($P \leq 4.8\text{x}10^{-6}$; Figure 6A). When the association analysis was repeated separately by family, five families provided some new 36 significant markers. All of them (66 SNPs including the whole-population and families) were spread in the SseLG18 with a hot

region around 9.5-10.9 Mb (Figure 6B). RR in this region was low (see above) with partially restricted RR associated with sex. Overall, 80.7% of significant markers using the whole population were preferentially heterozygous in males although penetrance was incomplete in most of them. This model is compatible with a nascent XY system. It should be noted that specific markers in family 4 had an expected high number of heterozygous *loci* in females.

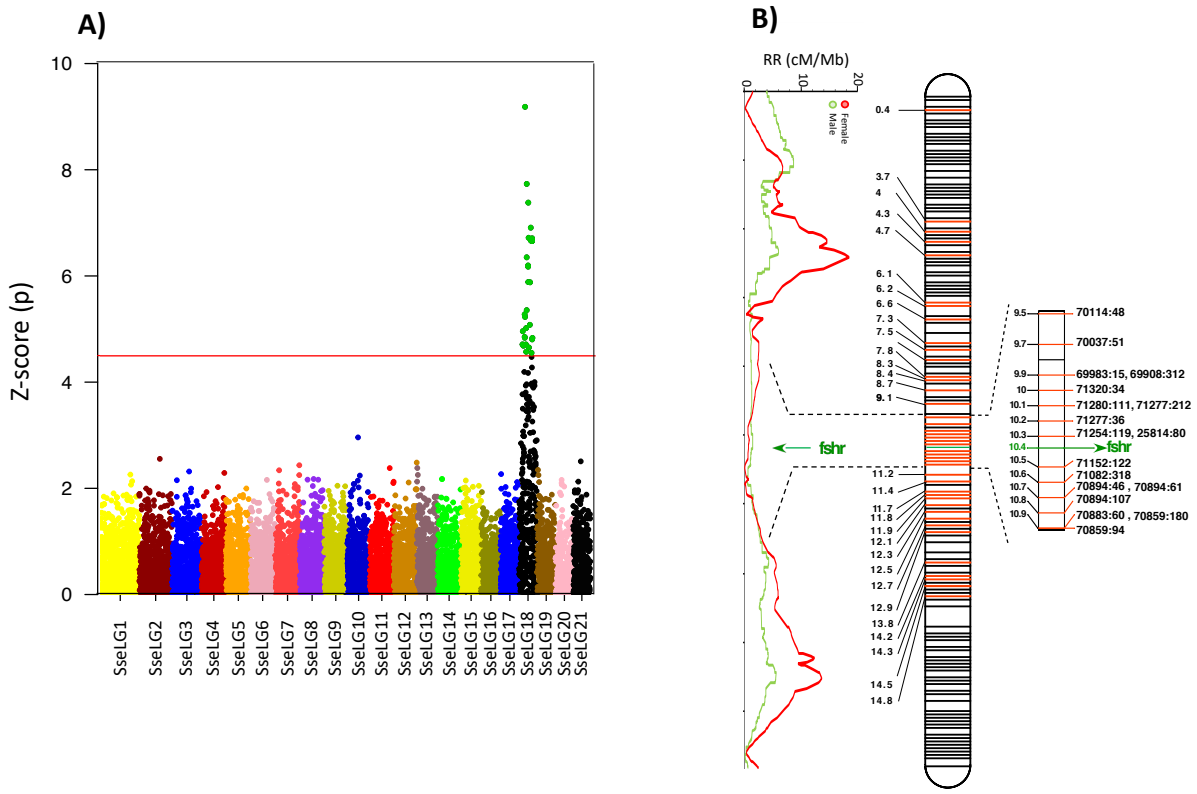


**Figure 6.** Sex-associated SNPs and RR landscape for males and females in SseLG18. A) Manhattan plot of GWAS results for sex-associated SNPs using seven families. Significant markers are indicated in green. The horizontal red line represents the Bonferroni significance threshold. B) Distribution of all 66 sex-associated significant markers using seven families and by family (in red) and RR (cM/Mb) landscape of males and females. A hot region from 9.5 to 10.9 Mb containing the candidate gene *fshr* is indicated on the right side. Physical positions of SseLG18 in Mb are indicated in black. Black lines indicate non-significant markers in SseLG18.

To detect candidate sex-related genes, the full-length transcriptome(Cordoba-Caballero et al. 2020) was blasted onto the SseLG18 and a total of 229 genes were positioned. The significant

SNPs were highly distributed through the pseudo-chromosome, but the follicle stimulating hormone receptor (*fshr*) gene just appeared located in the hot region revealing as a clear candidate gene for sex determination.

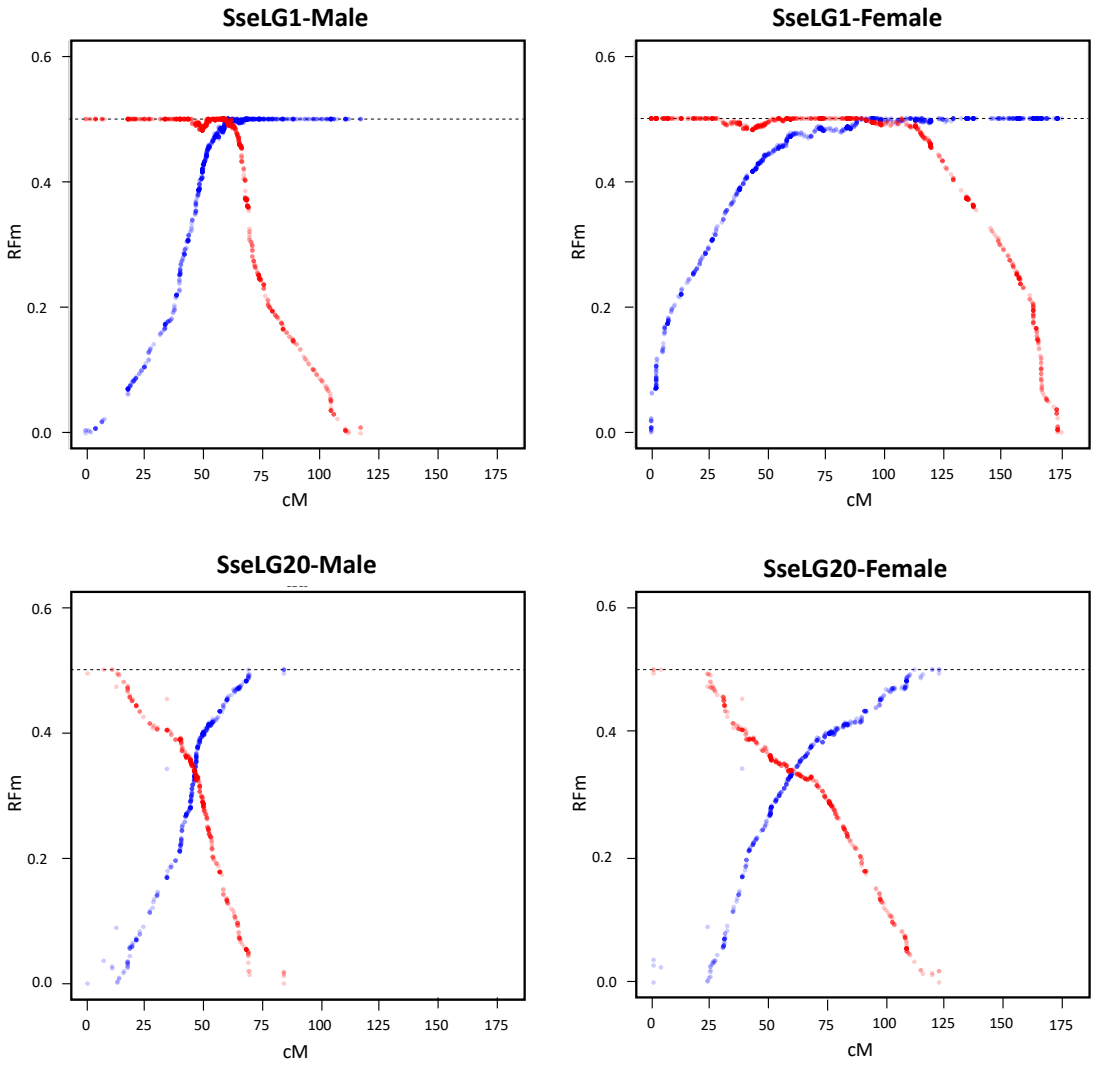


**Figure 7.** Plots illustrating the recombination frequency estimates (RFm) for intervals between markers along SseLG1 and SseLG20 in the male and female. For each LG, RFm was calculated from both chromosomal extremities (right: red circles; left: blue circles), using each of the two terminal markers as a reference starting point. The RFm plots of SseLG1 and SseLG20 show a classical metacentric and acrocentric pattern, respectively.

### 3.3.7- Interspecific chromosome rearrangements

An alignment of SseLGs pseudo-chromosomes with the chromosomes of three other Pleuronectiformes genomes (*Cynoglossus semilaevis*, *Scophthalmus maximus*, *Paralichthys olivaceus*) showed high similarity rates of and conserved macrosynteny level for fifteen out of 21 SseLGs (Figure 8). However, deviations from diagonal in the dot plot alignment indicated extensive intrachromosomal rearrangements among species. The three largest SseLGs appeared to be the result of total or partial chromosome fusions when compared with other flatfish genomes, and *S. maximus* seemed to be the flatfish species with the highest number of chromosome rearrangements between the four species compared. Genome comparisons using D-Genies (Cabanettes & Klopp 2018) indicated that the highest similarity was with *P. olivaceus* (no match 57.3%), followed by *S. maximus* (no match 59.6%), and *C. semilaevis* (no match 78.4%).

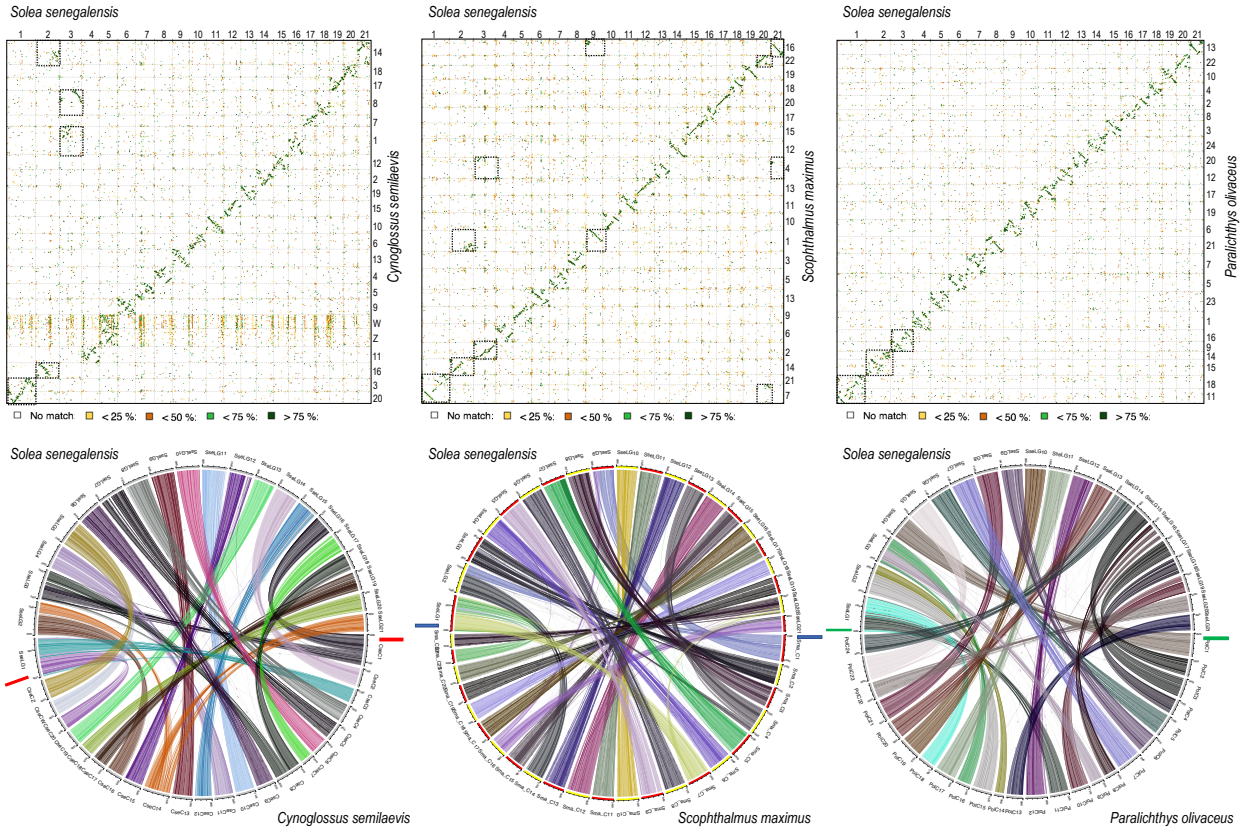


**Figure 8.** Chromosomal alignment and synteny analysis between flatfish genomes. Top panel, Dot plot comparison of 21 pseudo-chromosomes of *S. senegalensis* with the genomes of the flatfish *C. semilaevis* (left), *S. maximus* (center)

and *P. olivaceus* (right). Chromosome numbers or SseLGs are indicated. The chromosome fusions are boxed. Identity scale is indicated below. Bottom panel, syntenic comparison between flatfish genomes.

When the reduction of the number of chromosomes was explored three main Robertsonian fusions in the SseLG1 (Chr18-Chr11), SseLG2 (Chr14-Chr15) and SseLG3 (Chr9-Chr16) could explain the reduction from n = 24 in *P. olivaceus* to n = 21 in *S. senegalensis* (Figure 8). When compared to *S. maximus* (n = 22), the SseLG1 appeared as a fusion of Chr7 and Chr21. Moreover, translocations of regions from Chr1, Chr4, Chr7, Chr14 and Chr16 were also observed. In the case of *C semilaevis* with sexual chromosomes (ZW) and the same number of chromosome than *S. senegalensis,* a Robertsonian fusion in SseLG1 between Chr3-Chr20 was observed. Moreover, the SseLG3 appeared as a new chromosome resulting of the fission of Chr1 (mainly located in SseLG16) and Chr8 (mainly located in SseLG18). Two other major features in this species with respect to *S. senegalensis* were: i) a translocation of a Chr14 region to Chr16 to create the SseLG2; and ii) sexual ZW chromosomes appear concentrated in SseLG5 although high similar sequences are widely distributed throughout the genome. Comparison among all flatfish species (Figure 8) indicated that those chromosomal regions associated with SseLG2 and SseLG3 were mainly involved in the changes of karyotypes of the four Pleuronectiformes species whereas the SseLG1 arose as a lineage-specific fusion event.

# CHAPTER 4

## Development of whole-genome multiplex assays and construction of an integrated genetic map using SSR markers in Senegalese sole

The results of this chapter were published in: Guerrero-Cozar, I.; Perez-Garcia, C.; Benzekri, H., Sánchez, J.; Seoane, P.; Cruz, F.; Gut, M.; Zamorano, M.J.; Claros, M.G.; Manchado, M. Development of whole-genome multiplex assays and construction of an integrated genetic map using SSR markers in Senegalese sole. *Sci. Reports* 2020, *10*, 21905 DOI: 10.1038/s41598-020-78397-w

## 4.1- Introduction

Genomes are an essential source of markers required for ecological studies, breeding programs, traceability or functional studies. In the last years, the genomes of some commercially important flatfish belonging to the Cynoglossidae, Scophthalmidae, and Paralichthydae families were published indicating that overall, they are small and highly compact with sizes ranging between 470 and 584 Mb (Cerda & Manchado 2013; Chen et al. 2014a; Shao et al. 2017; Xu et al. 2020). These genomes have contributed to a better understanding of chromosome evolution in flatfish (Maroso *et al.* 2018), sex determination (Chen *et al.* 2014a) and the identification of mechanisms controlling metamorphosis (Shao *et al.* 2017) and growth performance (Robledo et al. 2017) with impact in aquaculture and stock population management. In Senegalese sole (*Solea senegalensis*), a preliminary draft of 600.3 Mb that fully covered the tongue sole (*Cynoglossus semilaevis*) genome was assembled (Manchado *et al.* 2016; Manchado *et al.* 2019). Although this assembly was still a bit fragmented (N50 of 85 kb), it became an useful tool to understand hybridization and introgression between *S. senegalensis* and *S. aegyptiaca* (Souissi et al. 2018) and for synteny analysis (Roman-Padilla et al. 2016; Carballo et al. 2019; Manchado *et al.* 2019). Nevertheless, an improvement of scaffolding and chromosome architecture is required for association studies, gene mapping and comparative genomics.

Genetic linkage maps and physical genomes provide complementary information that can be useful for the refinement of genome assemblies, the identification of genes associated with QTLs and cross-species synteny analysis (Cordoba et al. 2010; Portela-Bens *et al.* 2017). In Senegalese sole, a low-density genetic linkage map constructed using three gynogenetic families and 129 microsatellites (also known as simple sequence repeats, SSRs) markers was described (Molina-Luzon *et al.* 2015b). This map contained 27 linkage groups (LG) with an average density of 4.7 markers per LG that it was still a bit far away from the 21 chromosomes expected in *S. senegalensis*. Comparative synteny mapped these LGs through most of the chromosomes (except three) of *C. semilaevis* suggesting that some chromosome rearrangements could have occurred during evolution of these species (Manchado *et al.* 2019). Moreover, an integrated map using BAC clones and repetitive DNA families was developed using multiple fluorescence *in situ* hybridization that comprised 64 BACs mapped through all genome except in the submetacentric chromosome five (Garcia *et al.* 2019). Although Senegalese sole has not morphologically heteromorphic sex chromosomes, the largest metacentric chromosome was proposed as a proto-

sex chromosome originated from the fusion of two acrocentric chromosomes during flatfish evolution (Portela-Bens *et al.* 2017; Rodriguez *et al.* 2019).

Even though SNP markers have attracted the attention of researchers in the last years to construct high density genetic linkage maps and for genetic association studies (Wang *et al.* 2015b), the SSR markers still remain as highly popular markers due to their high variability, reproducibility, and their codominant inheritance (Sundaray et al. 2016; Lu et al. 2019). To maximize the use of SSR markers, whole-genome genotyping using SSR-based multiplex PCRs have become the most suitable strategy to save costs, labour time and reduce data processing. This methodological approach can make feasible the implementation in small- to medium-sized laboratories since it requires basic equipment with comparable results between laboratories (Lee-Montero *et al.* 2013; Zarouri *et al.* 2015). These whole-genome multiplex PCRs have been successfully applied to pedigree reconstruction in genetic breeding programs and QTLs identification (Garcia-Celdran et al. 2015b; Lee-Montero *et al.* 2015; Negrin-Baez et al. 2016; Carballo et al. 2020). However, *loci* multiplexing requires a tailor-made design of primers to be combined and amplified simultaneously avoiding primer dimer and preventing the overlapping of allelic ranges in those markers labelled with the same fluorophore colour. Hence, *in silico* analysis of genome SSR information followed by experimental validation of multiplex PCR assays is required.

Senegalese sole genome and transcriptome are rich mainly in SSRs with dinucleotide motif representing ~60 % of total SSRs, tetranucleotides only 5.2 % and pentanucleotides 2.4% (Benzekri et al. 2014; Garcia *et al.* 2019). Although SSRs with dinucleotide motifs have a higher allelic diversity than those with larger motifs, these latter are less prone to artefacts such as allelic dropout and stutters. Hence, scoring accuracy is very high reducing genotyping errors and making feasible data automation (Nater et al. 2009; Flores-Renteria & Krohn 2013). Genome analysis provides enough information for *in silico* analysis to select and combine high polymorphic SSR markers while they maintain an reliable and robust scoring for multiplex PCRs. The aim of this study was to: 1) provide *de novo* improved assembly of a female Senegalese sole based on long and short reads; 2) identify tetra- or pentanucleotide SSRs *in silico* and carry out a flatfish cross-species comparison to design whole-genome Multiplex PCRs; 3) validate all SSR *loci*, structure in multiplex PCRs according to allelic ranges (with up to 11-plex amplification) and optimize amplification conditions for whole genome mapping; 4) design supermultiplex PCRs containing the most polymorphic *loci* to sustain breeding genetic programs in this species in which offspring

is communally reared; and 5) integrate SSR markers available in Senegalese sole in a genetic linkage map and carry out a synteny analysis with the flatfish *C. semilaevis* to understand chromosome evolution.

## 4.2- Methods

### 4.2.1- Genome sequencing, assembly and characterization

SSR identification was carried out by *in silico* analysis of a previously published female genome based on Illumina short-reads (Manchado *et al.* 2016; Manchado *et al.* 2019). Both the contig (named as assembly_51k according to k-mer used) and the scaffolded (named as 85k genome according to N50) assemblies were used.

To increase the reliability of predicted SSR flanking regions, genome positioning and map distribution, a *de novo* female hybrid genome was also assembled using short and long reads. High molecular weight DNA was prepared from heparinized whole blood using the MagAttract HMW DNA kit (Qiagen). Main figures of Oxford nanopore Technology (ONT) (female code H2074515) and Illumina paired-end (PE300) reads (female code H150612; Bioproject PRJNA643826) are depicted in Table 1. Sequencing was carried out at the National Center for Genomic Analysis (CNAG, Barcelona, Spain). For the hybrid assembly, libraries were pre-processed to remove contaminants and low-quality sequences. Briefly, the Illumina PE300 library was screened using Kraken (v0.10.5-beta) (Wood & Salzberg 2014) and contaminants filtered out with the gem-mapper (Marco-Sola et al. 2012) (with ≤ 2% mismatches). In the case of ONT, data were base-called with Albacore v2.0.2 using the following criteria: base quality per read Q < 7, match to the control Sequence (lambda phage 3.5 kb), length less than 1 kb, or more than 40% low complexity sequence. Finally, POMOXIS v0.1.0 and Racon (Vaser et al. 2017) via all-vs-all alignment with minimap2 (Li 2018) were used to correct the reads before assembly. The hybrid genome assembly (named as LR-hybrid female genome) was carried using MaSuRCA v3.2.3 (Zimin *et al.* 2013; Zimin *et al.* 2017) to construct mega-reads that were finally assembled with CABOG v6.2 (Miller et al. 2008). Completeness was determined using Benchmarking Universal Single-Copy Orthologs (BUSCO, v3.0.2) (Simao *et al.* 2015; Waterhouse *et al.* 2018) containing 4,854 single-copy orthologs from actinopterygii_odb9. Genome scaffolds are available at Claros *et al.* (2020).

**Table 6**: Summary of input datasets for Illumina (PE300) and Oxford Nanopore Technologies (ONT) reads for LR hybrid female assembly.

| Library | Read length N50 (bp) | Fragment length (bp) | Total reads | Yield (Gb) | error r1 (%) | error r2 (%) | Sequencing coverage[2] |
|---|---|---|---|---|---|---|---|
| PE300 | 101 | 330 | 1,005,526 | 101.56 | 0.29 | 0.62 | 142.24 |
| ONT 1DSQ | 8,203 | - | 64,016 | 0.40 | 6.7 | | 0.56 |
| ONT MinION[1] | 10,802 | - | 1,311,044 | 9.38 | 17.6 | - | 12.57 |

[1]Information corresponding to the filtered 1D and 1D2 reads produced by five MinION runs. Error rate estimated as sum of mismatched, inserted bases and deleted bases divided by length of alignment of Oxford Nanopore Technologies (ONT) reads to the control sequence.
[2]Coverage estimates are calculated assuming a genome size of 714 Mb (C-value of *Solea solea*)

### 4.2.2- SSR screening, primer design and *in silico* genome mapping

SSR screening on the genomes was carried out using MISA (Microsatellite identification tool) and the parameters were those previously described (Beier et al. 2017). A total of 224 contigs from the the 85k genome larger than 20 kb and containing several SSRs were preselected and positioned onto the *C. semilaevis* genome by local blast analysis. Moreover, unigenes from Senegalese sole transcriptome (Benzekri *et al.* 2014) were positioned within each contig to identify gene content and sysnteny with *C. semilaevis*. A final set of putative 113 tetra- or pentanucleotide SSRs located in contigs from different chromosomes or separated at least 1 Mb apart within the same chromosome were selected. To validate chromosome positioning, these selected contigs were further mapped onto the LR-hybrid female genome and the scaffolds blasted onto *C. semilaevis* chromosomes.

The criteria followed for primer design were those previously described for multiplex PCR reactions (Sanchez et al. 2003; Lee-Montero *et al.* 2013). The range of amplicon sizes oscillated between 70 and 300 base pairs (bp). The primer quality and amplicon specificity were assessed by mapping sequences onto the *de novo* LR-hybrid female genome. A quality scale was established as follows: 1) high-specific (H-S) when they yielded a single specific amplicon and they mapped just in one position in the genome; 2) specific (S) when they yielded a single specific amplicon but at least one of the primers mapped between 2-10 (S* 2), 11-100 (S**) or >100 (S***) positions in the genome; 3) multiple (M) when the primers amplified different regions in the genome; and 4) no amplification (NA) when no amplicon could be predicted or the amplicon was larger than 300

bp. A similar strategy was pursued to evaluate the quality of the primers published by (Molina-Luzon *et al.* 2015a)

### 4.2.3- Fish samples and DNA isolation

To characterize the SSR markers, wild specimens of Senegalese sole captured in the Gulf of Cádiz (Spain) and incorporated to the aquaculture broodstocks of the company CUPIMAR (San Fernando, Cádiz, Spain) and IFAPA center El Toruño (El Puerto de Santa María, Cádiz, Spain) were used. Animals were sampled for blood (~0.5 ml) by puncturing in the caudal vein using a heparinized syringe, added heparin (100 mU) and kept at -20 ºC until use. Overall, the whole set of animals used in this study was 150 (79 breeders from CUPIMAR and 71 from IFAPA). To optimize the multiplex PCR assays, the 71 animals from IFAPA's broodstock structured in four tanks (n = 6, 21, 22, and 22 fish) were used. As we carried out several tests to adjust the primer conditions and validate amplifications, some samples were run out and the total individuals finally analyzed in each multiplex PCR assay was slighlty different (althout the four tanks were represented in all assays) and specifically indicated in each case. To validate the supermultiplex PCR assays and carry out the simulations, fish from CUPIMAR (n =79 distributed in four tanks) and IFAPA (n=13) was used.

Total DNA from heparinized blood (~25 µl) was isolated using Isolate II Genomic DNA Kit (Bioline). DNA samples were treated with RNase A (Bioline) following the manufacture's protocol. DNA was quantified spectrophotometrically using the Nanodrop ND-8000. Each microsatellite marker was tested in singlepex PCR to confirm amplification. PCR reactions were carried out in a 12.5 µl final volume containing 40 ng of DNA, 300 nM each of specific forward and reverse primers, and 6.25 µl of Platinum Multiplex PCR Master Mix, 2× (Thermofisher Scientific). The amplification protocol consisted of an initial denaturation at 95 °C for 10 min, followed by 30 cycles of 95 °C for 20 s, 59 °C for 1 min and 72 °C for 2 min, with a final extension of 72 °C for 10 min. PCR products were separated by capillary electrophoresis in an ABI3130 Genetic Analyzer (Applied Biosystems). Raw data obtained by capillary electrophoresis were transformed into allelic sizes using the GeneMapper v3.8 software (Thermofisher Scientific).

### 4.2.4- Multiplex PCRs optimization

SSRs were initially distributed in thirteen multiplex PCR assays (ranging 6 to 10-plex amplification. However, when markers were tested in singleplex, three of them did not amplify (SSeneg12220, SSeneg13367 and SSeneg3342) and two (SSeneg977 and SSeneg398) amplified a

multipeak patterning and they were removed from the original sets. Moreover, SSeneg3502 and SSeneg106 markers were excluded from the mutiplex PCRs due to overlapping allelic range with other markers or a low amplification efficiency. All Multiplex PCRs were performed in a final volume of 12.5 µl containing 1× Platinum Multiplex PCR Master Mix, 40 ng of template DNA and the primer concentrations were optimized to balance the fluorescent signal intensity. The PCR program is the same indicated above.

To validate the robustness of the whole-genome multiplex PCRs, an independent lab (University of Las Palmas de Gran Canaria, Spain) analyzed a subset of DNA samples from IFAPA's broodstock (total n = 60). The amplification conditions were similar to those indicated above except that Platinum Multiplex PCR Master Mix was replaced by KAPA2G Fast Multiplex PCR Kit (Kappa Biosystems_Sigma Aldrich). Electropherograms were analyzed using Genemapper (v.3.8) software (Applied Biosystems) and a kit of bin set was created for each multiplex PCR. A protocol for evaluation of genotyping reliability and *loci* scoring was performed (Lee-Montero *et al.* 2013). Briefly, the rate of errors or potential errors for each marker were determined after identifying ambiguous or unambiguous genotypes in the samples. The main genotyping errors were classified as inadequate peak heights out of optimal ratio (600-3,000 relative fluorescent units), unclear banding pattern or intermediate alleles that could not be read automatically using the bin set.

In order to design genotyping tools for parentage assignments in genetic breeding programs, a set of 40 SSR markers with the highest variability according to the polymorphic information content (PIC) was selected and rearranged in four new supermultiplex (SM) assays considering the fluorescent labelling and the allelic range (named as SMA, SMB, SMC and SMD).

### 4.2.5- Data analysis

Genetic diversity parameters (number of alleles (k)), observed (Ho) and expected (He) heterozygosities, allelic range, non-exclusion probabilities for pair parent (NE-PP) and null allele frequency were estimated using Cervus *v*3.0.3 (Kalinowski et al. 2007). The Hardy–Weinberg equilibrium (HW) at each locus was tested based on $\chi^2$ tests using GenAlEx *v*6.502 software (Peakall & Smouse 2012). The test for null allele presence was performed using Micro-checker v2.2.3 (Van Oosterhout et al. 2004). Parentage assignment was performed in PARFEX v1.0 using exclusion approach (Sekino & Kakehi 2012). This package was further used to calculate the

minimum marker set required for optimal parentage using the given data set. Markers were ranked according to PIC information and exclusion probability. In the case of SMA, a total of n = 92 specimens (48 females and 44 males; see "Fish samples" section) were analyzed. As the number of sole breeders in each tank oscillated between 13 and 25 specimens, simulations for supermultiplex SMB, SMC and SMD were carried out using a subset of animals (n = 15; 8 females and 7 males).

To construct the integrated SSR genetic map, the 108 SSR markers of this study and 121 out of 129 SSRs of the low density genetic linkage map available in Senegalese sole (Molina-Luzon *et al.* 2015b ) were positioned in the LR-hybrid female genome by local megablast analysis. Primers from eight markers in the previous map were excluded due to low quality mapping rates. Later, all scaffolds were anchored to the 21 linkage groups (LG) of a high-density SNP genetic linkage map generated using ddRAD from five full-sib families. Data about families, SNPs and full procedure to construct the SNP-based genetic linkage map will be published elsewhere. The relative genetic distances between makers were obtained from the anchored physical map and the integrated map was drawn using the software linkagemapview (Ouellette *et al.* 2018). For macrosynteny comparison, scaffolds bearing the SSRs were blasted onto the *C. semilaevis* chromosomes and positions compared to identify chromosomal rearrangements.

#### 4.2.6- Compliance with ethical standards.

All procedures were performed in accordance with Spanish national (RD 53/2013) and European Union legislation for animal care and experimentation (Directive 86\609\EU) and authorized by the Bioethics and Animal Welfare Committee of IFAPA and given the registration number 10/06/2016/101

## 4.3- Results

#### 4.3.1- Identification of SSRs for multiplex design and assessment of their genome distribution

SSR markers were identified by *in silico* analysis of repetitive motifs in the 85k genome (Manchado *et al.* 2016) based on Illumina short-reads. A first search for SSR markers selected a set of 224 contigs bigger than 20 kb and putatively located in different chromosomes or separated at least 1 Mb apart in the same chromosome. Average size of selected contigs was 118.7 kb and a

cross-species comparison with the genome of the flatfish *C. semilaevis* confirmed that they were widely distributed in all chromosomes (between 6 and 17 contigs by chromosome). The average number of SSR markers in each contig was 14.6, 5.3, 4.3 and 2.3 for di-, tri- tetra- and pentanucleotide repeat motifs, respectively. Using as reference this information, a subset of 113 contigs putatively distributed through the genome (minimum 5 scaffolds by chromosome) containing SSRs with tetra- or pentanucleotide repeat motifs was selected. The final set of SSRs selected for primer design included 103 tetranucleotides, 5 pentanucleotides and 5 compound markers containing at least two tetranucleotide SSRs separated by a spacer. Overall, GATA was the most abundant repeat motif in the selected markers (30 SSRs).

To assess the conservation of SSR flanking regions and the expected amplicon sizes as indicator of SSR quality for primer design, a *de novo* assembly based on Nanopore long-reads corrected with Illumina reads was used (LR-hybrid female genome). Raw sequencing data are indicated in Table 6. Expected coverage was 141x for Illumina PE300 library and 13.5x for Nanopore reads. The new assembly resulted in 6,482 contigs and 5,748 scaffolds with a total length of 607,976,531 bp and scaffold N50 of 340 kb. The estimated gene integrity was 96.2%. Overall, the marker density was 886.7 SSRs per megabase (Mb) and the dinucleotide repeats were the most abundant (52.4%) followed by tri- (12.5%), tetra- (4.0%) and pentanucleotides (1.1%). The C/A motif represented the 75% of dinucleotide repeats. To assess the quality of 113 selected markers, all designed primers were mapped onto the scaffolds of LR-hybrid female genome and classified into four categories (high-specific (H-S), specific (S), multiple, (M) and no amplification (NA)) according to locus-specificity, predicted amplification success and amplicon size. Primers of 74 markers mapped specifically in just one position and generated locus-specific PCR amplicons of expected size similiar to 85k genome, 34 markers had one primer of the pair with more than one mapping through the genome although the primer pair generated a locus-specific PCR product of expected size, 2 markers were not locus-specific and 3 markers failed to provide a PCR product due to amplicon size larger than expected or mapping on different scaffolds. After assessment primer quality, 108 markers were finally selected and arranged in multiplex PCRs. The wide distribution through the genome was validated by mapping scaffolds of the 85k and LR-hybrid female genomes onto the *C. semilaevis* chromosomes. Mapping results were highly consistent between assemblies showing only some conflicts for those contigs (only13) located in the sexual chromosomes (Z and W) of *C. semilaevis* that are absent in sole.

### 4.3.2- Whole-genome multiplex assays and genetic parameters

All SSR primers were designed to be amplified under similar conditions and hence they could be combined and ready for rearrangement between multiplex PCR assays depending on the labelling and allelic range. Before optimizing the multiplex reactions, all markers were tested in singleplex under the same amplification conditions.

The expected range of amplicon sizes for the complete set of SSR markers oscillated between 84 and 341 bp. Depending on the fluorescent labelling and the expected amplicon sizes, the 108 SSRs were distributed into 13 multiplex PCR assays (ranging from 6- and 10-plex). After amplifying markers in fish samples, some of them had to be rearranged in other multiplex PCRs due to allelic range overlapping or low amplification efficiency in the assays and two markers (SSeneg3502 and SSeneg106) could not be combined in any way and they were excluded. Hence, the final design comprised 106 SSR markers amplified in thirteen multiplex PCRs (from 6 to 10-plex).

Main genetic parameters associated with each marker are depicted in Table 7. For each multiplex, between 44 and 71 specimens were analyzed. The number of alleles ranged between 2 and 43 by *loci*. Moreover, 89 SSR markers were experimentally confirmed as tetranucleotide and 5 as pentanucleotide after analysing the repetition patterns in genotyped samples. However, 13 SSR markers followed an allelic series compatible with a dinucleotide repeat motif. A total of 34 markers deviated from HW. Micro-checker results identified 24 markers with a possible presence of null alleles that in most of the cases deviated from HW. The allelic range of *loci* sorted by fluorescence labelling are depicted in Figure 9. To test the robustness of the amplification and test the genetic variation of the markers, the thirteen PCR multiplex assays were run by an independent laboratory (ULPGC). Data comparison confirmed the genetic variability parameters, feasibility to amplify and consistent scoring of markers. Only 17 markers were deviated from HW. *Loci* quality scoring identified 11 markers with a bit stuttering, 4 markers allele dropout and only two intermediate alleles but all of them could be successfully read.

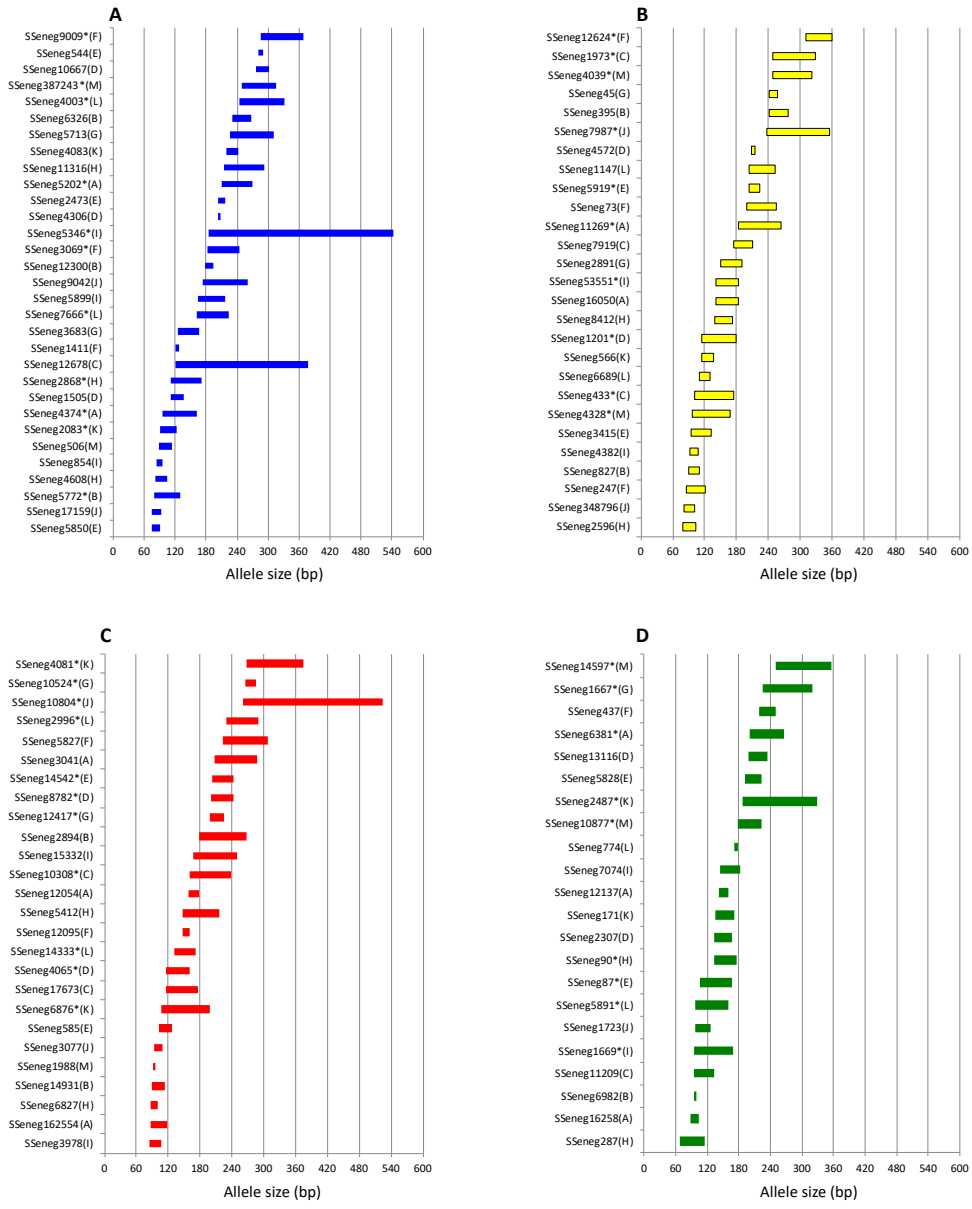


**Figure 9.** Allelic ranges of the 106 SSRs analysed in this study by fluorescence labelling (A-D). The name of the multiplex PCRs in which each marker is included is indicated between brackets. The asterisk indicates that the marker was selected to be included in the supermultiplex PCRs.

**Table 7. Genetic diversity estimates of 106 by multiplex PCRs (A-M).** Fluorescent labelling (B, blue; G, green; Y, yellow; R, red), repeat motif (Di, tetra or pentanucleoide), Number of samples (N), number of alleles (k), Allelic range, observed heterozygosity (Ho) and expected heterozygosity (He), polymorphic information content (PIC), non-exclusion probability of pair parent (NE-PP); null allele frequency (F(N)). Hardy–Weinberg equilibrium (HW;

*significant after bonferroni correction; ns, non-significant) and Null alleles as determined by micro-checker (yes, significant after bonferroni correction; ns, non-significant)

**MultiplexA**

| Locus | L | Motif | N | k | Range | Ho | He | PIC | NE-PP | F(N) | HW | NA & |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SSeneg4374 | B | Tetra | 61 | 10 | 96-162 | 0.53 | 0.72 | 0.69 | 0.28 | 0.16 | ns | ns |
| SSeneg5202 | B | Tetra | 63 | 16 | 210-270 | 0.89 | 0.88 | 0.86 | 0.09 | -0.01 | ns | ns |
| SSeneg16258 | G | Tetra | 63 | 4 | 88-104 | 0.48 | 0.51 | 0.47 | 0.54 | 0.02 | ns | ns |
| SSeneg12137 | G | Tetra | 63 | 7 | 141-159 | 0.62 | 0.64 | 0.61 | 0.37 | 0.00 | ns | ns |
| SSeneg6381 | G | Tetra | 63 | 18 | 200-266 | 0.71 | 0.87 | 0.85 | 0.10 | 0.10 | (*) | yes |
| SSeneg16050 | Y | Tetra | 63 | 10 | 142-184 | 0.49 | 0.62 | 0.59 | 0.38 | 0.12 | ns | yes |
| SSeneg11269 | Y | Di | 63 | 33 | 183-263 | 0.95 | 0.96 | 0.95 | 0.02 | 0.00 | ns | ns |
| SSeneg162554 | R | Tetra | 63 | 9 | 86-118 | 0.89 | 0.80 | 0.76 | 0.21 | -0.06 | ns | ns |
| SSeneg12054 | R | Tetra | 63 | 6 | 159-179 | 0.78 | 0.75 | 0.70 | 0.29 | -0.03 | ns | ns |
| SSeneg3041 | R | Tetra | 63 | 14 | 207-287 | 0.52 | 0.78 | 0.75 | 0.22 | 0.19 | * | yes |
| **MultiplexB** | | | | | | | | | | | | |
| SSeneg5772 | B | Tetra | 51 | 11 | 80-130 | 0.77 | 0.81 | 0.78 | 0.19 | 0.03 | (*) | ns |
| SSeneg12300 | B | Tetra | 51 | 5 | 177-193 | 0.67 | 0.61 | 0.53 | 0.51 | -0.06 | ns | ns |
| SSeneg6326 | B | Tetra | 51 | 9 | 231-267 | 0.82 | 0.80 | 0.77 | 0.19 | -0.02 | ns | ns |
| SSeneg6982 | G | Penta | 51 | 2 | 94-100 | 0.28 | 0.27 | 0.23 | 0.81 | -0.02 | ns | ns |
| SSeneg827 | Y | Tetra | 51 | 6 | 91-111 | 0.55 | 0.63 | 0.58 | 0.41 | 0.07 | ns | ns |
| SSeneg395 | Y | Penta | 51 | 8 | 241-276 | 0.80 | 0.82 | 0.79 | 0.18 | 0.01 | ns | ns |
| SSeneg14931 | R | Tetra | 51 | 7 | 89-113 | 0.69 | 0.77 | 0.73 | 0.26 | 0.06 | ns | ns |
| SSeneg2894 | R | Tetra | 51 | 7 | 178-268 | 0.43 | 0.70 | 0.64 | 0.37 | 0.24 | (*) | yes |
| MultiplexC | | | | | | | | | | | | |
| SSeneg12678 | B | Tetra | 46 | 29 | 121-377 | 0.37 | 0.96 | 0.95 | 0.02 | 0.44 | * | yes |
| SSeneg11209 | G | Tetra | 54 | 10 | 94-134 | 0.78 | 0.79 | 0.76 | 0.20 | 0.01 | (*) | ns |
| SSeneg433 | Y | Tetra | 54 | 6 | 101-174 | 0.56 | 0.62 | 0.57 | 0.44 | 0.05 | (*) | ns |
| SSeneg7919 | Y | Tetra | 53 | 9 | 174-210 | 0.53 | 0.79 | 0.75 | 0.23 | 0.2 | (*) | yes |
| SSeneg1973 | Y | Tetra | 53 | 23 | 249-329 | 0.94 | 0.93 | 0.92 | 0.04 | -0.01 | (*) | ns |
| SSeneg17673 | R | Tetra | 54 | 5 | 116-177 | 0.44 | 0.73 | 0.67 | 0.35 | 0.24 | * | yes |
| SSeneg10308 | R | Tetra | 54 | 13 | 161-239 | 0.87 | 0.90 | 0.88 | 0.08 | 0.01 | ns | ns |
| **MultiplexD** | | | | | | | | | | | | |
| SSeneg1505 | B | Tetra | 57 | 7 | 112-136 | 0.75 | 0.78 | 0.75 | 0.23 | 0.01 | ns | ns |
| SSeneg4306 | B | Tetra | 57 | 2 | 204-208 | 0.51 | 0.50 | 0.37 | 0.72 | -0.01 | ns | ns |
| SSeneg10667 | B | Tetra | 54 | 7 | 277-301 | 0.56 | 0.64 | 0.61 | 0.36 | 0.06 | ns | ns |
| SSeneg2307 | G | Tetra | 57 | 6 | 134-166 | 0.53 | 0.55 | 0.50 | 0.51 | 0.02 | (ns) | ns |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SSeneg13116 | G | Tetra | 57 | 7 | 199-235 | 0.63 | 0.72 | 0.69 | 0.28 | 0.06 | ns | ns |
| SSeneg1201 | Y | Penta | 57 | 13 | 115-180 | 0.40 | 0.83 | 0.81 | 0.16 | 0.35 | * | yes |
| SSeneg4572 | Y | Tetra | 57 | 2 | 207-215 | 0.26 | 0.48 | 0.36 | 0.73 | 0.29 | * | yes |
| SSeneg4065 | R | Tetra | 57 | 10 | 117-161 | 0.86 | 0.84 | 0.82 | 0.14 | -0.01 | ns | ns |
| SSeneg8782 | R | Tetra | 57 | 10 | 200-242 | 0.88 | 0.83 | 0.81 | 0.15 | -0.03 | ns | ns |
| **MultiplexE** | | | | | | | | | | | | |
| SSeneg5850 | B | Tetra | 50 | 4 | 74-92 | 0.62 | 0.62 | 0.55 | 0.49 | 0 | ns | ns |
| SSeneg2473 | B | Tetra | 50 | 4 | 204-216 | 0.70 | 0.56 | 0.46 | 0.61 | -0.12 | ns | ns |
| SSeneg544 | B | Tetra | 50 | 4 | 282-290 | 0.70 | 0.61 | 0.53 | 0.52 | -0.08 | ns | ns |
| SSeneg87 | G | Tetra | 50 | 12 | 106-166 | 0.64 | 0.67 | 0.62 | 0.37 | 0.02 | ns | ns |
| SSeneg5828 | G | Tetra | 49 | 7 | 192-224 | 0.55 | 0.48 | 0.46 | 0.52 | -0.11 | ns | ns |
| SSeneg3415 | Y | Tetra | 50 | 8 | 94-132 | 0.56 | 0.64 | 0.60 | 0.39 | 0.06 | (ns) | ns |
| SSeneg5919 | Y | Di | 50 | 8 | 204-224 | 0.66 | 0.74 | 0.69 | 0.31 | 0.05 | ns | ns |
| SSeneg585 | R | Tetra | 50 | 7 | 103-127 | 0.62 | 0.75 | 0.71 | 0.27 | 0.07 | (*) | yes |
| SSeneg14542 | R | Tetra | 49 | 8 | 202-244 | 0.67 | 0.76 | 0.71 | 0.29 | 0.06 | ns | ns |
| MultiplexF | | | | | | | | | | | | |
| SSeneg1411 | B | Tetra | 64 | 3 | 120-128 | 0.22 | 0.25 | 0.23 | 0.78 | 0.05 | ns | ns |
| SSeneg3069 | B | Tetra | 63 | 13 | 183-245 | 0.73 | 0.87 | 0.85 | 0.11 | 0.08 | (*) | ns |
| SSeneg9009 | B | Tetra | 64 | 19 | 286-368 | 0.89 | 0.93 | 0.92 | 0.04 | 0.02 | ns | ns |
| SSeneg437 | G | Tetra | 65 | 9 | 219-249 | 0.52 | 0.81 | 0.78 | 0.19 | 0.22 | ns | yes |
| SSeneg247 | Y | Tetra | 61 | 7 | 85-122 | 0.71 | 0.72 | 0.68 | 0.30 | -0.02 | ns | ns |
| SSeneg73 | Y | Di | 65 | 16 | 199-255 | 0.85 | 0.86 | 0.84 | 0.12 | 0.01 | ns | ns |
| SSeneg12624 | Y | Penta | 64 | 11 | 311-359 | 0.84 | 0.82 | 0.79 | 0.17 | -0.02 | ns | ns |
| SSeneg12095 | R | Tetra | 65 | 4 | 148-160 | 0.51 | 0.48 | 0.42 | 0.61 | -0.03 | ns | ns |
| SSeneg582 | R | Di | 62 | 18 | 224-308 | 0.94 | 0.87 | 0.85 | 0.11 | -0.05 | ns | ns |
| MultiplexG | | | | | | | | | | | | |
| SSeneg3683 | B | Tetra | 69 | 11 | 125-167 | 0.75 | 0.82 | 0.80 | 0.15 | 0.04 | ns | ns |
| SSeneg5713 | B | Di | 65 | 21 | 227-311 | 0.86 | 0.88 | 0.87 | 0.08 | 0.01 | ns | ns |
| SSeneg1667 | G | Di | 69 | 24 | 225-319 | 0.80 | 0.89 | 0.88 | 0.07 | 0.05 | (ns) | yes |
| SSeneg2891 | Y | Tetra | 68 | 9 | 150-190 | 0.65 | 0.82 | 0.79 | 0.19 | 0.11 | * | yes |
| SSeneg45 | Y | Tetra | 69 | 5 | 242-258 | 0.59 | 0.65 | 0.59 | 0.44 | 0.04 | (*) | ns |
| SSeneg12417 | R | Di | 69 | 9 | 199-225 | 0.86 | 0.78 | 0.74 | 0.25 | -0.06 | (ns) | ns |
| SSeneg10524 | R | Tetra | 69 | 7 | 266-286 | 0.75 | 0.71 | 0.66 | 0.34 | -0.05 | ns | ns |
| MultiplexH | | | | | | | | | | | | |
| SSeneg4608 | B | Tetra | 71 | 4 | 82-104 | 0.13 | 0.16 | 0.15 | 0.86 | 0.14 | ns | ns |
| SSeneg2868 | B | Tetra | 71 | 9 | 112-172 | 0.78 | 0.83 | 0.80 | 0.17 | 0.03 | (*) | ns |
| SSeneg11316 | B | Tetra | 71 | 10 | 214-292 | 0.72 | 0.71 | 0.67 | 0.33 | 0.00 | (*) | ns |
| SSeneg287 | G | Tetra | 71 | 7 | 68-114 | 0.55 | 0.52 | 0.47 | 0.55 | -0.05 | * | ns |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SSeneg90 | G | Tetra | 71 | 13 | 133-175 | 0.93 | 0.85 | 0.84 | 0.13 | -0.05 | ns | ns |
| SSeneg2596 | Y | Tetra | 71 | 5 | 78-104 | 0.38 | 0.40 | 0.35 | 0.68 | 0.00 | * | ns |
| SSeneg8412 | Y | Tetra | 71 | 8 | 138-172 | 0.41 | 0.47 | 0.43 | 0.57 | 0.05 | (*) | ns |
| SSeneg6827 | R | Tetra | 71 | 4 | 88-100 | 0.32 | 0.33 | 0.30 | 0.71 | -0.02 | ns | ns |
| SSeneg5412 | R | Tetra | 71 | 7 | 148-216 | 0.41 | 0.46 | 0.43 | 0.57 | 0.05 | * | ns |
| MultiplexI | | | | | | | | | | | | |
| SSeneg854 | B | Di | 69 | 6 | 85-95 | 0.52 | 0.60 | 0.53 | 0.50 | 0.07 | ns | ns |
| SSeneg5899 | B | Tetra | 69 | 5 | 164-216 | 0.41 | 0.47 | 0.43 | 0.58 | 0.06 | ns | ns |
| SSeneg5346 | B | Di | 68 | 43 | 184-542 | 0.87 | 0.95 | 0.94 | 0.02 | 0.04 | (*) | ns |
| SSeneg1669 | G | Tetra | 69 | 16 | 94-168 | 0.75 | 0.82 | 0.80 | 0.15 | 0.05 | ns | ns |
| SSeneg7074 | G | Tetra | 69 | 6 | 144-182 | 0.64 | 0.76 | 0.71 | 0.30 | 0.08 | ns | yes |
| SSeneg4382 | Y | Tetra | 64 | 5 | 92-108 | 0.22 | 0.42 | 0.37 | 0.65 | 0.31 | * | yes |
| SSeneg53551 | Y | Tetra | 67 | 8 | 142-184 | 0.72 | 0.72 | 0.68 | 0.30 | 0.00 | ns | ns |
| SSeneg3978 | R | Tetra | 69 | 7 | 84-108 | 0.67 | 0.68 | 0.64 | 0.34 | 0.00 | ns | ns |
| SSeneg15332 | R | Tetra | 68 | 19 | 168-250 | 0.91 | 0.89 | 0.87 | 0.08 | -0.02 | ns | ns |
| MultiplexJ | | | | | | | | | | | | |
| SSeneg17159 | B | Tetra | 58 | 5 | 75-93 | 0.48 | 0.47 | 0.42 | 0.59 | -0.03 | ns | ns |
| SSeneg9042 | B | Tetra | 56 | 19 | 174-260 | 0.77 | 0.86 | 0.83 | 0.12 | 0.06 | ns | ns |
| SSeneg1723 | G | Tetra | 58 | 7 | 97-127 | 0.55 | 0.51 | 0.47 | 0.53 | -0.08 | ns | ns |
| SSeneg348796 | Y | Tetra | 58 | 6 | 81-101 | 0.78 | 0.67 | 0.61 | 0.41 | -0.09 | (ns) | ns |
| SSeneg7987 | Y | Di | 58 | 32 | 238-354 | 0.88 | 0.94 | 0.93 | 0.03 | 0.03 | (*) | yes |
| SSeneg3077 | R | Tetra | 58 | 4 | 94-110 | 0.40 | 0.36 | 0.33 | 0.68 | -0.07 | (ns) | ns |
| SSeneg10804 | R | Tetra | 54 | 23 | 261-525 | 0.93 | 0.87 | 0.85 | 0.10 | -0.04 | ns | ns |
| MultiplexK | | | | | | | | | | | | |
| SSeneg2083 | B | Tetra | 62 | 9 | 92-124 | 0.69 | 0.65 | 0.61 | 0.36 | -0.05 | ns | ns |
| SSeneg4083 | B | Tetra | 63 | 6 | 220-242 | 0.56 | 0.58 | 0.54 | 0.43 | 0.00 | (ns) | ns |
| SSeneg171 | G | Tetra | 63 | 7 | 136-172 | 0.78 | 0.75 | 0.71 | 0.28 | -0.03 | ns | ns |
| SSeneg2487 | G | Tetra | 50 | 26 | 188-328 | 0.98 | 0.95 | 0.93 | 0.03 | -0.02 | * | ns |
| SSeneg566 | Y | Tetra | 63 | 7 | 114-136 | 0.84 | 0.77 | 0.73 | 0.27 | -0.05 | ns | ns |
| SSeneg6876 | R | Tetra | 63 | 21 | 108-198 | 0.94 | 0.91 | 0.90 | 0.06 | -0.02 | ns | ns |
| SSeneg4081 | R | Tetra | 61 | 19 | 268-374 | 0.90 | 0.88 | 0.87 | 0.08 | -0.02 | ns | ns |
| MultiplexL | | | | | | | | | | | | |
| SSeneg7666 | B | Di | 46 | 21 | 162-224 | 0.89 | 0.92 | 0.90 | 0.05 | 0.01 | ns | ns |
| SSeneg4003 | B | Di | 46 | 21 | 244-332 | 0.89 | 0.93 | 0.91 | 0.05 | 0.01 | ns | ns |
| SSeneg5891 | G | Tetra | 46 | 12 | 97-159 | 0.76 | 0.78 | 0.75 | 0.22 | 0.01 | (ns) | ns |
| SSeneg774 | G | Tetra | 46 | 4 | 172-178 | 0.17 | 0.27 | 0.26 | 0.75 | 0.26 | * | yes |
| SSeneg6689 | Y | Tetra | 44 | 5 | 111-131 | 0.11 | 0.41 | 0.38 | 0.61 | 0.55 | (*) | yes |
| SSeneg1147 | Y | Tetra | 46 | 14 | 204-252 | 0.80 | 0.91 | 0.89 | 0.06 | 0.06 | ns | yes |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SSeneg14333 | R | Tetra | 46 | 8 | 132-172 | 0.37 | 0.83 | 0.79 | 0.18 | 0.38 | * | yes |
| SSeneg2996 | R | Tetra | 45 | 14 | 229-291 | 0.64 | 0.90 | 0.88 | 0.07 | 0.16 | (*) | yes |
| MultiplexM | | | | | | | | | | | | |
| SSeneg506 | B | Tetra | 63 | 6 | 88-114 | 0.22 | 0.66 | 0.60 | 0.43 | 0.49 | * | yes |
| SSeneg387243 | B | Tetra | 62 | 17 | 250-316 | 0.86 | 0.87 | 0.85 | 0.11 | 0.01 | ns | ns |
| SSeneg10877 | G | Tetra | 63 | 12 | 177-223 | 0.71 | 0.80 | 0.77 | 0.19 | 0.04 | * | ns |
| SSeneg14597 | G | Tetra | 62 | 13 | 250-356 | 0.76 | 0.90 | 0.89 | 0.07 | 0.08 | * | yes |
| SSeneg4328 | Y | Tetra | 63 | 16 | 96-168 | 0.92 | 0.92 | 0.90 | 0.06 | -0.01 | (ns) | ns |
| SSeneg4039 | Y | Di | 60 | 26 | 248-322 | 0.43 | 0.91 | 0.90 | 0.05 | 0.36 | * | yes |
| SSeneg1988 | R | Tetra | 62 | 2 | 91-95 | 0.02 | 0.02 | 0.02 | 0.98 | 0.00 | ns | na |

To identify the genes close to the SSRs, the contigs selected for primer design were compared with Senegalese sole transcriptome and *C. semilaevis* genome. The analysis indicated a high degree of gene synteny conservation (higher than 90% in most multiplex PCRs) between *S. senegalensis* transcripts and *C. semilaevis* genes. Some of genes identified are of interest for aquaculture due to their role the role in immune response (toll-like receptor 3, interleukin-27 subunit beta, chemokine-like receptor 1, C-type mannose receptor 2 isoform X1), hormonal signalling (thyroid hormone receptor alpha-B, retinoic acid receptor RXR-alpha, retinol dehydrogenase 10, retinol dehydrogenase 8), antioxidant defences (superoxide dismutase [Cu-Zn]) or larval survival (high choriolytic enzyme 1), epigenetics (betaine--homocysteine S-methyltransferase 1), reproduction (Prostaglandin E synthase 3) or sensing (taste receptor type 1 member 1).

### 4.3.3- Design of supermultiplex for parentage assignment

To design high variable PCR multiplex assays (named as supermultiplex) suitable for pedigree reconstruction in breeding programs, a subset of 40 out of 106 markers was selected according to their allelic range and genetic variability markers and they were rearranged in four supermultiplex assays (referred from SMA, SMB, SMC and SMD) ranging from 8- to 11-plex. Allelic ranges are depicted in Figure 10. As average, PIC information in the four supermultiplex ranged between 0.79-0.82 and 73% of markers had a PIC value higher than 0.8 and 89% higher than 0.7. In total, motifs of 9 markers were dinucleotide, 29 tetranucleotide and 2 pentanucleotide. According to the synteny analysis these markers were positioned in 17 out of 21 chromosomes.

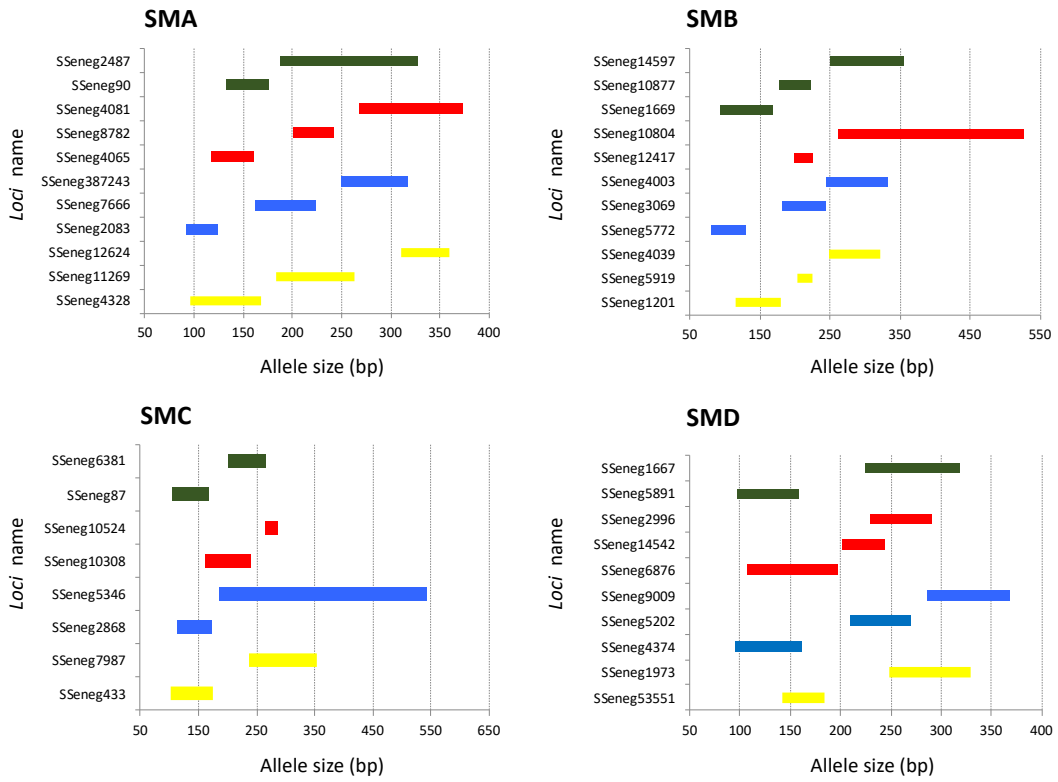


**Figure 10.** Allelic ranges of the 40 SSRs selected for the supermultiplex (SM) PCRs. The markers are shown by SM(A-D).

In order to validate the usefulness of the four supermultiplex for parentage assignment in sole, they were tested using different set of parents and offspring. In the case of SMA, an offspring set of 100 individuals and 92 putative parents from 4 different broodstocks (48 females and 44 males) were 100% assigned using to a single parent pair without observing null allele mismatches. For SMB, SMC and SMD, a broodstock tank of 15 parents was characterized and 5 offspring were 100% assigned to a single pair without mismatches. Ranking markers using PIC resulted in accumulative success rate higher than 99% with 7, 5, 4 and 3 markers in SMA, SMB, SMC and SMD, respectively ( Figure 11).

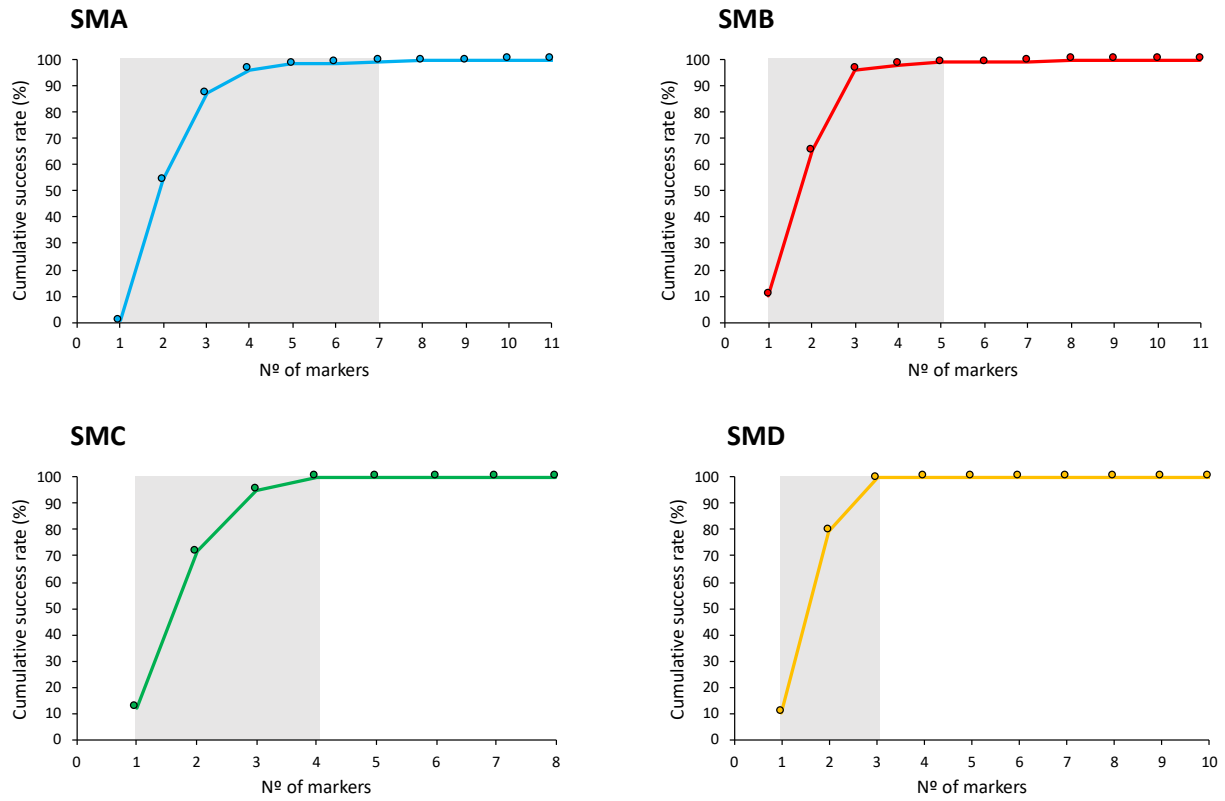


**Figure 11.** Cumulative success rate for parentage assignment based on exclusion with markers ranked on PIC value. The grey area indicates the *loci* required to reach more than 99% probability of assigning a correct parent–offspring relationship. SMA n = 92 parents; SMB, n = 15 parents; SMC, n = 15; SMD, n = 15.

### 4.3.4- Construction of an integrated genetic map and synteny analysis

To construct the integrated genetic map, 121 out of 129 SSRs reported by Molina-Luzon *et al.* (2015b) were succesfully mapped onto the LR-hybrid female genome. Overall, a total of 229 SSRs (108 of this study+121 previously published) were located in genome scaffolds anchored to the 21 linkage groups (SseLGs) of a recenlty high-density SNP genetic linkage map built in the lab that matches with the expected number of chromosomes *S. senegalensis*. The number of markers per LG ranged from 4 located in SseLG13 to 19 in SseLG07 (Table 8; Figure 12). Eight markers were located in unplaced scaffolds. Interestingly, marker distribution in the SseLGs was highly conincident with LGs of Molina-Luzon *et al.* (2015b). Only those markers from LG1 were split into the SseLG6 and SseLG19 probably due to a misarrangement in the previous map since these markers moved as two blocks between SseLGs.

**Table 8.** SSR distribution. Markers are groups by the 21 linkage groups (SseLG) of the high-density SNP genetic map. The number of SSRs of this study and those from Low-density (LD) genetic linkage map (Molina-Luzon *et al.* 2015a) are indicated. The location of markers in *C. semilaevis* genome by blasting the scaffold containing the SSR marker and the LG in the LD genetic map are indicated.

| High density SNP map | SSR Markers | | | *Cynoglossus* Chromosomes | LD Genetic map |
|---|---|---|---|---|---|
| | This study | LD genetic map | Total | | |
| SseLG01 | 11 | 3 | 14 | chr3,chr20 | LG21,LG27 |
| SseLG02 | 6 | 6 | 12 | chr14,chr16 | LG17,LG18,LG25 |
| SseLG03 | 5 | 5 | 10 | chr1, chr8, chrZ | LG7 |
| SseLG04 | 4 | 8 | 12 | chr11, chrZ | LG2 |
| SseLG05 | 8 | 7 | 15 | chrZ | LG4 |
| SseLG06 | 7 | 10 | 17 | chr9 | LG1 |
| SseLG07 | 4 | 15 | 19 | chr5 | LG3,LG26 |
| SseLG08 | 5 | 4 | 9 | chr4 | LG22,LG24 |
| SseLG09 | 5 | 5 | 10 | chr13 | LG16,LG20 |
| SseLG10 | 4 | 6 | 10 | chr6 | LG6 |
| SseLG11 | 3 | 5 | 8 | chr10 | LG10 |
| SseLG12 | 6 | 8 | 14 | chr15 | LG13,LG23 |
| SseLG13 | 4 | 0 | 4 | chr19 | -- |
| SseLG14 | 4 | 8 | 12 | chr2 | LG8 |
| SseLG15 | 5 | 4 | 9 | chr12 | LG12 |
| SseLG16 | 4 | 5 | 9 | chr1 | LG15 |
| SseLG17 | 4 | 6 | 10 | chr7 | LG11 |
| SseLG18 | 5 | 2 | 7 | chr8 | LG19 |
| SseLG19 | 4 | 7 | 11 | chr17 | LG1,LG14 |
| SseLG20 | 4 | 3 | 7 | chr18 | LG5 |
| SseLG21 | 3 | 4 | 7 | chr14 | LG9 |
| Unplaced | 3 | 5 | 8 | -- | |
| Total | 108 | 126 | 234 | | |

Macrosynteny analysis bewteen *S. senegalensis* and *C. semilaevis* chromosomes demonstrated that 17 SseLGs of *S. senegalensis* matched perfectly with different chromosomes of *C. semilaevis*

(Table 8). Only four chromosomes in *S. senegalensis* appeared as chromosomal rearrangements of C. *semilaevis* and the sequences of Z chromosome were dispersed through the SseLG3, SseLG4 and SseLG5. The SseLG1 appeared as a fusion of chromosomes 3 and 20 of *C. semilaevis*. Moreover, some rearrangements were observed for SseLG2 that included the chromosome 16 and part of 14, the SseLG3 that grouped regions of chromosomes 1, 8 and Z and the SseLG4 that combined the chromosome 11 and regions of Z.

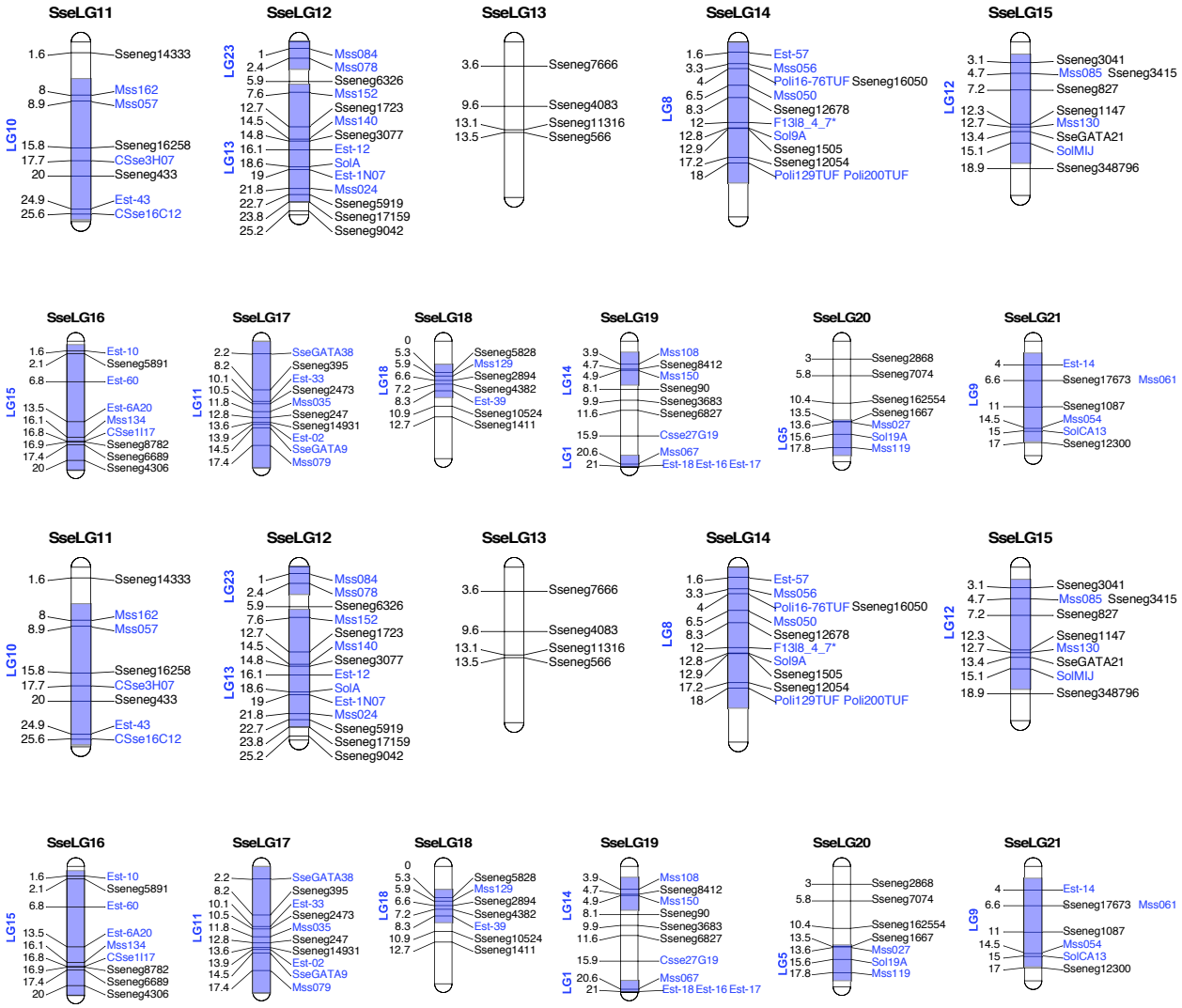

**Figure 12**. Integrated SSR genetic map of Senegalese sole (*S. senegalensis*). SseLG refer to the linkage groups according the high-density SNP genetic map. Genetic distance is indicated on the left. SSRs of this study are indicate in black and those from (Molina-Luzon *et al.* 2015a) in blue. The LGs previously assigned to these markers are shaded and indicated on the left.

# CHAPTER 5

## Genetic parameter estimates and identification of SNPs associated with growth traits in Senegalese sole

The results of this chapter were published in: Guerrero-Cozar, I.; Jimenez-Fernandez, E.; Berbel, C.; Cordoba-Caballero, J.; Claros, M.G.; Zerolo, R.; Manchado, M. Genetic parameter estimates and identification of SNPs associated with growth traits in Senegalese sole. *Aquaculture* 2021, *539*, 736665 DOI: 10.1016/j.aquaculture.2021.736665

## 5.1- Introduction

The Senegalese sole (*Solea senegalensis*) is one the most valuable flatfish in Southern Europe aquaculture. Its production has grown exponentially in the last decade due to significant improvements in larval rearing, optimization of dietary requirements and the use of recirculation technologies (RAS) for on-growing (Manchado *et al.* 2016; Morais *et al.* 2016; Manchado *et al.* 2019). However, reproductive dysfunction of breeders reared in captivity and the understanding of the courtship behavior required for a successful spawning still persist as two major bottlenecks for larval production (Fatsini *et al.* 2016; Fatsini *et al.* 2017; Fatsini *et al.* 2020). Hormonal therapies revealed as unsuccessful to release fertilized eggs (Agulleiro *et al.* 2006) and the low volume of sperm production (<130 µl) makes impractical the use of *in vitro* techniques as a routine procedure in the hatcheries (Chauvigne *et al.* 2017). However, in the last years, environmental control techniques based on thermocycles applied to wild males and hatchery-produced females have become a useful strategy to circumvent at least partially some of larval production limitations (Martin *et al.* 2019). This new approach can be used for the design of breeding programs based on mass spawning and the selection of best-ranked females.

A sustainable aquaculture is dependent on genetic breeding programs that select the best breeders to produce high-quality offspring. Currently, production of most economically important marine species in Europe are supported by selection schemes for relevant traits related to growth, disease resistant, morphology and flesh quality (Janssen et al. 2017). These programs have been designed taking into account the reproduction characteristics and the industrial production models of each species. A mass-spawning model followed by pedigree reconstruction using microsatellites and BLUP analysis was used in gilthead seabream (*Sparus aurata*) and common sole (*Solea solea*) (Brown et al. 2005; Navarro et al. 2009; Blonk *et al.* 2010b; Blonk *et al.* 2010c; Lee-Montero *et al.* 2013; Lee-Montero *et al.* 2015; Carballo *et al.* 2020). A major issue that these programs have to deal with is the skewed familial contributions and the high variance of family sizes. In gilthead seabream, these effects have been minimized by synchronizing egg release by photoperiod followed by mixing egg batches from different broodstock tanks through four consecutive days (Carballo *et al.* 2020). Following this strategy, heritabilities for weight at different ages (from hatchery to harvest), flesh quality, disease resistance against bacteria and virus and skeletal deformations were estimated (Garcia-Celdran et al. 2015a; Garcia-Celdran et al. 2015c; Lee-Montero *et al.* 2015; Garcia-Celdran et al. 2016; Aslam et al. 2018; Carballo *et al.* 2020). Interestingly, a significant genotype×production system interaction was reported in the close

species *S. solea* demonstrating that in soles exist families with different growth performance in RAS (artificial environment) or ponds (natural environment) (Mas-Muñoz *et al.* 2013). In Senegalese sole, there is still no genetic estimates for growth traits. However, recent advances in methodologies for juvenile tagging (Carballo et al. 2018), genotyping using multiplex PCR for parentage assignment (Guerrero-Cozar *et al.* 2020), and the synchronization of spawns using thermoperiod control (Martin et al. 2014; Martin *et al.* 2019) make possible to investigate genetic parameters under industrial conditions in RAS.

Breeding programs in aquaculture benefit of cost-effective sequencing technologies to implement marker-assisted selection (MAS) and genomic selection. The identification of genomic regions associated with a trait of interest (Genome Wide Association Studies;GWAS) are becoming popular in aquaculture and markers associated with growth (Kyriakis et al. 2019), pigmentation (Bertolini et al. 2020) disease resistance (Palaiokostas et al. 2016) or sex (Purcell et al. 2018) have been reported. Although several methodological approaches and genotyping platforms have been used in marine cultivated species, recent studies have demonstrated that that low-density single nucleotide polymorphism (SNP) panels are a cost-effective solution for broadening the impact of genomic selection in aquaculture. However, this approach require non-random SNP selection to increase prediction accuracy (Kriaridou et al. 2020). One source of genetic markers is the RNA sequencing (RNA-seq) (Brouard et al. 2019; Espinosa et al. 2020), most of which are linked to specific genes and suitable for quantitative PCR (qPCR). The use of these markers in low-density arrays (OpenArray® technology) has been successfully applied in marker-assisted selection (MAS) for plant breeding (Chagne et al. 2019) and fine-mapping for disease diagnosis (Verbeek et al. 2012; Gutierrez-Camino et al. 2018).

The aim of this study was to estimate heritabilities and genetic correlations for four growth traits both at the beginning of on-growing period in RAS (~400 days) and at harvest (~800 d) using offspring of a commercial Senegalese sole broodstock. Data for two traits were validated by *in situ* measurement and digital image analysis. Moreover, a SNP-based array was designed and validated using wild fish. An association analysis using low-density arrays was carried out using four families with different growth rates. The results obtained will be useful to design breeding program schemes to enhance the Senegalese sole aquaculture.

## 5.2- Material and methods

### 5.2.1- Animals

Genetic families were created by mass-spawning using a wild broodstock (n=150 animals) distributed into nine tanks (ranging from 6 to 26 breeders) with a 1:1 sex ratio. Mass spawning was synchronized by thermoperiod manipulation as previously reported (Martin *et al.* 2014). Animals were sample for blood (~0.5 mL) by puncturing in the caudal vein using a heparinized syringe, adding heparin (100 mU) and keeping at -20 °C until use.

Since not all broodstock tanks responded to each thermocycle treatment, seven evaluation batches were created with eggs from at least three tanks from July 2014 to Nov 2015 (Table 9). To maximize family representation, eggs in each batch were proportionally mixed by considering the total volume of eggs by tank. Moreover, eggs from one broodstock were represented in all batches to make easier data normalization and convergence. Each evaluation batch was always managed as an experimental unit from larvae to harvest under commercial procedures and fish were never graded. For evaluation, a subset of 200 and 550 animals proportional to the estimated number of families in each batch (Table 9) was randomly selected and intraperitoneally tagged as previously reported (Carballo *et al.* 2018).

Animals were sampled before entering the final on-growing period in recirculation aquaculture systems (RAS; ~400 days) and at harvest (~800 d). Since total number of tagged fish in the evaluation batches represented between 4 and 11% of total population, only those tagged fish identified after one pass through the FISH Reader (Zeus, Trovan) (n = 1 843) were sampled at 400 d. Later at harvest (800 d), all animals (n = 2 171) were *in situ* sacrificed using slurring ice following commercial techniques and 60 animals of each batch were kept alive as future breeders. Sacrificed animals were taken a piece of caudal fin that was preserved in 99% alcohol and alive fish were sampled for blood as indicated above for DNA isolation. Moreover, sacrificed animals were dissected and alive fish sampled to record the sex and the presence of white nodules compatible with amoebic disease. Body weight (W), standard length (from mouth to beginning of caudal fin;SL) and width (Wi) were *in situ* measured for all fish. Moreover, animals were individually photographed using a Canon EOs1300D camera following the methodology previously established in PROGENSA® (Navarro et al. 2016). Image analysis was carried out using the Fiji 2.0.0-rc-69/1.52p and standard length (SLI), width (WiI) and total area (A) were

measured. All procedures were authorized by the Bioethics and Animal Welfare Committee of IFAPA and given the registration number 10/06/2016/101.

### 5.2.2- DNA isolation and Parentage Assignment

Total DNA from caudal fin (30 mg) or blood (20 µL) was isolated using Isolate II Genomic DNA Kit treated with RNase A (Bioline, London, UK) using the manufacturer's instructions. DNA was quantified spectrophotometrically using the Nanodrop ND-8000. Genotyping was carried out using 11 *loci* in a supermultiplex PCR (Guerrero-Cozar *et al.* 2020) that were run on an ABI3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Genotypes were collected using Genemapperv3.8 (Applied Biosystems, Foster City, CA, USA) and parentage assignment was determined using the exclusion method with Vitassign v8.2.1 (Vandeputte et al. 2006).

### 5.2.3- Genetic analysis

Before carrying out the genetic analyses, main factors were identified by ANOVA analysis using SPSS v.23 (SPSS, Chicago, IL, USA). Weight and area were transformed using cubic square root and square root, respectively to fit normality. Sex, amoebic disease and evaluation batch factors were found to be significant for all traits, hence, they were included as fixed factors in the model. Genetic parameters were estimated at the beginning of on-growing in RAS (~400 d) and at harvest (~800 d). Genetic estimates for heritability and correlations were calculated using trivariate animal models fitted by restricted maximum likelihood (REML) in WOMBAT (Meyer 2007): $y = X\beta + Zu + e$, where y is the observed trait, β, is the vector for fixed factor (sex, batch and amoebic lesions), u is the vector for animal random factor and e is the error. The age was initially tested as covariable but it was removed from the model since no effect on genetic estimates was observed. The adjusted breeding values were estimated using BLUP and later used for association tests.

### 5.2.4- OpenArray® design and methodology

For fish genotyping, a set of 60 SNPs was predicted after mapping 30 Illumina RNA-seq libraries (Benzekri *et al.* 2014; Cordoba-Caballero *et al.* 2020) onto a genome draft (Manchado *et al.* 2019). SNPs were identified with *Varscan mpileup2snp* (Koboldt et al. 2012) with a minimal coverage of 65 reads per position and a minimal amount of 15 reads supporting the SNP. As general criteria for selecting putative candidate sequences for primers and probes design, no ambiguities should exist into the 10 nucleotides (nt) upstream or downstream of identified SNP marker and the total number of ambiguities in the 600 nt surrounding the SNP should not be higher than 5. To position

the markers in the Senegalese sole genome, they were mapped in the scaffolds of a female genome (Claros *et al.* 2020; Guerrero-Cozar *et al.* 2020) and the anchored megascaffolds (SseLG) to a high-density genetic map of *S. senegalensis*. Markers were drawn in the genome using the software linkagemapview (Ouellette *et al.* 2018). As we were also interested in the identification of markers associated with sex, the sequences were mapped onto the flatfish *Cynoglossus semilaevis* genome to identify markers located in sexual chromosomes ZW and 14.

Primers and probes were designed using the on-line Custom Taqman® Assay Design Tool (www.thermofisher.com). Genotyping was carried out using a custom TaqMan® OpenArray® Genotyping Plate with 60 assays. Samples (2.5 µL of DNA sample normalized at 40ng $\mu l^{-1}$) were mixed with an equal volume of TaqMan® OpenArray® Genotyping Master Mix and each subarray was loaded into the OpenArray® plates with the OpenArray® AccuFill™ System according to the manufacturer's protocols. Genotyping plates were run in the QuantStudio™ 12 K Flex Real-Time PCR System (Thermo Fisher Scientific). The samples were amplified using the thermal cycling conditions established manufacturer's protocol.

To test if markers were polymorphic, wild breeders from the broodstock population (n = 164; female 83 and male 81) were selected. Results were analysed using the QuantStudio™ 12K Flex software and Thermo Fisher cloud. As the assays were newly designed, each assay was manually examined by viewing the real-time trace and the endpoint call. Data were exported as a matrix of genotypic calls for each individual sample. Hardy-Weinberg equilibrium (HWE) was tested using the SNPstats (Sole et al. 2006). Seven assays with a call rate lower than 90% were excluded from analysis (AN3267M, ANMF2WF, ANNKWGD, ANPRP2A, AN2XDMN, ANEPWZV, ANRWJK9). Only three markers appeared as monomorphic (ANFVRKU, ANKA9CK, ANKA9CJ) and one marker was not in HWE (ANAAFAZ) in whole population data set and it was excluded from the association analysis.

### 5.2.5- Association analysis

Four families (n = 279) with different adjusted weight at harvest (800 d) were selected for association analysis. Family 1 (n = 87) and 2 (n = 64) were considered as fast growing (FG-1) families (average adjusted weight 442.8 ± 19.7 g and 382.0 ± 9.6 g, respectively). The family 3 (n = 50) and 4 (n = 70) were slow-growing (SG) and half-sibs (average weight of 262.8 ± 6.2 g and 229.5 ± 4.3 g, respectively) (Figure 13). When the subset of animals also sampled at 400 d was

compared, only family 1 but not family 2 had a higher weight than families 3 and 4 at this age (Figure 13).

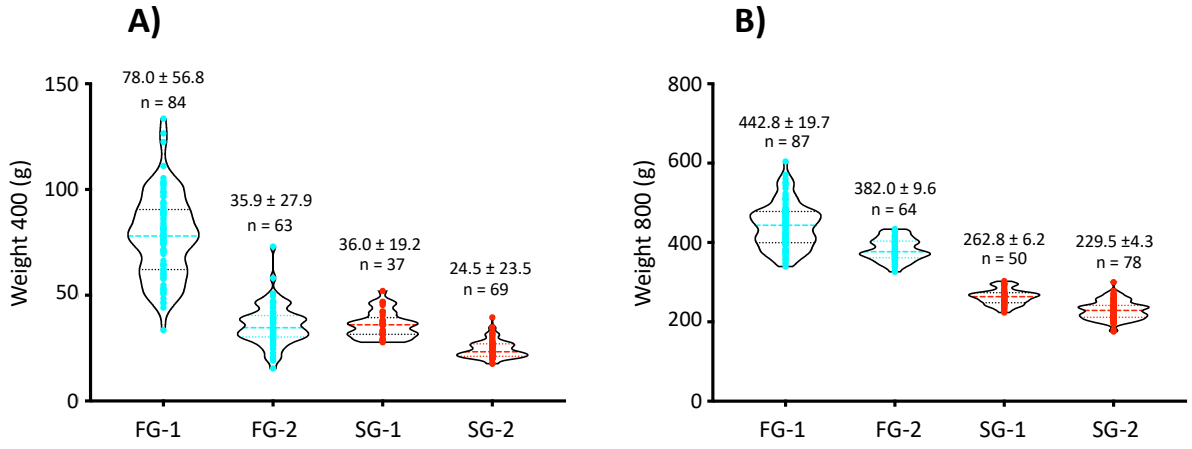


**Figure 13**. Selected families for association study. Families were selected by their differences in adjusted weight at harvest: two fast-growing (FG-1 and FG-2; blue) and two slow-growing (SG-1 and SG-2; red) families. Weight at 400 (A) and 800 d (B). Average weight ± SD at 400 (A) and 800 d (B). and number of individuals analysed for in each family are indicated. The number of individuals in each family at 400 d is lower than 800 d since not all animals were sampled to reduce handling effects.

To identify SNPs associated with sex, wild fish and animals from the four families were analyzed using a logistic mixed model (multi-step) approach as implemented in the R package GENABEL (v1.8-0) (Aulchenko *et al.* 2007) that was the best fit model for binary traits (Female= 0 and Male= 1). Growth marker-trait associations using adjusted phenotypic traits for the four families were tested using TASSEL software *v*5 (Bradbury et al. 2007). SNPs were filtered using a minimum allele frequency (MAF) of 0.05. In order to control false associations, population structure (Q) and/or relatedness (K) between individuals were taken into account in the general linear model (GLM) and the mixed linear model (MLM). The Q matrix based on principal component analysis and the kinship (K) matrix were calculated using TASSEL. Four statistical models were tested: naïve-model (GLM without any correction for population structure); Q-model (GLM with Q-matrix as correction for population structure); K-model (MLM with K-matrix as correction for kinship relationship structure). QK-model (MLM with Q-matrix and K-matrix as correction for population structure and kinship relationships). Fitness of different GWAS models for all traits was evaluated using quantile (Q-Q) plots of the observed *vs* expected -log10(*p*) values which should follow a uniform distribution under the null hypothesis. Association tests and Q-Q plots

were further validated using GWASpoly R package (Rosyara *et al.* 2016) under the QK-model in which K was constructed on DAPC technique. Bonferroni's correction (with genome-wide $\alpha$ = 0.05) was used for establishing a *P*-value detection threshold for statistical significance.

## 5.3- Results

### 5.3.1- Phenotype data

Seven fish batches (Table 9) from overall 150 breeders (distributed in nine tanks, sex ratio 1:1) were evaluated after synchronizing mass spawning. To reduce environmental variability, larvae and juveniles of each batch were always managed as a single unit and cultivated under industrial conditions from larval mouth opening to harvest. For genetic evaluation, a small subset of specimens (Table 9) was randomly selected and intraperitoneally tagged as previously reported (Carballo *et al.* 2018). Fish weight at tagging ranged between 1.7 ± 0.9 and 5.8 ± 4.5g (Table 9). All batches were evaluated at the beginning of on-growing stage in RAS (~400 d) and at harvest (~800 d). As tagged animals were cultivated under commercial conditions in large tanks (4-11% of whole batch population) not all tagged fish was sampled at 400 d to minimize handling effects and maximize survival. The total number of fish evaluated in each batch at 400 and 800 d is indicated in Table 9. Significant differences in growth traits between batches at both ages were observed (Figure 14, Table 9). The average weight of each batch at 400 d ranged between 18.2 ± 14.9 g (batch 3; age 397 dph) and 60.9 ± 44.0 g (batch 7; age 414 dph). SL oscillated between 10.4 ± 2.3 and 14.6 ± 3.5 cm and Wi between 3.8 ± 1.0 and 6.0 ± 1.7 cm. At 800 d, weight oscillated between 157.4 ± 103.1 and 425.3 ± 182.8, SL between 20.6 ± 4.5 and 27.4 ± 3.7 and Wi between 8.1 ± 2.1 and 12.0 ± 2.0 for batch 3 and 6, respectively.

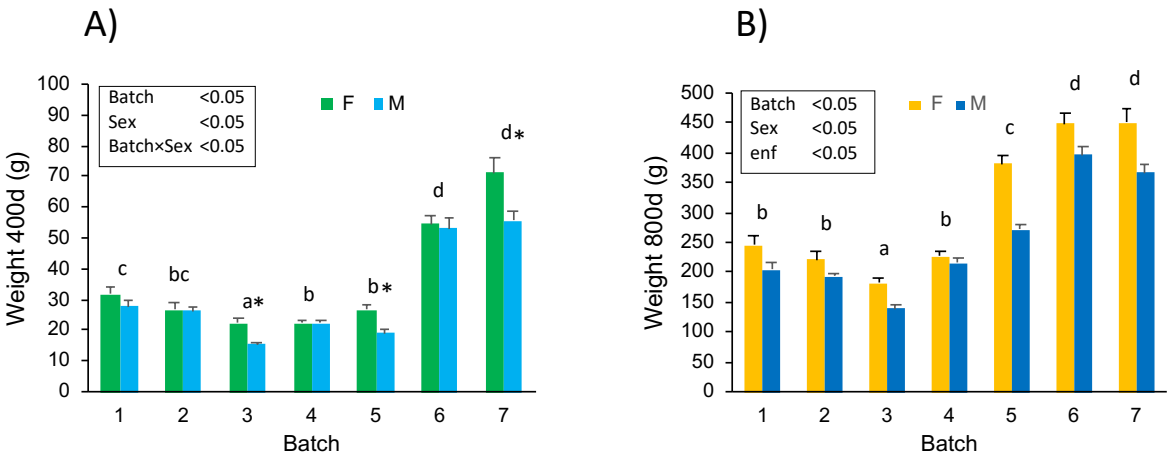


**Figure 14.** Weight of soles in each evaluation batch by sex at 400 (A) and 800 d (B). ANOVA results are indicated in the square. Letters denote significant differences between batches and asterisks significant differences between sexes (M: male; F: female) for each batch when exist interaction.

In addition to direct biometric measures, all sampled animals were photographed for digital image analysis to estimate standard length (SLI), width (WiI) and total area (A). Data are presented in Table 9. SLI and WiI values were similar to the corresponding traits directly measured on fish. Total area oscillated between 44.2 ± 21.0 and 66.1 ± 32.4 $cm^2$ at 400 d and between 128.7 ± 57.7 and 246.0 ± 71.6 $cm^2$ at 800 d.

**Table 9.** Growth traits for Senegalese sole for different batches. The birth date (birth), number of tagged fish (n), age at tagging (Aget), weight at tagging (Wt) and % males (M) of each batch are shown. For samplings at 400 and 800 d, number of fish sampled (n), age, weight (W), standard length (L) and width (Wi) as *in situ* determined, standard length (SLI), width (WiI) and total area (A) from image analysis are indicated. The number of families in each batch (nF) is also shown. Letters indicate statistically significant differences between batches for each batch.

| Batch | Birth | %M | Aget | Wt | n400 | Age | W400 | SL400 | Wi400 | SLI400 | WiI400 | A400 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Jul_14 | 61.7 | 278 | 5.3±3.1 | 137 | 446 | 29.2±16.4$^{c}$ | 12.1±2.1$^{c}$ | 4.8±1.1$^{c}$ | 12.4±2.2$^{c}$ | 4.8±1.1$^{c}$ | 50.4±19.8$^{d}$ |
| 2 | Oct_14 | 69.1 | 239 | 2.9±1.5 | 287 | 415 | 26.4±19.2$^{bc}$ | 11.6±2.3$^{bc}$ | 4.4±1.0$^{b}$ | 11.9±2.4$^{c}$ | 4.5±1.0$^{b}$ | 44.2±21.0$^{c}$ |
| 3 | Mar_15 | 58.0 | 243 | 5.8±4.5 | 273 | 397 | 18.2±14.9$^{a}$ | 10.4±2.3$^{a}$ | 3.8±1.0$^{a}$ | 10.2±2.3$^{a}$ | 3.7±1.0$^{a}$ | 31.2±15.9$^{a}$ |
| 4 | May_15 | 43.9 | 193 | 3.8±2.8 | 422 | 399 | 22.3±14.3$^{b}$ | 11.3±2.1$^{b}$ | 4.3±1.0$^{b}$ | 11.1±2.1$^{b}$ | 4.2±0.9$^{b}$ | 37.3±15.3$^{b}$ |
| 5 | Jun_15 | 49.0 | 171 | 1.8±0.8 | 229 | 398 | 22.9±14.1$^{b}$ | 11.3±2.0$^{b}$ | 4.4±0.9$^{b}$ | 11.1±2.0$^{b}$ | 4.3±0.9$^{b}$ | 38.1±15.0$^{b}$ |
| 6 | Sep_15 | 43.3 | 150 | 2.2±1.0 | 234 | 395 | 53.9±34.6$^{d}$ | 14.3±3.0$^{d}$ | 5.8±1.4$^{d}$ | 13.8±2.9$^{d}$ | 5.6±1.4$^{d}$ | 62.0±26.7$^{e}$ |
| 7 | Nov_15 | 65.1 | 154 | 1.7±0.9 | 261 | 414 | 60.9±44.0$^{d}$ | 14.6±3.5$^{d}$ | 6.0±1.7$^{d}$ | 14.3±3.5$^{d}$ | 5.8±1.6$^{d}$ | 66.1±32.4$^{e}$ |

| Batch | nF | n800 | Age | W800 | SL800 | Wi800 | SLI800 | WiI800 | A800 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 11 | 167 | 861 | 219.9±123.0$^{b}$ | 22.1±3.8$^{b}$ | 9.5±2.0$^{c}$ | 22.2±3.9$^{b}$ | 9.3±2.0$^{b}$ | 159.7±59.9$^{b}$ |
| 2 | 23 | 408 | 844 | 201.2±129.8$^{b}$ | 22.4±4.3$^{b}$ | 8.9±2.0$^{b}$ | 22.5±4.3$^{b}$ | 8.8±1.9$^{b}$ | 153.0±63.6$^{b}$ |
| 3 | 14 | 333 | 733 | 157.4±103.1$^{a}$ | 20.6±4.5$^{a}$ | 8.1±2.1$^{a}$ | 20.3±4.4$^{a}$ | 7.9±2.0$^{a}$ | 128.7±57.7$^{a}$ |
| 4 | 17 | 490 | 789 | 221.9±113.9$^{b}$ | 22.7±3.8$^{b}$ | 9.2±1.9$^{bc}$ | 22.2±3.8$^{b}$ | 9.0±1.9$^{b}$ | 158.0±55.9$^{b}$ |
| 5 | 12 | 259 | 817 | 327.3±146.1$^{c}$ | 25.6±3.9$^{c}$ | 10.7±1.9$^{d}$ | 25.1±3.9$^{c}$ | 10.4±1.9$^{c}$ | 206.5±64.9$^{c}$ |
| 6 | 13 | 245 | 780 | 425.3±182.8$^{d}$ | 27.4±3.7$^{cd}$ | 12.0±2.0$^{f}$ | 27.9±3.7$^{d}$ | 11.8±1.9$^{e}$ | 246.0±71.6$^{d}$ |
| 7 | 22 | 269 | 806 | 395.6±206.2$^{d}$ | 26.4±5.0$^{cd}$ | 11.4±2.4$^{e}$ | 25.9±4.8$^{c}$ | 11.0±2.3$^{d}$ | 221.4±80.9$^{c}$ |

As sex and disease symptoms are two important factors that modulate growth in sole, they were recorded for each animal at harvest. ANOVA analysis showed that males were statistically smaller than females for all growth traits ($P < 0.05$). Average weight of males was 8.2% smaller than females at 400 d (34.4 vs 30.7 g). Later at 800 d, these differences between males and females were more pronounced ($P < 0.05$) for the six growth traits. Males were as average 22.7% smaller than females (240.6 vs 295.3 g).

Overall, males were more abundant than females (55.3% of sampled animals) in the tested population. In batches 1, 2, 3 and 7, male percentages ranged between 58.0 and 69.1 % while in batches 4-6 between 43.3 and 49.0%. When sex ratios were analyzed by family a high variation was observed with male proportions oscillating between 16% and 90% ( Figure 15).

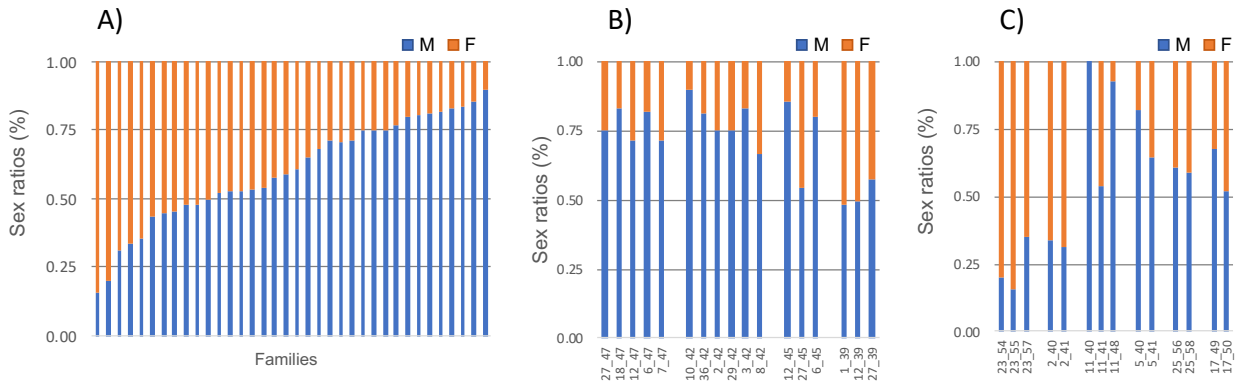


**Figure 15.** Male (M: blue) and female (F: orange) ratios. A) Sex ratios in families with n>10 individuals. B) Sex ratios in maternal-half sib families (mothers 39, 45, 42 and 47). C) Sex ratios in paternal-half-sib families (fathers 2, 5, 22, 17, 23, 25). The numbers indicate the father code followed by mother code.

A comparison of sex ratios in four maternal half-sib families indicated that male percentages were always higher than females (52-78% of males) (Figure 15). In contrast, paternal half-sib families produced offspring with a low (24-32%) or a high (60-82%) proportion of males even after crossing with the same females (codes 40 and 41, Figure 15) in the same batch. These families with an inverted male:female ratio were more represented in batches 4-6.

In some fish, some nodule lesions were observed in liver and/or intestine. These lesions were associated with an amoebic infection of genus Endolimax, very common in RAS systems. The incidence of animals with these lesions ranged between 0.6% in batch 3 and 21% in batch 4. Animal with internal lesions were as average 54.6% lower than healthy fish (179.7 vs 277.8 g).

### 5.3.2- Parentage assignment

Offspring was genotyped using 11 *loci* in a supermultiplex PCR. Parentage assignment was carried out using the exclusion method with a maximal tolerance of two errors. The 98.3% of specimens were successfully assigned to single parent pair. The number of breeders that contributed offspring was 68 (38 males and 30 females). A high bias in offspring contribution was detected since 9 fathers and 8 mothers contributed more than 100 individuals each one that overall represented 62.1 and 71.1% of total population. The total number of families was 70 with an average of 31.0 descendants per family (ranging between 1 and 163). The number of families per batch ranged between 11 (from 8 males and 9 females) and 23 (from 14 males and 10 females) for batch 1 and

2, respectively. One breeder was represented in the seven batches, 4 breeders in 6 batches, 1 breeder in 5 batches, 8 breeders in 4 batches and the remaining 54 breeders in three or less batches.

### 5.3.3- Genetic estimates

Heritabilities and correlations for growth traits at 400 and 800 days are depicted in Table 10. Heritabilities were higher at 400 than 800 d for the all traits. Wi had the highest heritability (0.643 and 0.500 at 400 d and 800 d, respectively) followed by weight (0.609 and 0.463), area (0.596 and 0.456) and SL (0.593 and 0.425). Heritabilities for SLI and WiI as determined by digital image analysis were similar to those directly measured on the animals.

**Table 10**. Heritabilities (diagonal), phenotypic correlations (below the diagonal) and genetic correlations (above the diagonal), with ± standard error) for growth traits at 400 and 800 d. weight (W), standard length (L) and width (Wi) as *in situ* determined, standard length (SLI), width (WiI) and total area (A) from image analysis are indicated.

| | W400 | SL400 | Wi400 | SLI400 | WiI400 | A400 |
|---|---|---|---|---|---|---|
| W400 | **0.609±0.108** | 0.995±0.002 | 0.995±0.002 | 0.993±0.003 | 0.995±0.002 | 0.997±0.001 |
| SL400 | 0.993±0.001 | **0.593±0.107** | 0.988±0.004 | 1.000±0.000 | 0.988±0.004 | 0.996±0.001 |
| Wi400 | 0.990±0.001 | 0.981±0.002 | **0.643±0.110** | 0.986±0.005 | 0.999±0.000 | 0.996±0.002 |
| SLI400 | 0.990±0.001 | 0.997±0.000 | 0.978±0.002 | **0.568±0.105** | 0.986±0.005 | 0.995±0.002 |
| WiI400 | 0.991±0.001 | 0.982±0.002 | 0.993±0.001 | 0.982±0.002 | **0.631±0.109** | 0.997±0.001 |
| A400 | 0.992±0.001 | 0.989±0.001 | 0.986±0.001 | 0.990±0.001 | 0.989±0.001 | **0.596±0.107** |
| | W800 | SL800 | Wi800 | SLI800 | WiI800 | A800 |
| W800 | **0.463±0.096** | 0.982±0.007 | 0.977±0.008 | 0.982±0.007 | 0.977±0.008 | 0.995±0.002 |
| SL800 | 0.982±0.002 | **0.425±0.091** | 0.944±0.020 | 1.000±0.000 | 0.944±0.02 | 0.983±0.006 |
| Wi800 | 0.982±0.002 | 0.958±0.004 | **0.500±0.099** | 0.944±0.020 | 1.000±0.000 | 0.987±0.005 |
| SLI800 | 0.982±0.002 | 0.996±0.000 | 0.959±0.004 | **0.424±0.091** | 0.943±0.02 | 0.984±0.006 |
| WiI800 | 0.985±0.002 | 0.960±0.004 | 0.994±0.000 | 0.961±0.004 | **0.499±0.099** | 0.985±0.006 |
| A800 | 0.994±0.001 | 0.988±0.001 | 0.985±0.001 | 0.988±0.001 | 0.988±0.001 | **0.456±0.094** |

Genetic correlations between growth traits at 400 d were higher than 0.98. At 800, genetic correlations were a bit lower but still very high (>0.94). Phenotypic correlations were higher than 0.99 at 400 d and higher than 0.96 at 800 d. As a subset of n = 1 843 was measured at both 400 and 800d, correlations between both ages were also determined. Genetic correlation between 400 and 800 ranged between 0.824 and 0.875 and phenotypic correlations between 0.766 and 0.807 (Table 11).

**Table 11**. Genetic (top) and phenotypic correlations (down) (± standard error) between growth traits at 400 and 800 d. Weight (W); standard length (SL); width (Wi); Total area (A)

| | W800 | SL800 | Wi800 | A800 |
|---|---|---|---|---|
| W400 | 0.850±0.053 | 0.832±0.050 | 0.856±0.049 | 0.861±0.049 |
| SL400 | 0.824±0.060 | 0.835±0.057 | 0.828±0.058 | 0.837±0.056 |
| Wi400 | 0.852±0.052 | 0.844±0.056 | 0.875±0.044 | 0.867±0.047 |
| A400 | 0.848±0.053 | 0.847±0.054 | 0.857±0.049 | 0.861±0.048 |
| | | | | |
| W400 | 0.786±0.018 | 0.781±0.019 | 0.786±0.019 | 0.784±0.018 |
| SL400 | 0.785±0.025 | 0.783±0.018 | 0.766±0.019 | 0.774±0.019 |
| Wi400 | 0.796±0.018 | 0.787±0.019 | 0.807±0.017 | 0.797±0.018 |
| A400 | 0.796±0.017 | 0.797±0.017 | 0.801±0.017 | 0.796±0.017 |

### 5.3.4- OpenArray validation

To identify SNP markers associated with growth traits, an openarray chip containing 60 SNPs markers was designed after mapping RNA-seq information using as reference a female genome draft (Manchado *et al.* 2019; Claros *et al.* 2020). The markers were positioned into a high-density SNP genetic linkage map indicating that a high number of markers were located in SseLG02, SseLG05 and SseLG21 that overall covered 17 out of twenty-one SseLGs (Figure 16). A comparison between species indicated most of the selected SNPs were putatively located in chromosomes Z, W and 14 of *C. semilaevis*. Those markers located in the chromosomes ZW of *C. semilaevis* appeared dispersed in several SseLG (SseLG01, SseLG04, SseLG08SseLG05, SseLG07, SseLG09, SseLG16, SseLG17, SseLG19SseLG20) and those SNPs located in chromosome 14 in two SseLGs (SseLG02 and SseLG21).

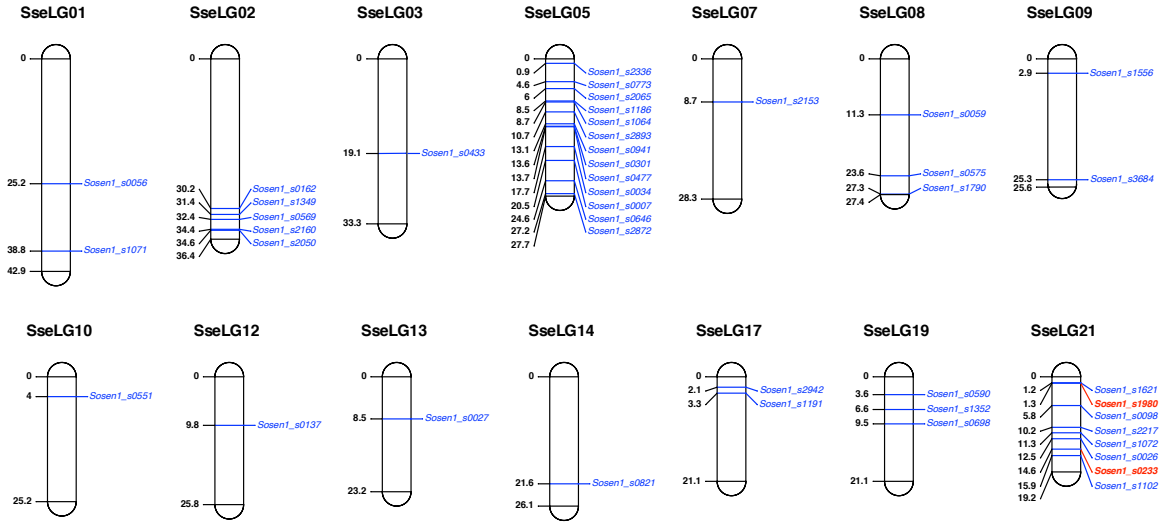


**Figure 16.** Distribution of SNP markers in the genome of *S. senegalensis*. Markers (in blue) are sorted by linkage groups (SseLG). Numbers on the left denote distance in cM. Significant markers in the association analysis are indicated in red.

The SNP array was validated using 164 wild soles (81 males and 83 females). A total of 7 markers (11.7%) failed to amplify and three markers were monomorphic. Moreover, one marker was not in HWE both in males and females and it was excluded from the association study. Finally, a total of 49 markers were selected with a call rate for the individuals higher than 95%, calculated as the proportion of SNPs giving a successful genotypic call for each individual.

### 5.3.5- Marker association with growth traits.

For the association study, four families with 50 or more individuals were selected by their weight at harvest. Two families were referred to as fast-growing (FG-1 and FG-2) with an average adjusted weight of 442.8 and 382.0 g (Figure 13). The other two families were named as slow-growing (SG-1 and SG-2) and the weight was 262.8 and 229.5 g. Weights of the individuals did not overlap between FG and SG families. When weight of those specimens also sampled at 400 d was compared, such differences were only observable for FG-1 (Figure 13).

A total of 47 polymorphic assays were detected with overall had genotyping call rates higher than 94%. PCA analysis identified two full-sib and two half-sib families (Figure 17). Association analyses were performed using four models: GLM naïve, GLM with Q matrices as covariate, MLM with K matrices as covariate and MLM (Q+K). The quantile-quantile plots indicated that MLM (K) and MLM (Q+K) models were significantly better than the GLM naïve and GLM (Q)

models. The MLM (Q+K) model was selected for the analysis as implemented in Tassel and GWASPoly programs.

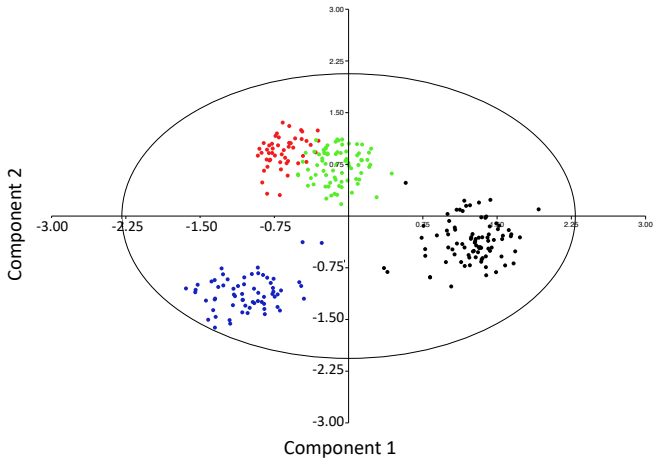


**Figure 17.** Family structure using a principal component analysis (PC1 and PC2) based on 47 SNPs from 279 animals. The first 2 components explain 37.4 and 30.7% of variance, respectively. Individuals of FG families are indicated in black and blue and those of SG families in red and green.

Two significant SNPs (Sosen_s1980 and Sosen_s0233) associated with growth traits after Bonferroni's correction ($P = 0.0011$; $\alpha = 0.05$) were detected at 400 and 800 d (Figure 18, Table 12).

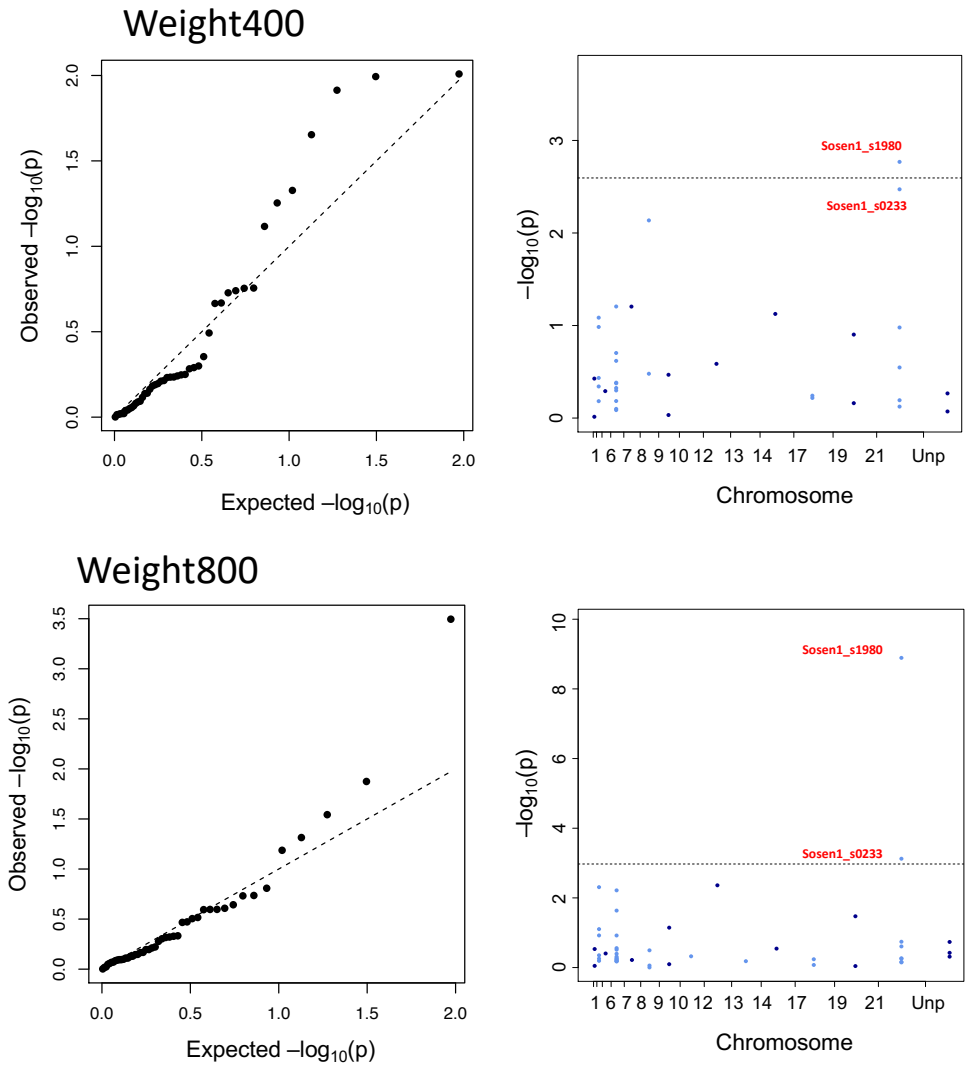


**Figure 18.** Quantile-quantile (QQ) and Manhattan plots of association study for weight at 400 d and 800 d using MLM (Q+K) model. The two significant SNPs at 800 d are indicated.

Sosen_s1980 was located in the general transcription factor 3C polypeptide 4 (*gtf3c4*) and it was significant for the four traits at both ages but association scores were stronger at 800 than 400 d. At 800 d, this marker explained from 11.89 to 14.77% of the variation. At 400 d, these values ranged between 5.16 and 7.98%. The scatter violin plots (Figure 18) show the effects of SNP alleles on weight at 400 and 800 d. Animals with TT and TG were weightier (54 and 130% at 400 d; 23 and 61% at 800 d) than GG (Figure 18).

**Table 12**. Significance levels for markers Sosen_s1980 and Sosen_s0233 from MLM(Q+K) analyses for growth traits at 400 and 800 d. Additive (add) and dominant (dom) *P*-values as determined by TASSEL (left) and GWASpoly (right) programs are shown. $R^2$ indicates proportion of phenotypic variation as determined by TASSEL. ns, not-significant. P-values are indicated in scientific notation.

| Marker | Trait | add *P*-value | dom *P*-value | $R^2$ (%) |
|---|---|---|---|---|
| Sosen_s1980 | W400 | ns/7.8E-06 | ns/ns | 8.0 |
| | SL400 | ns/8.8E-05 | ns/ns | 5.7 |
| | Wi400 | 1.4E-04/ns | ns/3.9E-05 | 5.4 |
| | A400 | ns/1.6E-04 | ns/8.1E-04 | 5.2 |
| | W800 | 2.6E-08/3.2E-04 | 7.3E-06/6.9E-10 | 12.5 |
| | SL800 | 4.2E-08/2.9E-04 | 1.5E-05/7.6E-10 | 11.9 |
| | Wi800 | 3.2E-08/6.5E-04 | 1.5E-07/1.7E-11 | 14.8 |
| | A800 | 3.7E-08/3.1E-04 | 6.2E-06/2.8E-10 | 12.6 |
| Sosen_s0233 | W400 | ns/ns | 4.7E-04/ns | 4.3 |
| | SL400 | ns/ns | 1.6E-04/1.1E-03 | 5.1 |
| | Wi400 | ns/ns | 6.1E-05/4.3E-04 | 5.9 |
| | A400 | ns/ns | 1.7E-04/1.3E-03 | 5.0 |
| | W800 | ns/ns | 3.9E-04/ns | 4.1 |
| | SL800 | ns/ns | 5.2E-04/ns | 4.0 |
| | Wi800 | ns/ns | ns/ns | ns |
| | A800 | ns/ns | 6.7E-04/ns | 3.9 |

The marker Sosen_s0233 was located within the mitochondrial fission process protein 1 (*mtfp1*) and it was significant for all traits except width at 800 d and only explained between 3.89 and 5.85% of variation. Animals with CC were weightier than TT and TC (104 and 131% at 400 d; 25 and 61% at 800 d, respectively) (Figure 19). No significant association with sex (male or female) using wild fish (n = 164) or the four families was observed (data not shown).

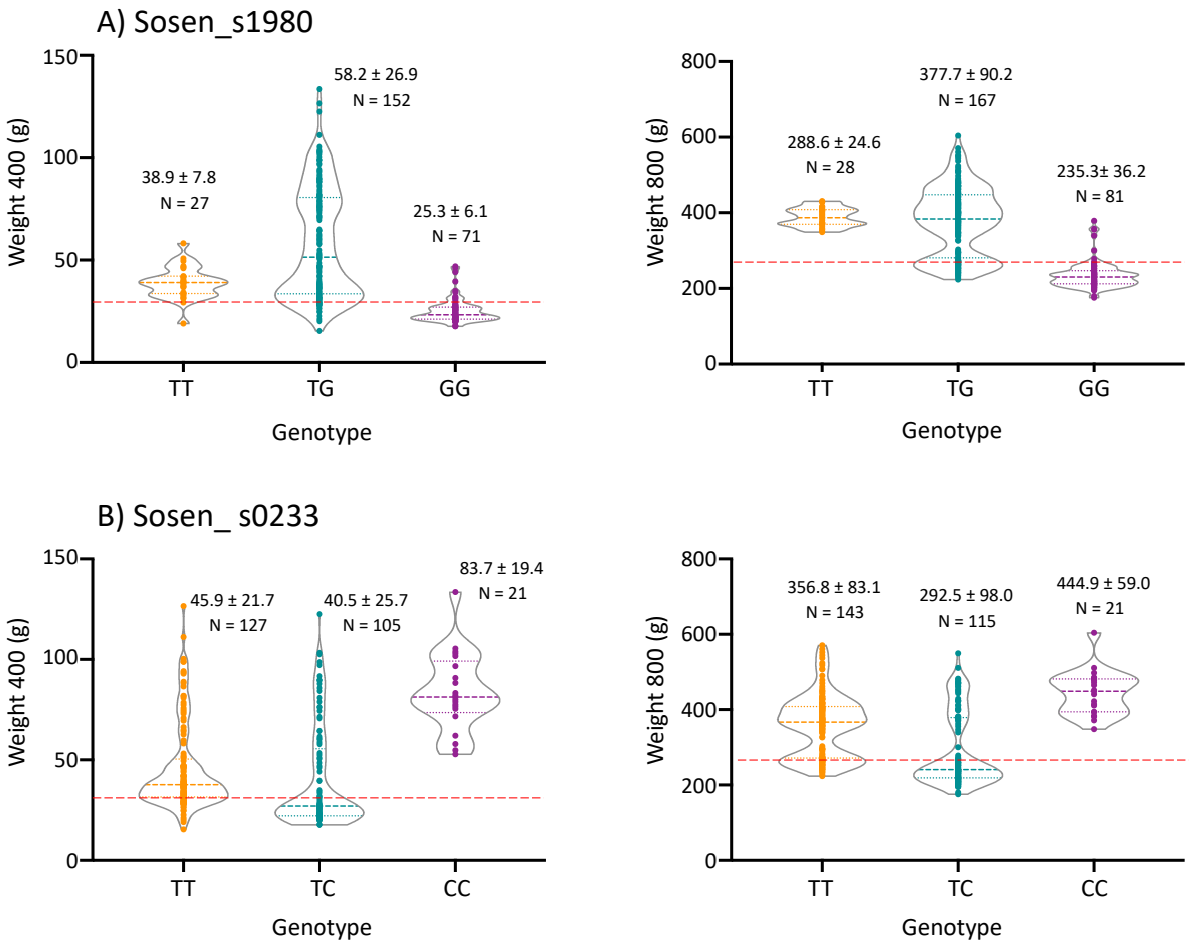


**Figure 19**. Validation of Sosen_s1980 and Sosen_s0233 markers for weight at 400 and 800. The adjusted weight average ± SD for each genotype and number of individuals analyzed are indicated. The red dashline represents the average population mean.

# CHAPTER 6

## Genetic estimates for growth and shape-related traits in the flatfish Senegalese sole

The results of this chapter were published in: Guerrero-Cozar, I.; Jimenez-Fernandez, E.; Berbel, C.; Espinosa, E.; Claros, M.G.; Zerolo, R.; Manchado, M. Genetic Estimates for Growth and Shape-Related Traits in the Flatfish Senegalese Sole. *Animals* 2021, 11, 1206 DOI: 10.3390/ani11051206

## 6.1- Introduction

Flatfish is a general name for a diverse group of highly appreciated species worldwide both in fisheries and aquaculture. They are morphologically unique among fishes due to their body asymmetry that is acquired after the migration of one eye to the opposite side and the cranium remodelling in early larval stages. This process, known as metamorphosis, also entails a drastic reorganization of abdominal cavity, skin pigmentation patterns and development of sensory structures for the adaptation to a bottom-dwelling mode of life. As consequence, the new flattened bodies acquire species-specific shapes for swimming and camouflage capabilities as adaptive mechanisms to specific ecological niches (Akkaynak *et al.* 2017). In a general way, flatfish species from families Bothidae, Cynoglossidae, Poecilopsettidae and Soleidae are characterized by oblong bodies with shorter jaws and long dorsal and anal fins than families Citharidae, Paralichthyidae, Pleuronectidae or Scophthalmidae among others (Black & Berendzen 2020). Although flatfish shape is slightly modified with the age and size, the species-specific morphological features are well identified by consumers and are usually important criteria in commercial decisions and price of fresh marketed products. Due to the high relevance of external morphology on commercialization, the production of high-quality shaped fish is really important in aquaculture to enhance consumers' awareness and support their perception of fish aquaculture products (Reinders et al. 2016).

Senegalese sole (*Solea senegalensis*) is a marine flatfish of high economic value whose aquaculture in rapidly growing in Southern Europe. The shape of this right-eyed flatfish is well-recognized by the lanceolate bodies, short jaws, and long fins. However, this species exhibits a high plasticity of the skeletal components such as the vertebral number that oscillates between 44 and 48 (mode = 45) with 8-9 in abdominal region, 34-35 in the caudal region and 3-4 in the caudal complex (de Azevedo *et al.* 2017; Fernandez *et al.* 2017). Moreover, this species is highly prone to vertebral abnormalities and other skeletal malformations that can reach even more than 70 % of individuals in cultured populations, most of them corresponding to vertebral fusions in the caudal region and deformities in the caudal complex (Gavaia et al. 2009; Dionisio *et al.* 2012; Losada et al. 2014; de Azevedo *et al.* 2017; de Azevedo et al. 2019a). Most of these malformations are usually externally unnoticed or they have a moderate effect on gross phenotypic morphology (approximately 46 % of animals with vertebral deformities were categorized as normal) (de Azevedo *et al.* 2017). However, this plasticity and high incidence of malformations can shift the body ellipticity with impact in the quality of the marketable product, hence, it is very important to

identify the phenotypic and genetic determination of main morphological traits and the association with other productive parameters.

Nutritional factors and environmental conditions have been identified as two major modulators of morphological features and malformations in Senegalese sole. High levels of vitamin A increase the mean number of vertebrae and the malformation rates in vertebrae and caudal fin (Fernández *et al.* 2009). Moreover, a high stocking density (29.8 kg $m^{-2}$) shifts the relative body proportions toward a wider head and a shorten caudal region with an enlarged peduncle (Ambrosio *et al.* 2008). A high temperature (>18°C) during larval rearing also increases vertebral anomalies in the caudal region and caudal complex although the effects on external morphology were not evaluated (Dionisio *et al.* 2012). In the closely related species *Solea solea*, the body ellipticity measured using image analysis was proposed as an optimal trait to assess the quality of external sole shape (Blonk *et al.* 2010c). This trait showed a moderate heritability (0.34±0.11) and a moderate and negative genetic correlation with body weight highlighting the importance of controlling for this trait to maintain high-quality shaped fish in genetic breeding programs (Blonk *et al.* 2010c). This study aimed at estimating the genetic and phenotypic parameters for growth and shape-related traits at two important stages in the production cycle of Senegalese sole, before entering growth-out in recirculation aquaculture systems (RAS) (~400 d) and at harvest (~800 d). Weight, standard length, three body heights (at the pectoral fin, maximal and in the peduncle), their relative ratios and body ellipticity were evaluated as quality indicators of sole shape. Heritability estimates and genetic and phenotypic correlations at both ages are provided. The data provided are highly relevant in genetic breeding programs.

## 6.2- Materials and Methods

### 6.2.1- Animals

Broodstock used to produced families comprised 150 wild specimens approx. 8-years old caught in salt marshes from the Gulf of Cadiz (Spain). They were fed with frozen feed including mussels, small squids and polychaeta worms (Seabait Ltd., UK) in alternative days. Mass spawning strategy to create the families was previously described (Guerrero-Cozar *et al.* 2021). Briefly, spawning was synchronized by thermoperiod control (Martin *et al.* 2014). Due to the courtship behavior of sole (Fatsini *et al.* 2020), it is not easy to achieve that all the breeder tanks (n=9) respond simultaneously in the same thermocycle. Hence, with the objective to increase number of families

in the population on evaluation, seven evaluation batches (EB) obtained after different thermocycles were created by mixing proportionally the volume of eggs from each tank that contributed offspring in each thermal treatment. To facilitate the data comparison and convergence, the offspring of a breeder tank (n=6) were always included in all EB. Larval rearing and weaning protocols for each EB were those previously described (Cañavate & Fernandez-Diaz 1999; Roman-Padilla et al. 2017) and each EB was always managed as a unit until harvest without any grading.

For genetic evaluation, fish (ranging from 200 to 550 specimens per EB) were intraperitoneally tagged with ages ranging between 150- and 278-days post-hatch (dph) as previously reported (Carballo *et al.* 2018; Guerrero-Cozar *et al.* 2021). Later, fish were phenotypically *in vivo* evaluated at ~400 d (ranging from 395 to 446 dph) before entering in the growth-out period in RAS and at harvest age ~800 d (ranging between 733 and 861 dph). No intermediate samplings were carried out to follow standard production practices and minimize animal handling and stress. Cumulative mortality between ages was lower than 5% and a total of 1 840 fish (EB1=136; EB2=289; EB3= 273; EB4= 420; EB5= 229; EB6=234; EB7=259) sampled at both ages was considered in this study. Information about the full dataset and culture conditions was previously reported (Guerrero-Cozar *et al.* 2021). Fish were individually weighted (W) using Gram FC-200 and taken a photograph using a Canon EOs1300D camera following the methodology previously established in PROGENSA® (Navarro *et al.* 2016). Image analysis was carried out using the Fiji 2.0.0-rc-69/1.52p and standard length (SL), body height at the insertion of the pectoral fin (BHP), body maximum height (BMH) and caudal peduncle height (CPH) were measured (Figure 20). The two ratios between heights (BMH to BHP and BMH to CPH) and ellipticity [(SL-BMH)/(SL+BMH)] (Blonk *et al.* 2010c) were calculated. At harvest, fish were sacrificed using slurring ice following commercial techniques and 60 specimens of each batch were kept alive as future breeders. Sacrificed fish were taken a piece of caudal fin that was preserved in 99% alcohol and alive fish were sampled for blood by puncturing in the caudal vein using a heparinized syringe, adding heparin (100 mU) and keeping at -20 °C until use. All fish were sexed and the presence of white nodules compatible with amoebic disease (AD) were recorded.

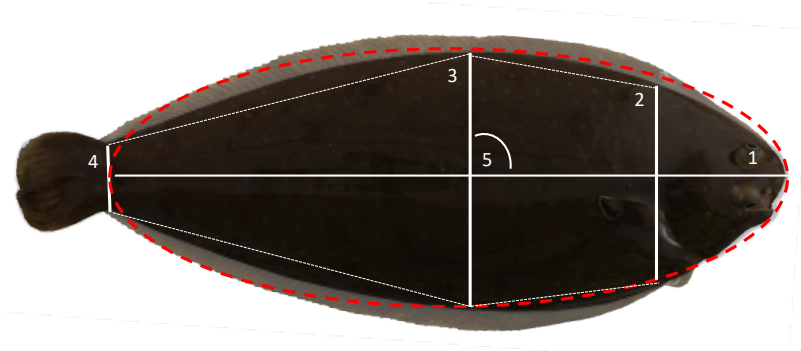


**Figure 20.** Shape measurements: 1: standard length (SL); 2: body height at the insertion of the pectoral fin (BHP); 3: body maximum height (BMH); 4: caudal peduncle height (CPH); 5: Ellipticity [(SL-BMH)/(SL+BMH)]. A theoretical ellipse fitting the horizontal axis from the mouth tip to the peduncle centre and the vertical axis to BMH in indicated in red dashed line.

### 6.2.2- DNA isolation and Parentage Assignment

DNA isolation from blood (broodstock and non-sacrificed offspring) or caudal fin (slaughtered F1; 30 mg) was carried out using the Isolate II genomic DNA kit (Bioline, London, UK) following the manufacturer's instructions. DNA was quantified using a Nanodrop ND-8000 and quality was evaluated by agarose gene electrophoresis. Genotyping of breeders and offspring was carried out using a 11-*loci* supermultiplex PCR (Guerrero-Cozar *et al.* 2020) on an ABI3130 sequencer (Applied Biosystems, USA) and genotypes were collected using Genemapperv3.8 (Applied Biosystems, USA). Finally, parentage assignment was performed with Vitassign v8.2.1 (Vandeputte *et al.* 2006) following the allelic exclusion method. Assignment rates to a single parent pair was 100%. A total number of 71 families from 37 males and 30 females were evaluated. The number of families per batch ranged from 11 (EB1 and EB5) and 23 (EB7). Offspring of seven males and six females were represented in four or more EB.

### 6.2.3- Statistical analysis and genetic parameters

All data were tested for normality and homogeneity of variance using SPSS v.23 (SPSS, Chicago, IL, USA). Weight at 400 and 800 d were cube square root and square root transformed, respectively to fit normality. ANOVA analysis using the GLM procedure was carried out using the gender, EB, and AD as fixed factors. To test the effect of age (evaluation of traits between 400 and 800 d), a repeated measures ANOVA was carried out for each trait using the same fixed factors. Regression analysis and slope significance testing were carried out with Prism 9.0

(Graphpad Software Inc.). Genetic estimates of heritability and correlations were calculated using restricted maximum likelihood adjusted linear mixed models (REML) in WOMBAT (Meyer 2007): $y = X\beta + Zu + e$, where y is the observed trait, β, is the fixed factor vector (gender, EB and AD), u is the animal random factor vector and e is the error.

## 6.3- Results

### 6.3.1- Phenotypic data for growth traits

The phenotypic mean ± SE of growth traits (weight and SL) at 400 and 800 d are depicted in Table 13. Mean weight and SL were 32.4 ± 29.2 g and 12.00 ± 2.87 cm at 400 d and 264.9 ± 171.9 g and 23.35 ± 4.79 cm at 800 d. Statistical ANOVA analysis showed statistically significant differences associated with the gender, EB and AD (Figure 21) for both traits. Estimated marginal means indicated that the females appeared as average 16.1 % heavier and 2.8 % longer than males at 400 d and 12.2% heavier and 2.5% longer than males at 800 d (Figure 21). A significant interaction genderxEB was observed at both ages. Also, significant differences associated with the EB ($P$ <0.05) were found that ranged between 18.9 and 63.3 g at 400 d for EB3 and EB7, respectively and between 126.8 and 376.7 g at 800 d for EB3 and EB6, respectively (Figure 21). A total of 15.3% of evaluated fish at harvest had nodules compatible with amoebic lesions in the gut and/or liver. Fish without hepatic or intestinal amoebic lesions at 800 d were significantly heavier (44.9%) than infected fish (Figure 21). A repeated measures ANOVA analysis revealed significant agexEB and agexAD interactions for weight and length gain and agexgender for weight gain. Tendencies for the different three fixed factors and levels are depicted in Figure 21. A regression weight-length analysis for gender at both ages showed that the coefficients of determination ($R^2$) were ≥0.95 with slopes between 3.32-3.34 (not statistically significant).

**Table 13**. Phenotypic data for growth traits (weight and SL), heights (BHP, BMH and CPH), height ratios (BMH/BHP and BMH/CPH) and ellipticity at 400 (A) and 800 d (B). Overall mean±standard error and by gender are shown. The number (n) of soles evaluated at each age are indicated.

| 400 days n=1 840 | Male (n=1 007) | Female (n=833) | Mean |
|---|---|---|---|
| Weight | 30.7±28.0 | 34.4±30.4 | 32.4±29.2 |
| SL | 11.85±2.87 | 12.18±2.86 | 12.00±2.87 |
| BHP | 3.84±1.00 | 4.00±1.02 | 3.91±1.01 |
| BMH | 4.54±1.33 | 4.71±1.34 | 4.62±1.34 |
| CPH | 1.13±0.34 | 1.10±0.34 | 1.11±0.34 |
| BMH/BHP | 0.45±0.03 | 0.45±0.03 | 0.45±0.03 |
| BMH/CPH | 1.17±0.06 | 1.17±0.06 | 1.17±0.05 |
| Ellipticity | 4.165±0.358 | 4.183±0.335 | 4.173±0.347 |
| 800 days n=1 840 | Male (n=1 007) | Female (n=833) | Mean |
| Weight | 244.0±153.0 | 290.3±189.3 | 264.9±171.9 |
| SL | 22.91±4.64 | 23.88±4.91 | 23.35±4.79 |
| BHP | 7.51±1.66 | 8.01±1.85 | 7.74±1.77 |
| BMH | 9.24±2.22 | 9.87±2.46 | 9.53±2.35 |
| CPH | 2.55±0.66 | 2.68±0.67 | 2.61±0.67 |
| BMH/ BHP | 0.426±0.027 | 0.419±0.028 | 0.424±0.028 |
| BMH/CPH | 1.225±0.054 | 1.227±0.050 | 1.226±0.052 |
| Ellipticity | 3.650±0.325 | 3.699±0.320 | 3.673±0.323 |

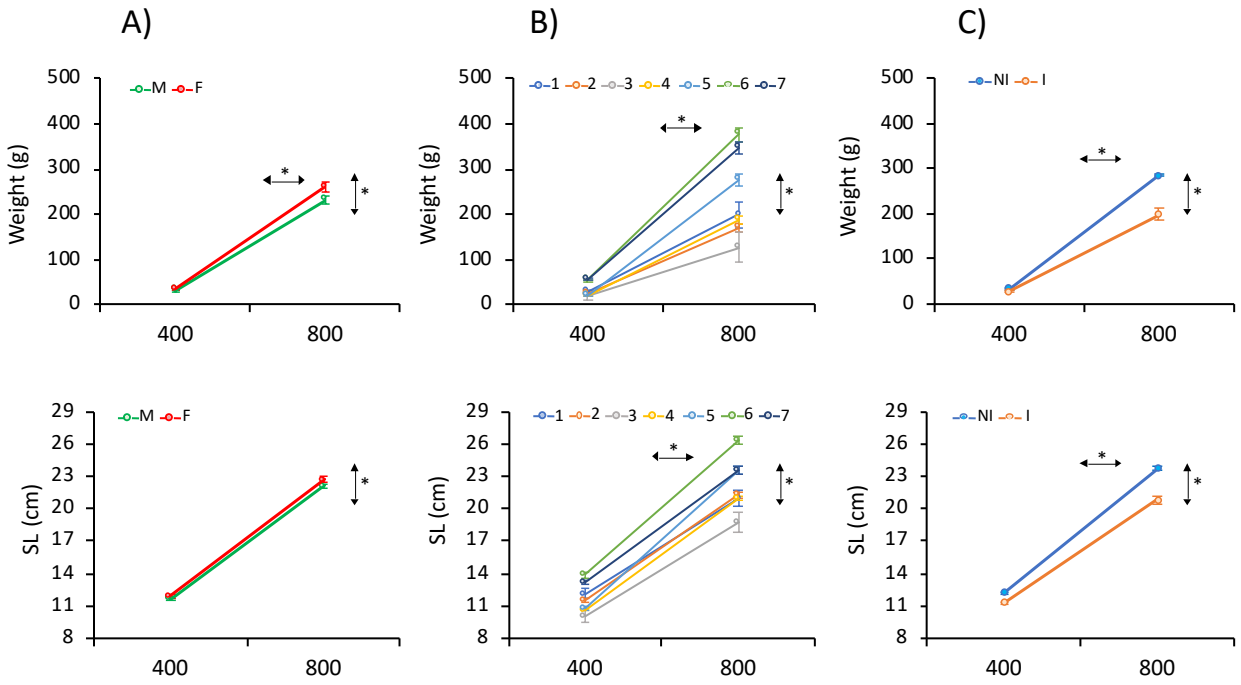


**Figure 21.** Estimated marginal means for weight and SL as determined by repeated-measures ANOVA at 400 and 800 d for the A) gender (Male: M; Female: F), B) EB (1-7) and C) AD (Infected: I; Non-Infected: NI). The asterisks on the horizontal or vertical arrows denote if within- or between-subjects were significant.

### 6.3.2- Phenotypic data for height traits

Due to the flatten morphology of sole, the body height at the insertion of pectoral fin (BHP), body maximum height (BMH) and caudal peduncle height (CPH) and the two ratios BMH/BHP and BMH/CPH were determined both at 400 and 800 d (Table 13; Figure 20). Mean BHPs were 3.91 ± 1.01 and 7.74 ± 1.77 cm at 400 and 800 d, respectively, the BMHs 4.62 ± 1.34 and 9.53 ± 2.35 cm, respectively, and the CPHs 1.11±0.34 and 2.61 ± 0.67 cm, respectively. The three height traits showed statistically significant differences associated with the gender and EB at both ages and AD at 800 d (Figure 22). Females and non-infected soles had higher heights than males and infected fish. As average, heights in females were 4.6, 4.5 and 4.3% higher than in males and the non-infected fish 10.2 11.2, 11.9% higher than in infected fish for BHP, BMH and CPH, respectively. Moreover, the EB6 and EB3 showed the largest and lowest heights, respectively. A longitudinal analysis to determine the height gain from 400 to 800 d using repeated measures ANOVA demonstrated significant interactions agexgender, agexEB and agexAD (Figure 22).

A regression analysis of CPH and BHP on BMH indicated a stronger association between BHP and BMH ($R^2$>0.97) than CPH and BMH ($R^2$>0.86). Moreover, slopes for males were statistically significant smaller than females at 800 d in both ages.

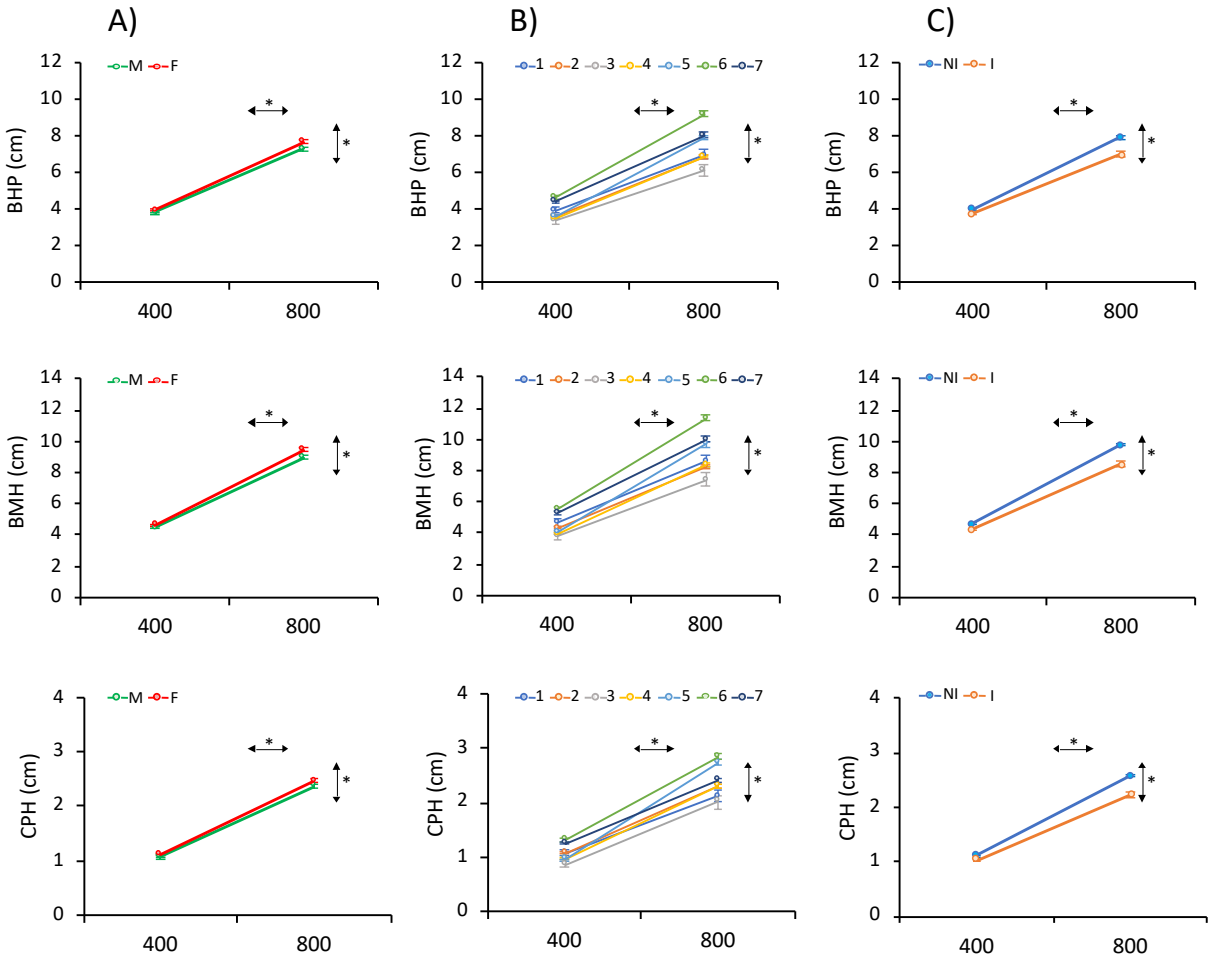


**Figure 22.** Estimated marginal means for height at the pectoral fin base (BHP), body maximum height (BMH) and caudal peduncle height (CPH) as determined by repeated-measures ANOVA at 400 and 800 d for the A) gender (Male: M; Female: F), B) EB (1-7) and C) AD (Infected: I; Non-Infected: NI). The asterisks on the horizontal or vertical arrows denote if within- or between- subjects were significant.

With respect to the BMH/CPH and BMH/BHP ratios, BMH/CPH significantly reduced and the BMH/BHP increased with the age from 400 to 800 d (Figure 23). A significant effect of the EB on both ratios at 400 and 800 d was detected. Nevertheless, gender effect was only significant for BMH/BHP at 400 d. The longitudinal analysis only identified a significant interaction agexEB. In

the repeated-measures ANOVA a significant between-subject effect of AD for BMH/BHP was also found.

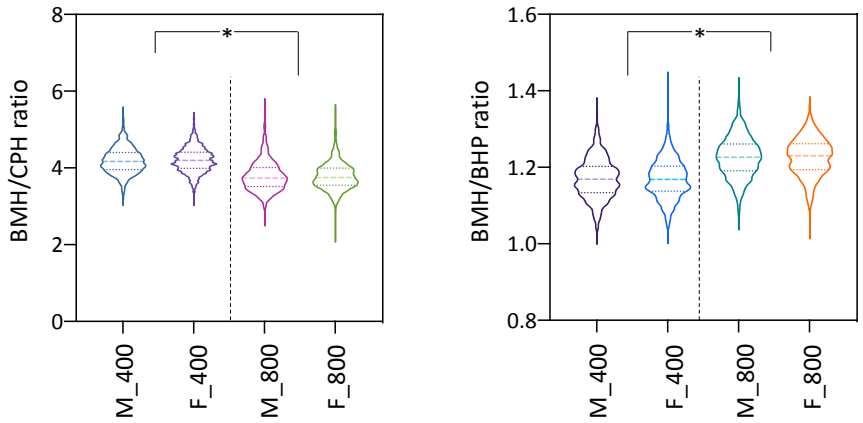


**Figure 23.** Violin plots for BMH/CPH and BMH/BHP ratios. Data for males (M) and females (F) at both 400 and 800 d are indicated. The asterisk denotes statistically significant differences between ages.

### 6.3.3- Phenotypic data for ellipticity

Mean ellipticity was 0.449 ± 0.025 at 400 d and 0.422 ± 0.029 at 800 d (Table 13). The distribution of ellipticity at both shapes is shown in Figure 24. Values ranged from 0.32 to 0.52 at 400 d and between 0.24 and 0.51 at 800 d. ANOVA analysis indicated statistically significant differences associated with the gender and EB at both ages and with AD at 800 d (Figure 25). Males and infected fish were more elliptic than females (1.0 and 2.3% higher at 400 and 800d, respectively) and non-infected fish (1.4% higher) (Table 13; Figure 25). The longitudinal analysis demonstrated a significant interaction agexgender and agexEB during the cultivation period in RAS (Figure 25). As ellipticity was significantly and negatively correlated with weight ($R^2$ ranging 0.362-0.443), an ANCOVA analysis using the weight as covariate was carried out and significant differences associated with the EB and gender at both ages were still observable. An analysis of ellipticity by weight classes indicated that females were statistically less elliptic than males in class 0-10 g at 400 d and classes 0-100, 300-400, 400-500 and >600 g at 800 d.

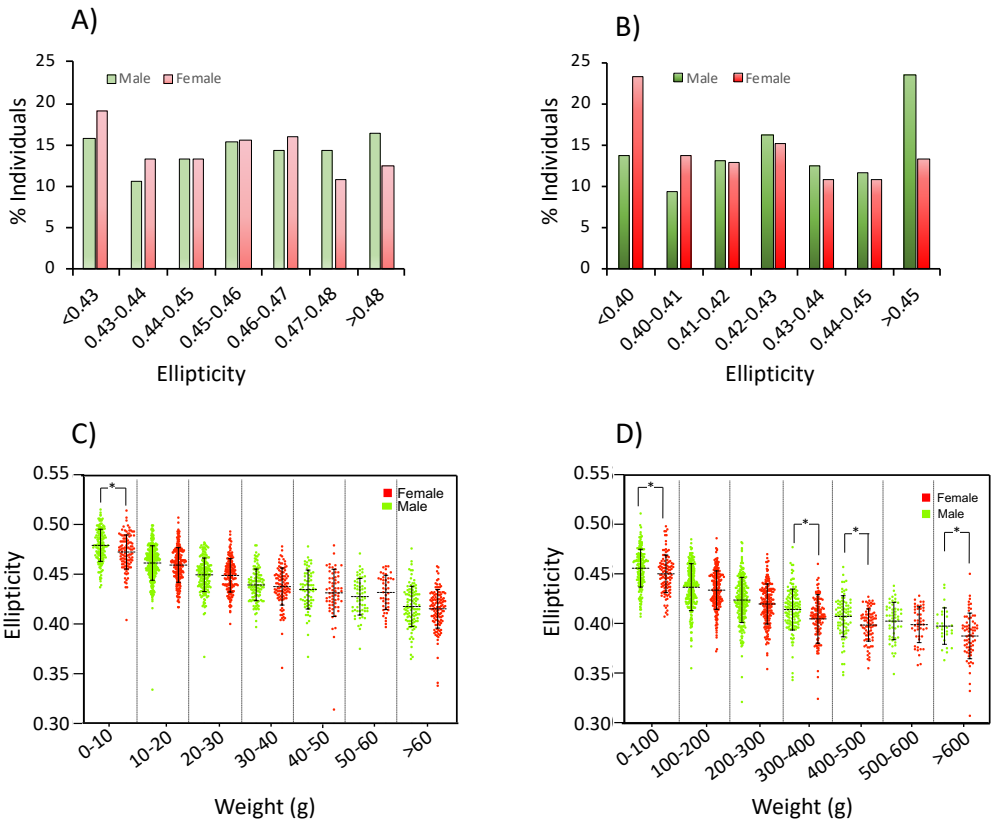


**Figure 24.** Distribution of ellipticity classes and by weight. Panels A (400 d) and B (800 d) show the frequency of males (green) and females (red) by ellipticity classes. Panels C (400 d) and D (800 d) show the ellipticity scatterplot by weight classes and gender. The asterisks denote statistically significant differences between gender in a weight class.

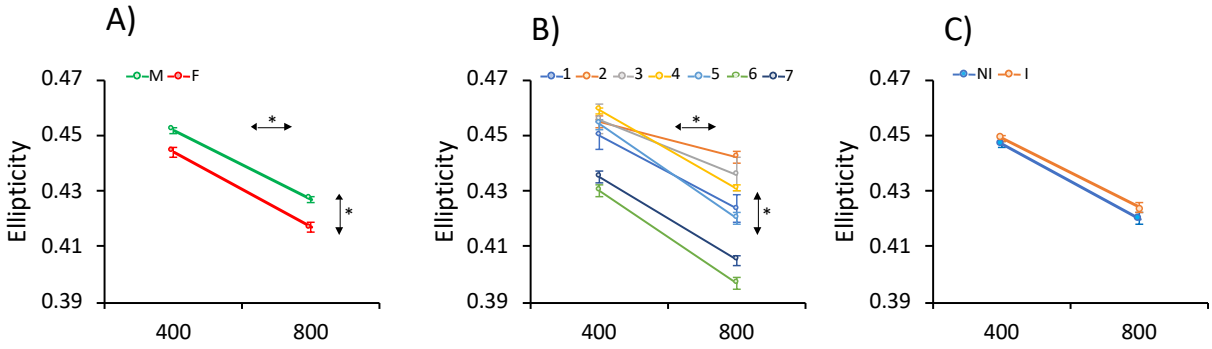


**Figure 25.** Estimated marginal means for ellipticity as determined by repeated-measures ANOVA at 400 and 800 d for the A) gender (Male: M; Female: F), B) EB (1-7) and C) AD (Infected: I; Non-Infected: NI). The asterisks on the horizontal or vertical arrows denote if within- or between-subjects were significant

### 6.3.4- Genetic estimates

#### *6.3.4.1- Heritability*

Heritabilities and correlations for growth and shape-related traits at 400 and 800 d are depicted in Table 14. Heritabilities were higher for all the traits (except BMH/CPH) at 400 than 800 d. Heritability estimates for weight, SL, the three heights and ellipticity were high or very high at both ages. They ranged between 0.567 and 0.774 at 400 d and between 0.433 and 0.735 at 800 d for ellipticity and SL, respectively. The ratios BMH/BHP and BMH/CPH had low or moderate heritability values (0.270-0.303 at 400 d and 0.144-0.306 at 800 d).

#### *6.3.4.2- Genetic correlations*

Genetic correlations between growth and heigh traits were very high both at 400 and 800 d (>0.95). The ratio BMH/BHP had moderate-high genetic correlations with growth and heigh traits that were higher at 400 d (0.858-0.881) than 800 d (0.412-0.612). The genetic correlations of BMH/CPH were low (<0.28). The ellipticity had negative and high genetic correlations with growth and heigh traits ranging from -0.724 to -0.828 at 400 d and from -0.509 and -0.733 at 800 d and a negative and low with heigh ratios (Table 14).

Genetic and phenotypic correlations between both ages are depicted in Table 15. Ellipticity (0.912) had the highest genetic correlation when the same traits were compared at 400 and 800 d followed by growth and height traits (average 0.825 and 0.874). The lowest values were between height ratios (0.663-0.687).

**Table 14**. Heritabilities (diagonal), phenotypic correlations (below the diagonal) and genetic correlations (above the diagonal) for growth traits (weight and SL), heights (BHP, BMH and CPH), height ratios (BMH/BHP and BMH/CPH) and ellipticity (E) at 400 d (top) and 800 d (bottom).

| 400 d | W | SL | BHP | BMH | CPH | BMH/BHP | BMH/CPH | E |
|---|---|---|---|---|---|---|---|---|
| W | **0.625±0.109** | 0.991±0.004 | 0.988±0.004 | 0.992±0.003 | 0.990±0.005 | 0.874±0.057 | 0.161±0.189 | -0.768±0.073 |
| SL | 0.983±0.002 | **0.567±0.104** | 0.981±0.007 | 0.986±0.005 | 0.984±0.007 | 0.881±0.054 | 0.167±0.189 | -0.724±0.085 |
| BHP | 0.981±0.002 | 0.976±0.002 | **0.623±0.110** | 0.991±0.001 | 0.948±0.005 | 0.858±0.064 | 0.284±0.180 | -0.738±0.025 |
| BMH | 0.988±0.001 | 0.982±0.002 | 0.999±0.001 | **0.621±0.109** | 0.955±0.004 | 0.878±0.055 | 0.247±0.183 | -0.828±0.056 |
| CPH | 0.952±0.004 | 0.953±0.004 | 0.974±0.010 | 0.979±0.008 | **0.576±0.105** | 0.881±0.055 | 0.044±0.193 | -0.749±0.079 |
| BMH/BHP | 0.591±0.024 | 0.61±0.023 | 0.528±0.028 | 0.622±0.023 | 0.587±0.024 | **0.270±0.069** | 0.102±0.035 | -0.521±0.034 |
| BMH/CPH | 0.062±0.045 | 0.047±0.043 | 0.09±0.044 | 0.094±0.044 | -0.191±0.044 | 0.076±0.204 | **0.303±0.076** | -0.254±0.045 |
| E | -0.673±0.031 | -0.284±0.033 | -0.838±0.054 | -0.750±0.025 | -0.677±0.031 | -0.662±0.115 | -0.487±0.153 | **0.774±0.117** |

| 800d | W | SL | BHP | BMH | CPH | BMH/BHP | BMH/CPH | E |
|---|---|---|---|---|---|---|---|---|
| W | **0.486±0.099** | 0.983±0.007 | 0.974±0.01 | 0.978±0.008 | 0.983±0.008 | 0.546±0.162 | 0.016±0.198 | -0.608±0.115 |
| SL | 0.975±0.002 | **0.433±0.094** | 0.961±0.015 | 0.957±0.016 | 0.964±0.014 | 0.412±0.183 | 0.011±0.198 | -0.509±0.137 |
| BHP | 0.967±0.003 | 0.948±0.005 | **0.549±0.105** | 0.996±0.002 | 0.953±0.018 | 0.506±0.177 | 0.181±0.193 | -0.703±0.092 |
| BMH | 0.982±0.002 | 0.962±0.004 | 0.983±0.001 | **0.515±0.102** | 0.961±0.016 | 0.586±0.155 | 0.182±0.192 | -0.733±0.085 |
| CPH | 0.928±0.005 | 0.918±0.006 | 0.911±0.007 | 0.926±0.006 | **0.463±0.097** | 0.612±0.151 | -0.073±0.195 | -0.586±0.123 |
| BMH/BHP | 0.479±0.027 | 0.486±0.027 | 0.345±0.031 | 0.508±0.026 | 0.474±0.026 | **0.144±0.046** | 0.178±0.211 | -0.389±0.031 |
| BMH/CPH | -0.003±0.042 | -0.015±0.04 | 0.046±0.043 | 0.050±0.043 | -0.314±0.040 | 0.048±0.029 | **0.306±0.075** | -0.217±0.044 |
| E | -0.557±0.037 | -0.447±0.042 | -0.644±0.033 | -0.662±0.030 | -0.548±0.038 | -0.719±0.117 | -0.534±0.144 | **0.735±0.115** |

**Table 15**. Genetic (top) and phenotypic correlations between 400 (left) and 800d (right) for growth traits (weight and SL) heights (BHP, BMH and CPH), height ratios (BMH/BHP and BMH/CPH) and ellipticity (E).

| | | 800 d | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Genetic | W | SL | BHP | BMH | CPH | BMH/BHP | BMH/CPH | E |
| 400 d | W | **0.843±0.054** | 0.831±0.060 | 0.832±0.057 | 0.838±0.055 | 0.849±0.054 | 0.509±0.169 | 0.041±0.194 | -0.554±0.122 |
| | SL | 0.828±0.058 | **0.837±0.057** | 0.813±0.062 | 0.817±0.061 | 0.826±0.057 | 0.509±0.161 | 0.049±0.194 | -0.511±0.134 |
| | BHP | 0.868±0.047 | 0.859±0.052 | **0.874±0.045** | 0.876±0.044 | 0.853±0.053 | 0.524±0.166 | 0.213±0.186 | -0.637±0.105 |
| | BMH | 0.853±0.051 | 0.856±0.050 | 0.862±0.048 | **0.870±0.046** | 0.846±0.055 | 0.542±0.162 | 0.180±0.188 | -0.618±0.109 |
| | CPH | 0.814±0.063 | 0.821±0.062 | 0.786±0.070 | 0.811±0.063 | **0.825±0.062** | 0.523±0.167 | 0.025±0.195 | -0.533±0.126 |
| | BMH/BHP | 0.633±0.123 | 0.604±0.133 | 0.575±0.141 | 0.653±0.121 | 0.668±0.120 | **0.687±0.139** | 0.038±0.206 | -0.541±0.141 |
| | BMH/CPH | 0.296±0.184 | 0.275±0.186 | 0.408±0.167 | 0.372±0.172 | 0.194±0.193 | -0.009±0.229 | **0.663±0.140** | -0.442±0.161 |
| | E | -0.762±0.079 | -0.712±0.095 | -0.849±0.054 | -0.858±0.051 | -0.733±0.088 | -0.601±0.151 | -0.492±0.152 | **0.912±0.032** |

| | Phenotypic | W | SL | BHP | BMH | CPH | BMH/BHP | BMH/CPH | E |
|---|---|---|---|---|---|---|---|---|---|
| 400 d | W | **0.786±0.018** | 0.765±0.019 | 0.778±0.021 | 0.783±0.019 | 0.740±0.021 | 0.337±0.031 | 0.015±0.044 | -0.474±0.046 |
| | SL | 0.791±0.018 | **0.790±0.017** | 0.781±0.020 | 0.787±0.019 | 0.746±0.020 | 0.358±0.028 | 0.013±0.039 | -0.430±0.046 |
| | BHP | 0.798±0.017 | 0.777±0.019 | **0.809±0.017** | 0.809±0.016 | 0.752±0.020 | 0.328±0.032 | 0.059±0.044 | -0.536±0.041 |
| | BMH | 0.798±0.018 | 0.796±0.017 | 0.803±0.018 | **0.810±0.017** | 0.751±0.020 | 0.354±0.031 | 0.056±0.044 | -0.533±0.042 |
| | CPH | 0.761±0.020 | 0.765±0.020 | 0.748±0.022 | 0.764±0.020 | **0.739±0.021** | 0.336±0.031 | -0.036±0.042 | -0.464±0.045 |
| | BMH/BHP | 0.478±0.030 | 0.474±0.027 | 0.313±0.031 | 0.496±0.029 | 0.458±0.030 | **0.385±0.026** | 0.032±0.035 | -0.362±0.038 |
| | BMH/CPH | 0.081±0.043 | 0.072±0.041 | 0.130±0.044 | 0.118±0.043 | -0.001±0.042 | 0.005±0.031 | **0.306±0.031** | -0.197±0.045 |
| | E | -0.615±0.035 | -0.521±0.037 | -0.672±0.029 | 0.001±0.028 | -0.574±0.035 | -0.274±0.036 | -0.175±0.047 | **0.797±0.021** |

# CHAPTER 7

# GENERAL DISCUSSION

The general strategy pursued in this thesis is depicted in Figure 26. Firstly, a high-density SNP genetic map was built from ddRAD sequencing that was also used to study recombination rate landscape between sexes. Secondly, a *de novo* hybrid genome assembly was generated using a combination of Illumina and Nanopore long-reads and main features were determined. Finally, both the genetic and physical maps were anchored and integrated into 21 linkage groups (SseLGs) and genomic information sorted in pseudochromosomes. The new integrated physical map represents a valuable genetic tool that was further used for an association study using familial information to identify sex-associated markers. In addition to SNP markers, this integrated map was used for searching and identifying SSR markers distributed through the genome. A total of 108 markers were validated as polymorphic and 106 were combined in thirteen SSR multiplex assays suitable for genetic studies. Four supermultiplex of high variability were optimized to be used routinely in pedigree analysis in breeding programs. The identified SSR markers together with other SSRs previously described in sole (Molina-Luzon *et al.* 2015a) were also used to generate an integrated SSR genetic map with 229 markers. Both maps generated in this thesis were used to carry out a synteny analysis with other flatfish species and decipher some clues about the chromosome evolution between flatfish.

Using the genomic information generated in this thesis, a low-density DNA chip with 49 polymorphic assays was validated and used for the identification of growth-associated markers. Moreover, SSR multiplex PCR assays were used to investigate genetic estimates for growth and morphology quality related traits. The results obtained in this thesis and discussed below support effective breeding programmes can be successfully applied in Senegalese sole to boost industrial production. The discussion will be structured according with main papers published in this thesis and rearranged according to the general structure of this document.

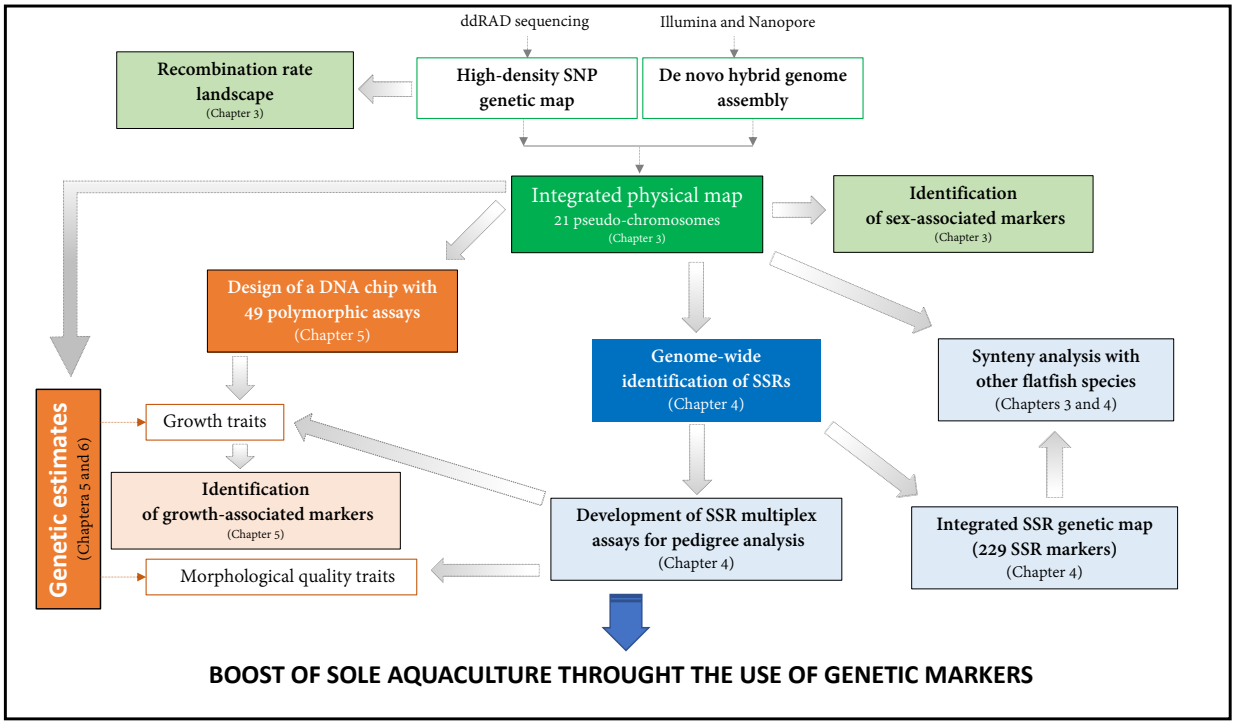


**Figure 26**. General scheme of the current PhD thesis.

## 7.1- Genome assembly and main features in Senegalese sole (*Solea senegalensis*)

Genome assemblies and genetic linkage maps provide complementary information that can be integrated to produce high-quality physical maps. The resulting accurate chromosome assemblies are suitable to investigate genome evolution and species diversification, the genetic architecture of QTLs and the regulation of targeted genome regions. In this PhD thesis, a *de novo* hybrid assembly for a male (chapter 3) and female (chapter 3 and 4) soles and a high-density SNP map (chapter 3) were generated and combined to provide a polished draft assembly of 21 pseudo-chromosomes, as described in chapter 3. A fragmented genome for a female sole was previously reported (Manchado *et al.* 2016) (N50=85 kb, 600.3 MB long). This assembly was improved in this thesis by integrating Nanopore and Illumina reads, resulting in 5,748 contigs with N50 = 339.9 kb and 608 Mb long as previously reported. In the case of the male assembly, a lower number of contigs (3,403) was obtained with a higher N50 (512.7 kb) and confirmed that the genome size of sole is around 609 Mb. This genome size is similar or even a bit larger than other flatfish (Chen et al. 2014b; Figueras et al. 2016; Shao *et al.* 2017; Xu *et al.* 2020). A dot-plot alignment analysis indicated a high similarity between male and female genome assemblies perfected aligned along

the diagonal (Figure 2) with a completeness similar to other high-quality fish assemblies (>95.5% complete genes) (Ge et al. 2019; Yang et al. 2019; Xu *et al.* 2020).

Male genome characterization identified 50,133 transcripts and 27,175 protein-coding that agrees with the number of predicted transcripts in a recently assembled informative transcriptome (Cordoba-Caballero *et al.* 2020). Moreover, a small subset of unique genes was identified in both sexes with a high overrepresentation of cell-cycle regulation and regulation of transcription categories (including mainly the histones H3.2 and H4) in the female. In mammals, unique histone variants are specifically expressed in spermatogenic cells (Hoghoughi et al. 2018). Moreover, expansion of histone multigene clusters in scleractinians was associated with sexually dimorphic expression of some variants playing a role in the control of gene expression in female and male germ cells during gametogenesis (Chiu et al. 2020). In sole, at least two loci of canonical histones in the largest metacentric chromosome SseLG1 linked to dmrt1, a key determination gene in other flatfish, were reported in sole (Chen *et al.* 2014b; Portela-Bens *et al.* 2017). This chromosome arose after a Robertsonian fusion and intense reorganization events (Garcia-Angulo *et al.* 2018) that could have birth to new histone clusters under purifying selection (Rooney et al. 2002). Although we cannot exclude that some differences in the number of histone copies between both genomes could be attributed to individual variation, one plausible hypothesis is that some of these histone clusters could have subfunctionalizated and acquired a role in gametogenesis in a sex-specific manner. This hypothesis is supported by the identification of a rxra-like receptor also represented in such GO categories able to mediate the masculinizing effects of females mediated by its ligand TBT in rockfish females (Zhang et al. 2013).

## 7.2- Genetic and physical map anchoring, and *in silico* comparisons of integrated map.

*De novo* assembled male genome was used as reference to map the ddRAD sequences and construct a high-density genetic map. The sole consensus map size and the number of high-quality SNP markers used (Figure 3; Table 4) were similar to those reported for turbot (2,622.09 cM) (Wang *et al.* 2015b) and flounder (3,497.29 cM) (Shao et al. 2015) although with a higher density of markers (only 6,647 and 12,712 SNPs in turbot and flounder, respectively). Most importantly, markers were distributed into 21 SseLGs (3 metacentric pairs, 2 submetacentric pairs, 4 subtelocentric pairs and 12 acrocentric pairs) (Merlo et al. 2017), that match with the haploid karyotype (2n = 42) of the species (Vega et al. 2002). Until now, a low density genetic maps with

129 microsatellites were reported in Senegalese sole (Molina-Luzon *et al.* 2015a; Guerrero-Cozar *et al.* 2020) Moreover, a cytogenetic map was also published although the number of BACs did not still cover all chromosomes (Portela-Bens *et al.* 2017; Garcia *et al.* 2019). This new high-density SNP map (Figure 3) thus represents a key step forward for future genomic studies and QTL identification with respect the current information available until now in this species.

Although hybrid assemblies using long and short sequences reads reduce genome fragmentation and increase the average scaffold sizes as observed in this study, most of *de novo* genome assemblies still do not reach chromosome-level with the expected number of chromosomes due to, among other factors, the repetitive fraction of the genome. To get around this limitation, information of genome-wide physical maps and dense genetic linkage maps can be integrated to assign chromosomal locations to sequence contigs (Mascher & Stein 2014). This anchoring can also remove assembly artifacts and position misplaced scaffolds to increase the contiguity of the assembled scaffolds. In chapter 3, the high-density SNP genetic map was used to anchor, sort and refine the assembled contigs. Overall, 89.9% of the genome assembly could be anchored to 21 pseudo-chromosomes and a total of 102 contigs were removed or split to separate positions in SseLGs. A similar strategy was followed in turbot using 31 families that allowed for the rearrangement of 20% of the genome assembly (Maroso *et al.* 2018). A comparison between male and female demonstrated a high co-linearity between our physical map and female scaffolds (only 5.53% mismatch). Although 10.1% of genome information remained as unplaced, the anchored physical map is essential for gene association analysis, synteny and cross-species studies and targeted genome resequencing. Further studies will be required to accurately anchor the remaining 61.9 Mb unanchored regions to their position in the genome. We exploited the information from this integrated physical map for the design of a low-density chip for Senegalese sole containing 60 probes for SNPs, of which only 49 polymorphic probes remained after validation with wild sole distributed in 17 SseLG as mentioned in chapter 5, obtaining a versatile and useful tool for QTL identification.

## 7.3- Recombination rates and sex-associated markers

In addition, the provision of a high-density SNP genetic map anchored to the physical map allowed us to carry out a recombination rate study and a genome-wide association analysis of sex-linked SNPs in chapter 3. It is well-known that the genome-wide RR differs between males and females (heterochiasmy) and that the recombination landscape also varies along chromosomes. In animals and plants, females tend to have higher RR than males, which in turn result in larger map lengths

(Stapley et al. 2017; Sardell et al. 2018; Sardell & Kirkpatrick 2020). In our study, map was longer in the female than in the male (2,698.4 vs 2,036.6 cM; ratio 1.32). Assessment of sex-specific RR indicated a female-biased heterochiasmy across all SseLGs, with an average RR of 3.02 in male vs 4.51 cM/Mb in female. Four species of Pleuronectidae also exhibited wide heterochiasmy through all chromosomes similarly to sole with some intervals of male- and female-restricted meiotic recombination (Edvardsen et al. 2020). However, such differences in RR between males and females are not fully conserved in flatfish when map size is considered. Female maps are larger in turbot (1.36 times) and halibut (1.07 times) (Bouza *et al.* 2007; Reid *et al.* 2007; Ruan et al. 2010), this is not the case of flounder or tongue sole with slightly larger maps in males (1.03-1.09 times) (Castaño-Sanchez et al. 2010; Song et al. 2012a; Shao *et al.* 2015). *C. semilaevi*s is the only flatfish known with heteromorphic sex chromosomes (ZZ/ZW) that has been described in several mammals, birds and insects as a cause for an arrest of recombination in the heterogametic sex (XY males or ZW females). This could explain a shift in the direction of heterochiasmy (Stapley *et al.* 2017).

In addition to such differences in overall RR between sexes, the chromosomal recombination landscapes also differed between male and female according to typical patterns. In fish, it has been shown that recombination occurs at higher frequencies near telomeres in males while the distribution is quite more uniform or elevated near centromeres in females (Sardell & Kirkpatrick 2020). In stickleback fish, it has been demonstrated that centromeres and telomeres have little or no effect on recombination in females, however, in males, the recombination rates are suppressed near the centromeres and hence crossovers localize mainly at the ends of long arms in acrocentric chromosomes (Sardell *et al.* 2018). This feature seems to be conserved in sole since RR were also more frequent toward the end of males SseLGs compared to females (Figure 5).

Heterochiasmy is considered a major force that guides the evolution of genetic sex determination systems and speciation (Kitano et al. 2009; Edvardsen *et al.* 2020). Normally, genome regions with very low RR are associated with sex-determining regions in young sex chromosome systems and sex-linked traits such as pigmentation (Wright et al. 2017). In Atlantic halibut, the sex determining gene gsdf is located in a region of chromosome 13 with restricted male and female RR (Edvardsen *et al.* 2020). In *S. senegalensis*, 30 significant sex-associated SNPs (66 if we consider the SNPs of separated families) were distributed throughout the SseLG18 with very low RR hot region (Figure 6). The shift and crossing between male and female RR suggest sex-specific

restricted meiotic recombination events and that heterochiasmy might be involved in nascent sex chromosome system. (Chapter 3)

Most of SNP markers in the whole-population were heterozygous in males suggesting an XX/XY system. However, it should be noted high levels of incomplete penetrance in the families analysed. The fact that this proportion was even inverted in specific markers of F4 indicates a high effect of environmental factors on sex determination. The temperature seems to be a major factor that modifies sex ratios during larval development generating skewed populations of neomales and neofemales (Blanco-Vives *et al.* 2011; Viñas et al. 2012). Familial sex ratios in sole reported in chapter 5 to oscillate from 16 up to 90% males supporting a high impact of environmental factors to modulate sex differentiation and sex population ratios (Guerrero-Cozar *et al.* 2021).

After analyzing the hot region in SseLG18, the fshr appeared as a putative candidate for sex determination. The fshr locus was recently associated with male sex in flatfhead grey mullet with an incomplete penetrance as observed in sole (Ferraresso et al. 2021). These authors proposed that fshr might act as a proxy for the genetic transduction of environmental factors such as temperature Under this hypothesis, sex determination would not rely on a single genetic cascade but a continuum of environmental and genetic factors. In sole, fshr was mainly expressed in testis (Chauvigne et al. 2010). The Fshr together with StAR are expressed in the steroidogenic Leydig cells and Fshr act as a promiscuous receptor that mediates the steroidogenic activity induced by both FSH and LH (Chauvigne et al. 2012; Chauvigne et al. 2014). This double action supports a prolonged spermatogenesis and spermatid availability within the testis throughout the year mediated by FSH and the differentiation of spermatids into spermatozoa and subsequent spermiation mediated by LH (Chauvigne *et al.* 2012). Functional studies are needed to validate this putative candidate.

## 7.4- Identification of SSR markers through the genome and design of Multiplex PCR assays

In chapter 4, we took advantage of a 85k genome draft (Manchado *et al.* 2016) and the *de novo* female and male hybrid genomes done in this thesis to identify and characterize SSR markers through the genome. The SSRs are highly abundant in the genome of vertebrates although their use has been limited by the knowledge of flanking regions suitable for primer design. Some authors considered as alternative the cross-species amplification of highly conserved SSRs (Funes et al. 2004; Castro et al. 2006; Molina-Luzon *et al.* 2015a). Recently, a study in Senegalese sole based

on the 1.1% of the genome information estimated a high density of SSRs (675 per Mb) with dinucleotide SSRs representing overall 59.7% (Garcia *et al.* 2019). This assembly had a high-quality gene representativity (completeness was 96.2% similar to previous flatfish assemblies) (Xu *et al.* 2020) with the marker density of 886.7 SSRs per megabase. Previous cytogenetic analyses demonstrated that most of di- and tetranucleotides appear widely distributed in subtelomeric position of metacentric, submetacentric and acrocentric chromosomes (Garcia *et al.* 2019) and hence both of them were considered suitable for primer design and multiplex amplification in this thesis.

Whole-genome mapping requires high-throughput strategies to save consumables, labour costs and reduce the processing and analysis times. PCR multiplex assays have been successfully developed in seabream (Negrin-Baez et al. 2015; Negrin-Baez *et al.* 2016) and grapevine (Zarouri *et al.* 2015) for QTLs identification and pedigree reconstruction. In chapter 4 of this thesis, thirteen PCR multiplex assays comprising 108 markers widespread in the genome were optimized. Although previous studies in sole have reported microsatellite markers derived from EST or SSR-enriched libraries (Funes *et al.* 2004; Chen et al. 2008; Molina-Luzon et al. 2012) only three of them considered SSR multiplexing (from 4 to 8-pex) (Castro *et al.* 2006; Porta *et al.* 2006; De La Herran et al. 2008).

Tetra- and pentanucleotides predicted motifs were initially selected for multiplex PCRs although finally some of them (12%) followed a dinucleotide allelic series. It has been demonstrated that SSRs with dinucleotide motifs have a higher variability but more prone to genotyping errors than those with larger motifs (Nater *et al.* 2009; Zalapa et al. 2012)., was observed that the average number of alleles per locus was 10.9 ranging from 2 to 43 in accordance with previous SSR markers in Senegalese sole (Funes *et al.* 2004; Chen *et al.* 2008; Molina-Luzon *et al.* 2012; Molina-Luzon *et al.* 2015a). As expected, the dinucleotide markers showed a higher variability (average PIC 0.84) than tetra- (0.65) and pentanucleotides (0.66). Moreover, scoring accuracy was estimated using a standardized methodology to identify potential errors in the electropherograms (Lee-Montero *et al.* 2013) indicating only a small set of markers (17) with stuttering, allele dropout or intermediate alleles, ~16% of total markers. In seabream, the percentage of loci with some of these errors was similar although with higher rates of intermediate alleles (Lee-Montero *et al.* 2013). It should be indicated that stutter peaks have a low effect to assign loci size in tetranucleotides as observed by a double validation across two independent labs reaching similar values in genetic diversity parameters.

The use of genetic tools to infer genealogies is a demand for genetic breeding programs in mass-spawning species such as Senegalese sole. Due to the economic value of these species, the optimization of genotyping tools for parental assignment in a feasible, accurate and cost-effective way is a requirement. Moreover, the loss in variability that occurs in subsequent selection cycles makes necessary a minimal number of markers to sustain the program through some generations. Both the number of loci and their heterozygosity level may influence the power of markers for parentage exclusion approaches (Labuschagne et al. 2015). In chapter 4, a total of 40 high variable and genome widespread markers were selected according to PIC and combined in four supermultiplex (7 to 11-pex). Assignment simulations indicated that a subset of 7, 5, 4 and 3 markers were able to assign 99% offspring with SMA (11-pex), SMB (11-pex), SMC (8-pex) or SMD (10-pex), respectively. Moreover, a real testing using SMA to genotype 92 parents accurately allocated all 100 parent–offspring relationships. All these data indicate that these supermultiplex can be transferred to the industry as standards for pedigree reconstruction to support a long-term use for genetic breeding selection.

## 7.5- Integration of SSR markers

In addition to their use in the design of new multiplex PCRs, the 108 SSRs identified in this thesis along with 121 SSRs previously published in Senegalese sole (Molina-Luzon *et al.* 2015a), were used to build a new integrated genetic map with 229 SSR markers. Using the high-density SNP genetic map as reference, the whole set of SSR markers were distributed in 21 LGs. This new SSR genetic map improve the current low density genetic linkage map available for this species (Molina-Luzon *et al.* 2015a) and confirmed that the LGs from the previous genetic map clustered perfectly within the SseLGs after anchoring the LR-hybrid male genome and the high density genetic map (Figure 12 and Table 8). Only LG1 was split into two SseLGs that might be due to an error in the consensus between gynogenetic families. The new genome information provided facilitates the integration with SNP markers and the redesign of some SSR primers in the map to construct new multiplexes that improve the genome coverage.

## 7.6- Flatfish Synteny analysis

Both the integrated SSR and anchored physical maps represented key genomic tools very useful for the cross-species synteny and genomic comparison studies carried out in chapters 3 and 4. Flatfish genome comparisons have demonstrated a high degree of conservation at macrosynteny level (Bouza et al. 2012; Maroso *et al.* 2018; Garcia-Angulo et al. 2019), moreover, chromosome

fusions and translocations have occurred frequently during flatfish evolution shaping the number of chromosomes from n = 24 pairs in Japanese flounder to n = 20 autosome pairs and one sexual chromosome pair in *C. semilaevis*. A synteny comparison of SseLGs with different flatfish genomes indicated that there was a one-to-one correspondence for 15 chromosomes, with some lineage-specific rearrangements (Figure 8), in our data, deviations from diagonal unlike in the comparison between male and female are indicative of this intense internal reorganization across species. The three SseLGs (SseLG1, SseLG2 and SseLG3) deserve special attention as they can provide an evolutionary framework to understand the history of chromosome fusions and fissions that shaped the karyotypes in flatfish. The SseLG1, predicted as a metacentric chromosome by the analysis of recombination frequency (Figure 6), was previously identified by cross-species genomic comparison as the largest metacentric chromosome in Senegalese sole suggesting it may be a proto-sexual chromosome (Portela-Bens *et al.* 2017; Garcia-Angulo *et al.* 2018). Our data support the hypothesis that this chromosome has primarily emerged by a lineage-specific Robertsonian fusion, since the homologs in other flatfish maintained their integrity across evolution. A complex series of events including small chromosomal translocations and rearrangements, fusions, and pericentric inversions would explain the current gene content and organization (Garcia-Angulo *et al.* 2018). Unlike SseLG1, the SseLG2 and SseLG3 contain those chromosomes whose remodelling have shaped the karyotypes in flatfish from n=24 in *P. olivaceus* to 22 *S. maximus* and 21 in *S. senegalensis* and *C. semilaevis*. A fusion model envisaged suggests a small number of chromosomes in the older lineage Paralichthyidae (9,14 and 16) (Shi *et al.* 2018) that combined with other chromosomes in a lineage-specific way could explain the major rearrangement events that shaped the karyotype in this species. Most interestingly, the high remodelling of sexual ZW chromosomes (part of the Z chromosome was syntenic with SseLG05 (table 8) and the remaining ZW sequences were found spread throughout the genome), that was also previously assessed by a scaffold mapping strategy (Manchado 2019) suggests that a shift in the sex determining system might have occurred in Senegalese sole. In fact, a sex determination XX-XY system was proposed in this thesis and other previous studies (Viñas *et al.* 2012; Manchado 2019) with the female as homogametic sex. Although the SseLG01 has been proposed as a sex proto-chromosome due to the location of some key sex-determining genes and repetitive sequences (Portela-Bens *et al.* 2017; Rodriguez *et al.* 2019), the spreading of Z/W sequences through the genome as seen in chapter 4 and the identification of 30 sex-linked markers within a

hot region with low recombination rates in SseLG18 indicate that a further experimental validation is required to identify a putative major locus for sex determination.

## 7.7- Genetic estimates for growth and morphology quality

The availability of the different tools developed during this thesis allowed us to study in chapter 5 and 6, the genetic determination of commercially important traits such as growth and morphology quality in Senegalese sole. Genetic research in Senegalese sole has been hampered until now by the lack of full-control on reproduction success. Nevertheless, the recent advances in thermoperiod manipulation for spawning synchronization have made feasible the design of genetic programs (Martin *et al.* 2019). In addition to reproductive success, an important aspect is that this highly plastic taxonomic group transforms during development from a bilateral symmetry to an asymmetric high-specialized flatten body. Evolutionary studies have demonstrated that different ecological traits act as a driver of body shape in flatfish acquiring a wide range of body depth, jaw length and fin length (Black & Berendzen 2020). Hence, flatfish families can be identified by specific shapes and morphological features that should be carefully preserved in aquaculture to maintain consumer acceptance and commercial value. In the case of Senegalese sole, body shape is expected to be highly elliptic and lanceolate with short jaws and long dorsal and anal fins that contrast with the shape of most pleuronectids or scophthalmids with deeper bodies, longer jaws and short dorsal and anal fins. However, several reports that deal with morphological traits in Senegalese sole in aquaculture reported high rates of malformations that in most cases have not a severe impact on external gross morphology (Dionisio *et al.* 2012; Losada *et al.* 2014; de Azevedo *et al.* 2017; de Azevedo et al. 2019b).

In chapter 5 and 6, we have genetically evaluated a commercial broodstock distributed in nine tanks of Senegalese sole using a mass spawning methodology followed by molecular pedigree reconstruction as previously reported for other marine species (Navarro *et al.* 2009; Garcia-Celdran *et al.* 2015c; Lee-Montero *et al.* 2015), and investigated for first time in Senegalese sole the phenotypic and genetic variation associated with growth and morphology-related traits under industrial conditions in RAS. The evaluation was carried out before entering in RAS (~400 d) and at harvest (~800 d). This period appears as critical in sole production since RAS is a technology very different of natural ponds in which soles inhabit and a genotype×environment interaction was previously demonstrated in the close species *S. solea* (Mas-Muñoz *et al.* 2013). Although tagging, stocking density, water temperature, type of tanks or feed were common to all EBs, this factor had

an important effect on growth and shape-related traits after the RAS growth-out phase. This is in line with the results obtained in chapter 6, in which a longitudinal analysis (since the same subset of tagged soles was analysed at both ages) was carried out showed different tendencies in RAS even between EBs with very similar genetic structure and age at sampling suggesting that some additional factors such as social interactions or differences in the actual flow-through dynamics could also play a key role in the evaluated traits. In addition to the EB, gender also had an important effect on growth and shape-related traits.

In Senegalese sole, females grow faster than males and significant differences in weight are observable from young juveniles (females 13.6% heavier than males) (Carballo *et al.* 2018) to harvest size (19-32% heavier). These differences are even more evident at high stocking densities (Sanchez *et al.* 2010). In the study carried out in chapter 5, females were significantly 8.2% heavier than males at the beginning of RAS and 22.7% at harvest. In spite of the interest of cultivating female-enriched populations, sole populations cultured under standard production conditions at 20 ºC are normally skewed toward males that usually represent around 60-67% of whole populations (Sanchez *et al.* 2010; Blanco-Vives *et al.* 2011; Viñas *et al.* 2012; Carballo *et al.* 2018). This skewed abundance of males in sole has been associated with epigenetic effects mediated by environmental temperature that induces masculinization (Blanco-Vives *et al.* 2011; Viñas *et al.* 2012). In the study, in which a high number of families is represented, the average percentage of males in the population was 55.3% (ranging from 43.4 to 69.1% between batches). Larval rearing was carried out under commercial conditions using constant temperature (~20ºC) that could explain this slightly higher proportion of males especially in some batches (1, 2, 3 and 7). However, it should be noted a high variation in sex ratios associated with genetic families (% males ranging from 16 up to 90%; Figure 15). It was striking that when maternal half-sibs were compared, all of them contributed a higher proportion of males in similar percentages. In contrast, when paternal half-sibs were compared (mainly from batches 4-6), the families could be enriched in females or males. These data indicate that epigenetic effects act differently on fathers than on mothers and that males could be the heterogametic sex able to skew population toward neomales or neofemales. These data together with the results obtained in chapter 3 are agree with XX-XY sex determining system proposed in *S. senegalens*is using gynogenetic families although modulated by other genetic or environmental factors, such autosomal genes or temperature (Molina-Luzon et al. 2015c).

Multiplex assays assigned 98.1% offspring (2,171 specimens) to a single parent pair supporting the high assignation rates previously mentioned in chapter 4. Heritabilities for all growth traits were higher at 400 d than 800 d and values ranged between 0.568 and 0.643 at 400 d and between 0.424 and 0.500 at 800 d (Table 10). These heritabilities were higher than those observed in *S. solea* cultivated in RAS at harvest (0.23-0.25) (Blonk *et al.* 2010b; Blonk *et al.* 2010c) and in gilthead seabream (0.34-0.40 at 509-689d) (Navarro *et al.* 2009; Lee-Montero *et al.* 2015). The fast growth rates of *S. senegalensis* and the adaptation to high stocking densities and handling under different production systems (flow-throw or RAS) explain these high genetic estimates. The slight reduction in heritabilities observed from 400 to 800 d could be due to sexual maturation effects that mask the effects associated with growth potential. (Dupont-Nivet et al. 2010). Interestingly the high genetic correlation between 400-800 d for growth traits support that growth parameters estimated in juvenile stages before RAS could be used as a good predictor of growth performance later at harvest.

Genetic correlations between all growth traits were very high and positive. High correlations between weight, length and width are routinely reported (Vandeputte et al. 2008; Navarro *et al.* 2009; Blonk *et al.* 2010c; Lee-Montero *et al.* 2015). It should be noted that in this study the body width was the trait with the highest heritability at both ages. This is important since this trait has very high genetic correlations with other growth traits, it is important to control shape quality, and it can be measured during fieldwork conditions or from images, hence, this could be a good candidate to be used for genetic selection as alternative to weight in Senegalese sole breeding programs. In the case of gilthead seabream, length was also suggested as more adequate than weight due to higher heritabilities and lower coefficients of variation (Navarro *et al.* 2009; Lee-Montero *et al.* 2015). It should be note that genetic correlations between ages at 400 and 800 d were still high. Normally, genetic correlations between growth traits at different ages are usually low (0.3-0.5) when long time periods are considered (hatchery ~120-150 d vs harvest ~500-800 d) (Vandeputte *et al.* 2008; Navarro *et al.* 2009) and they increase to ~0.8 or more when the time gap is smaller as occurs in this study (Vandeputte *et al.* 2008; Navarro *et al.* 2009; Lee-Montero *et al.* 2015). Interestingly, genetic estimates and standard errors for SL, width and total area measured directly on fish and after image analysis were almost identical for equivalent traits confirming that the image analysis was a feasible approach to measure these growth traits reducing the times during sampling and fish stress (Blonk *et al.* 2010c; Navarro *et al.* 2016).

In the morphological study reported in chapter 6, we demonstrate that females are less elliptic than males even after correcting by weight. These differences were more evident at harvest probably due the ovary maturation increasing the abdominal cavity that in turn reduces ellipticity. A regression analysis between heights also evidenced a small change in the slopes by gender that was not clearly observable when height ratios were analysed indicating that compensatory mechanisms could modify the relative body proportions. In addition to gender effects, the presence of amoebic nodules in liver or intestine at harvest also influenced growth and shape-related traits. This parasite accumulates mainly in the intestinal mucosa and later spreads to some different tissues (Constenla & Padros 2010). Although mortality is scarce, this study demonstrate that non-infected fish was 44.9% heavier than infected fish. Moreover, the infected fish were slightly less elliptic at harvest even after correcting by weight and changed the ratio BMH/BHP due to the excess of nodules that in some cases distort the size of abdominal cavity.

The ellipticity of the sagittal plan was proposed as the best trait to measure shape quality in sole since this trait could be easily derived from direct measures on fish (using body height and body length) or by image analysis fitting theoretical ellipses with a similar precision (Blonk *et al.* 2010c). This theoretical assumption is based on the expected elongated body shape of soleids that differentiate from pleuronectids or scophthalmids. Although this trait is highly influenced by body size and bigger fish tend to be rounder, the ellipse fitting still remains as a good predictor of shape for soles. Our ellipticity data confirmed a major effect of weight on ellipticity distribution with bigger values at 400 than 800 d and a progressively reduction with bigger weight class sizes (Figure 24). Moreover, as indicated above, females were rounder than males ever after correcting by weight due to the increase of size abdominal for sexual maturation. In yellowtail flounder, females had relatively deeper abdomens and larger heads than males (Cadrin & Silva 2005). However, these differences associated with gender were not observed in *S. solea* although these authors did not follow a longitudinal approach or provide information about gonad development that could explain such differences. The significant effects of EB conditions as indicated above on the ellipticity trajectory also denote the importance of culture conditions on shape and the relevance to control this important feature to maintain high shape-quality standards for fish commercialization.

Shape predictors such as ellipticity had very high heritabilities (>0.74) at both ages. These values were quite higher than those obtained in *S. solea* (0.34) (Blonk *et al.* 2010c) or *Oreochromis niloticus* 0.12-0.45 (Omasaki et al. 2016; Mengistu et al. 2020). Although some nutritional,

management and culture conditions were reported as regulators of meristic characters and malformation rates in sole (Ambrosio *et al.* 2008; Fernández *et al.* 2009; Dionisio *et al.* 2012), our results indicate a high additive genetic component on the external shape-related traits evaluated in this study. Although malformations could exist (they were not evaluated in this study), most of them would have a low impact on gross morphology as previously indicated (de Azevedo *et al.* 2017) evidencing a high genetic component for ellipticity. Nevertheless, further studies are required to associate the shape-traits with the skeletal characteristics in order to understand the main causes behind the ellipticity range. Moreover, the high genetic correlations (0.91) between both ages confirm that those genetic factors controlling shape are already acting in juveniles and hence selection could also carry out in juveniles.

It should be noted that ellipticity was dependent on fish size. In *S. solea*, a moderate negative genetic correlation (rg = -0.44) between ellipticity and weight was reported (Blonk *et al.* 2010c). Similarly, in chapter 6, a high and negative genetic correlation between both traits at 400 (-0.768) and 800 d (-0.608) was determined indicating that fish reaching a bigger size were also rounder. These negative correlations should be carefully considered if selection for increased weight at harvest is carried since less elliptic fish will be produced. Since most of soles are sold in fresh markets, ellipticity was proposed as a correction factor for weight-targeted selection breeding programs (Blonk *et al.* 2010c). A combined selection index setting a zero change in shape reduced 9.9-13.8% the response to harvest weight that is assumable to preserve a high-quality shape standard (Blonk *et al.* 2010c). However, no correction would be necessary if finally, industry moves toward transformed seafood products that it is one of the most promising markets for flatfish.

The three heights showed a positive and very high genetic correlation with growth traits (>0.95). However, the heritabilities for the two height ratios were low-moderate and the BMH/CPH had very low genetic correlations with growth and height traits. This latter ratio is strongly related to the swimming speed and performance (Fisher & Hogan 2007; Assumpção et al. 2012). A deep caudal peduncle provides the fish a superior ability to accelerate and power for propulsion allowing it to reach a high swimming speed and efficiency (Assumpção *et al.* 2012). Soles are usually very sedentary in the tanks and they do not require high water columns since they are passive feeders in the tank bottom. So, it is not expected to exist a high selection pressure on swimming efficiency in the RAS although fish should adapt to water currents in the tanks. However, this ratio seems to be useful to refine a lanceolate shape toward a more theoretical elliptic one that fits better to the

sole body structure. The peduncle is usually considered as the caudal reference point for body length since caudal fin is highly variable in size and morphology and the BMH/CPH increases with age (Figure 23). High pronounced ratios (due to higher BMH or lower CPH) are associated with very high lanceolate shapes that deviate from symmetrical body ellipse giving arise to turbot-like morphologies. The low heritability for this trait could be due the benthic way of life and the sensitive of caudal complex to traumatisms and malformations that in turn can remodel the peduncle. The low genetic correlations with other ellipticity and growth traits indicate that this trait can provide new relevant information to genetic selection index to preserve a sole high-quality shape.

## 7.8- Design of low-density DNA chip and identification of growth-associated markers.

The study of genetic estimates for growth traits was completed with the identification of genetic SNP markers significantly associated with these traits using the low-density chip designed in chapter 5. We exploited the information from several genomic resources available in Senegalese sole for the design of a low-density chip with 49 markers and the study in 4 families with highly different breeding values at harvest. Although the arrays were originally intended to identify sex-related markers and most of the markers were located in sexual chromosomes ZW and chromosome 14 of *C. semilaevis*, we failed to find any association with sex in Senegalese sole.

Two markers were significantly associated with adjusted growth traits at both 400 and 800 d. Growth is a polygenic trait controlled by many genes spread through the chromosomes involved in cell growth, cell proliferation, cell cycle, lipid metabolism, proteolytic activities, chromatin modification, and developmental processes (Ali et al. 2020). The Sosen1_s1980 is located in the general transcription factor 3C polypeptide 4 (gtf3c4, also known as TFIIIC90). This gene is responsible for the recruitment of RNA polymerase III and initiating of tRNA transcription in eukaryotes and it has also been reported to possess histone acetyltransferase activity (HAT) in vitro (Hsieh et al. 1999; Kundu et al. 1999; Trisciuoglio et al. 2018). This HAT activity is important for chromatin relaxation and activation of gene transcription. In rat the gtf3c4 is located within a region where localize a QTL rat associated with body weight (Casiro et al. 2017). Several SNPs associated with body weight gain in rainbow trout were located in genes related to chromatin modification, and developmental processes (Ali *et al.* 2020). Moreover, HAT genes were potentially involved in cotton growth and development, fiber-related traits, and plant response to

the environment (Imran et al. 2019). The other significant marker is the Sosen1_s0233 encodes for mitochondrial fission process protein 1 (MTFP1, also known as MTP18) that plays an essential role for maintaining mitochondrial integrity and it is highly expressed in organs enriched with mitochondria such as heart and skeletal muscles and whose regulation of mitochondrial physiology is essential for maintenance of muscle mass and function (Tondera et al. 2004; Aung et al. 2017).

The results obtained in this thesis provide a source of genomic information and valuable genetic tools with which to design effective genetic selection programmes to boost the production of aquaculture sole.

# CHAPTER 8

# CONCLUSIONS

1. A high-density SNP genetic map and a de novo sole genome assembly were generated. Both genetic and physical map were anchored to generate 21 pseudochromosomes matching with the expected chromosome number of this species. *In silico* analysis demonstrated that females have a larger genetic map due to differences in recombination rates observing chromosome-specific recombination landscapes. A comparative genome analysis between different flatfishes showed a high conservation of chromosomal synteny and identified lineage-specific Robertsonian fusions and several other rearrangements that explain changes in chromosome number in the karyotype through flatfish evolution.

2. A genome-wide association study identified 30 sex-linked markers within a hot region with low recombination rates in SseLG18. However, an incomplete penetrance of sex markers with males as the heterogametic sex was determined. The follicle-stimulating hormone receptor gene was identified as a putative candidate for sex determination in sole.

3. A wide set of 108 SSR polymorphic markers were identified by *in silico* analysis. They were widely distributed throughout the Senegalese sole genome. For amplification, thirteen multiplex PCR assays (with up to 10-plex) were designed and the amplification conditions were optimized with a high-quality score. A subset of 40 high polymorphic markers were selected to optimize four supermultiplex PCRs (with up to 11-plex) for pedigree analysis. Theoretical exclusion probabilities and real parentage allocation tests using parent–offspring information confirmed their robustness and effectiveness for parental assignment.

4. A new integrated genetic map containing 229 SSRs distributed in 21 SseLGs was created by *in silico* genome analysis. Predicted positioning of SSR markers within the genome highly matched with a published genetic map. Synteny analysis with respect to other flatfish identified new fusions and rearrangements of telocentric chromosomes that gave arise to metacentric and submetacentric chromosomes in Senegalese sole.

5. High heritabilities were estimated for growth traits in Senegalese sole at two important stages in the production cycle. Moreover, high genetic correlations were determined between growth traits and sampling ages. Phenotypic data confirmed that females grew faster than males

existing a high variation of sex ratios by family. All these data are highly relevant to improve growth in aquaculture industry and indicate epigenetic effects acting on the offspring

6. A low-density array was designed and validated containing 49 probes for SNPs distributed in 17 SseLGs. A pilot study using this DNA array to identify markers associated with growth traits at 400d and 800 d in fast- and slow-growing families identified two significant markers in the general transcription factor 3C polypeptide 4 and the mitochondrial fission process protein 1.

7. Genetic estimates for morphology quality traits and their association with growth traits were determined. High or very high heritabilities were found for growth traits, body heights and ellipticity while they low-moderate values were estimated for BMH/BHP and BMH/CPH ratios. The negative and medium-high genetic correlations of ellipticity with growth traits and heights indicated that fish selected for bigger size will become rounder. The low genetic correlations of BMH/CPH with all traits tested demonstrated that this trait provides complementary information to ellipticity for a better fitting to the expected lanceolate body morphology of sole. All these data are highly relevant for breeding programs.

# CHAPTER 9

# BIBLIOGRAPHY

1. Agulleiro M.J., Anguis V., Cañavate J.P., Martínez-Rodríguez G., Mylonas C.C. & Cerdà J. (2006) Induction of spawning of captive-reared Senegal sole (*Solea senegalensis*) using different administration methods for gonadotropin releasing hormone agonist. Aquaculture 257, 511-24.
2. Akkaynak D., Siemann L.A., Barbosa A. & Mathger L.M. (2017) Changeable camouflage: how well can flounder resemble the colour and spatial scale of substrates in their natural habitats? R Soc Open Sci 4, 160824.
3. Ali A., Al-Tobasei R., Lourenco D., Leeds T., Kenney B. & Salem M. (2020) Genome-wide identification of loci associated with growth in rainbow trout. BMC Genomics 21, 209.
4. Ambrosio P.P., Costa C., Sánchez P. & Flos R. (2008) Stocking density and its influence on shape of Senegalese sole adults. Aquacult Int 16, 333-43.
5. APROMAR (2020) *La Acuicultura en España*, Cadiz, España.
6. Aslam M.L., Carraro R., Bestin A., Cariou S., Sonesson A.K., Bruant J.S., Haffray P., Bargelloni L. & Meuwissen T.H.E. (2018) Genetics of resistance to photobacteriosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. BMC Genet 19, 43.
7. Assumpção L., Makrakis M.C., Makrakis S., Wagner R., Silva P.S., Lima A.F. & Kashiwaqui E.A.L. (2012) The use of morphometric analysis to predict the swimming efficiency of two Neotropical long-distance migratory species in fish passage. Neotrop. Ichthyol. 10, 797-804.
8. Aulchenko Y.S., Ripke S., Isaacs A. & van Duijn C.M. (2007) GenABEL: an R library for genome-wide association analysis. Bioinformatics 23, 1294-6.
9. Aung L.H.H., Li R., Prabhakar B.S. & Li P. (2017) Knockdown of Mtfp1 can minimize doxorubicin cardiotoxicity by inhibiting Dnm1l-mediated mitochondrial fission. J Cell Mol Med 21, 3394-404.
10. Azevedo M.F.C., Oliveira C., Pardo B.G., Martinez P. & Foresti F. (2007) Cytogenetic characterization of six species of flatfishes with comments to karyotype differentiation patterns in Pleuronectiformes (Teleostei). J. Fish Biol. 70, 1-15.
11. Beier S., Thiel T., Munch T., Scholz U. & Mascher M. (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics 33, 2583-5.
12. Benzekri H., Armesto P., Cousin X., Rovira M., Crespo D., Merlo M.A., Mazurais D., Bautista R., Guerrero-Fernandez D., Fernandez-Pozo N., Ponce M., Infante C., Zambonino J.L., Nidelet S., Gut M., Rebordinos L., Planas J.V., Begout M.L., Claros M.G. & Manchado M. (2014) *De novo* assembly, characterization and functional annotation of Senegalese sole (*Solea senegalensis*) and common sole (*Solea solea*) transcriptomes: integration in a database and design of a microarray. BMC Genomics 15, 952.
13. Berglund B. (2008) Genetic improvement of dairy cow reproductive performance. Reprod Domest Anim 43 Suppl 2, 89-95.
14. Bertolini F., Ribani A., Capoccioni F., Buttazzoni L., Utzeri V.J., Bovo S., Schiavo G., Caggiano M., Fontanesi L. & Rothschild M.F. (2020) Identification of a major locus determining a pigmentation defect in cultivated gilthead seabream (Sparus aurata). Anim Genet 51, 319-23.
15. Black C.R. & Berendzen P.B. (2020) Shared ecological traits influence shape of the skeleton in flatfishes (Pleuronectiformes). PeerJ 8, e8919.
16. Blanco-Vives B., Vera L.M., Ramos J., Bayarri M.J., Mananos E. & Sanchez-Vazquez F.J. (2011) Exposure of larvae to daily thermocycles affects gonad development, sex ratio, and sexual steroids in *Solea senegalensis*, kaup. J Exp Zool A Ecol Genet Physiol 315, 162-9.

17. Blonk R.J., Komen H., Kamstra A. & van Arendonk J.A. (2010a) Effects of grading on heritability estimates under commercial conditions: A case study with common sole, *Solea solea*. Aquaculture 300, 43-9.
18. Blonk R.J., Komen H., Kamstra A. & van Arendonk J.A. (2010b) Estimating breeding values with molecular relatedness and reconstructed pedigrees in natural mating populations of common sole, *Solea solea*. Genetics 184, 213-9.
19. Blonk R.J., Komen H., Tenghe A., Kamstra A. & van Arendonk J.A.M. (2010c) Heritability of shape in common sole, *Solea solea*, estimated from image analysis data. Aquaculture 307, 6-11.
20. Bouza C., Hermida M., Pardo B.G., Fernandez C., Fortes G.G., Castro J., Sanchez L., Presa P., Perez M., Sanjuan A., de Carlos A., Alvarez-Dios J.A., Ezcurra S., Cal R.M., Piferrer F. & Martinez P. (2007) A microsatellite genetic map of the turbot (*Scophthalmus maximus*). Genetics 177, 2457-67.
21. Bouza C., Hermida M., Pardo B.G., Vera M., Fernandez C., de la Herran R., Navajas-Perez R., Alvarez-Dios J.A., Gomez-Tato A. & Martinez P. (2012) An Expressed Sequence Tag (EST)-enriched genetic map of turbot (*Scophthalmus maximus*): a useful framework for comparative genomics across model and farmed teleosts. BMC Genet 13, 54.
22. Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y. & Buckler E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23, 2633-5.
23. Brouard J.S., Schenkel F., Marete A. & Bissonnette N. (2019) The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments. J Anim Sci Biotechnol 10, 44.
24. Brown R.C., Woolliams J.A. & McAndrew B.J. (2005) Factors influencing effective population size in commercial populations of gilthead seabream, *Sparus aurata*. Aquaculture 247, 219–25.
25. Cabanettes F. & Klopp C. (2018) D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ 6, e4958.
26. Cabrita E., Soares F., Beirao J., Garcia-Lopez A., Martinez-Rodriguez G. & Dinis M.T. (2011) Endocrine and milt response of Senegalese sole, *Solea senegalensis*, males maintained in captivity. Theriogenology 75, 1-9.
27. Cadrin S.X. & Silva V.M. (2005) Morphometric variation of yellowtail flounder. ICES J. Mar. Sci 62, 683–94.
28. Cañavate J.P. & Fernandez-Diaz C. (1999) Influence of co-feeding larvae with live and inert diets on weaning the sole *Solea senegalensis* onto commercial dry feeds. Aquaculture 174, 255-63.
29. Carazo I., Chereguini O., Martin I., Huntingford F. & Duncan N. (2016) Reproductive ethogram and mate selection in captive wild Senegalese sole (*Solea senegalensis*). Span. J. Agric. Res. 14, e0401.
30. Carballo C., Berbel C., Guerrero-Cozar I., Jimenez-Fernandez E., Cousin X., Bégout M.L. & Manchado M. (2018) Evaluation of different tags on survival, growth and stress response in the flatfish Senegalese sole. Aquaculture 494, 10-8.
31. Carballo C., Chronopoulou E.G., Letsiou S., Spanidi E., Gardikis K., Labrou N.E. & Manchado M. (2019) Genomic and phylogenetic analysis of choriolysins, and biological activity of hatching liquid in the flatfish Senegalese sole. PLoS One 14, e0225666.
32. Carballo C., Shin H.S., Berbel C., Zamorano M.J., Borrego J.J., Armero E., Afonso J.M. & Manchado M. (2020) Heritability estimates and genetic correlation for gowth traits and LCDV susceptibility in gilthead sea bream (*Sparus aurata*). Fishes 5, 2.

33. Casiro S., Velez-Irizarry D., Ernst C.W., Raney N.E., Bates R.O., Charles M.G. & Steibel J.P. (2017) Genome-wide association study in an F2 Duroc x Pietrain resource population for economically important meat quality and carcass traits. J Anim Sci 95, 545-58.
34. Castaño-Sanchez C., Fuji K., Ozaki A., Hasegawa O., Sakamoto T., Morishima K., Nakayama I., Fujiwara A., Masaoka T., Okamoto H., Hayashida K., Tagami M., Kawai J., Hayashizaki Y. & Okamoto N. (2010) A second generation genetic linkage map of Japanese flounder (*Paralichthys olivaceus*). BMC Genomics 11, 554.
35. Castro J., Pino A., Hermida M., Bouza C., Riaza A., Ferreiro I., Sánchez L. & Martínez P. (2006) A microsatellite marker tool for parentage analysis in Senegal sole (*Solea senegalensis*): Genotyping errors, null alleles and conformance to theoretical assumptions. Aquaculture, 1194-203.
36. Catchen J., Amores A. & Bassham S. (2020) Chromonomer: A tool set for repairing and enhancing assembled genomes through integration of genetic maps and conserved synteny. G3 (Bethesda).
37. Cerda J. & Manchado M. (2013) Advances in genomics for flatfish aquaculture. Genes Nutr 8, 5-17.
38. Chagne D., Vanderzande S., Kirk C., Profitt N., Weskett R., Gardiner S.E., Peace C.P., Volz R.K. & Bassil N.V. (2019) Validation of SNP markers for fruit quality and disease resistance loci in apple (Malus x domestica Borkh.) using the OpenArray(R) platform. Hortic Res 6, 30.
39. Chauvigne F., Fatsini E., Duncan N., Olle J., Zanuy S., Gomez A. & Cerda J. (2016) Plasma levels of follicle-stimulating and luteinizing hormones during the reproductive cycle of wild and cultured Senegalese sole (*Solea senegalensis*). Comp Biochem Physiol A Mol Integr Physiol 191, 35-43.
40. Chauvigne F., Olle J., Gonzalez W., Duncan N., Gimenez I. & Cerda J. (2017) Toward developing recombinant gonadotropin-based hormone therapies for increasing fertility in the flatfish Senegalese sole. PLoS One 12, e0174387.
41. Chauvigne F., Tingaud-Sequeira A., Agulleiro M.J., Calusinska M., Gomez A., Finn R.N. & Cerda J. (2010) Functional and evolutionary analysis of flatfish gonadotropin receptors reveals cladal- and lineage-level divergence of the teleost glycoprotein receptor family. Biol Reprod 82, 1088-102.
42. Chauvigne F., Verdura S., Mazon M.J., Duncan N., Zanuy S., Gomez A. & Cerda J. (2012) Follicle-stimulating hormone and luteinizing hormone mediate the androgenic pathway in Leydig cells of an evolutionary advanced teleost. Biol Reprod 87, 35.
43. Chauvigne F., Zapater C., Gasol J.M. & Cerda J. (2014) Germ-line activation of the luteinizing hormone receptor directly drives spermiogenesis in a nonmammalian vertebrate. Proc Natl Acad Sci U S A 111, 1427-32.
44. 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., Wu K., Yang C., Zhou Q., Liao X., Yang L., Hu Q., Zhang J., Meng L., Jin L., Tian Y., Lian J., Yang J., Miao G., Liu S., Liang Z., Yan F., Li Y., Sun B., Zhang H., Zhang J., Zhu Y., Du M., Zhao Y., Schartl M., Tang Q. & Wang J. (2014a) Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nat. Genet. 46, 253-60.
45. 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., Wu K., Yang C., Zhou Q., Liao X., Yang L., Hu Q., Zhang J., Meng L., Jin L., Tian Y., Lian J., Yang J., Miao G., Liu S., Liang Z., Yan F., Li Y., Sun B., Zhang H., Zhang J., Zhu

Y., Du M., Zhao Y., Schartl M., Tang Q. & Wang J. (2014b) Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nat Genet.
46. Chen S.-L., Shao C.-W., Xu G.-B., Liao X.-L. & Tian Y.-S. (2008) Development of 15 novel dinucleotide microsatellite markers in the Senegalese sole *Solea senegalensis*. Fisheries Sci. 74, 1357-9.
47. Chiu Y.L., Shikina S., Yoshioka Y., Shinzato C. & Chang C.F. (2020) De novo transcriptome assembly from the gonads of a scleractinian coral, Euphyllia ancora: molecular mechanisms underlying scleractinian gametogenesis. BMC Genomics 21, 732.
48. Claros M.G., Seoane P. & Manchado M. (2020) Sequences and annotations of a provisional genome draft of a Senegalese sole female. figshare https://doi.org/10.6084/m9.figshare.12472100.v1.
49. Cobcroft J.M. B.S.C. (2013) Skeletal malformations in Australian marine finfish hatcheries. Aquaculture 396–399, 51-8.
50. Conesa A., Gotz S., Garcia-Gomez J.M., Terol J., Talon M. & Robles M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674-6.
51. Constenla M. & Padros F. (2010) Histopathological and ultrastructural studies on a novel pathological condition in *Solea senegalensis*. Dis Aquat Organ 90, 191-6.
52. Cordoba J.M., Chavarro C., Schlueter J.A., Jackson S.A. & Blair M.W. (2010) Integration of physical and genetic maps of common bean through BAC-derived microsatellite markers. BMC Genomics 11, 436.
53. Cordoba-Caballero J., Seoane P., Jabato F.M., Perkins J.R., Manchado M. & Claros M.G. (2020) An improved de novo assembling and polishing of Solea senegalensis transcriptome shed light on retinoic acid signalling in larvae. Sci Rep 10, 20654.
54. Davoli R. & Braglia S. (2008) Molecular approaches in pig breeding to improve meat quality. Briefings in Functional Genomics 6, 313-21.
55. de Azevedo A.M., Losada A.P., Barreiro A., Barreiro J.D., Ferreiro I., Riaza A., Vazquez S. & Quiroga M.I. (2017) Skeletal anomalies in reared Senegalese sole Solea senegalensis juveniles: a radiographic approach. Dis Aquat Organ 124, 117-29.
56. de Azevedo A.M., Losada A.P., Barreiro A., Vazquez S. & Quiroga M.I. (2019a) Skeletal Anomalies in Senegalese Sole ( Solea senegalensis), an Anosteocytic Boned Flatfish Species. Vet Pathol 56, 307-16.
57. de Azevedo A.M., Losada A.P., Barreiro A., Vazquez S. & Quiroga M.I. (2019b) Skeletal Anomalies in Senegalese Sole (Solea senegalensis), an Anosteocytic Boned Flatfish Species. Vet Pathol 56, 307-16.
58. De Donato M., Peters S.O., Mitchell S.E., Hussain T. & Imumorin I.G. (2013) Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. PLoS One 8, e62137.
59. De La Herran R., Robles F., Navas J.I., Hamman-Khalifa A.M., Herrera M., Hachero I., Mora M.J., Ruiz-Rejon C., Garrido-Ramos M. & Ruiz-Rejon M. (2008) A highly accurate, single PCR reaction for parentage assignment in Senegal sole based on eight informative microsatellite loci. Aquaculture Res. 39, 1169–74.
60. Dinis M.T., Ribeiro L., Soares F. & Sarasquete C. (1999) A review on the cultivation potential of Solea senegalensis in Spain and in Portugal. Aquaculture 176, 27–38.
61. Dionisio G., Campos C., Valente L.M.P., Conceicao L.E.C., Cancela M.L. & Gavaia P.J. (2012) Effect of egg incubation temperature on the occurrence of skeletal deformities in *Solea senegalensis*. J Appl Ichthyol. 28, 471-6.

62. Dong L., Xiao S., Chen J., Wan L. & Wang Z. (2016) Genomic Selection Using Extreme Phenotypes and Pre-Selection of SNPs in Large Yellow Croaker (*Larimichthys crocea*). Mar Biotechnol (NY) 18, 575-83.
63. Ducrocq V., Besbes B. & Protais M. (2000) Genetic improvement of laying hens viability using survival analysis. Genet Sel Evol 32, 23-40.
64. Dupont-Nivet M., Chevassus B., Mauger S., Haffray P. & Vandeputte M. (2010) Side effects of sexual maturation on heritability estimates in rainbow trout (*Oncorhynchus mykiss*). Aquacult Res. 41, e878-e80.
65. Duran C., Edwards D. & Batley J. (2009) Genetic maps and the use of synteny. Methods Mol Biol 513, 41-55.
66. Edvardsen R., Wallerman O., Furmanek T., Kleppe L., Jern P., Wallberg A., Kjærner-Semb E., Mæhle S., Olausson S., Sundström E., Harboe T., Mangor-Jensen R., Møgster M., Perrichon P., Norberg B. & Rubin C. (2020) Heterochiasmy facilitated the establishment of gsdf as a novel sex determining gene in Atlantic halibut.
67. Eisbrenner W.D., Botwright N., Cook M., Davidson E.A., Dominik S., Elliott N.G., Henshall J., Jones S.L., Kube P.D., Lubieniecki K.P., Peng S. & Davidson W.S. (2014) Evidence for multiple sex-determining loci in Tasmanian Atlantic salmon (*Salmo salar*). Heredity 113, 86-92.
68. Espinosa E., Arroyo M., Larrosa R., Manchado M., Claros M.G. & Bautista R. (2020) Micro-variations from RNA-seq experiments for non-model organisms. In: Bioinformatics and Biomedical Engineering. IWBBIO 2020 (ed. by I. Rojas, O. Valenzuela, F. Rojas, L. Herrera & F. Ortuño). Springer, Cham.
69. FAO (2020) *El estado mundial de la pesca y la acuicultura 2020. La sostenibilidad en acción*, Roma.
70. Fatsini E., Bautista R., Manchado M. & Duncan N.J. (2016) Transcriptomic profiles of the upper olfactory rosette in cultured and wild Senegalese sole (Solea senegalensis) males. Comp Biochem Physiol Part D Genomics Proteomics 20, 125-35.
71. Fatsini E., Carazo I., Chauvigne F., Manchado M., Cerda J., Hubbard P.C. & Duncan N.J. (2017) Olfactory sensitivity of the marine flatfish Solea senegalensis to conspecific body fluids. J Exp Biol 220, 2057-65.
72. Fatsini E., Gonzalez W., Ibarra-Zatarain Z., Napuchi J. & Duncan N. (2020) The presence of wild Senegalese sole breeders improves courtship and reproductive success in cultured conspecifics. Aquaculture 519, 734922.
73. Fernandez I., Ortiz-Delgado J.B., Darias M.J., Hontoria F., Andree K.B., Manchado M., Sarasquete C. & Gisbert E. (2017) Vitamin A Affects Flatfish Development in a Thyroid Hormone Signaling and Metamorphic Stage Dependent Manner. Front Physiol 8, 458.
74. Fernández I., Pimentel M.S., Ortiz Delgado J.B., Hontoria F., Sarasquete C., Estévez A., Zambonino Infante J.L. & Gisbert E. (2009) Effect of dietary vitamin A on Senegalese sole (*Solea senegalensis*) skeletogenesis and larval quality. Aquaculture 295, 250-65.
75. Ferraresso S., Bargelloni L., Babbucci M., Cannas R., Follesa M.C., Carugati L., Melis R., Cau A., Koutrakis M., Sapounidis A., Crosetti D. & Patarnello T. (2021) fshr: a fish sex-determining locus shows variable incomplete penetrance across flathead grey mullet populations. iScience 24, 10186.
76. Fierst J.L. (2015) Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. Front Genet 6, 220.
77. Figueras A., Robledo D., Corvelo A., Hermida M., Pereiro P., Rubiolo J.A., Gomez-Garrido J., Carrete L., Bello X., Gut M., Gut I.G., Marcet-Houben M., Forn-Cuni G., Galan B., Garcia J.L., Abal-Fabeiro J.L., Pardo B.G., Taboada X., Fernandez C., Vlasova A.,

Hermoso-Pulido A., Guigo R., Alvarez-Dios J.A., Gomez-Tato A., Vinas A., Maside X., Gabaldon T., Novoa B., Bouza C., Alioto T. & Martinez P. (2016) Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): a fish adapted to demersal life. DNA Res 23, 181-92.

78. Fisher R. & Hogan J.D. (2007) Morphological predictors of swimming speed: a case study of pre-settlement juvenile coral reef fishes. J Exp Biol 210, 2436-43.
79. Flores-Renteria L. & Krohn A. (2013) Scoring microsatellite loci. Methods Mol Biol 1006, 319-36.
80. Forne I., Castellana B., Marin-Juez R., Cerda J., Abian J. & Planas J.V. (2011) Transcriptional and proteomic profiling of flatfish (Solea senegalensis) spermatogenesis. Proteomics 11, 2195-211.
81. Fu B., Liu H., Yu X. & Tong J. (2016) A high-density genetic map and growth related QTL mapping in bighead carp (*Hypophthalmichthys nobilis*). Sci Rep 6, 28679.
82. Funes V., Zuasti E., Catanese G., Infante C. & Manchado M. (2004) Isolation and characterization of ten microsatellite loci for Senegal sole (*Solea senegalensis* Kaup). Mol Ecol Notes 4, 339-41.
83. Garcia E., Cross I., Portela-Bens S., Rodriguez M.E., Garcia-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. Sci Rep 9, 17146.
84. Garcia-Angulo A., Merlo M.A., Portela-Bens S., Rodriguez M.E., Garcia E., Al-Rikabi A., Liehr T. & Rebordinos L. (2018) Evidence for a Robertsonian fusion in *Solea senegalensis* (Kaup, 1858) revealed by zoo-FISH and comparative genome analysis. BMC Genomics 19, 818.
85. Garcia-Angulo A., Merlo M.A., Rodriguez M.E., Portela-Bens S., Liehr T. & Rebordinos L. (2019) Genome and phylogenetic analysis of genes involved in the immune system of *Solea senegalensis* - Potential applications in aquaculture. Front Genet 10, 529.
86. Garcia-Celdran M., Cutáková Z., Ramis G., Estévez A., Manchado M., Navarro A., Maria-Dolores E., Peñalver J., Sanchez J.A. & Armero E. (2016) Estimates of heritabilities and genetic correlations of skeletal deformities and uninflated swimbladder in a reared gilthead sea bream (*Sparus aurata* L.) juvenile population sourced from three broodstocks along the Spanish coasts. Aquaculture 464, 601-8.
87. Garcia-Celdran M., Ramis G., Manchado M., Estévez A., Afonso J.M. & Armero E. (2015a) Estimates of heritabilities and genetic correlations of carcass quality traits in a reared gilthead sea bream (*Sparus aurata* L.) population sourced from three broodstocks along the Spanish coasts. Aquaculture 446, 175-80.
88. Garcia-Celdran M., Ramis G., Manchado M., Estevez A., Afonso J.M., Maria-Dolores E., Peñalver J. & Armero E. (2015b) Estimates of heritabilities and genetic correlations of growth and external skeletal deformities at different ages in a reared gilthead sea bream (*Sparus aurata* L.) population sourced from three broodstocks along the Spanish coasts. Aquaculture 445, 33-41.
89. Garcia-Celdran M., Ramis G., Manchado M., Estévez A., Navarro A. & Armero E. (2015c) Estimates of heritabilities and genetic correlations of raw flesh quality traits in a reared gilthead sea bream (*Sparus aurata* L.) population sourced from broodstocks along the Spanish coasts. Aquaculture 446, 181-6.
90. Gavaia P., Domingues S., Engrola S., Drake P., Sarasquete C., Dinis M.T. & Cancela M.L. (2009) Comparing skeletal development of wild and hatchery-reared Senegalese sole

(Solea senegalensis, Kaup 1858): evaluation in larval and postlarval stages. Aquacult Res. 40, 1585–93.
91. Gavaia P.J., Dinis M.T. & Cancela M.L. (2002) Osteological development and abnormalities of the vertebral column and caudal skeleton in larval and juvenile stages of hatchery-reared Senegal sole (*Solea senegalensis*). Aquaculture 211, 305- 23.
92. Ge H., Lin K., Shen M., Wu S., Wang Y., Zhang Z., Wang Z., Zhang Y., Huang Z., Zhou C., Lin Q., Wu J., Liu L., Hu J., Huang Z. & Zheng L. (2019) De novo assembly of a chromosome-level reference genome of red-spotted grouper (*Epinephelus akaara*) using nanopore sequencing and Hi-C. Mol. Ecol. Resour. 19, 1461-9.
93. Gjedrem T. (2000) Genetic improvement of cold-water fish species. Aquaculture Research 31, 25–33.
94. Gjedrem T. (2005a) *Selection and Breeding Programs in Aquaculture*. Springer Netherlands.
95. Gjedrem T. (2005b) Status and scope of aquaculture. In: Selection and Breeding Programs in Aquaculture (ed. by T. Gjedrem). Springer Netherlands.
96. Gjedrem T. (2012) Genetic improvement for the development of efficient global aquaculture: A personal opinion review. Aquaculture 344-349, 12-22.
97. Goddard M.E. & Hayes B.J. (2007) Genomic selection. J Anim Breed Genet 124, 323-30.
98. Goodwin S., Gurtowski J., Ethe-Sayers S., Deshpande P., Schatz M.C. & McCombie W.R. (2015) Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome Res. 25, 1750-6.
99. Grabherr M.G., Russell P., Meyer M., Mauceli E., Alfoldi J., Di Palma F. & Lindblad-Toh K. (2010) Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics 26, 1145-51.
100. Guerrero-Cozar I., Jimenez-Fernandez E., Berbel C., Cordoba-Caballero J., Claros M.G., Zerolo R. & Manchado M. (2021) Genetic parameter estimates and identification of SNPs associated with growth traits in Senegalese sole. Aquaculture 539, 736665.
101. Guerrero-Cozar I., Perez-Garcia C., Benzekri H., Sánchez J., Seoane P., Cruz F., Gut M., Zamorano M.J., Claros M.G. & Manchado M. (2020) Development of whole-genome multiplex assays and construction of an integrated genetic map using SSR markers in Senegalese sole. Sci. Reports 10, 21905.
102. Gutierrez-Camino A., Martin-Guerrero I., Dolzan V., Jazbec J., Carbone-Baneres A., Garcia de Andoin N., Sastre A., Astigarraga I., Navajas A. & Garcia-Orad A. (2018) Involvement of SNPs in miR-3117 and miR-3689d2 in childhood acute lymphoblastic leukemia risk. Oncotarget 9, 22907-14.
103. Guzman J.M., Rubio M., Ortiz-Delgado J.B., Klenke U., Kight K., Cross I., Sanchez-Ramos I., Riaza A., Rebordinos L., Sarasquete C., Zohar Y. & Mananos E.L. (2009) Comparative gene expression of gonadotropins (FSH and LH) and peptide levels of gonadotropin-releasing hormones (GnRHs) in the pituitary of wild and cultured Senegalese sole (Solea senegalensis) broodstocks. Comp Biochem Physiol A Mol Integr Physiol 153, 266-77.
104. Henson J., Tischler G. & Ning Z. (2012) Next-generation sequencing and large genome assemblies. Pharmacogenomics 13, 901-15.
105. Hoghoughi N., Barral S., Vargas A., Rousseaux S. & Khochbin S. (2018) Histone variants: essential actors in male genome programming. J Biochem 163, 97-103.
106. Houston R.D., Taggart J.B., Cezard T., Bekaert M., Lowe N.R., Downing A., Talbot R., Bishop S.C., Archibald A.L., Bron J.E., Penman D.J., Davassi A., Brew F., Tinch A.E.,

Gharbi K. & Hamilton A. (2014) Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). BMC Genomics 15, 90.
107. Hsieh Y.J., Kundu T.K., Wang Z., Kovelman R. & Roeder R.G. (1999) The TFIIIC90 subunit of TFIIIC interacts with multiple components of the RNA polymerase III machinery and contains a histone-specific acetyltransferase activity. Mol Cell Biol 19, 7697-704.
108. Imran M., Shafiq S., Farooq M.A., Naeem M.K., Widemann E., Bakhsh A., Jensen K.B. & Wang R.R. (2019) Comparative genome-wide analysis and expression profiling of histone acetyltransferase (HAT) gene family in response to hormonal applications, metal and abiotic stresses in cotton. Int J Mol Sci 20.
109. Imsland A.K., Foss A., Conceiçao L.E.C., Dinis M.T., Delbare D., Schram E., Kamstra A., Rema P. & White P. (2003) A review of the culture potential of *Solea solea* and *S. senegalensis*. Rev. Fish Biol. Fish. 13, 379-407.
110. Isorna E., Obregon M.J., Calvo R.M., Vazquez R., Pendon C., Falcon J. & Munoz-Cueto J.A. (2009) Iodothyronine deiodinases and thyroid hormone receptors regulation during flatfish (Solea senegalensis) metamorphosis. J Exp Zool B Mol Dev Evol 312B, 231-46.
111. Janssen K., Chavanne H., Berentsen P. & Komen H. (2017) Impact of selective breeding on European aquaculture. Aquaculture 472, 8-16.
112. Kalinowski S.T., Taper M.L. & Marshall T.C. (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16, 1099-106.
113. Kitano J., Ross J.A., Mori S., Kume M., Jones F.C., Chan Y.F., Absher D.M., Grimwood J., Schmutz J., Myers R.M., Kingsley D.M. & Peichel C.L. (2009) A role for a neo-sex chromosome in stickleback speciation. Nature 461, 1079-83.
114. Klaren P.H., Wunderink Y.S., Yufera M., Mancera J.M. & Flik G. (2008) The thyroid gland and thyroid hormones in Senegalese sole (Solea senegalensis) during early development and metamorphosis. Gen Comp Endocrinol 155, 686-94.
115. Koboldt D.C., Zhang Q., Larson D.E., Shen D., McLellan M.D., Lin L., Miller C.A., Mardis E.R., Ding L. & Wilson R.K. (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22, 568-76.
116. Kriaridou C., Tsairidou S., Houston R.D. & Robledo D. (2020) Genomic prediction using low density marker panels in aquaculture: performance across species, traits, and genotyping platforms. Front Genet 11, 124.
117. Kundu T.K., Wang Z. & Roeder R.G. (1999) Human TFIIIC relieves chromatin-mediated repression of RNA polymerase III transcription and contains an intrinsic histone acetyltransferase activity. Mol Cell Biol 19, 1605-15.
118. Kyriakis D., Kanterakis A., Manousaki T., Tsakogiannis A., Tsagris M., Tsamardinos I., Papaharisis L., Chatziplis D., Potamias G. & Tsigenopoulos C.S. (2019) Scanning of Genetic Variants and Genetic Mapping of Phenotypic Traits in Gilthead Sea Bream Through ddRAD Sequencing. Front Genet 10, 675.
119. Labuschagne C., Nupen L., Kotze A., Grobler P.J. & Dalton D.L. (2015) Assessment of microsatellite and SNP markers for parentage assignment in *ex situ* African Penguin (*Spheniscus demersus*) populations. Ecol. Evol. 5, 4389-99.
120. Laghari M.Y., Lashari P., Zhang X., Xu P., Narejo N.T., Xin B., Zhang Y. & Sun X. (2015) QTL mapping for economically important traits of common carp (*Cyprinus carpio* L.). J Appl Genet 56, 65-75.

121. Laghari M.Y., Lashari P., Zhang Y. & Sun X. (2014) Identification of Quantitative Trait Loci (QTLs) in Aquaculture Species. Reviews in Fisheries Science & Aquaculture 22, 221-38.
122. Lee-Montero I., Navarro A., Borrell Y., Garcia-Celdran M., Martin N., Negrin-Baez D., Blanco G., Armero E., Berbel C., Zamorano M.J., Sanchez J.J., Estevez A., Ramis G., Manchado M. & Afonso J.M. (2013) Development of the first standardised panel of two new microsatellite multiplex PCRs for gilthead seabream (*Sparus aurata* L.). Anim Genet 44, 533-46.
123. Lee-Montero I., Navarro A., Negrin-Baez D., Zamorano M.J., Borrell Pichs Y.J., Berbel C., Sanchez J.A., Garcia-Celdran M., Manchado M., Estevez A., Armero E. & Afonso J.M. (2015) Genetic parameters and genotype-environment interactions for skeleton deformities and growth traits at different ages on gilthead seabream (*Sparus aurata* L.) in four Spanish regions. Anim Genet 46, 164-74.
124. Lerceteau-Köhler E. & Weiss S. (2006) Development of a multiplex PCR microsatellite assay in brown trout Salmo trutta, and its potential application for the genus. Aquaculture 258, 641-5.
125. Li H. (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094-100.
126. Li R., Zhang W., Lu J., Zhang Z., Mu C., Song W., Migaud H., Wang C. & Bekaert M. (2020) The Whole-Genome Sequencing and Hybrid Assembly of Mytilus coruscus. Front Genet 11, 440.
127. Liao X., Li M., Zou Y., Wu F.-X., Yi P. & Wang J. (2019) Current challenges and solutions of de novo assembly. Quantitative Biology 7, 90-109.
128. Limborg M.T., McKinney G.J., Seeb L.W. & Seeb J.E. (2016) Recombination patterns reveal information about centromere location on linkage maps. Mol. Ecol. Resour. 16, 655-61.
129. Liu Z.J. & Cordes F.J. (2004) DNA marker technology and their applications in aquaculture genetics. Aquaculture 238, , 1-37.
130. Losada A.P., de Azevedo A.M., Barreiro A., Barreiro J.D., Ferreiro I., Riaza A., Quiroga M.I. & Vazquez S. (2014) Skeletal malformations in Senegalese sole (*Solea senegalensis* Kaup, 1858): gross morphology and radiographic correlation. J. Appl. Ichthyol. 30, 804–8.
131. Loukovitis D., Batargias C., Sarropoulou E., Apostolidis A.P., Kotoulas G., Magoulas A., Tsigenopoulos C.S. & Chatziplis D. (2013) Quantitative trait loci affecting morphology traits in gilthead seabream (Sparus aurata L.). Anim Genet 44, 480-3.
132. Lu Q., Hong Y., Li S., Liu H., Li H., Zhang J., Lan H., Liu H., Li X., Wen S., Zhou G., Varshney R.K., Jiang H., Chen X. & Liang X. (2019) Genome-wide identification of microsatellite markers from cultivated peanut (*Arachis hypogaea* L.). BMC Genomics 20, 799.
133. Manchado M., Infante C., Asensio E., Planas J.V. & Canavate J.P. (2008) Thyroid hormones down-regulate thyrotropin beta subunit and thyroglobulin during metamorphosis in the flatfish Senegalese sole (Solea senegalensis Kaup). Gen Comp Endocrinol 155, 447-55.
134. Manchado M., Planas J.V., Cousin X., Rebordinos L. & Claros M.G. (2016) Current status in other finfish species: Description of current genomic resources for the gilthead seabream (*Sparus aurata*) and soles (*Solea senegalensis* and *Solea solea)*. In: Genomics in Aquaculture (ed. by S. Mackenzie & S. Jentoft), pp. 195–221. Elsevier, Amsterdam.

135. Manchado M., Planas J.V., Cousin X., Rebordinos L. & Claros M.G. (2019) Genetic and genomic characterization of soles. In: *The Biology of Sole* (ed. by J. Munoz-Cueto, E. Mañanós-Sánchez & F. Sánchez-Vázquez), pp. 375-94. CDC Press., Boca Raton.
136. Manchado M., Planas, J.V., Cousin, X., Rebordinos, L., Claros, M.G. (2019) Genetic and genomic characterization of soles. In: The Biology of Sole (ed. by J. Muñoz-Cueto, Mañanós-Sánchez, E., Sánchez-Vázquez, F. (Eds.)), pp. 375-94. CDC Press, Boca Raton.
137. Marco-Sola S., Sammeth M., Guigo R. & Ribeca P. (2012) The GEM mapper: fast, accurate and versatile alignment by filtration. Nat Methods 9, 1185-8.
138. Maroso F., Hermida M., Millan A., Blanco A., Saura M., Fernandez A., Dalla Rovere G., Bargelloni L., Cabaleiro S., Villanueva B., Bouza C. & Martinez P. (2018) Highly dense linkage maps from 31 full-sibling families of turbot (*Scophthalmus maximus*) provide insights into recombination patterns and chromosome rearrangements throughout a newly refined genome assembly. DNA Res 25, 439-50.
139. Martin I., Carazo I., Rasines I., Rodriguez C., Fernandez R., Martinez P., Norambuena P., Chereguini O. & Duncan N. (2019) Reproductive performance of captive Senegalese sole, *Solea senegalensis*, according to the origin (wild or cultured) and gender. Span. J. Agric. Res. 17, 2171-9292.
140. Martin I., Rasines I., Gomez M., Rodriguez C., Martinez P. & Chereguini O. (2014) Evolution of egg production and parental contribution in Senegalese sole, *Solea senegalensis*, during four consecutive spawning seasons. Aquaculture 424-425, 45-52.
141. Martinez P., Vinas A.M., Sanchez L., Diaz N., Ribas L. & Piferrer F. (2014) Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture. Front Genet 5, 340.
142. Mas-Muñoz J., Blonk R.J., Schrama J.W., van Arendonk J. & Komen H. (2013) Genotype by environment interaction for growth of sole (*Solea solea*) reared in an intensive aquaculture system and in a semi-natural environment. Aquaculture 410-411, 230–5.
143. Mascher M. & Stein N. (2014) Genetic anchoring of whole-genome shotgun assemblies. Front Genet 5, 208.
144. Massault C., Bovenhuis H., Haley C. & de Koning D.J. (2008) QTL mapping designs for aquaculture. Aquaculture 285, 23-9.
145. Massault C., Hellemans B., Louro B., Batargias C., Van Houdt J.K., Canario A., Volckaert F.A., Bovenhuis H., Haley C. & de Koning D.J. (2010) QTL for body weight, morphometric traits and stress response in European sea bass Dicentrarchus labrax. Anim Genet 41, 337-45.
146. Mengistu S.B., Mulder H.A., Benzie J.A.H., Khaw H.L., Megens H., Trinh T.Q. & Komen H. (2020) Genotype by environment interaction between aerated and non-aerated ponds and the impact of aeration on genetic parameters in Nile tilapia (Oreochromis niloticus). Aquaculture 529, 735704.
147. Merlo M.A., Iziga R., Portela-Bens S., Cross I., Kosyakova N., Liehr T., Manchado M. & Rebordinos L. (2017) Analysis of the histone cluster in Senegalese sole (Solea senegalensis): evidence for a divergent evolution of two canonical histone clusters. Genome 60, 441-53.
148. Metzker M.L. (2010) Sequencing technologies - the next generation. Nat. Rev. Genet. 11, 31-46.
149. Meyer K. (2007) WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). J Zhejiang Univ Sci B 8, 815-21.

150. Miller J.R., Delcher A.L., Koren S., Venter E., Walenz B.P., Brownley A., Johnson J., Li K., Mobarry C. & Sutton G. (2008) Aggressive assembly of pyrosequencing reads with mates. Bioinformatics 24, 2818-24.
151. Molina-Luzon M.J., Hermida M., Navajas-Perez R., Robles F., Navas J.I., Ruiz-Rejon C., Bouza C., Martinez P. & de la Herran R. (2015a) First haploid genetic map based on microsatellite markers in Senegalese sole (Solea senegalensis, Kaup 1858). Mar Biotechnol (NY) 17, 8-22.
152. Molina-Luzon M.J., Hermida M., Navajas-Perez R., Robles F., Navas J.I., Ruiz-Rejon C., Bouza C., Martinez P. & de la Herran R. (2015b) First haploid genetic map based on microsatellite markers in Senegalese sole (*Solea senegalensis*, Kaup 1858). Mar. Biotechnol (NY) 17, 8-22.
153. Molina-Luzon M.J., Lopez J.R., Navajas-Perez R., Robles F., Ruiz-Rejon C. & De La Herran R. (2012) Validation and comparison of microsatellite markers derived from Senegalese sole (*Solea senegalensis*, Kaup) genomic and expressed sequence tags libraries. Mol. Ecol. Resour. 12, 956-66.
154. Molina-Luzon M.J., Lopez J.R., Robles F., Navajas-Perez R., Ruiz-Rejon C., De la Herran R. & Navas J.I. (2015c) Chromosomal manipulation in Senegalese sole (*Solea senegalensis* Kaup, 1858): induction of triploidy and gynogenesis. J Appl Genet 56, 77-84.
155. 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. (2016) New developments and biological insights into the farming of *Solea senegalensis* reinforcing its aquaculture potential. Rev Aquacult 6, 1-37.
156. Nater A., Kopps A.M. & Krutzen M. (2009) New polymorphic tetranucleotide microsatellites improve scoring accuracy in the bottlenose dolphin *Tursiops aduncus*. Mol. Ecol. Resour. 9, 531-4.
157. Navarro A., Badilla R., Zamorano M.J., Pasamontes V., Hildebrandt S., Sánchez J.J. & Afonso J.M. (2008) Development of two new microsatellite multiplex PCRs for three sparid species: Gilthead seabream (*Sparus auratus* L.), red porgy (*Pagrus pagrus* L.) and redbanded seabream (*P. auriga*, Valenciennes, 1843) and their application to paternity studies. Aquaculture 285, 30-3.
158. Navarro A., Lee-Montero I., Santana D., Henríquez P., Ferrer M.A., Morales A., Soula M., Badilla R., Negrin-Baez D., Zamorano M.J. & Afonso J.M. (2016) IMAFISH_ML: A fully-automated image analysis software for assessing fish morphometric traits on gilthead seabream (*Sparus aurata* L.), meagre (*Argyrosomus regius*) and red porgy (*Pagrus pagrus*). Comput Electron Agr. 121, 66-73.
159. Navarro A., Oliva V., Zamorano M.J., Ginés R., Izquierdo M.S., Astorga N. & Afonso J.M. (2006) Evaluation of PIT system as method to tag fingerlings of gilthead seabream (*Sparus auratus* L.): effects on growth, mortality and tag loss. Aquaculture 257, 309–15.
160. Navarro A., Zamorano M.J., Hildebrandt S., Ginés R., Aguilera C. & Afonso J.M. (2009) Estimates of heritabilities and genetic correlations for growth and carcass traits in gilthead seabream (*Sparus auratus* L.), under industrial conditions. Aquaculture 289, 225-30.
161. Negrin-Baez D., Negrin-Baez D., Rodriguez-Ramilo S.T., Afonso J.M. & Zamorano M.J. (2016) Identification of quantitative trait loci associated with the skeletal deformity LSK complex in gilthead seabream (*Sparus aurata* L.). Mar. Biotechnol (NY) 18, 98-106.
162. Negrin-Baez D., Navarro A., Lee-Montero I., Afonso J.M., Sánchez J.J., Elalfy I.S., Manchado M., Sánchez J.A., García-Celdrán M. & Zamorano M.J. (2015) A set of 13 multiplex PCRs of specific microsatellite markers as a tool for QTL detection in gilthead seabream (*Sparus aurata* L.). Aquac Res. 46, 45-58.

163. Nguinkal J.A., Brunner R.M., Verleih M., Rebl A., de Los Rios-Perez L., Schafer N., Hadlich F., Stueken M., Wittenburg D. & Goldammer T. (2019) The First Highly Contiguous Genome Assembly of Pikeperch (*Sander lucioperca*), an Emerging Aquaculture Species in Europe. Genes (Basel) 10.
164. Nguyen N.H., Rastas P.M.A., Premachandra H.K.A. & Knibb W. (2018) First high-density linkage map and single nucleotide polymorphisms significantly associated with traits of economic importance in Yellowtail Kingfish *Seriola lalandi*. Front Genet 9, 127.
165. Novel P., Porta J.M., Porta J., Bejar J. & Alvarez M.C. (2010) PCR multiplex tool with 10 microsatellites for the European seabass (*Dicentrarchus labrax*) — Applications in genetic differentiation of populations and parental assignment. Aquaculture 308, S34-S8.
166. Odegard J., Moen T., Santi N., Korsvoll S.A., Kjoglum S. & Meuwissen T.H. (2014) Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). Front Genet 5, 402.
167. Omasaki S.K., Charo-Karisa H., Kahi A.K. & Komen H. (2016) Genotype by environment interaction for harvest weight, growth rateand shape between monosex and mixed sex Nile tilapia (*Oreochromis niloticus*). Aquaculture 458, 75-81.
168. Ouellette L.A., Reid R.W., Blanchard S.G. & Brouwer C.R. (2018) LinkageMapView-rendering high-resolution linkage and QTL maps. Bioinformatics 34, 306-7.
169. Palaiokostas C., Bekaert M., Davie A., Cowan M.E., Oral M., Taggart J.B., Gharbi K., McAndrew B.J., Penman D.J. & Migaud H. (2013a) Mapping the sex determination locus in the Atlantic halibut (Hippoglossus hippoglossus) using RAD sequencing. BMC Genomics 14, 566.
170. Palaiokostas C., Bekaert M., Khan M.G., Taggart J.B., Gharbi K., McAndrew B.J. & Penman D.J. (2013b) Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) Using RAD sequencing. PLoS One 8, e68389.
171. Palaiokostas C., Ferraresso S., Franch R., Houston R.D. & Bargelloni L. (2016) Genomic Prediction of Resistance to Pasteurellosis in Gilthead Sea Bream (Sparus aurata) Using 2b-RAD Sequencing. G3 (Bethesda) 6, 3693-700.
172. Palti Y., Genet C., Gao G., Hu Y., You F.M., Boussaha M., Rexroad C.E. & Luo M.-C. (2012) A Second Generation Integrated Map of the Rainbow Trout (*Oncorhynchus mykiss*) Genome: Analysis of Conserved Synteny with Model Fish Genomes. Marine Biotechnology 14, 343-57.
173. Peakall R. & Smouse P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics 28, 2537-9.
174. Peterson B.K., Weber J.N., Kay E.H., Fisher H.S. & Hoekstra H.E. (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7, e37135.
175. Pop M. (2009) Genome assembly reborn: recent computational challenges. Briefings in Bioinformatics 10, 354-66.
176. Porta J., Porta J.M., Martínez-Rodríguez G. & Álvarez M.C. (2006) Development of a microsatellite multiplex PCR for Senegalese sole (*Solea senegalensis*) and its application to broodstock management. Aquaculture 256, 159-66.
177. Portela-Bens S., Merlo M.A., Rodriguez M.E., Cross I., Manchado M., Kosyakova N., Liehr T. & Rebordinos L. (2017) Integrated gene mapping and synteny studies give insights into the evolution of a sex proto-chromosome in Solea senegalensis. Chromosoma 126, 261-77.

178. Purcell C.M., Seetharam A.S., Snodgrass O., Ortega-Garcia S., Hyde J.R. & Severin A.J. (2018) Insights into teleost sex determination from the *Seriola dorsalis* genome assembly. BMC Genomics 19, 31.
179. Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., Maller J., Sklar P., de Bakker P.I., Daly M.J. & Sham P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81, 559-75.
180. Rastas P. (2017) Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. Bioinformatics 33, 3726-32.
181. Rastas P. (2020) Lep-Anchor: automated construction of linkage map anchored haploid genomes. Bioinformatics 36, 2359-64.
182. Reid D.P., Smith C.A., Rommens M., Blanchard B., Martin-Robichaud D. & Reith M. (2007) A Genetic linkage map of Atlantic halibut (*Hippoglossus hippoglossus* L.). Genetics 177, 1193-205.
183. Reinders M.J., Banovi M., Guerrero L. & Krystallis A. (2016) Consumer perceptions of farmed fish: A cross-national segmentation in five European countries. British Food J. 118, 2581-97.
184. Rezvoy C., Charif D., Gueguen L. & Marais G.A. (2007) MareyMap: an R-based tool with graphical interface for estimating recombination rates. Bioinformatics 23, 2188-9.
185. Riesco M.F., Valcarce D.G., Martinez-Vazquez J.M., Martin I., Calderon-Garcia A.A., Gonzalez-Nunez V. & Robles V. (2019) Male reproductive dysfunction in Solea senegalensis: new insights into an unsolved question. Reprod Fertil Dev 31, 1104-15.
186. Robledo D., Palaiokostas C., Bargelloni L., Martinez P. & Houston R. (2018) Applications of genotyping by sequencing in aquaculture breeding and genetics. Rev Aquac 10, 670-82.
187. Robledo D., Rubiolo J.A., Cabaleiro S., Martinez P. & Bouza C. (2017) Differential gene expression and SNP association between fast- and slow-growing turbot (*Scophthalmus maximus*). Sci Rep 7, 12105.
188. Rochette N.C. & Catchen J.M. (2017) Deriving genotypes from RAD-seq short-read data using Stacks. Nat Protoc 12, 2640-59.
189. Rodriguez M.E., Molina B., Merlo M.A., Arias-Perez A., Portela-Bens S., Garcia-Angulo A., Cross I., Liehr T. & Rebordinos L. (2019) Evolution of the Proto Sex-Chromosome in *Solea senegalensis*. Int J Mol Sci 20.
190. Roman-Padilla J., Rodriguez-Rua A., Claros M.G., Hachero-Cruzado I. & Manchado M. (2016) Genomic characterization and expression analysis of four apolipoprotein A-IV paralogs in Senegalese sole (*Solea senegalensis* Kaup). Comp Biochem Physiol B Biochem Mol Biol 191, 84-98.
191. Roman-Padilla J., Rodríguez-Rua A., Ponce M., Manchado M. & Hachero-Cruzado I. (2017) Effects of dietary lipid profile on larval performance and lipid management in Senegalese sole. Aquaculture 468, Part 1, 80-93.
192. Rooney A.P., Piontkivska H. & Nei M. (2002) Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. Mol Biol Evol 19, 68-75.
193. Rosyara U.R., De Jong W.S., Douches D.S. & Endelman J.B. (2016) Software for Genome-Wide Association Studies in Autopolyploids and Its Application to Potato. Plant Genome 9.
194. Ruan X., Wang W., Kong J., Yu F. & Huang X. (2010) Genetic linkage mapping of turbot (*Scophthalmus maximus* L.) using microsatellitemarkers and its application in QTL analysis. Aquaculture 308, 89-100.

195. Sanchez J.J., Borsting C., Hallenberg C., Buchard A., Hernandez A. & Morling N. (2003) Multiplex PCR and minisequencing of SNPs--a model with 35 Y chromosome SNPs. Forensic Sci Int 137, 74-84.
196. Sanchez P., Ambrosio P.P. & Flos R. (2010) Stocking density and sex influence individual growth of Senegalese sole (*Solea senegalensis*). Aquaculture 300, 93–101.
197. Sardell J.M., Cheng C., Dagilis A.J., Ishikawa A., Kitano J., Peichel C.L. & Kirkpatrick M. (2018) Sex differences in recombination in Sticklebacks. G3 (Bethesda) 8, 1971-83.
198. Sardell J.M. & Kirkpatrick M. (2020) Sex differences in the recombination landscape. Am Nat 195, 361-79.
199. Scarbrough J.R., Cowles D.L. & Carter R.L. (2002) Microsatellites in Shrimp Species. In: Modern Approaches to the Study of Crustacea (ed. by E. Escobar-Briones & F. Alvarez), pp. 291-9. Springer US, Boston, MA.
200. Sekino M. & Kakehi S. (2012) PARFEX v1.0: an EXCEL-based software package for parentage allocation. Conserv. Genet. Resour. 4, 275–8.
201. Shao C., Bao B., Xie Z., Chen X., Li B., Jia X., Yao Q., Orti G., Li W., Li X., Hamre K., Xu J., Wang L., Chen F., Tian Y., Schreiber A.M., Wang N., Wei F., Zhang J., Dong Z., Gao L., Gai J., Sakamoto T., Mo S., Chen W., Shi Q., Li H., Xiu Y., Li Y., Xu W., Shi Z., Zhang G., Power D.M., Wang Q., Schartl M. & Chen S. (2017) The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. Nat. Genet. 49, 119-24.
202. Shao C., Niu Y., Rastas P., Liu Y., Xie Z., Li H., Wang L., Jiang Y., Tai S., Tian Y., Sakamoto T. & Chen S. (2015) Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (*Paralichthys olivaceus*): applications to QTL mapping of Vibrio anguillarum disease resistance and comparative genomic analysis. DNA Res 22, 161-70.
203. Shi W., Chen S., Kong X., Si L., Gong L., Zhang Y. & Yu H. (2018) Flatfish monophyly refereed by the relationship of Psettodes in Carangimorphariae. BMC Genomics 19, 400.
204. Simao F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V. & Zdobnov E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210-2.
205. Simpson J.T. & Durbin R. (2012) Efficient de novo assembly of large genomes using compressed data structures. Genome Res 22, 549-56.
206. Sole X., Guino E., Valls J., Iniesta R. & Moreno V. (2006) SNPStats: a web tool for the analysis of association studies. Bioinformatics 22, 1928-9.
207. Sonesson A.K. & Meuwissen T.H. (2009) Testing strategies for genomic selection in aquaculture breeding programs. Genet Sel Evol 41, 37.
208. Song W., Li Y., Zhao Y., Liu Y., Niu Y., Pang R., Miao G., Liao X., Shao C., Gao F. & Chen S. (2012a) Construction of a high-density microsatellite genetic linkage map and mapping of sexual and growth-related traits in half-smooth tongue sole (*Cynoglossus semilaevis*). PLoS One 7, e52097.
209. Song W., Pang R., Niu Y., Gao F., Zhao Y., Zhang J., Sun J., Shao C., Liao X., Wang L., Tian Y. & Chen S. (2012b) Construction of high-density genetic linkage maps and mapping of growth-related quantitative trail loci in the Japanese flounder (*Paralichthys olivaceus*). PLoS One 7, e50404.
210. Souissi A., Bonhomme F., Manchado M., Bahri-Sfar L. & Gagnaire P.A. (2018) Genomic and geographic footprints of differential introgression between two divergent fish species (Solea spp.). Heredity (Edinb) 121, 579-93.

211. Srivastava S., Puri D., Garapati H.S., Dhawan J. & Mishra R.K. (2013) Vertebrate GAGA factor associated insulator elements demarcate homeotic genes in the HOX clusters. Epigenetics Chromatin 6, 8.
212. Stapley J., Feulner P.G.D., Johnston S.E., Santure A.W. & Smadja C.M. (2017) Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philos Trans R Soc Lond B Biol Sci 372.
213. Sundaray J.K., Rasal K.D., Chakrapani V., Swain P., Kumar D., Ninawe A.S., Nandi S. & Jayasankar P. (2016) Simple sequence repeats (SSRs) markers in fish genomic research and their acceleration via next-generation sequencing and computational approaches. Aquac Int. 24, 1089-102.
214. Tondera D., Santel A., Schwarzer R., Dames S., Giese K., Klippel A. & Kaufmann J. (2004) Knockdown of MTP18, a novel phosphatidylinositol 3-kinase-dependent protein, affects mitochondrial morphology and induces apoptosis. J Biol Chem 279, 31544-55.
215. Toro M.A. & López-Fanjúl C. (2007) Diseño de programa de mejora genética en acuicultura. In: Genética y Genómica en Acuicultura (ed. by C.S.d.I. Científicas), pp. 185-211. J.Espinosa. Coord. P. Martínez & A. Figueras., Madrid.
216. Trisciuoglio D., Di Martile M. & Del Bufalo D. (2018) Emerging Role of Histone Acetyltransferase in Stem Cells and Cancer. Stem Cells Int 2018, 8908751.
217. Van Oosterhout C., Hutchinson W.F., Wills D.P.M. & Shipley P. (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes 4, 535-8.
218. Vandeputte M., Gagnaire P.A. & Allal F. (2019) The European sea bass: a key marine fish model in the wild and in aquaculture. Anim Genet 50, 195-206.
219. Vandeputte M., Kocour M., Mauger S., Rodina M., Launay A., Gela D., Dupont-Nivet M., Hulak M. & Linhart O. (2008) Genetic variation for growth at one and two summers of age in the common carp (*Cyprinus carpio* L.): heritability estimates and response to selection. Aquaculture 277, 7-13.
220. Vandeputte M., Mauger S. & Dupont-Nivet M. (2006) An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. Mol Ecol Notes 6, 265-7.
221. Vaser R., Sovic I., Nagarajan N. & Sikic M. (2017) Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27, 737-46.
222. Vega L., Díaz E., Cross I. & Rebordinos L. (2002) Caracterizaciones citogenética e isoenzimática del lenguado *Solea senegalensis* Kaup, 1858. Boletín Inst Español Oceanogr. 18, 1-6.
223. Verbeek E.C., Bakker I.M., Bevova M.R., Bochdanovits Z., Rizzu P., Sondervan D., Willemsen G., de Geus E.J., Smit J.H., Penninx B.W., Boomsma D.I., Hoogendijk W.J. & Heutink P. (2012) A fine-mapping study of 7 top scoring genes from a GWAS for major depressive disorder. PLoS One 7, e37384.
224. Vignal A., Milan D., SanCristobal M. & Eggen A. (2002) A review on SNP and other types of molecular markers and their use in animal genetics. Genet Sel Evol 34, 275-305.
225. Viñas J., Asensio E., J.P. C. & Piferrer F. (2012) Gonadal sex differentiation in the Senegalese sole (*Solea senegalensis*) and first data on the experimental manipulation of its sex ratios. Aquaculture, 384-6.
226. Wang L., Wan Z.Y., Bai B., Huang S.Q., Chua E., Lee M., Pang H.Y., Wen Y.F., Liu P., Liu F., Sun F., Lin G., Ye B.Q. & Yue G.H. (2015a) Construction of a high-density linkage map and fine mapping of QTL for growth in Asian seabass. Sci Rep 5, 16358.

227. Wang W., Hu Y., Ma Y., Xu L., Guan J. & Kong J. (2015b) High-density genetic linkage mapping in turbot (*Scophthalmus maximus* L.) based on SNP markers and major sex- and growth-related regions detection. PLoS One 10, e0120410.
228. Waterhouse R.M., Seppey M., Simao F.A., Manni M., Ioannidis P., Klioutchnikov G., Kriventseva E.V. & Zdobnov E.M. (2018) BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol Biol Evol 35, 543-8.
229. Wood D.E. & Salzberg S.L. (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15, R46.
230. Wright A.E., Darolti I., Bloch N.I., Oostra V., Sandkam B., Buechel S.D., Kolm N., Breden F., Vicoso B. & Mank J.E. (2017) Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. Nat Commun 8, 14251.
231. Xu X.W., Shao C.W., Xu H., Zhou Q., You F., Wang N., Li W.L., Li M. & Chen S.L. (2020) Draft genomes of female and male turbot *Scophthalmus maximus*. Sci Data 7, 90.
232. Yang L., Wang Y., Wang T., Duan S., Dong Y., Zhang Y. & He S. (2019) A Chromosome-Scale Reference Assembly of a Tibetan Loach, Triplophysa siluroides. Front Genet 10, 991.
233. Yu Y., Ouyang Y. & Yao W. (2018) shinyCircos: an R/Shiny application for interactive creation of Circos plot. Bioinformatics 34, 1229-31.
234. Yue G.H. (2013) Recent advances of genome mapping and marker-assisted selection in aquaculture. FISH and FISHERIES 15, 376-96.
235. Zalapa J.E., Cuevas H., Zhu H., Steffan S., Senalik D., Zeldin E., McCown B., Harbut R. & Simon P. (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am J Bot 99, 193-208.
236. Zarouri B., Vargas A.M., Gaforio L., Aller M., de Andres M.T. & Cabezas J.A. (2015) Whole-genome genotyping of grape using a panel of microsatellite multiplex PCRs. Tree Genet. Genomes 11.
237. Zenger K.R., Khatkar M.S., Jones D.B., Khalilisamani N., Jerry D.R. & Raadsma H.W. (2019) Genomic Selection in Aquaculture: Application, Limitations and Opportunities With Special Reference to Marine Shrimp and Pearl Oysters. Frontiers in Genetics 9.
238. Zhang J., Chu W. & Fu G. (2009) DNA microarray technology and its application in fish biology and aquaculture. Front. Biol. 4, 305-13.
239. Zhang J., Zuo Z., Zhu W., Sun P. & Wang C. (2013) Sex-different effects of tributyltin on brain aromatase, estrogen receptor and retinoid X receptor gene expression in rockfish (Sebastiscus marmoratus). Mar Environ Res 90, 113-8.
240. Zimin A.V., Marcais G., Puiu D., Roberts M., Salzberg S.L. & Yorke J.A. (2013) The MaSuRCA genome assembler. Bioinformatics 29, 2669-77.
241. Zimin A.V., Puiu D., Luo M.C., Zhu T., Koren S., Marcais G., Yorke J.A., Dvorak J. & Salzberg S.L. (2017) Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Res. 27, 787-92.

# ANEXO I
# PUBLICATIONS

scientific reports

Check for updates

OPEN

# Chromosome anchoring in Senegalese sole (*Solea senegalensis*) reveals sex-associated markers and genome rearrangements in flatfish

Israel Guerrero-Cózar[1], Jessica Gomez-Garrido[2], Concha Berbel[1], Juan F. Martinez-Blanch[3], Tyler Alioto[2,4], M. Gonzalo Claros[5,6,7,8], Pierre-Alexandre Gagnaire[9] & Manuel Manchado[1,10]✉

**The integration of physical and high-density genetic maps is a very useful approach to achieve chromosome-level genome assemblies. Here, the genome of a male Senegalese sole (*Solea senegalensis*) was de novo assembled and the contigs were anchored to a high-quality genetic map for chromosome-level scaffolding. Hybrid assembled genome was 609.3 Mb long and contained 3403 contigs with a N50 of 513 kb. The linkage map was constructed using 16,287 informative SNPs derived from ddRAD sequencing in 327 sole individuals from five families. Markers were assigned to 21 linkage groups with an average number of 21.9 markers per megabase. The anchoring of the physical to the genetic map positioned 1563 contigs into 21 pseudo-chromosomes covering 548.6 Mb. Comparison of genetic and physical distances indicated that the average genome-wide recombination rate was 0.23 cM/Mb and the female-to-male ratio 1.49 (female map length: 2,698.4 cM, male: 2,036.6 cM). Genomic recombination landscapes were different between sexes with crossovers mainly concentrated toward the telomeres in males while they were more uniformly distributed in females. A GWAS analysis using seven families identified 30 significant sex-associated SNP markers located in linkage group 18. The follicle-stimulating hormone receptor appeared as the most promising locus associated with sex within a region with very low recombination rates. An incomplete penetrance of sex markers with males as the heterogametic sex was determined. An interspecific comparison with other Pleuronectiformes genomes identified a high sequence similarity between homologous chromosomes, and several chromosomal rearrangements including a lineage-specific Robertsonian fusion in *S. senegalensis*.**

Genetic maps represent essential tools for genomic research in aquaculture. Originally, linkage mapping studies were mainly based on microsatellite (SSR) and AFLP markers[1,2]; nevertheless, they recently reached a milestone with the development of genotyping methods based on cost-effective massive parallel sequencing. The genomic revolution has made single-nucleotide polymorphisms (SNPs) very popular, opening up access to a simple

[1]IFAPA Centro El Toruño, Junta de Andalucía, Camino Tiro Pichón s/n, 11500 El Puerto de Santa María, Cádiz, Spain. [2]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain. [3]Biopolis S.L.-ADM, Parc Cientific Universidad De Valencia, Edif. 2, C/ Catedrático Agustín Escardino Benlloch, 9, 46980 Paterna, Spain. [4]Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain. [5]Department of Molecular Biology and Biochemistry, Universidad de Málaga, 29071 Málaga, Spain. [6]CIBER de Enfermedades Raras (CIBERER), 29071 Málaga, Spain. [7]Institute of Biomedical Research in Málaga (IBIMA), IBIMA-RARE, 29010 Málaga, Spain. [8]Instituto de Hortofruticultura Subtropical Y Mediterránea (IHSM-UMA-CSIC), 29010 Málaga, Spain. [9]ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France. [10]Crecimiento Azul, Centro IFAPA El Toruño, Unidad Asociada al CSIC, El Puerto de Santa María, Spain. ✉email: manuel.manchado@juntadeandalucia.es

biallelic marker with a wide distribution and high abundance across the genome. As consequence, an increasing number of high-density genetic maps is nowadays reported in non-model organisms including aquaculture fish[3,4]. These maps have proven to be useful to provide new clues on genome evolution and speciation between closely related lineages, and to unravel the genetic architecture of both simple Mendelian and complex quantitative traits in many fish species, thus facilitating marker-assisted selection in aquaculture[5,6]. More recently, a new application of high-density linkage maps as backbones to anchor de novo genome assemblies into pseudo-chromosomes has become more widespread[7,8]. Although long-read sequences have significantly enhanced the average size of scaffolds in de novo assembled genomes[9], the total number of scaffolds are still far beyond the expected number of chromosomes. The large arrays of repeated sequences and the degree of conservation for some tandem repeats families widely distributed across the genome still remain a major obstacle for most de novo assembly algorithms, resulting in fragmented scaffolds or even misassembled sequences within chimeric contigs. Linkage maps thus provide highly valuable tools to anchor physical maps into pseudo-chromosomes, while enabling the identification of chimeric or misassembled contigs towards enhancing the quality of new genome assemblies[7].

Flatfish (Pleuronectiformes) is an attractive group of fish that have long been investigated due to the drastic morphological, physiological and behavioural remodelling changes that occur during metamorphosis from a pelagic larva to a benthic juvenile stage. Several flatfish species are worldwide exploited in fisheries and aquaculture, thus representing an important resource for human consumption. This taxonomic group diverged from carangimorphs in the early Paleocene, and underwent a major diversification in the middle Paleocene[10]. Cytogenetic studies have suggested that the Pleuronectiformes ancestor should have 2n = 48 chromosomes in agreement with the most frequent number of chromosomes found in the sister clade Carangidae, and in the most deep-branching flatfish families (Pleuronectidae and Paralichthyidae)[11]. However, the number of chromosomes in flatfish encompasses a wide range varying from 2n = 26 to 2n = 50[11,12]. An intense cascade of Robertsonian rearrangements and pericentromeric inversions seems to have shaped flatfish genome evolution, especially reducing the chromosome number in most recently diverged families of Soleidae, Cynoglossidae and Achiridae[11]. A recent comparison of the turbot genome with other fish assemblies clearly pointed out the high degree of conserved synteny across chromosomes in Pleuronectiformes, although with high rates of intrachromosomal reorganisations. Moreover, some chromosome fusions identified through comparative mapping are thought to have given arise to a new karyotype organization in turbot[3]. Hence, integrated genetic and physical maps are important genomic resources to understand chromosome evolution in flatfish.

The Senegalese sole is an important flatfish in aquaculture and fisheries. A genetic linkage map based on 129 SSRs grouped into 27 linkage groups (LG) was previously reported[13]. Moreover, an integrated map using BAC clones and repetitive DNA families was also developed using a multiple fluorescence in situ hybridization (mFISH) technique with at least one BAC mapped to each chromosome arm[14]. This cytogenetic study evidenced a lack of heteromorphic sex chromosomes and identified the largest metacentric chromosome to result from a Robertsonian fusion of two acrocentric chromosomes during flatfish evolution[15,16]. Moreover, a preliminary draft genome sequence of a female Senegalese sole was reported (600.3 Mb, N50 of 85 kb), and then further improved with a hybrid assembly using Nanopore and Illumina reads (608 Mb long, N50 of 340 kb)[17,18]. This genome information was used to design whole-genome multiplex PCR and create a new integrated SSR map with 234 markers. Nevertheless, further efforts are required to better assemble and anchor scaffolds onto the 21 expected chromosomes, and to better understand the genomic architecture of sex-determination.

The aim of this study was to: (1) generate an improved de novo assembly of a male Senegalese sole based on a combination of long and short read sequencing; (2) build a high-density genetic map using ddRAD markers; (3) anchor the physical to the genetic map in order to (4) improve the scaffolding of the reference genome assembly; (5) estimate genome-wide variation in recombination rates; and (6) carry out GWAS analysis to identify sex-associated markers and intra- and interspecific comparative mapping to better understand the evolutionary history of chromosome rearrangements in flatfish.

## Material and methods

**Animals.** Soles used for the preparation of ddRAD libraries and sequencing were selected from the genetic breeding program carried out by the IFAPA in collaboration with a commercial aquaculture company (CUPIMAR S.A.). Production of families used in this study, genotyping and parentage assignment were previously published[19,20]. Five families (three full-sib and two maternal half-sib families) containing between 48 and 96 individuals per family (total n = 356) were selected to construct the genetic linkage map (Table 1). Moreover, seven families with sex ratios close to 1:1 were selected for genome-wide association analysis (GWAS). Average weight and length of each family are depicted in Table 1. As genotyping of parents was also required to build the genetic map, five fathers and three mothers involved in family production were sampled for blood by puncturing in the caudal vein using a heparinized syringe, adding heparin (100 mU) and keeping at − 20 °C until use. To obtain high-molecular weight genomic DNA for genome sequencing, a wild male from the broodstock (weight higher than 2 kg; code Sse05_10M) was sampled for blood as indicated above.

All procedures were authorized by the Bioethics and Animal Welfare Committee of IFAPA and given the registration number 10/06/2016/101 by the National authorities for regulation of animal care and experimentation. The study was carried out in compliance with the ARRIVE guidelines and all procedures were performed in accordance with Spanish national (RD 53/2013) and European Union legislation for animal care and experimentation (Directive 86\609\EU).

**Genome sequencing and assembly.** Methods for genome sequencing and assembly are fully described in "Supplementary method". Briefly, high-molecular weight genomic DNA was prepared from heparinized whole blood using the MagAttract HMW DNA kit (Qiagen). Once confirmed quality, four libraries were pre-

| Family name | Use | Parents | Weight | Length | n | nQ | Final |
|---|---|---|---|---|---|---|---|
| Fam1 | LM/A | F1/M1 | 161.6 ± 94.3 | 20.6 ± 4.0 | 76 | 76 | 73 |
| Fam2 | LM | F2/M2 | 244.5 ± 157.8 | 22.7 ± 4.4 | 95 | 95 | 90 |
| Fam3 | A | F3/M3 | 219.3 ± 95.9 | 22.4 ± 3.5 | 68 | 67 | 65 |
| Fam4 | A | F4/M4 | 460.8 ± 195.4 | 27.8 ± 4.1 | 99 | 79 | 77 |
| Fam5 | LM/A | F5/M5 | 216.2 ± 67.1 | 22.5 ± 2.3 | 48 | 48 | 47 |
| Fam6 | LM/A | F6/M5 | 345.5 ± 136.2 | 25.6 ± 3.4 | 71 | 65 | 63 |
| Fam7 | LM/A | F7/M2 | 540.4 ± 211.3 | 28.6 ± 3.6 | 66 | 62 | 54 |
| Fam8 | A | F8/M1 | 129.8 ± 72.7 | 19.5 ± 3.9 | 76 | 73 | 73 |
| TotalLM | | | | | 356 | 346 | 327 |
| TotalA | | | | | 504 | 470 | 452 |

**Table 1.** Families used to construct the genetic linkage map (LM) and association study (A). Father (F) and Mother (M) of each family, the average weight and standard length at age 800 days and the number of specimens originally selected for analysis (n) are indicated. Moreover, the number of animals that passed that DNA quality analysis (nQ) and the final number of animals that passed after checking for Mendelian errors.

pared for sequencing using the Oxford nanopore Technology (ONT) MinION platform. Overall, 19.2 Gb of genome information was generated with an average read length of 4.3 kb. In parallel, the same sample was also sequenced in a NextSeq550 sequencer (Illumina, USA) that overall generated 43 Gb of sequence from 143 million reads (average length 147 nt). The main features of the libraries used during the genome assembly are presented in Supplementary Table S1. The raw read data were deposited to the NCBI Sequence Read Archive (SRA) under accession number SAMN16809702. The hybrid genome assembly was carried out using MaSuRCA *v*3.2.3[21,22] with the Illumina libraries (57.3 × coverage) and the error-corrected Nanopore reads (25.5x). The LR-hybrid assembly was characterized for completeness using Benchmarking Universal Single-Copy Orthologs (BUSCO*v*3.0.2)[23,24] containing 4,854 single-copy orthologs from actinopterygii_odb9.

**ddRAD-seq library preparation and sequencing.** Genomic DNA from the caudal fin (offspring) or whole blood (parents) were purified using the Isolate II Genomic DNA Kit (Bioline). DNA was sent to the company LifeSequencing S.L. (Valencia, Spain) and a total of 346 samples were selected for library construction (Table 1). Libraries were constructed based on the protocol described by Peterson et al.[25] using the EcoRI/NcoI enzyme combination that generated as average 24,874 SNPs per sample. Pools of libraries were loaded on a Novaseq 6000 sequencer (Illumina), following the manufacturer's instructions and the specifications mentioned above. The total number of reads generated for each library are indicated in Supplementary Table S2.

**Genetic linkage map and scaffold anchoring.** Illumina reads were processed using Stacks v2.3e[26] as indicated in "Supplementary method". To construct the map, SNPs were filtered using Plink v1.9[27] to remove markers that segregated with Mendelian errors in more than 10% of individuals. Moreover, those individuals with more than 5% of markers with Mendelian errors were removed (Supplementary Fig. S1). The final SNP dataset contained 40,041 markers from 327 individuals (Table 1) and 8 parents that were imported in LepMap3[7]. The SNPs were assigned to 21 linkage groups (named as SseLGs) corresponding to the expected number of chromosomes (2n = 42) using the "SeparateChromosomes" module. A LOD threshold of 11 and a size limit of 200 were selected as the most adequate parameters to keep an optimal number of markers grouped in the expected number of SseLGs (Fig. 1A,B). Module JoinSingles2 was run to assign additional single SNPs to existing SseLG using decreasing LOD score iterations from 10 to 5 (Fig. 1B). Finally, the genetic distances between markers on each SseLG was calculated with the OrderMarkers2 module (male, female, sex average (SA)) using the Kosambi mapping function. The resulting genetic map was visualized using the software linkagemapview[28]. Scaffolds anchoring was carried out using the Lep-Anchor program following the author's recommendation[29] and indicated in "Supplementary method".

**Genome annotation.** Genome annotation was performed by combining alignments of *Danio rerio*, *S. maximus* and *S. semilaevis* proteins, RNAseq from several tissues and developmental stages alignments and ab initio gene predictions. Annotation process is described in "Supplementary method" with a higher detail. Functional annotation was performed on the male annotated proteins with Blast2GO[30]. After performing an alignment-based strategy to determine equivalences between female and male genomes (see "Supplementary method"), the female proteins inherited the functional annotation of their male equivalences. Next, functional annotation was performed in the female genes that remained unannotated after this step. Gene Ontology (GO) enrichment was carried out with topGO in those genes that were unique to one of the genomes (Supplementary Table S3).

**Recombination rates, association analyses and cross-species comparisons.** Recombination rate variation along the genome was evaluated by comparing the consensus linkage map for both sexes and SA and the physical map of each pseudo-chromosome using MareyMap[31]. The cumulative recombination frequency (RFm) along LGs was used to infer the chromosome type as previously described[32]. GWAS analysis were carried

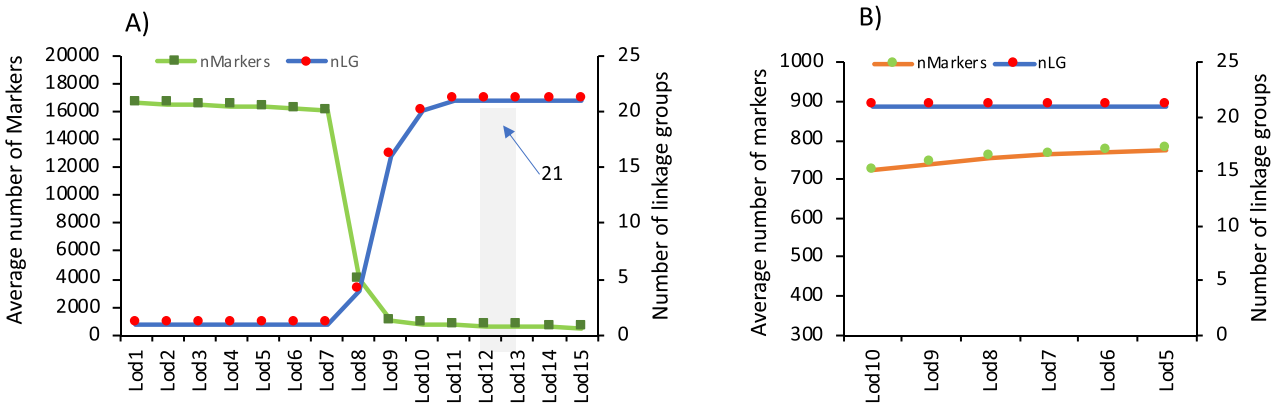


**Figure 1.** Selection of LOD score limit (Lod) to construct genetic map in LepMap3. (**A**) The average of number of markers (nMarkers) positioned in linkage groups (left Y axis) and the number of linkage groups (nLG; right Y axis) for Lod values from 1 to 15 as implemented in the "SeparateChromosomes" module. Lod11 (shaded) indicates the value selected that grouped the markers in 21 LGs. (**B**) Average number of markers recovered and added to the 21 LGs using decreasing LOD score iterations from 10 to 5 in the JoinSingles2 module.

out with seven families (Table 1) using a logistic mixed model (multi-step) approach as implemented in the R package GENABEL (v1.8–0)[33] for binary traits (Female = 0 and Male = 1). A highly detailed analysis of synteny across flatfish is beyond the scope of this study, but a chromosome alignment analysis was carried out to identify chromosomal rearrangements in flatfish using D-Genies[34]. We then used the SatsumaSynteny to compute whole-genome synteny blocks[35] that were later represented using Shinycircos[36].

## Results

**Male genome assembly and annotation.** A de novo hybrid genome for a male sole was assembled using a combination of Illumina and Nanopore long-reads. Main features about the total number of input reads used for each sequencing platform, the average read length and quality and total sequencing information used in the assembly are indicated in Supplementary Table S1. The hybrid assembly draft sequence was generated using MaSuRCA and later refined with Pilon to correct bases, mis-assemblies and filling gaps. Main statistics about the assembly are depicted in Supplementary Table S4. The new assembly consists of 3,403 contigs with a total length of 609,359,514 bp, and a N50 of 513 kb. Overall, 49.4% of contigs had a size longer than 50 kb and the largest fragment was 4.5 Mb long. The estimated gene integrity, as determined by BUSCO analysis, revealed 97.0% completeness. For comparison purposes, the assembly statistics for a recent female genome draft of *S. senegalensis*[20,20] are also shown in Supplementary Table S4. Both genome assemblies had a similar size (608–610 Mb) although the newly assembled male genome had longer contigs with higher N50 values. A dot-plot alignment using the scaffolds of both genomes indicated that with 92.8% of genomic information highly similar (> 75%) and only 5.3% had no similarity (average similarity 94%) (Fig. 2).

Assembly annotation statistics are depicted in Table 2. The number of protein-coding genes in the male assembly (27,175) was slightly lower than in the female (28,988) but with a longer mean length (7.4 vs 6.7 kb). The estimated percentages of annotated transcripts (69.4–72.1%) and gene density (45.03–47.68) were similar between both assemblies. Around 85% of the annotated genes in each assembly had an equivalent gene in the other assembly. However, a few genes were only present in one of the genomes (unique genes). Some of these gene differences might be due to genome heterozygosity and repeat content or even sex-specific genes. A GO enrichment analysis using these unique genes indicated that categories related to the cell-cycle regulation and regulation of transcription, involving canonical histones H3.2 and H4 and retinoid X receptor alpha (*rxra*), were highly significantly overrepresented in the female (p-value < $10^{-3}$). Mapping of these two histone genes on female assembly showed that they were co-localized in five scaffolds (Sosen1_s0284, Sosen1_s0324, Sosen1_s1454, Sosen1_s1522, Sosen1_s1726), four of which clustered in SseLG1 and one in SseLG16. In male, the most significant enriched categories for unique genes were skeletal system development and morphogenesis although with *P*-values > 0.001 (Supplementary Table S3). Some short, single-exonic unique genes might be the result of scaffold splitting or annotation processes. The non-coding gene annotation resulted in 23,822 female and 21,123 male transcripts, respectively. From these, 6,549 and 6,007 female and male transcripts were long non-coding RNAs (lncRNAs) and the rest short non-coding RNAs.

**ddRAD sequencing and SNP detection for genetic linkage map.** Three full-sib and two half-sib families consisting of 47 to 95 individuals were used for ddRAD analysis (Table 1). The total number of paired-end reads generated for each family ranged between 280,609,738 (F5) and 398,313,256 (F2) with an average length of 150 nt (Table 3). The average number of reads per individual in each family varied between 6,444,752 (F1) and 11,692,072 (F5) (Table 3 and Supplementary Table S2). For parents, the average number of reads was 8,847,913.

The new assembled male genome was used as reference to map the ddRAD reads. The average fraction of primary alignments onto this reference genome ranged between 88.04 (F6) and 89.71% (F2). An average of 10.5% of reads had insufficient mapping qualities or excessively soft-clipped primary alignments while less than 0.34%

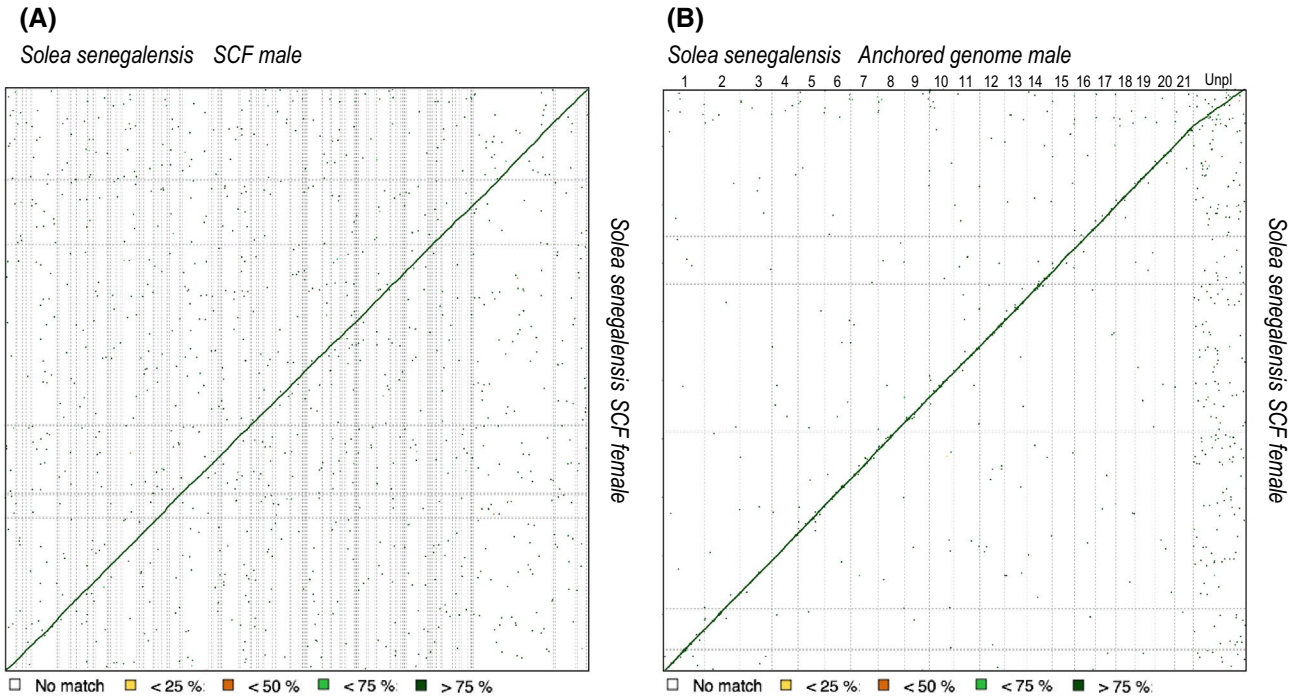


**Figure 2.** Dot plot comparison of scaffolds (SCF) assembled (**A**) or 21 pseudo-chromosomes (**B**) in the male with respect to SCF in the female. Scale is indicated below.

| | Male | Female[#] |
|---|---|---|
| Repeat content | 23.55% | 23.41% |
| Number of protein-coding genes | 27,175 | 28,988 |
| Median gene length (bp) | 7,368 | 6,721 |
| Number of transcripts | 50,133 | 51,844 |
| Number of exons | 303,132 | 307,753 |
| Number of coding exons | 284,414 | 288,788 |
| Coding GC content | 52.67% | 52.57% |
| Median UTR length (bp) | 1,231 | 1,222 |
| Median intron length (bp) | 388 | 371 |
| Exons/transcript | 11.88 | 11,53 |
| Transcripts/gene | 1.84 | 1.79 |
| Multi-exonic transcripts | 0.956 | 0.941 |
| Gene density (gene/Mb) | 45.026 | 47.679 |
| Functionally annotated transcripts | 36,130 (72.1%) | 35,999 (69.4%) |
| Unique genes | 3,806 (14%) | 4,643 (16%) |
| non-conding RNAs | 21,123 | 23,822 |

**Table 2.** Summary annotation statistics for male and female assemblies. Annotation pipeline is described with more details in "Supplementary method". [#]Sequence deposited in figshare https://doi.org/10.6084/m9.figshare.12472100.v1.

were unmapped. A total of 199,188 ddRAD *loci* were reconstructed with an average number of loci per sample ranging between 23,828 (F1) and 30,550 (F7) and a mean insert length of 330.7 bp. The effective coverage per sample was 193.3 ± 110.4 (ranging from 146 to 242 between families) and the estimated mean number of sites per locus was 242.8 (Table 3).

**Construction of a linkage genetic map and anchoring to physical map.** To construct the genetic map, only those SNPs detectable in at least 80% of samples with a coverage of 10 reads per sample were considered. Moreover, SNPs with a significant deviation from Mendelian segregation were also removed (a total of 2,439 markers, 5.7% SNPs). By family, the number of markers with Mendelian errors ranged from 1.5 to 1.7% (Supplementary Fig. S1). Moreover, those animals with markers that had more than 5% of Mendelian errors (19

| | n | Total reads family | Av. raw reads | Av. reads stacks | PA (%) | Unmapped | loci | mean cov | n_gts |
|---|---|---|---|---|---|---|---|---|---|
| F1 | 76 | 244,900,564 | 6,444,752 | 6,215,911 | 88.23 | 0.34% | 23,828 | 146 | 22,040 |
| F2 | 95 | 398,313,256 | 8,385,542 | 8,090,267 | 89.71 | 0.33% | 24,978 | 190 | 22,823 |
| F3 | 67 | 226,072,540 | 6,649,192 | 6,090,258 | 86.20 | 0.32% | 26,068 | 132 | 24,054 |
| F4 | 79 | 248,271,546 | 6,130,162 | 5,972,512 | 87.74 | 0.33% | 25,525 | 135 | 23,157 |
| F5 | 48 | 280,609,738 | 11,692,072 | 11,384,985 | 88.13 | 0.33% | 30,005 | 237 | 27,011 |
| F6 | 65 | 363,499,961 | 11,184,614 | 10,899,007 | 88.04 | 0.31% | 27,742 | 242 | 24,883 |
| F7 | 62 | 337,573,225 | 10,889,459 | 10,627,007 | 88.93 | 0.34% | 30,550 | 226 | 26,773 |
| F8 | 73 | 447,768,745 | 12,267,637 | 11,674,383 | 89.42 | 0.33% | 28,002 | 260 | 25,371 |
| Parents | 8 | 39,815,609 | 8,847,913 | 8,323,338 | 86.08 | 0.36% | 17,632 | 242 | 15,898 |

**Table 3.** Main statistics of ddRAD libraries, mapping and SNP detection. The total number of individuals analysed (n), the total reads per family, the average number of paired-end reads per individual, the average number reads used by stacks, the % of primary alignment and unmapped reads, number of *loci*, effective coverage, and number of genotypes (n_gts).

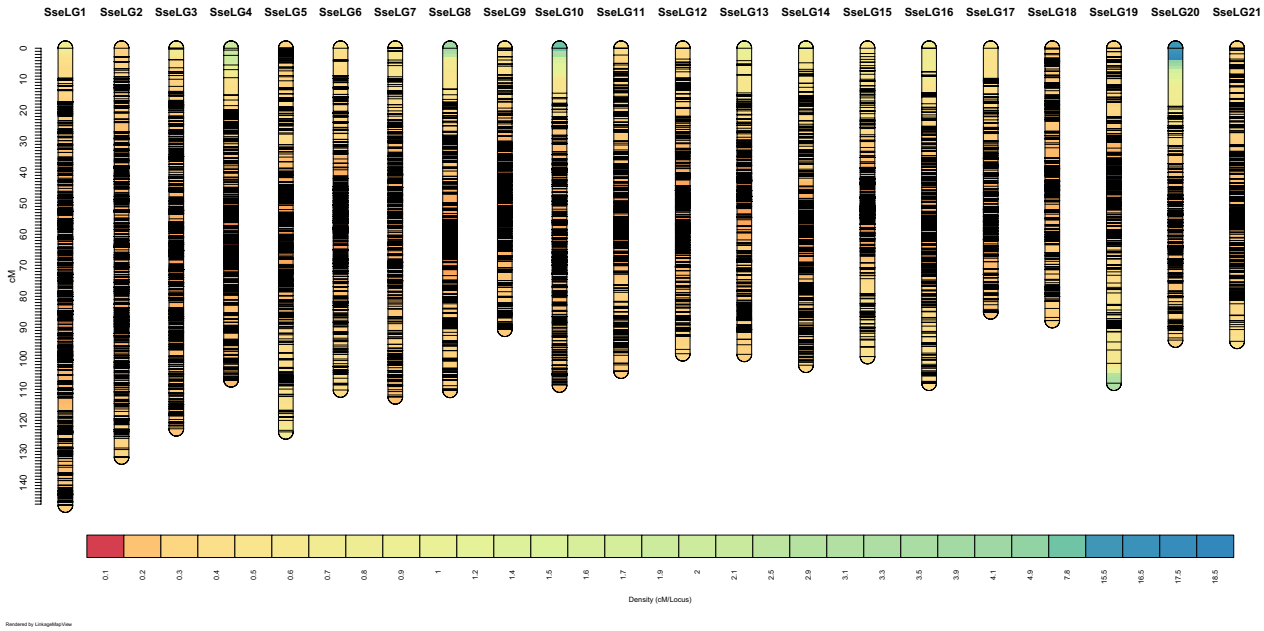


**Figure 3.** Genetic distance (cM) and SNP distribution across 21 linkage groups (SseLG) of the Senegalese sole.

specimens) were also removed. Overall, the final dataset contained 40,041 SNPs segregating in eight parents and their 327 offspring.

For linkage analysis, the ParentCall2 module retained only 16,287 informative markers after checking for segregation distortion ($P < 0.05$). Markers grouped into 21 SseLGs (via the SeparateChromosomes2 module) with a LOD = 11 (Fig. 1), which is consistent with the number of chromosomes in *S. senegalensis*. Each SseLG contained between 530 and 1,337 markers with an average number of 21.9 markers per Mb (Fig. 3, Table 4 "Anchoring genetic map and physical map"). In total, the genetic map allowed the anchoring and positioning of 1,665 out of 3,403 total contigs, ranging between 50 to 129 contigs in each SseLG. The genome sequence positioned on the linkage map was larger (746.3 bp) than the assembly size, mainly due to the presence of chimeric contigs (n = 133) positioned in various chromosomes.

**Rescaffolding of reference genome with the genetic map.** SNP marker information was further used for fine-scale correction of genome contigs to build 21 pseudo-chromosomes. After masking the repetitive sequences, the contigs were orientated and sorted within each SseLG (Table 4 "Genome re-scaffolding"). The total number of positioned contigs reduced from 1,665 to 1,563. Lep-anchor corrected the contig errors removing six contigs, splitting another 105 into two fragments, 20 in three fragments, and two in more than four fragments. After these corrections, the total number of markers assigned to the SseLGs decreased by 1.3% (16,075 SNPs) and 212 markers were moved to unplaced with an average density of 10.3 markers per contig. After these corrections, 548.6 Mb out of the 610.4 Mb total assembly length (89.9%) were assigned to the 21 SseLGs and only 61.9 Mb remained as unanchored (Table 4). The total map length was 2,408.1 cM, SseLG1 was the largest group (42,924,012 bp and 147.3 cM) and SseLG4 showed the highest marker density per megabase (33.1). The

| | Anchoring genetic map and physical map | | | | | Genome re-scaffolding | | | | | | Marker refining | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Length (bp) | nMar | nCont | ACL | M/Mb | Length(bp) | NM | nCont | ACL | L(cM) | M/Mb | NMar | Mb/cM | M/Mb |
| **1** | 59,220,137 | 1,337 | 129 | 459,071 | 22.6 | 42,924,012 | 1,323 | 124 | 343,392 | 147.3 | 30.8 | 1,296 | 0.29 | 30.2 |
| **2** | 42,658,310 | 1,054 | 91 | 468,773 | 24.7 | 36,396,255 | 1,046 | 88 | 413,594 | 131.8 | 28.7 | 1,032 | 0.28 | 28.4 |
| **3** | 47,587,809 | 1,015 | 85 | 559,857 | 21.3 | 33,319,822 | 1,006 | 80 | 416,498 | 136.6 | 30.2 | 978 | 0.24 | 29.4 |
| **4** | 42,630,187 | 920 | 83 | 513,617 | 21.6 | 27,129,084 | 899 | 73 | 366,609 | 106.9 | 33.1 | 885 | 0.25 | 32.6 |
| **5** | 32,366,427 | 891 | 86 | 376,354 | 27.5 | 27,692,037 | 872 | 78 | 350,532 | 142.5 | 31.5 | 811 | 0.19 | 29.3 |
| **6** | 34,539,569 | 864 | 80 | 431,745 | 25.0 | 26,866,643 | 860 | 77 | 348,917 | 114.0 | 32.0 | 832 | 0.24 | 31.0 |
| **7** | 36,891,773 | 849 | 87 | 424,043 | 23.0 | 28,334,760 | 836 | 77 | 367,984 | 133.8 | 29.5 | 795 | 0.21 | 28.1 |
| **8** | 36,615,909 | 784 | 86 | 425,766 | 21.4 | 27,361,452 | 769 | 82 | 333,676 | 119.3 | 28.1 | 756 | 0.23 | 27.6 |
| **9** | 32,328,246 | 804 | 65 | 497,358 | 24.9 | 25,679,769 | 802 | 63 | 407,615 | 105.1 | 31.2 | 765 | 0.24 | 29.8 |
| **10** | 35,518,751 | 768 | 88 | 403,622 | 21.6 | 25,170,845 | 762 | 84 | 299,653 | 113.7 | 30.3 | 748 | 0.22 | 29.7 |
| **11** | 37,595,336 | 780 | 99 | 379,751 | 20.7 | 26,846,769 | 769 | 93 | 288,675 | 126.2 | 28.6 | 732 | 0.21 | 27.3 |
| **12** | 37,197,923 | 763 | 80 | 464,974 | 20.5 | 25,840,656 | 752 | 77 | 335,593 | 98.5 | 29.1 | 731 | 0.26 | 28.3 |
| **13** | 34,656,556 | 665 | 50 | 693,131 | 19.2 | 23,154,965 | 658 | 48 | 482,395 | 98.7 | 28.4 | 637 | 0.24 | 27.5 |
| **14** | 33,597,656 | 668 | 76 | 442,074 | 19.9 | 26,091,242 | 665 | 74 | 352,584 | 109.5 | 25.5 | 637 | 0.24 | 24.4 |
| **15** | 36,416,189 | 644 | 66 | 551,760 | 17.7 | 22,903,974 | 632 | 59 | 388,203 | 113.1 | 27.6 | 601 | 0.20 | 26.2 |
| **16** | 26,721,177 | 630 | 58 | 460,710 | 23.6 | 21,637,702 | 618 | 52 | 416,110 | 108.0 | 28.6 | 602 | 0.20 | 27.8 |
| **17** | 30,251,165 | 616 | 79 | 382,926 | 20.4 | 21,095,432 | 610 | 75 | 277,572 | 103.3 | 28.9 | 563 | 0.20 | 26.7 |
| **18** | 24,300,965 | 587 | 62 | 391,951 | 24.2 | 19,718,726 | 577 | 57 | 345,943 | 87.8 | 29.3 | 561 | 0.23 | 28.5 |
| **19** | 36,478,108 | 584 | 75 | 486,375 | 16.0 | 21,051,312 | 575 | 70 | 296,497 | 108.0 | 27.3 | 562 | 0.20 | 26.7 |
| **20** | 24,034,263 | 534 | 62 | 387,649 | 22.2 | 20,166,255 | 530 | 62 | 325,262 | 105.7 | 26.3 | 497 | 0.19 | 24.6 |
| **21** | 24,720,343 | 530 | 78 | 316,928 | 21.4 | 19,202,697 | 514 | 70 | 270,461 | 98.3 | 26.8 | 490 | 0.20 | 25.5 |
| **ST** | 746,326,799 | 16,287 | 1,665 | 453,259 | 21.9 | 548,584,409 | 16,075 | 1,563 | 349,640 | 2,408.1 | 29.3 | 15,511 | 0.23 | 28.3 |
| Not-anchored | | | 1,738 | | | 61,859,804 | 212 | 1,840 | | | | 776 | | |
| **Total** | 746,326,799 | 16,287 | 3,403 | 453,259 | 21.9 | 610,444,213 | 16,287 | 3,403 | | | | 16,287 | | |

**Table 4.** Information for anchored physical map (LepMap3 step), after genome re-scaffolding (Lep-anchor3 step) and after removal of markers with discrepancies between genetic and physical maps (MareyMap step). The physical (bp) and genetic (cM) length of each linkage group, number of markers (nMar), number of contigs (nCon), average contig length (ACL), marker density density (markers per megabase; M/Mb) and the ratio physical to genetic length (Mb/cM) for sex-average genetic-physical map are indicated.

average marker interval reached 0.155 cM. A further refining of anchored markers was carried out through the comparison of physical and genetic distance in MareyMap. The average genome-wide recombination rate (RR) was 4.35 cM/Mb (ranging between 3.45 and 5.26 cM/Mb among chromosomes) (Table 4 "Marker refining"). An alignment of the anchored and refined reference male genome with the scaffolds of the female assembly (Fig. 2B) slightly increased to 93.2% the regions with more than 75% similarity and provided a clear sequence alignment in the diagonal with only dispersion in unplaced scaffolds.

**Analysis of recombination rates.** Consensus genetic maps for female and male were 2,698.4 cM (15,022 markers) and 2,036.6 cM (15,390 markers), respectively. These differences in map size were observable for the 21 SseLGs (Fig. 4A and Table 5). Overall, the female-to-male ratio (F:M) for genetic distances was 1.32, ranging from 1.08 (SseLG15) to 1.77 (SseLG5) (Table 5). The genetic map length of chromosomes was highly positively correlated with their physical length in both males ($r = 0.43$) and females ($r = 0.60$) (Fig. 4B). The average genome-wide RR was estimated $3.02 \pm 0.37$ cM/Mb in males and $4.51 \pm 0.57$ cM/Mb in females (Table 5). The overall female-to-male ratio (F: M) for RR was 1.49, ranging from 1.43 to 1.90 across chromosomes. In the case of males, SseLG12 showed the lowest (2.47 cM/Mb) and SseLG16 the highest (3.60) mean RR values. In females, SseLG4 had the lowest (3.57 cM/Mb) and SseLG5 the highest (5.65 cM/Mb) mean RR values.

The local RR value as estimated by the relative distance to the nearest telomere was clearly different between males and females. High RR values were mainly concentrated close to the telomeres in males (Fig. 5A), while they were more uniformly distributed in females with higher RR being found around 15% of the distance to the nearest telomere (Fig. 5B). This was illustrated by contrasted chromosomal RR landscapes between males and females, as shown Fig. 5C,D for SseLG1 (landscape for all SseLGs are represented in the Supplementary Fig. S2 for males and Supplementary Fig. S3 for females). We detected some regions within SseLGs (i.e. 5, 11, 13, 14, 15, 18) with very low RR. In the case of SsseLG18, partially restricted male or female RR was detected in the region comprised between 9.5 and 10.9 Mb. This region had very low RR in males (1.2) and females (0.6) compared with average SseLG18 (3.0 and 4.9 RR, respectively). Cumulative RR crossed between both sexes around chromosomal position 10 Mb with female RR closed to zero in 10.8–10.9 Mb (Fig. 6, Supplementary Fig. S2 and S3). Moreover, recombination frequencies were used to describe and classify chromosome morphologies.

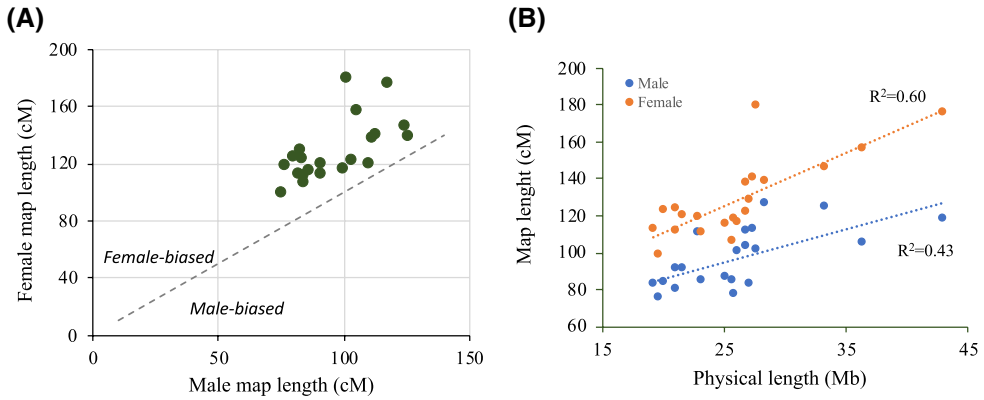


**Figure 4.** Comparison of male and female genetic maps. (**A**) Male *vs* female linkage groups lengths (cM) for the 21 Senegalese sole chromosomes. All chromosomes exhibit female-biased recombination. (**B**) Correlation between recombination map and physical map lengths in both males (blue) and females (orange). The determination coefficient $R^2$ is shown separately for each sex.

| | Male genetic map | | | | Female genetic map | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | nMar | L(cM) | Mb/cM | M/Mb | nMar | Length (cM) | Mb/cM | M/Mb | F:M (cM) | MRR | FRR | F/M (RR) |
| **1** | 1,297 | 117.7 | 0.37 | 30.2 | 1,254 | 175.7 | 0.24 | 29.2 | 1.49 | 2.56 | 4.05 | 1.58 |
| **2** | 1,027 | 105.5 | 0.35 | 28.2 | 998 | 156.2 | 0.23 | 27.4 | 1.48 | 2.64 | 4.09 | 1.55 |
| **3** | 976 | 124.9 | 0.27 | 29.3 | 962 | 145.9 | 0.23 | 28.9 | 1.17 | 3.05 | 4.15 | 1.36 |
| **4** | 881 | 83.3 | 0.33 | 32.5 | 868 | 128.8 | 0.21 | 32.0 | 1.55 | 2.58 | 3.57 | 1.38 |
| **5** | 811 | 101.4 | 0.27 | 29.3 | 811 | 179.5 | 0.15 | 29.3 | 1.77 | 3.38 | 5.65 | 1.67 |
| **6** | 833 | 103.6 | 0.26 | 31 | 814 | 122.2 | 0.22 | 30.3 | 1.18 | 3.24 | 4.50 | 1.39 |
| **7** | 786 | 126.3 | 0.22 | 27.7 | 777 | 138.2 | 0.21 | 27.4 | 1.09 | 2.73 | 4.75 | 1.74 |
| **8** | 737 | 112.9 | 0.24 | 26.9 | 758 | 140 | 0.20 | 27.7 | 1.24 | 3.15 | 3.94 | 1.25 |
| **9** | 757 | 84.8 | 0.30 | 29.5 | 762 | 106.4 | 0.24 | 29.7 | 1.25 | 2.78 | 4.12 | 1.48 |
| **10** | 732 | 86.6 | 0.29 | 29.1 | 713 | 115 | 0.22 | 28.3 | 1.33 | 3.50 | 4.50 | 1.28 |
| **11** | 722 | 111.8 | 0.24 | 26.9 | 724 | 137.6 | 0.20 | 27.0 | 1.23 | 3.16 | 3.85 | 1.22 |
| **12** | 709 | 77.3 | 0.33 | 27.4 | 677 | 118.2 | 0.22 | 26.2 | 1.53 | 2.47 | 4.70 | 1.90 |
| **13** | 628 | 84.6 | 0.27 | 27.1 | 613 | 110.7 | 0.21 | 26.5 | 1.31 | 2.76 | 4.15 | 1.50 |
| **14** | 645 | 100.3 | 0.26 | 24.7 | 608 | 116.4 | 0.22 | 23.3 | 1.16 | 2.99 | 4.10 | 1.37 |
| **15** | 609 | 110.5 | 0.21 | 26.6 | 574 | 119.3 | 0.19 | 25.1 | 1.08 | 2.64 | 4.41 | 1.67 |
| **16** | 575 | 91.6 | 0.24 | 26.6 | 580 | 119.6 | 0.18 | 26.8 | 1.31 | 3.60 | 5.15 | 1.43 |
| **17** | 585 | 80.1 | 0.26 | 27.7 | 540 | 123.7 | 0.17 | 25.6 | 1.54 | 3.38 | 5.17 | 1.53 |
| **18** | 552 | 75.4 | 0.26 | 28 | 542 | 98.5 | 0.20 | 27.5 | 1.31 | 3.05 | 4.87 | 1.60 |
| **19** | 555 | 91.2 | 0.23 | 26.4 | 543 | 111.8 | 0.19 | 25.8 | 1.23 | 3.58 | 5.33 | 1.49 |
| **20** | 502 | 84.1 | 0.24 | 24.9 | 458 | 122.7 | 0.16 | 22.7 | 1.46 | 2.64 | 4.26 | 1.61 |
| **21** | 471 | 82.7 | 0.23 | 24.5 | 446 | 112.1 | 0.17 | 23.2 | 1.36 | 3.47 | 5.38 | 1.55 |
| **ST** | 15,390 | 2,036.6 | 0.27 | 28.1 | 15,022 | 2,698.4 | 0.20 | 27.4 | 1.32 | 3.02 | 4.51 | 1.49 |
| **NA** | 897 | | | | 1,265 | | | | | | | |
| **Total** | 16,287 | | | | 16,287 | | | | | | | |

**Table 5.** Refined genetic maps for male (M) and female (F). The genetic (cM) length of each linkage group, number of markers (nMar), the ratio physical to genetic length (Mb/cM), marker density (markers per megabase; M/Mb), the F:M ratio of genetic map length, the recombination rates (RR) in both sexes and the F:M ratio of RR are indicated.

Figure 7 depicts the typical RFm plots for an acrocentric (SseLG20) and a metacentric (SseLG1) chromosome (for all SseLG see Supplementary Fig. S4).

**Association analyses for sex.** To identify genome regions associated with sex, a GWAS analysis was carried using seven families (Table 1) and a total of 10 426 markers. Data for RAD-seq data and markers are indicated in Table 3. The results showed 30 markers significantly associated with sex after bonferroni correction using seven families ($P \leq 4.8 \times 10^{-6}$; Fig. 6A and Supplementary Table S5). When the association analysis was

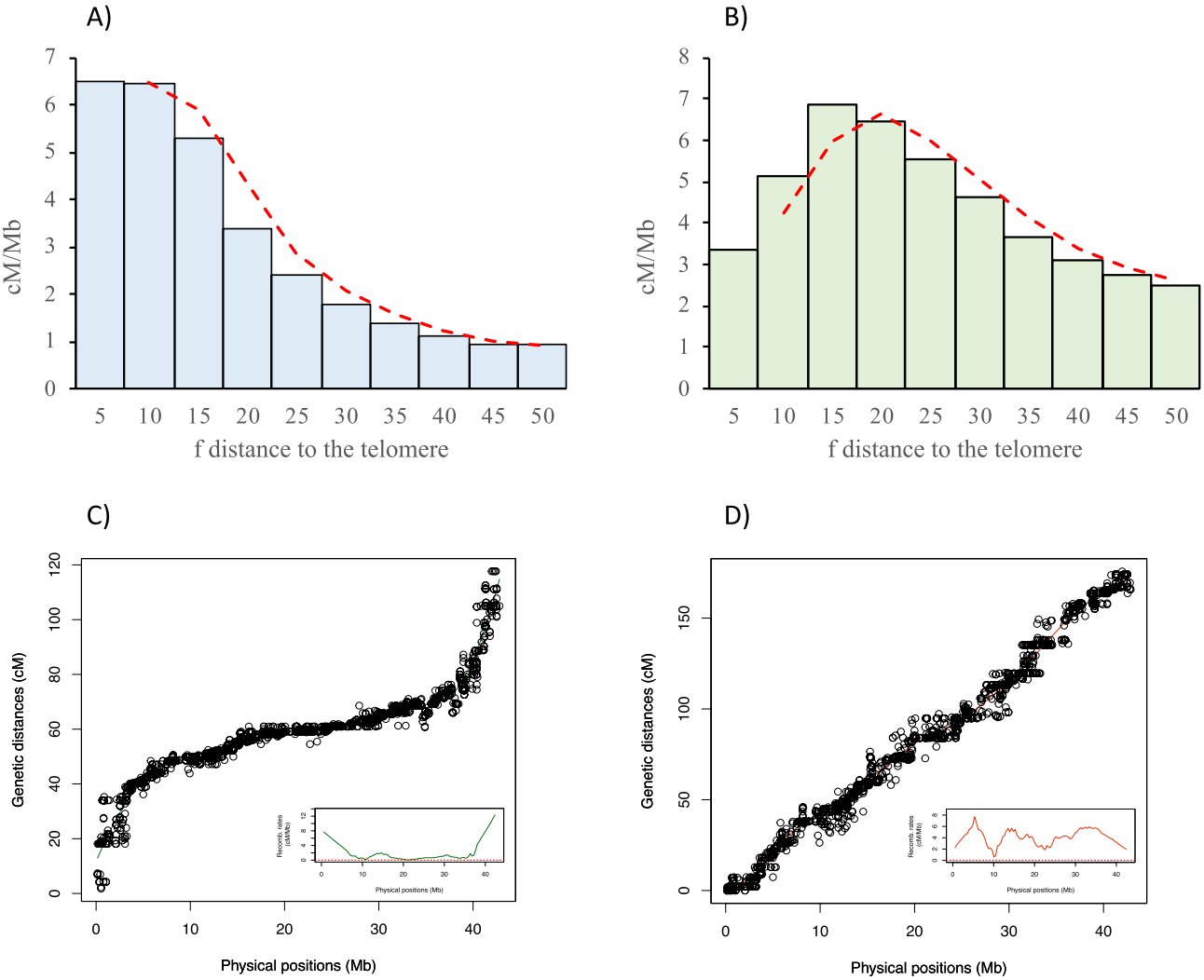


**Figure 5.** Recombination landscape averaged across linkage groups for (**A**) male and (**B**) female. The recombination rates (cM/Mb) and the relative distance from the nearest telomere scaled by the chromosome length (f) is represented. The red dashed line indicates the observed tendency. Panels (**C**,**D**) show the relationship between physical and genetic distances for SseLG1 in male and female, respectively. The square inside the panels (**C**,**D**) show the specific recombination landscape. The complete information for all SseLGs is shown in Supplementary Fig. S2 and S3.

repeated separately by family, five families provided some new 36 significant markers (Supplementary Table S5). All of them (66 SNPs including the whole-population and families) were spread in the SseLG18 with a hot region around 9.5–10.9 Mb (Fig. 6B). RR in this region was low (see above) with partially restricted RR associated with sex. Overall, 80.7% of significant markers using the whole population were preferentially heterozygous in males although penetrance was incomplete in most of them. This model is compatible with a nascent XY system. It should be noted that specific markers in family 4 had an expected high number of heterozygous *loci* in females.

To detect candidate sex-related genes, the full-length transcriptome[38] was blasted onto the SseLG18 (Supplementary Fig. S5) and a total of 229 genes were positioned. The significant SNPs were highly distributed through the pseudo-chromosome, but the follicle stimulating hormone receptor (*fshr*) gene just appeared located in the hot region revealing as a clear candidate gene for sex determination.

**Interspecific chromosome rearrangements.** An alignment of SseLGs pseudo-chromosomes with the chromosomes of three other Pleuronectiformes genomes (*Cynoglossus semilaevis*, *Scophthalmus maximus*, *Paralichthys olivaceus*) showed high similarity rates of and conserved macrosynteny level for fifteen out of 21 SseLGs (Fig. 8 and Supplementary Table S6). However, deviations from diagonal in the dot plot alignment indicated extensive intrachromosomal rearrangements among species. The three largest SseLGs appeared to be the result of total or partial chromosome fusions when compared with other flatfish genomes (Supplementary Fig. S6 and S7), and *S. maximus* seemed to be the flatfish species with the highest number of chromosome rearrangements between the four species compared. Genome comparisons using D-Genies[34] indicated that the highest similarity was with *P. olivaceus* (no match 57.3%), followed by *S. maximus* (no match 59.6%), and *C. semilaevis* (no match 78.4%).

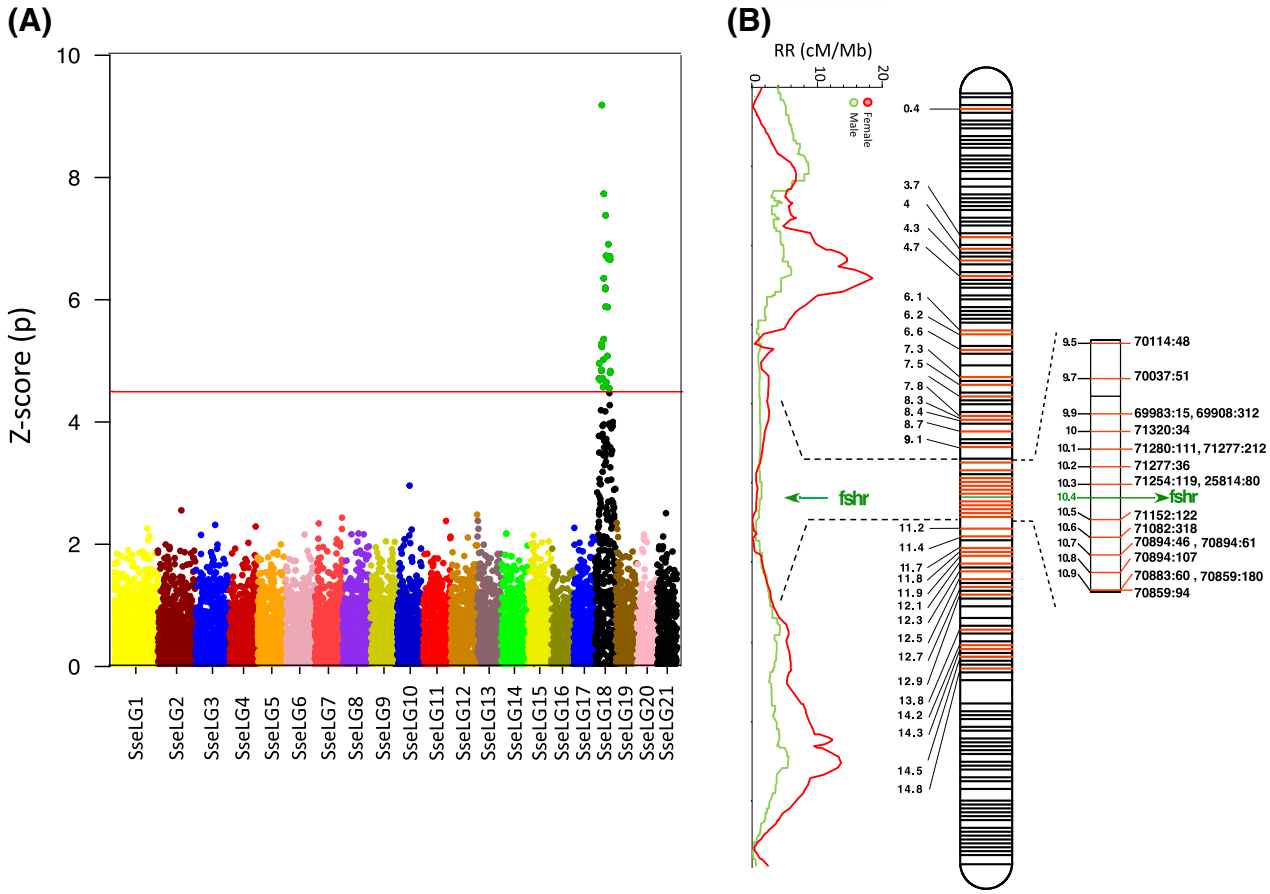


**Figure 6.** Sex-associated SNPs and RR landscape for males and females in SseLG18. (**A**) Manhattan plot of GWAS results for sex-associated SNPs using seven families. Significant markers are indicated in green. The horizontal red line represents the Bonferroni significance threshold. (**B**) Distribution of all 66 sex-associated significant markers using seven families and by family (in red, Supplementary Table S5) and RR (cM/Mb) landscape of males and females. A hot region from 9.5 to 10.9 Mb containing the candidate gene *fshr* is indicated on the right side. Physical positions of SseLG18 in Mb are indicated in black. Black lines indicate non-significant markers in SseLG18.

When the reduction of the number of chromosomes was explored three main Robertsonian fusions in the SseLG1 (Chr18-Chr11), SseLG2 (Chr14-Chr15) and SseLG3 (Chr9-Chr16) could explain the reduction from n = 24 in *P. olivaceus* to n = 21 in *S. senegalensis* (Fig. 7, Supplementary Fig. S6 and S7 and Supplementary Table S6). When compared to *S. maximus* (n = 22), the SseLG1 appeared as a fusion of Chr7 and Chr21. Moreover, translocations of regions from Chr1, Chr4, Chr7, Chr14 and Chr16 were also observed. In the case of *C semilaevis* with sexual chromosomes (ZW) and the same number of chromosome than *S. senegalensis,* a Robertsonian fusion in SseLG1 between Chr3–Chr20 was observed. Moreover, the SseLG3 appeared as a new chromosome resulting of the fission of Chr1 (mainly located in SseLG16) and Chr8 (mainly located in SseLG18). Two other major features in this species with respect to *S. senegalensis* were: (i) a translocation of a Chr14 region to Chr16 to create the SseLG2; and (ii) sexual ZW chromosomes appear concentrated in SseLG5 although high similar sequences are widely distributed throughout the genome. Comparison among all flatfish species (Fig. 7, Supplementary Fig. S6 and S7, Supplementary Table S6) indicated that those chromosomal regions associated with SseLG2 and SseLG3 were mainly involved in the changes of karyotypes of the four Pleuronectiformes species whereas the SseLG1 arose as a lineage-specific fusion event.

## Discussion

Genome assemblies and genetic linkage maps provide complementary information that can be integrated to produce high-quality physical maps. The resulting accurate chromosome assemblies are suitable to investigate genome evolution and species diversification, the genetic architecture of QTLs and the regulation of targeted genome regions. In this study, a de novo hybrid assembly for a male sole and a high-density SNP map were generated and combined to provide a polished draft assembly of 21 pseudo-chromosomes. A genome for a female sole was previously reported[17] although it was highly fragmented (N50 = 85 kb, 600.3 MB long). Later, this assembly was improved by integrating Nanopore and Illumina reads, resulting in 5,748 contigs with N50 = 339.9 kb and 608 Mb long[20] (Supplementary Table S4). In this study, the newly obtained male assembly has a lower number of contigs (3,403) and higher N50 (512.7 kb) and confirmed that the genome size of sole is around 609 Mb. This

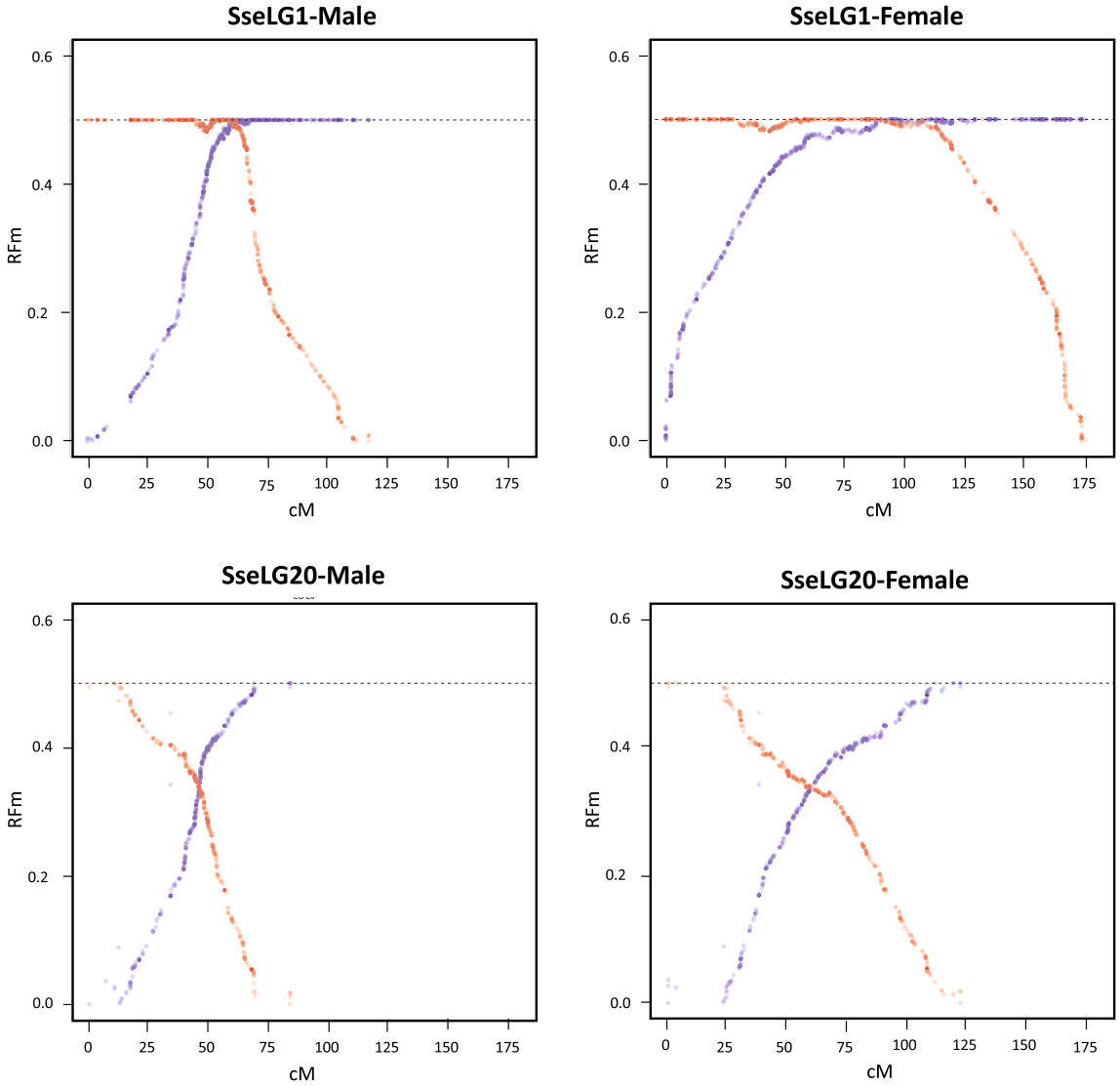


**Figure 7.** Plots illustrating the recombination frequency estimates (RFm) for intervals between markers along SseLG1 and SseLG20 in the male and female. For each LG, RFm was calculated from both chromosomal extremities (right: red circles; left: blue circles), using each of the two terminal markers as a reference starting point. The RFm plots of SseLG1 and SseLG20 show a classical metacentric and acrocentric pattern, respectively. The RFm plots of all SseLGs are illustrated in Supplementary Fig. S4.

genome size is similar or even a bit larger than other flatfish[39–42]. A dot-plot alignment analysis indicated a high similarity between male and female genome assemblies perfected aligned along the diagonal (Fig. 2) with a completeness similar to other high-quality fish assemblies (> 95.5% complete genes)[40,43,44].

Male genome characterization identified 50,133 transcripts and 27,175 protein-coding that agrees with the number of predicted transcripts in a recently assembled informative transcriptome[38]. Moreover, a small subset of unique genes was identified in both sexes with a high overrepresentation of cell-cycle regulation and regulation of transcription categories (including mainly the histones H3.2 and H4) in the female. In mammals, unique histone variants are specifically expressed in spermatogenic cells[45]. Moreover, expansion of histone multigene clusters in scleractinians was associated with sexually dimorphic expression of some variants playing a role in the control of gene expression in female and male germ cells during gametogenesis[46]. In sole, at least two *loci* of canonical histones in the largest metacentric chromosome SseLG1 linked to *dmrt1*, a key determination gene in other flatfish, were reported in sole[16,39,47]. This chromosome arose after a Robertsonian fusion and intense reorganization events[12] that could have birth to new histone clusters under purifying selection[48]. Although we cannot exclude that some differences in the number of histone copies between both genomes could be attributed to individual variation, one plausible hypothesis is that some of these histone clusters could have subfunctionalizated and acquired a role in gametogenesis in a sex-specific manner. This hypothesis is supported by the identification of a *rxra*-like receptor also represented in such GO categories able to mediate the masculinizing effects of females mediated by its ligand TBT in rockfish females [49].

De novo assembled male genome was used as reference to map the ddRAD sequences and construct a high-density genetic map. The sole consensus map size and the number of high-quality markers used (Fig. 3; Table 4)

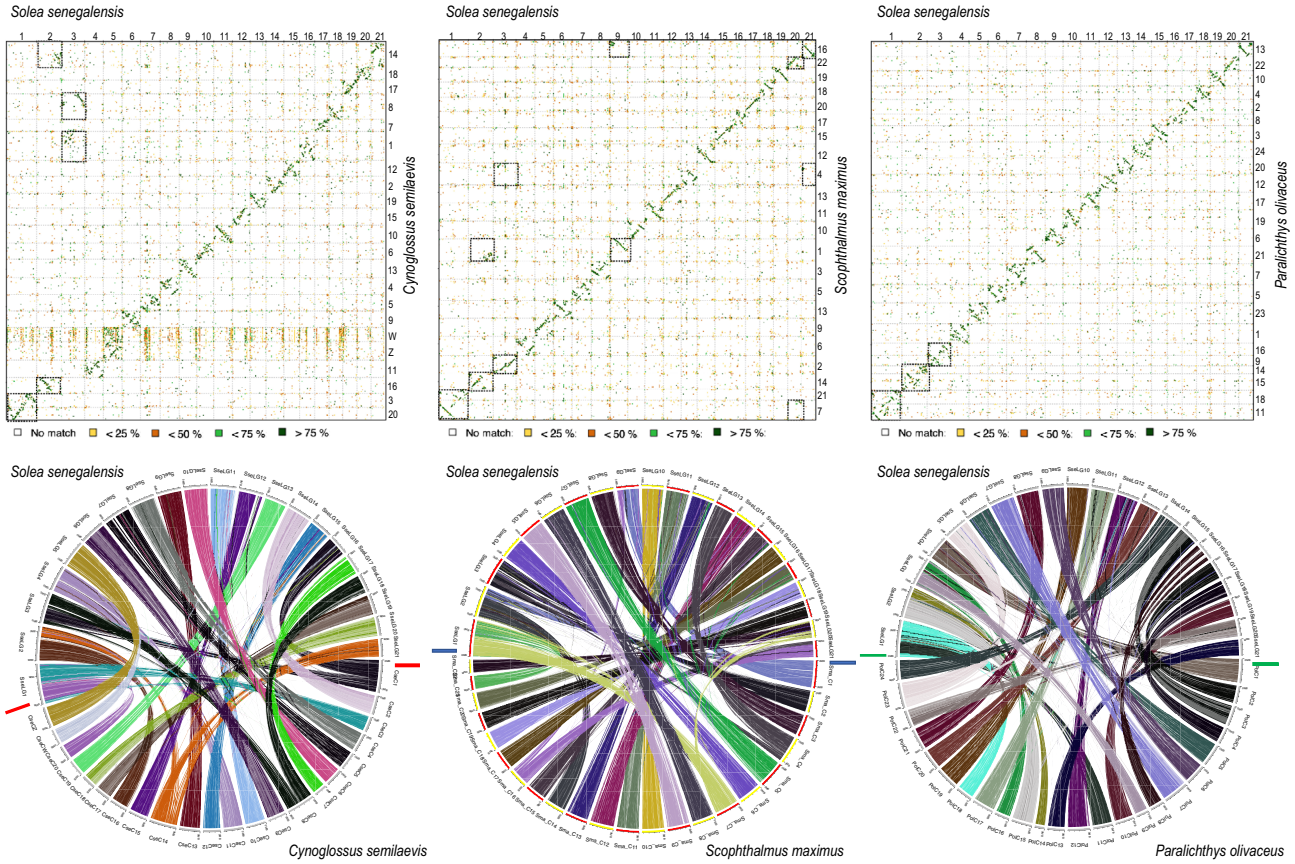


**Figure 8.** Chromosomal alignment and synteny analysis between flatfish genomes. Top panel, Dot plot comparison of 21 pseudo-chromosomes of *S. senegalensis* with the genomes of the flatfish *C. semilaevis* (left), *S. maximus* (center) and *P. olivaceus* (right). Chromosome numbers or SseLGs are indicated. The chromosome fusions are boxed. Identity scale is indicated below. Bottom panel, syntenic comparison between flatfish genomes.

were similar to those reported for turbot (2,622.09 cM)[6] and flounder (3,497.29 cM)[50] although with a higher density of markers (only 6,647 and 12,712 SNPs in turbot and flounder, respectively). Most importantly, markers were distributed into 21 SseLGs that match with the haploid karyotype (2n = 42) of the species[51]. Until now, two genetic maps with 129–229 microsatellites were reported in Senegalese sole[13,20] Moreover, a cytogenetic map was also published although the number of BACs did not still cover all chromosomes[14,16]. This new high-density SNP map (Fig. 3) thus represents a key step forward for future genomic studies and QTL identification with respect the current information available until now in this species.

Although hybrid assemblies using long and short sequences reads reduce genome fragmentation and increase the average scaffold sizes as observed in this study, most of de novo genome assemblies still do not reach chromosome-level with the expected number of chromosomes due to, among other factors, the repetitive fraction of the genome. To get around this limitation, information of genome-wide physical maps and dense genetic linkage maps can be integrated to assign chromosomal locations to sequence contigs[52]. This anchoring can also remove assembly artifacts and position misplaced scaffolds to increase the contiguity of the assembled scaffolds. In this study, the high-density SNP genetic map was used to anchor, sort and refine the assembled contigs. Overall, 89.9% of the genome assembly could be anchored to 21 pseudo-chromosomes and a total of 102 contigs were removed or split to separate positions in SseLGs. A similar strategy was followed in turbot using 31 families that allowed for the rearrangement of 20% of the genome assembly[3]. A comparison between male and female demonstrated a high co-linearity between our physical map and female scaffolds (only 5.53% mismatch). Although 10.1% of genome information remained as unplaced, the anchored physical map is essential for gene association analysis, synteny and cross-species studies and targeted genome resequencing. Further studies will be required to accurately anchor the remaining 61.9 Mb unanchored regions to their position in the genome.

It is well-known that the genome-wide RR differs between males and females (heterochiasmy) and that the recombination landscape also varies along chromosomes. In animals and plants, females tend to have higher RR than males, which in turn result in larger map lengths[53–55]. In our study, map was longer in the female than in the male (2,698.4 *vs* 2,036.6 cM; ratio 1.32). Assessment of sex-specific RR indicated a female-biased heterochiasmy across all SseLGs, with an average RR of 3.02 in male *vs* 4.51 cM/Mb in female. Four species of Pleuronectidae also exhibited wide heterochiasmy through all chromosomes similarly to sole with some intervals of male- and female-restricted meiotic recombination[56]. However, such differences in RR between males and females are not fully conserved in flatfish when map size is considered. Female maps are larger in turbot (1.36 times) and halibut

(1.07 times)[1,2,57], this is not the case of flounder or tongue sole with slightly larger maps in males (1.03–1.09 times)[50,58,59]. *C. semilaevis* is the only flatfish known with heteromorphic sex chromosomes (ZZ/ZW) that has been described in several mammals, birds and insects as a cause for an arrest of recombination in the heterogametic sex (XY males or ZW females). This could explain a shift in the direction of heterochiasmy[53].

In addition to such differences in overall RR between sexes, the chromosomal recombination landscapes also differed between male and female according to typical patterns. In fish, it has been shown that recombination occurs at higher frequencies near telomeres in males while the distribution is quite more uniform or elevated near centromeres in females[54]. In stickleback fish, it has been demonstrated that centromeres and telomeres have little or no effect on recombination in females, however, in males, the recombination rates are suppressed near the centromeres and hence crossovers localize mainly at the ends of long arms in acrocentric chromosomes[55]. This feature seems to be conserved in sole since RR were also more frequent toward the end of males SseLGs compared to females (Fig. 5).

Heterochiasmy is considered a major force that guides the evolution of genetic sex determination systems and speciation[56,60]. Normally, genome regions with very low RR are associated with sex-determining regions in young sex chromosome systems and sex-linked traits such as pigmentation[61]. In Atlantic halibut, the sex determining gene *gsdf* is located in a region of chromosome 13 with restricted male and female RR[56]. In *S. senegalensis*, 30 significant sex-associated SNPs (66 if we consider the SNPs of separated families) were distributed throughout the SseLG18 with very low RR hot region (Fig. 6 and Supplementary Fig. S2 and S3). The shift and crossing between male and female RR suggest sex-specific restricted meiotic recombination events and that heterochiasmy might be involved in nascent sex chromosome system.

Most of SNP markers in the whole-population were heterozygous in males suggesting an XX/XY system. However, it should be noted high levels of incomplete penetrance in the families analysed (Supplementary Table S5). The fact that this proportion was even inverted in specific markers of F4 indicates a high effect of environmental factors on sex determination. The temperature seems to be a major factor that modifies sex ratios during larval development generating skewed populations of neomales and neofemales[62,63]. Familial sex ratios in sole were reported to oscillate from 16 up to 90% males supporting a high impact of environmental factors to modulate sex differentiation and sex population ratios[19].

After analyzing the hot region in SseLG18, the *fshr* appeared as a putative candidate for sex determination. The *fshr* locus was recently associated with male sex in flathead grey mullet with an incomplete penetrance as observed in sole[64]. These authors proposed that *fshr* might act as a proxy for the genetic transduction of environmental factors such as temperature Under this hypothesis, sex determination would not rely on a single genetic cascade but a continuum of environmental and genetic factors. In sole, *fshr* was mainly expressed in testis[65]. The Fshr together with StAR are expressed in the steroidogenic Leydig cells and Fshr act as a promiscuous receptor that mediates the steroidogenic activity induced by both FSH and LH[66,67]. This double action supports a prolonged spermatogenesis and spermatid availability within the testis throughout the year mediated by FSH and the differentiation of spermatids into spermatozoa and subsequent spermiation mediated by LH[66]. Functional studies are needed to validate this putative candidate.

A synteny comparison of SseLGs with different flatfish genomes indicated that there was a one-to-one correspondence for 15 chromosomes, with some lineage-specific rearrangements (Fig. 8 and Supplementary Table S6). This correspondence between chromosomes was also confirmed when genome of turbot was compared with other flatfish demonstrating intrachromosomal rearrangements that shaped chromosome synteny and gene organization[3]. In our data, deviations from diagonal unlike in the comparison between male and female are indicative of this intense internal reorganization across species. The three SseLGs (SseLG1, SseLG2 and SseLG3) deserve special attention as they can provide an evolutionary framework to understand the history of chromosome fusions and fissions that shaped the karyotypes in flatfish. The SseLG1, predicted as a metacentric chromosome by the analysis of recombination frequency (Fig. 6), was previously identified by cross-species genomic comparison as the largest metacentric chromosome in Senegalese sole suggesting it may be a proto-sexual chromosome[12,16]. Our data support the hypothesis that this chromosome has primarily emerged by a lineage-specific Robertsonian fusion, since the homologs in other flatfish maintained their integrity across evolution (Supplementary Fig. S7). A complex series of events including small chromosomal translocations and rearrangements, fusions, and pericentric inversions would explain the current gene content and organization[12]. Unlike SseLG1, the SseLG2 and SseLG3 contain those chromosomes whose remodeling have shaped the karyotypes in flatfish from n = 24 in *P. olivaceus* to 22 *S. maximus* and 21 in *S. senegalensis and C. semilaevis*. A fusion model envisaged suggests a small number of chromosomes in the older lineage Paralichthyidae (9,14 and 16)[10] that combined with other chromosomes in a lineage-specific way could explain the major rearrangement events that shaped the karyotype in this species.

In conclusion, this study reports a new genome assembly for a male sole and a high-density SNP genetic map with 15,511 high-quality markers distributed in 21 linkage groups. The physical map was anchored to the consensus genetic map to generate 21 pseudo-chromosomes, in agreement with the number of chromosomes in this species. The larger map in females was the result of higher RR with distinct recombination landscape between sexes. Recombination frequencies were used to assess the putative morphology of SseLGs that will have to be validated by cytogenetic studies. A GWAS analysis identified 30 sex-associated markers, all located in SseLG18. A low recombining hot region hosted the putative candidate gene *fshr*. In silico comparison with other Pleuronectiformes genomes demonstrated a high conservation of chromosome synteny, although with much intrachromosomal reorganization. Moreover, these changes in karyotype chromosome number were associated with lineage-specific Robertsonian fusions (i.e. SseLG1 in *S. senegalensis*) and several other rearrangements that involved mainly three chromosomes in the ancestral lineage. The consistent physical and genetic maps reported in Senegalese sole represent a valuable genomic resource for functional and genome-wide association studies, and the identification of genomic processes involved in speciation.

Received: 21 November 2020; Accepted: 7 June 2021
Published online: 29 June 2021

## References


1. Bouza, C. *et al.* A microsatellite genetic map of the turbot (*Scophthalmus maximus*). *Genetics* **177**, 2457–2467. https://doi.org/10.1534/genetics.107.075416 (2007).
2. Reid, D. P. *et al.* A genetic linkage map of Atlantic halibut (*Hippoglossus hippoglossus* L.). *Genetics* **177**, 1193–1205. https://doi.org/10.1534/genetics.107.075374 (2007).
3. Maroso, F. *et al.* Highly dense linkage maps from 31 full-sibling families of turbot (*Scophthalmus maximus*) provide insights into recombination patterns and chromosome rearrangements throughout a newly refined genome assembly. *DNA Res.* **25**, 439–450. https://doi.org/10.1093/dnares/dsy015 (2018).
4. Nguyen, N. H., Rastas, P. M. A., Premachandra, H. K. A. & Knibb, W. First high-density linkage map and single nucleotide polymorphisms significantly associated with traits of economic importance in Yellowtail Kingfish *Seriola lalandi*. *Front. Genet.* **9**, 127. https://doi.org/10.3389/fgene.2018.00127 (2018).
5. Liu, Z. J. & Cordes, F. J. DNA marker technology and their applications in aquaculture genetics. *Aquaculture* **238**, 1–37. https://doi.org/10.1016/j.aquaculture.2004.05.027 (2004).
6. Wang, W. *et al.* High-density genetic linkage mapping in turbot (*Scophthalmus maximus* L.) based on SNP markers and major sex- and growth-related regions detection. *PLoS ONE* **10**, e0120410. https://doi.org/10.1371/journal.pone.0120410 (2015).
7. Rastas, P. Lep-MAP3: Robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* **33**, 3726–3732. https://doi.org/10.1093/bioinformatics/btx494 (2017).
8. Catchen, J., Amores, A. & Bassham, S. Chromonomer: A tool set for repairing and enhancing assembled genomes through integration of genetic maps and conserved synteny. *G3 Bethesda* https://doi.org/10.1534/g3.120.401485 (2020).
9. Goodwin, S. *et al.* Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res.* **25**, 1750–1756. https://doi.org/10.1101/gr.191395.115 (2015).
10. Shi, W. *et al.* Flatfish monophyly refereed by the relationship of Psettodes in Carangimorphariae. *BMC Genom.* **19**, 400. https://doi.org/10.1186/s12864-018-4788-5 (2018).
11. Azevedo, M. F. C., Oliveira, C., Pardo, B. G., Martinez, P. & Foresti, F. Cytogenetic characterization of six species of flatfishes with comments to karyotype differentiation patterns in Pleuronectiformes (Teleostei). *J. Fish Biol.* **70**, 1–15. https://doi.org/10.1111/j.1095-8649.2006.01287.x (2007).
12. Garcia-Angulo, A. *et al.* Evidence for a Robertsonian fusion in *Solea senegalensis* (Kaup, 1858) revealed by zoo-FISH and comparative genome analysis. *BMC Genom.* **19**, 818. https://doi.org/10.1186/s12864-018-5216-6 (2018).
13. Molina-Luzon, M. J. *et al.* First haploid genetic map based on microsatellite markers in Senegalese sole (*Solea senegalensis*, Kaup 1858). *Mar. Biotechnol (NY)* **17**, 8–22. https://doi.org/10.1007/s10126-014-9589-5 (2015).
14. Garcia, E. *et al.* Integrative genetic map of repetitive DNA in the sole *Solea senegalensis* genome shows a Rex transposon located in a proto-sex chromosome. *Sci. Rep.* **9**, 17146. https://doi.org/10.1038/s41598-019-53673-6 (2019).
15. Rodriguez, M. E. *et al.* Evolution of the proto sex-chromosome in *Solea senegalensis*. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms20205111 (2019).
16. Portela-Bens, S. *et al.* Integrated gene mapping and synteny studies give insights into the evolution of a sex proto-chromosome in *Solea senegalensis*. *Chromosoma* **126**, 261–277. https://doi.org/10.1007/s00412-016-0589-2 (2017).
17. Manchado, M., Planas, J. V., Cousin, X., Rebordinos, L. & Claros, M. G. In *Genomics in Aquaculture* (eds Mackenzie, S. & Jentoft, S.) 195–221 (Elsevier, 2016).
18. Manchado, M., Planas, J. V., Cousin, X., Rebordinos, L. & Claros, M. G. in *The Biology of Sole* (eds J. Muñoz-Cueto, E. Mañanós-Sánchez, & F. Sánchez-Vázquez) Ch. B-6.1, 375–394 (CDC Press, 2019).
19. Guerrero-Cozar, I. *et al.* Genetic parameter estimates and identification of SNPs associated with growth traits in Senegalese sole. *Aquaculture* **539**, 736665. https://doi.org/10.1016/j.aquaculture.2021.736665 (2021).
20. Guerrero-Cozar, I. *et al.* Development of whole-genome multiplex assays and construction of an integrated genetic map using SSR markers in Senegalese sole. *Sci. Rep.* **10**, 21905. https://doi.org/10.1038/s41598-020-78397-w (2020).
21. Zimin, A. V. *et al.* Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res.* **27**, 787–792. https://doi.org/10.1101/gr.213405.116 (2017).
22. Zimin, A. V. *et al.* The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669–2677. https://doi.org/10.1093/bioinformatics/btt476 (2013).
23. Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351 (2015).
24. Waterhouse, R. M. *et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548. https://doi.org/10.1093/molbev/msx319 (2018).
25. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135. https://doi.org/10.1371/journal.pone.0037135 (2012).
26. Rochette, N. C. & Catchen, J. M. Deriving genotypes from RAD-seq short-read data using Stacks. *Nat. Protoc.* **12**, 2640–2659. https://doi.org/10.1038/nprot.2017.123 (2017).
27. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575. https://doi.org/10.1086/519795 (2007).
28. Ouellette, L. A., Reid, R. W., Blanchard, S. G. & Brouwer, C. R. LinkageMapView-rendering high-resolution linkage and QTL maps. *Bioinformatics* **34**, 306–307. https://doi.org/10.1093/bioinformatics/btx576 (2018).
29. Rastas, P. Lep-Anchor: Automated construction of linkage map anchored haploid genomes. *Bioinformatics* **36**, 2359–2364. https://doi.org/10.1093/bioinformatics/btz978 (2020).
30. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676. https://doi.org/10.1093/bioinformatics/bti610 (2005).
31. Rezvoy, C., Charif, D., Gueguen, L. & Marais, G. A. MareyMap: an R-based tool with graphical interface for estimating recombination rates. *Bioinformatics* **23**, 2188–2189. https://doi.org/10.1093/bioinformatics/btm315 (2007).
32. Limborg, M. T., McKinney, G. J., Seeb, L. W. & Seeb, J. E. Recombination patterns reveal information about centromere location on linkage maps. *Mol. Ecol. Resour.* **16**, 655–661. https://doi.org/10.1111/1755-0998.12484 (2016).
33. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296. https://doi.org/10.1093/bioinformatics/btm108 (2007).
34. Cabanettes, F. & Klopp, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958. https://doi.org/10.7717/peerj.4958 (2018).

35. Grabherr, M. G. *et al.* Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* **26**, 1145–1151. https://doi.org/10.1093/bioinformatics/btq102 (2010).
36. Yu, Y., Ouyang, Y. & Yao, W. shinyCircos: an R/Shiny application for interactive creation of Circos plot. *Bioinformatics* **34**, 1229–1231. https://doi.org/10.1093/bioinformatics/btx763 (2018).
37. Claros, M. G., Seoane, P. & Manchado, M. Sequences and annotations of a provisional genome draft of a Senegalese sole female. figshare https://doi.org/10.6084/m9.figshare.12472100.v1. (2020).
38. Cordoba-Caballero, J. *et al.* An improved de novo assembling and polishing of Solea senegalensis transcriptome shed light on retinoic acid signalling in larvae. *Sci. Rep.* **10**, 20654. https://doi.org/10.1038/s41598-020-77201-z (2020).
39. Chen, S. *et al.* Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat. Genet.* **46**, 253–260. https://doi.org/10.1038/ng.2890 (2014).
40. Xu, X. W. *et al.* Draft genomes of female and male turbot *Scophthalmus maximus*. *Sci. Data* **7**, 90. https://doi.org/10.1038/s41597-020-0426-6 (2020).
41. Figueras, A. *et al.* Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): A fish adapted to demersal life. *DNA Res.* **23**, 181–192. https://doi.org/10.1093/dnares/dsw007 (2016).
42. Shao, C. *et al.* The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nat. Genet.* **49**, 119–124. https://doi.org/10.1038/ng.3732 (2017).
43. Ge, H. *et al.* De novo assembly of a chromosome-level reference genome of red-spotted grouper (*Epinephelus akaara*) using nanopore sequencing and Hi-C. *Mol. Ecol. Resour.* **19**, 1461–1469. https://doi.org/10.1111/1755-0998.13064 (2019).
44. Yang, L. *et al.* A chromosome-scale reference assembly of a tibetan loach, *Triplophysa siluroides*. *Front Genet.* **10**, 991. https://doi.org/10.3389/fgene.2019.00991 (2019).
45. Hoghoughi, N., Barral, S., Vargas, A., Rousseaux, S. & Khochbin, S. Histone variants: Essential actors in male genome programming. *J. Biochem.* **163**, 97–103. https://doi.org/10.1093/jb/mvx079 (2018).
46. Chiu, Y. L., Shikina, S., Yoshioka, Y., Shinzato, C. & De Chang, C. F. novo transcriptome assembly from the gonads of a scleractinian coral, Euphyllia ancora: molecular mechanisms underlying scleractinian gametogenesis. *BMC Genom.* **21**, 732. https://doi.org/10.1186/s12864-020-07113-9 (2020).
47. Merlo, M. A. *et al.* Analysis of the histone cluster in Senegalese sole (*Solea senegalensis*): Evidence for a divergent evolution of two canonical histone clusters. *Genome* **60**, 441–453. https://doi.org/10.1139/gen-2016-0143 (2017).
48. Rooney, A. P., Piontkivska, H. & Nei, M. Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. *Mol. Biol. Evol.* **19**, 68–75. https://doi.org/10.1093/oxfordjournals.molbev.a003983 (2002).
49. Zhang, J., Zuo, Z., Zhu, W., Sun, P. & Wang, C. Sex-different effects of tributyltin on brain aromatase, estrogen receptor and retinoid X receptor gene expression in rockfish (Sebastiscus marmoratus). *Mar. Environ. Res.* **90**, 113–118. https://doi.org/10.1016/j.marenvres.2013.06.004 (2013).
50. Shao, C. *et al.* Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (*Paralichthys olivaceus*): Applications to QTL mapping of Vibrio anguillarum disease resistance and comparative genomic analysis. *DNA Res.* **22**, 161–170. https://doi.org/10.1093/dnares/dsv001 (2015).
51. Vega, L., Díaz, E., Cross, I. & Rebordinos, L. Caracterizaciones citogenética e isoenzimática del lenguado *Solea senegalensis* Kaup, 1858. *Boletín Inst Español Oceanogr.* **18**, 1–6 (2002).
52. Mascher, M. & Stein, N. Genetic anchoring of whole-genome shotgun assemblies. *Front. Genet.* **5**, 208. https://doi.org/10.3389/fgene.2014.00208 (2014).
53. Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W. & Smadja, C. M. Variation in recombination frequency and distribution across eukaryotes: Patterns and processes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* https://doi.org/10.1098/rstb.2016.0455 (2017).
54. Sardell, J. M. & Kirkpatrick, M. Sex differences in the recombination landscape. *Am. Nat.* **195**, 361–379. https://doi.org/10.1086/704943 (2020).
55. Sardell, J. M. *et al.* Sex differences in recombination in Sticklebacks. *G3 (Bethesda)* **8**, 1971–1983. https://doi.org/10.1534/g3.118.200166 (2018).
56. Edvardsen, R. *et al.* Heterochiasmy facilitated the establishment of gsdf as a novel sex determining gene in Atlantic halibut. *biorxiv* https://doi.org/10.1101/2020.11.24.396218 (2020).
57. Ruan, X., Wang, W., Kong, J., Yu, F. & Huang, X. Genetic linkage mapping of turbot (*Scophthalmus maximus* L.) using microsatellitemarkers and its application in QTL analysis. *Aquaculture* **308**, 89–100. https://doi.org/10.1016/j.aquaculture.2010.08.010 (2010).
58. Song, W. *et al.* Construction of a high-density microsatellite genetic linkage map and mapping of sexual and growth-related traits in half-smooth tongue sole (*Cynoglossus semilaevis*). *PLoS ONE* **7**, e52097. https://doi.org/10.1371/journal.pone.0052097 (2012).
59. Castaño-Sanchez, C. *et al.* A second generation genetic linkage map of Japanese flounder (*Paralichthys olivaceus*). *BMC Genom.* **11**, 554. https://doi.org/10.1186/1471-2164-11-554 (2010).
60. Kitano, J. *et al.* A role for a neo-sex chromosome in stickleback speciation. *Nature* **461**, 1079–1083. https://doi.org/10.1038/nature08441 (2009).
61. Wright, A. E. *et al.* Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. *Nat. Commun.* **8**, 14251. https://doi.org/10.1038/ncomms14251 (2017).
62. Viñas, J., Asensio, E. & J.P., C. & Piferrer, F. ,. Gonadal sex differentiation in the Senegalese sole (*Solea senegalensis*) and first data on the experimental manipulation of its sex ratios. *Aquaculture* https://doi.org/10.1016/j.aquaculture.2012.12.012 (2012).
63. Blanco-Vives, B. *et al.* Exposure of larvae to daily thermocycles affects gonad development, sex ratio, and sexual steroids in *Solea senegalensis*, kaup. *J. Exp. Zool. A Ecol. Genet. Physiol.* **315**, 162–169. https://doi.org/10.1002/jez.664 (2011).
64. Ferraresso, S. *et al.* fshr: a fish sex-determining locus shows variable incomplete penetrance across flathead grey mullet populations. *iScience* **24**, 10186. https://doi.org/10.1016/j.isci.2020.101886 (2021).
65. Chauvigne, F. *et al.* Functional and evolutionary analysis of flatfish gonadotropin receptors reveals cladal- and lineage-level divergence of the teleost glycoprotein receptor family. *Biol. Reprod.* **82**, 1088–1102. https://doi.org/10.1095/biolreprod.109.082289 (2010).
66. Chauvigne, F. *et al.* Follicle-stimulating hormone and luteinizing hormone mediate the androgenic pathway in Leydig cells of an evolutionary advanced teleost. *Biol. Reprod.* **87**, 35. https://doi.org/10.1095/biolreprod.112.100784 (2012).
67. Chauvigne, F., Zapater, C., Gasol, J. M. & Cerda, J. Germ-line activation of the luteinizing hormone receptor directly drives spermiogenesis in a nonmammalian vertebrate. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 1427–1432. https://doi.org/10.1073/pnas.1317838111 (2014).


## Acknowledgements

This study was funded by project RTA2017-00054-C03-01 and RTA2017-00054-C03-funded from MCIU/AEI/FEDER, UE and cofunded 80% by Programa Operativo FEDER de Andalucía 2014-2020, project PP.AVA.AVA201601.9 SOLEALGAE. Moreover, the study has received funding from EU H2020 research and innovation program under grant agreement 817992 ERANET-BLUEBIO COFUND project PCI2020-111994 BestBrood/AEI/10.13039/501100011033. IGC is funded by a predoctoral fellowship from INIA. This work would not have been possible without the computer resources and the technical support provided by the Plataforma Andaluza

de Bioinformática of the University of Málaga and CNAG. We acknowledge the support of the Spanish Ministry of Science, Innovation and Universities to the EMBL partnership, the Centro de Excelencia Severo Ochoa and the CERCA Programme/Generalitat de Catalunya, the Spanish Ministry of Science and Innovation through the Instituto de Salud Carlos III, Generalitat de Catalunya through Departament de Salut and Departament d'Empresa i Coneixement and co-financing with funds from the European Regional Development Fund by the Spanish Ministry of Science and Innovation corresponding to the Programa Operativo FEDER Plurirregional de España (POPE) 2014-2020 and by the Secretaria d'Universitats i Recerca, Departament d'Empresa i Coneixement of the Generalitat de Catalunya corresponding to the Programa Operatiu FEDER de Catalunya 2014-2020.

## Author contributions

I.G.C.: Investigation, Data analysis, Data curation, Writing- Original draft preparation. In silico analysis. J.G.G.: Investigation, genome annotation. C.B.: Methodology, fish production. J.M.B.: Methodology. Genome sequencing. T.A.: Investigation, genome annotation. Writing—Review & Editing. M.G.C.: Conceptualization, Funding acquisition, Writing—Review & Editing. P.A.G.: Conceptualization, Investigation, Data analysis, Data curation, Writing—Review & Editing. M.M.: Resources, Conceptualization, Funding acquisition, Writing- Original draft preparation, Writing—Review & Editing.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-92601-5.

**Correspondence** and requests for materials should be addressed to M.M.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

scientific reports

Check for updates

OPEN

# Development of whole-genome multiplex assays and construction of an integrated genetic map using SSR markers in Senegalese sole

Israel Guerrero-Cózar[1], Cathaysa Perez-Garcia[2], Hicham Benzekri[3], J. J. Sánchez[4], Pedro Seoane[3], Fernando Cruz[5], Marta Gut[5], Maria Jesus Zamorano[2], M. Gonzalo Claros[3,6,7,8] & Manuel Manchado[1,9]✉

**The Senegalese sole (*Solea senegalensis*) is an economically important flatfish species. In this study, a genome draft was analyzed to identify microsatellite (SSR) markers for whole-genome genotyping. A subset of 224 contigs containing SSRs were preselected and validated by using a de novo female hybrid assembly. Overall, the SSR density in the genome was 886.7 markers per megabase of genomic sequences and the dinucleotide motif was the most abundant (52.4%). In silico comparison identified a set of 108 SSRs (with di-, tetra- or pentanucleotide motifs) widely distributed in the genome and suitable for primer design. A total of 106 markers were structured in thirteen multiplex PCR assays (with up to 10-plex) and the amplification conditions were optimized with a high-quality score. Main genetic diversity statistics and genotyping reliability were assessed. A subset of 40 high polymorphic markers were selected to optimize four supermultiplex PCRs (with up to 11-plex) for pedigree analysis. Theoretical exclusion probabilities and real parentage allocation tests using parent–offspring information confirmed their robustness and effectiveness for parental assignment. These new SSR markers were combined with previously published SSRs (in total 229 makers) to construct a new and improved integrated genetic map containing 21 linkage groups that matched with the expected number of chromosomes. Synteny analysis with respect to C. *semilaevis* provided new clues on chromosome evolution in flatfish and the formation of metacentric and submetacentric chromosomes in Senegalese sole.**

Genomes are an essential source of markers required for ecological studies, breeding programs, traceability or functional studies. In the last years, the genomes of some commercially important flatfish belonging to the Cynoglossidae, Scophthalmidae, and Paralichthydae families were published indicating that overall, they are small and highly compact with sizes ranging between 470 and 584 Mb[1–4]. These genomes have contributed to a better understanding of chromosome evolution in flatfish[5], sex determination[2] and the identification of mechanisms controlling metamorphosis[4] and growth performance[6] with impact in aquaculture and stock population management. In Senegalese sole (*Solea senegalensis*), a preliminary draft of 600.3 Mb that fully covered the tongue sole (*Cynoglossus semilaevis*) genome was assembled[7,8]. Although this assembly was still a bit fragmented (N50 of 85 kb), it became an useful tool to understand hybridization and introgression between *S. senegalensis* and *S. aegyptiaca*[9] and for synteny analysis[8,10,11]. Nevertheless, an improvement of scaffolding and chromosome architecture is required for association studies, gene mapping and comparative genomics.

[1]IFAPA Centro El Toruño, Junta de Andalucía, Camino Tiro Pichón s/n, 11500 El Puerto de Santa María, Cádiz, Spain. [2]Aquaculture Research Group (GIA), IU-ECOAQUA, Universidad de Las Palmas de Gran Canaria, Crta. Taliarte s/n, 35214 Telde, Spain. [3]Department of Molecular Biology and Biochemistry, Universidad de Málaga, 29071 Málaga, Spain. [4]Instituto Nacional de Toxicología Y Ciencias Forenses (INT), La Cuesta, La Laguna, 38320 Sta. Cruz de Tenerife, Spain. [5]CNAG-CRG, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology (BIST), Baldiri Reixac 4, 08028 Barcelona, Spain. [6]CIBER de Enfermedades Raras (CIBERER), 29071 Málaga, Spain. [7]Institute of Biomedical Research in Málaga (IBIMA), IBIMA-RARE, 29010 Málaga, Spain. [8]Instituto de Hortofruticultura Subtropical Y Mediterránea (IHSM-UMA-CSIC), 29010 Málaga, Spain. [9]"Crecimiento Azul", Centro IFAPA El Toruño, Unidad Asociada Al CSIC", El Puerto de Sta María, Spain. ✉email: manuel.manchado@juntadeandalucia.es

| Library | Read length N50 (bp) | Fragment length (bp) | Total reads | Yield (Gb) | Error r1 (%) | Error r2 (%) | Sequencing coverage[b] |
|---|---|---|---|---|---|---|---|
| PE300 | 101 | 330 | 1,005,526 | 101.56 | 0.29 | 0.62 | 142.24 |
| ONT 1DSQ | 8203 | – | 64,016 | 0.40 | 6.7 | | 0.56 |
| ONT MinION[a] | 10,802 | – | 1,311,044 | 9.38 | 17.6 | – | 12.57 |

**Table 1.** Summary of input datasets for Illumina (PE300) and Oxford Nanopore Technologies (ONT) reads for LR hybrid female assembly. [a]Information corresponding to the filtered 1D and 1D2 reads produced by five MinION runs. Error rate estimated as sum of mismatched, inserted bases and deleted bases divided by length of alignment of Oxford Nanopore Technologies (ONT) reads to the control sequence [b]Coverage estimates are calculated assuming a genome size of 714 Mb (C-value of *Solea solea*).

Genetic linkage maps and physical genomes provide complementary information that can be useful for the refinement of genome assemblies, the identification of genes associated with QTLs and cross-species synteny analysis[12,13]. In Senegalese sole, a low-density genetic linkage map constructed using three gynogenetic families and 129 microsatellites (also known as simple sequence repeats, SSRs) markers was described[14]. This map contained 27 linkage groups (LG) with an average density of 4.7 markers per LG that it was still a bit far away from the 21 chromosomes expected in *S. senegalensis*. Comparative synteny mapped these LGs through most of the chromosomes (except three) of *C. semilaevis* suggesting that some chromosome rearrangements could have occurred during evolution of these species[8]. Moreover, an integrated map using BAC clones and repetitive DNA families was developed using multiple fluorescence in situ hybridization that comprised 64 BACs mapped through all genome except in the submetacentric chromosome five[15]. Although Senegalese sole has not morphologically heteromorphic sex chromosomes, the largest metacentric chromosome was proposed as a proto-sex chromosome originated from the fusion of two acrocentric chromosomes during flatfish evolution[12,16].

Even though SNP markers have attracted the attention of researchers in the last years to construct high-density genetic linkage maps and for genetic association studies[17], the SSR markers still remain as highly popular markers due to their high variability, reproducibility, and their codominant inheritance[18,19]. To maximize the use of SSR markers, whole-genome genotyping using SSR-based multiplex PCRs have become the most suitable strategy to save costs, labour time and reduce data processing. This methodological approach can make feasible the implementation in small- to medium-sized laboratories since it requires basic equipment with comparable results between laboratories[20,21]. These whole-genome multiplex PCRs have been successfully applied to pedigree reconstruction in genetic breeding programs and QTLs identification[22–25]. However, *loci* multiplexing requires a tailor-made design of primers to be combined and amplified simultaneously avoiding primer dimer and preventing the overlapping of allelic ranges in those markers labelled with the same fluorophore colour. Hence, in silico analysis of genome SSR information followed by experimental validation of multiplex PCR assays is required.

Senegalese sole genome and transcriptome are rich mainly in SSRs with dinucleotide motif representing ~ 60% of total SSRs, tetranucleotides only 5.2% and pentanucleotides 2.4%[15,26]. Although SSRs with dinucleotide motifs have a higher allelic diversity than those with larger motifs, these latter are less prone to artefacts such as allelic dropout and stutters. Hence, scoring accuracy is very high reducing genotyping errors and making feasible data automation[27,28]. Genome analysis provides enough information for in silico analysis to select and combine high polymorphic SSR markers while they maintain an reliable and robust scoring for multiplex PCRs. The aim of this study was to: (1) provide de novo improved assembly of a female Senegalese sole based on long and short reads; (2) identify tetra- or pentanucleotide SSRs in silico and carry out a flatfish cross-species comparison to design whole-genome Multiplex PCRs; (3) validate all SSR *loci*, structure in multiplex PCRs according to allelic ranges (with up to 11-plex amplification) and optimize amplification conditions for whole genome mapping; (4) design supermultiplex PCRs containing the most polymorphic *loci* to sustain breeding genetic programs in this species in which offspring is communally reared; and (5) integrate SSR markers available in Senegalese sole in a genetic linkage map and carry out a synteny analysis with the flatfish *C. semilaevis* to understand chromosome evolution.

## Methods

**Genome sequencing, assembly and characterization.** SSR identification was carried out by in silico analysis of a previously published female genome based on Illumina short-reads[7,8]. Both the contig (named as assembly_51k according to k-mer used) and the scaffolded (named as 85 k genome according to N50) assemblies were used.

To increase the reliability of predicted SSR flanking regions, genome positioning and map distribution, a de novo female hybrid genome was also assembled using short and long reads. High molecular weight DNA was prepared from heparinized whole blood using the MagAttract HMW DNA kit (Qiagen). Main figures of Oxford nanopore Technology (ONT) (female code H2074515) and Illumina paired-end (PE300) reads (female code H150612; Bioproject PRJNA643826) are depicted in Table 1. Sequencing was carried out at the National Center for Genomic Analysis (CNAG, Barcelona, Spain). For the hybrid assembly, libraries libraries were pre-processed to remove contaminants and low-quality sequences. Briefly, the Illumina PE300 library was screened using Kraken (v0.10.5-beta)[29] and contaminants filtered out with the gem-mapper[30] (with ≤ 2% mismatches). In the case of ONT, data were base-called with Albacore v2.0.2 and reads meeting the following criteria were filtered out: base quality per read Q < 7, match to the control Sequence (lambda phage 3.5 kb), length less than 1 kb, or more than 40% low complexity sequence. Finally, POMOXIS v0.1.0 (https://github.com/nanoporetech/

pomoxis) and Racon[31] via all-vs-all alignment with minimap2[32] were used to correct the reads before assembly. The hybrid genome assembly (named as LR-hybrid female genome) was carried using MaSuRCA v3.2.3[33,34] to construct mega-reads that were finally assembled with CABOG v6.2[35]. Completeness was determined using Benchmarking Universal Single-Copy Orthologs (BUSCO, v3.0.2)[36,37] containing 4854 single-copy orthologs from actinopterygii_odb9. Genome scaffolds are available at Claros et al.[38].

**SSR screening, primer design and in silico genome mapping.** SSR screening on the genomes was carried out using MISA (Microsatellite identification tool) and the parameters were those previously described[39]. A total of 224 contigs from the the 85 k genome larger than 20 kb and containing several SSRs were preselected and positioned onto the *C. semilaevis* genome by local blast analysis (Supplementary Table S1 tab "Preselected_contigs"). Moreover, unigenes from Senegalese sole transcriptome[26] were positioned within each contig to identify gene content and sysnteny with *C. semilaevis*. A final set of putative 113 tetra- or pentanucleotide SSRs located in contigs from different chromosomes or separated at least 1 Mb apart within the same chromosome were selected (Supplementary Table S1 tab "Selected contigs"). To validate chromosome positioning, these selected contigs were further mapped onto the LR-hybrid female genome and the scaffolds blasted onto *C. semilaevis* chromosomes.

The criteria followed for primer design were those previously described for multiplex PCR reactions[21,40]. Primer sequences in each multiplex PCR assay and fluorophore labelling are depicted in Supplementary Table S2. The range of amplicon sizes oscillated between 70 and 300 base pairs (bp). The primer quality and amplicon specificity were assessed by mapping sequences onto the de novo LR-hybrid female genome (Supplementary Table S2, tab "PrimerMappingSSR"). A quality scale was established as follows: (1) high-specific (H–S) when they yielded a single specific amplicon and they mapped just in one position in the genome; (2) specific (S) when they yielded a single specific amplicon but at least one of the primers mapped between 2–10 (S* 2), 11–100 (S**) or > 100 (S***) positions in the genome; (3) multiple (M) when the primers amplified different regions in the genome; and (4) no amplification (NA) when no amplicon could be predicted or the amplicon was larger than 300 bp. A similar strategy was pursued to evaluate the quality of the primers published by Molina-Luzon, et al.[14] (Supplementary Table S2, tab "PrimerMappingLuzon").

**Fish samples and DNA isolation.** To characterize the SSR markers, wild specimens of Senegalese sole captured in the Gulf of Cádiz (Spain) and incorporated to the aquaculture broodstocks of the company CUPIMAR (San Fernando, Cádiz, Spain) and IFAPA center El Toruño (El Puerto de Santa María, Cádiz, Spain) were used. Animals were sampled for blood (~ 0.5 ml) by puncturing in the caudal vein using a heparinized syringe, added heparin (100 mU) and kept at − 20 °C until use. Overall, the whole set of animals used in this study was 150 (79 breeders from CUPIMAR and 71 from IFAPA). To optimize the multiplex PCR assays, the 71 animals from IFAPA's broodstock structured in four tanks (n = 6, 21, 22, and 22 fish) were used. As we carried out several tests to adjust the primer conditions and validate amplifications, some samples were run out and the total individuals finally analyzed in each multiplex PCR assay was slighlty different (althout the four tanks were represented in all assays) and specifically indicated in each case. To validate the supermultiplex PCR assays and carry out the simulations, fish from CUPIMAR (n = 79 distributed in four tanks) and IFAPA (n = 13) was used.

Total DNA from heparinized blood (~ 25 µl) was isolated using Isolate II Genomic DNA Kit (Bioline). DNA samples were treated with RNase A (Bioline) following the manufacture's protocol. DNA was quantified spectrophotometrically using the Nanodrop ND-8000. Each microsatellite marker was tested in singlepex PCR to confirm amplification. PCR reactions were carried out in a 12.5 µl final volume containing 40 ng of DNA, 300 nM each of specific forward and reverse primers, and 6.25 µl of Platinum Multiplex PCR Master Mix, 2 × (Thermofisher Scientific). The amplification protocol consisted of an initial denaturation at 95 °C for 10 min, followed by 30 cycles of 95 °C for 20 s, 59 °C for 1 min and 72 °C for 2 min, with a final extension of 72 °C for 10 min. PCR products were separated by capillary electrophoresis in an ABI3130 Genetic Analyzer (Applied Biosystems). Raw data obtained by capillary electrophoresis were transformed into allelic sizes using the GeneMapper v3.8 software (Thermofisher Scientific).

**Multiplex PCRs optimization.** SSRs were initially distributed in thirteen multiplex PCR assays (ranging 6 to 10-plex amplification) (Supplementary Table S2 tab "InitialMultiplexDesign". However, when markers were tested in singleplex, three of them did not amplify (SSeneg12220, SSeneg13367 and SSeneg3342) and two (SSeneg977 and SSeneg398) amplified a multipeak patterning and they were removed from the original sets. Moreover, SSeneg3502 and SSeneg106 markers were excluded from the mutiplex PCRs due to overlapping allelic range with other markers or a low amplification efficiency. The final thirteen multiplex PCR sets (named from A to M) are indicated in Supplementary Table S2 (tab "FinalMultiplex"). All Multiplex PCRs were performed in a final volume of 12.5 µl containing 1 × Platinum Multiplex PCR Master Mix, 40 ng of template DNA and the primer concentrations indicated in Supplementary Table S2 (tab "Primer amounts") that were optimized to balance the fluorescent signal intensity. The PCR program is the same indicated above and the final electropherograms obtained for each Multiplex set are shown in Supplementary Fig. S1.

To validate the robustness of the whole-genome multiplex PCRs, an independent lab (University of Las Palmas de Gran Canaria, Spain) analyzed a subset of DNA samples from IFAPA's broodstock (total n = 60). The specific number of samples analyzed for each *locus* in the multiplex PCRs is indicated in Supplementary Table S3. The amplification conditions were similar to those indicated above except that Platinum Multiplex PCR Master Mix was replaced by KAPA2G Fast Multiplex PCR Kit (Kappa Biosystems_Sigma Aldrich). Electropherograms were analyzed using Genemapper (v.3.8) software (Applied Biosystems) and a kit of bin set was created for each multiplex PCR. A protocol for evaluation of genotyping reliability and *loci* scoring was performed[21]. Briefly, the

rate of errors or potential errors for each marker were determined after identifying ambiguous or unambiguous genotypes in the samples. The main genotyping errors were classified as inadequate peak heights out of optimal ratio (600–3000 relative fluorescent units), unclear banding pattern or intermediate alleles that could not be read automatically using the bin set.

In order to design genotyping tools for parentage assignments in genetic breeding programs, a set of 40 SSR markers with the highest variability according to the polymorphic information content (PIC) was selected and rearranged in four new supermultiplex (SM) assays considering the fluorescent labelling and the allelic range (named as SMA, SMB, SMC and SMD). PCR amplification conditions were those described above and the primer cocktails optimized to balance peak signals are indicated in Supplementary Table S2 Tab "Primer amounts".

**Data analysis.** Genetic diversity parameters (number of alleles (k)), observed (Ho) and expected (He) heterozygosities, allelic range, non-exclusion probabilities for pair parent (NE-PP) and null allele frequency were estimated using Cervus *v*3.0.3[41]. The Hardy–Weinberg equilibrium (HW) at each locus was tested based on $\chi^2$ tests using GenAlEx *v*6.502 software[42]. The test for null allele presence was performed using Micro-checker v2.2.3[43]. Parentage assignment was performed in PARFEX v1.0 using exclusion approach[44]. This package was further used to calculate the minimum marker set required for optimal parentage using the given data set. Markers were ranked according to PIC information and exclusion probability. In the case of SMA, a total of n = 92 specimens (48 females and 44 males; see "Fish samples" section) were analyzed. As the number of sole breeders in each tank oscillated between 13 and 25 specimens, simulations for supermultiplex SMB, SMC and SMD were carried out using a subset of animals (n = 15; 8 females and 7 males).

To construct the integrated SSR genetic map, the 108 SSR markers of this study and 121 out of 129 SSRs of the low density genetic linkage map available in Senegalese sole[14] were positioned in the LR-hybrid female genome by local megablast analysis. Primers from eight markers in the previous map were excluded due to low quality mapping rates (Supplementary Table S2 tab "PrimerMappingLuzon"). Later, all scaffolds were anchored to the 21 linkage groups (LG) of a high-density SNP genetic linkage map generated using ddRAD from five full-sib families. Data about families, SNPs and full procedure to construct the SNP-based genetic linkage map will be published elsewhere. The relative genetic distances between makers were obtained from the anchored physical map and the integrated map was drawn using the software linkagemapview[45]. For macrosynteny comparison, scaffolds bearing the SSRs were blasted onto the *C. semilaevis* chromosomes and positions compared to identify chromosomal rearrangements.

**Compliance with ethical standards.** All procedures were performed in accordance with Spanish national (RD 53/2013) and European Union legislation for animal care and experimentation (Directive 86\609\EU) and authorized by the Bioethics and Animal Welfare Committee of IFAPA and given the registration number 10/06/2016/101.

## Results

**Identification of SSRs for multiplex design and assessment of their genome distribution.** SSR markers were identified by in silico analysis of repetitive motifs in the 85 k genome[7] based on Illumina short-reads. A first search for SSR markers selected a set of 224 contigs bigger than 20 kb and putatively located in different chromosomes or separated at least 1 Mb apart in the same chromosome. Average size of selected contigs was 118.7 kb and a cross-species comparison with the genome of the flatfish *C. semilaevis* confirmed that they were widely distributed in all chromosomes (between 6 and 17 contigs by chromosome; Supplementary Table S1 tab "Preselection"). The average number of SSR markers in each contig was 14.6, 5.3, 4.3 and 2.3 for di-, tri- tetra- and pentanucleotide repeat motifs, respectively. Using as reference this information, a subset of 113 contigs putatively distributed through the genome (minimum 5 scaffolds by chromosome) containing SSRs with tetra- or pentanucleotide repeat motifs was selected (Supplementary Table S1 "Selected_contigs"). The final set of SSRs selected for primer design included 103 tetranucleotides, 5 pentanucleotides and 5 compound markers containing at least two tetranucleotide SSRs separated by a spacer (Supplementary Table S2 tab "InitialMultiplexDesign"). Overall, GATA was the most abundant repeat motif in the selected markers (30 SSRs).

To assess the conservation of SSR flanking regions and the expected amplicon sizes as indicator of SSR quality for primer design, a de novo assembly based on Nanopore long-reads corrected with Illumina reads was used (LR-hybrid female genome). Raw sequencing data are indicated in Table 1. Expected coverage was 141 × for Illumina PE300 library and 13.5 × for Nanopore reads. The new assembly resulted in 6,482 contigs and 5,748 scaffolds with a total length of 607,976,531 bp and scaffold N50 of 340 kb. The estimated gene integrity was 96.2%. Overall, the marker density was 886.7 SSRs per megabase (Mb) and the dinucleotide repeats were the most abundant (52.4%) followed by tri- (12.5%), tetra- (4.0%) and pentanucleotides (1.1%) (Supplementary Table S1, tab "SSR_genome"). The C/A motif represented the 75% of dinucleotide repeats. To assess the quality of 113 selected markers, all designed primers were mapped onto the scaffolds of LR-hybrid female genome and classified into four categories (high-specific (H–S), specific (S), multiple, (M) and no amplification (NA)) according to locus-specificity, predicted amplification success and amplicon size (Supplementary Table S2, tab "PrimerMappingSSR"). Primers of 74 markers mapped specifically in just one position and generated locus-specific PCR amplicons of expected size similiar to 85 k genome, 34 markers had one primer of the pair with more than one mapping through the genome although the primer pair generated a locus-specific PCR product of expected size, 2 markers were not locus-specific and 3 markers failed to provide a PCR product due to amplicon size larger than expected or mapping on different scaffolds (Supplementary Table S2 tab "PrimerMappingSSR"). After assessment primer quality, 108 markers were finally selected and arranged in multiplex PCRs. The wide distribution through the genome was validated by mapping scaffolds of the 85 k and LR-hybrid female genomes onto the *C. semilaevis* chromosomes

(Supplementary Table S1 and Table S2). Mapping results were highly consistent between assemblies showing only some conflicts for those contigs (only13) located in the sexual chromosomes (Z and W) of *C. semilaevis* that are absent in sole.

**Whole-genome multiplex assays and genetic parameters.** All SSR primers were designed to be amplified under similar conditions and hence they could be combined and ready for rearrangement between multiplex PCR assays depending on the labelling and allelic range. Before optimizing the multiplex reactions, all markers were tested in singleplex under the same amplification conditions.

The expected range of amplicon sizes for the complete set of SSR markers oscillated between 84 and 341 bp. Depending on the fluorescent labelling and the expected amplicon sizes, the 108 SSRs were distributed into 13 multiplex PCR assays (ranging from 6- and 10-plex) (Supplementary Table S2, tab "InitialMultiplexDesign"). After amplifying markers in fish samples, some of them had to be rearranged in other multiplex PCRs due to allelic range overlapping or low amplification efficiency in the assays and two markers (SSeneg3502 and SSeneg106) could not be combined in any way and they were excluded. Hence, the final design comprised 106 SSR markers amplified in thirteen multiplex PCRs (from 6 to 10-plex) (Supplementary Table S2 tab "FinalMultiplex"). Electropherograms obtained for each PCR multiplex assay and markers are shown in Supplementary Fig. S1.

Main genetic parameters associated with each marker are depicted in Table 2. For each multiplex, between 44 and 71 specimens were analyzed. The number of alleles ranged between 2 and 43 by *loci*. Moreover, 89 SSR markers were experimentally confirmed as tetranucleotide and 5 as pentanucleotide after analysing the repetition patterns in genotyped samples. However, 13 SSR markers followed an allelic series compatible with a dinucleotide repeat motif. A total of 34 markers deviated from HW. Micro-checker results identified 24 markers with a possible presence of null alleles that in most of the cases deviated from HW. The allelic range of *loci* sorted by fluorescence labelling are depicted in Fig. 1. To test the robustness of the amplification and test the genetic variation of the markers, the thirteen PCR multiplex assays were run by an independent laboratory (ULPGC). Data comparison confirmed the genetic variability parameters, feasibility to amplify and consistent scoring of markers. Only 17 markers deviated from HW (Supplementary Table S3). *Loci* quality scoring identified 11 markers with a bit stuttering, 4 markers allele dropout and only two intermediate alleles but all of them could be successfully read.

To identify the genes close to the SSRs, the contigs selected for primer design were compared with Senegalese sole transcriptome and *C. semilaevis* genome. The analysis indicated a high degree of gene synteny conservation (higher than 90% in most multiplex PCRs) between *S. senegalensis* transcripts and *C. semilaevis* genes (Supplementary Table S4). Some of genes identified are of interest for aquaculture due to their role the role in immune response (toll-like receptor 3, interleukin-27 subunit beta, chemokine-like receptor 1, C-type mannose receptor 2 isoform X1), hormonal signalling (thyroid hormone receptor alpha-B, retinoic acid receptor RXR-alpha, retinol dehydrogenase 10, retinol dehydrogenase 8), antioxidant defences (superoxide dismutase [Cu–Zn]) or larval survival (high choriolytic enzyme 1), epigenetics (betaine–homocysteine S-methyltransferase 1), reproduction (Prostaglandin E synthase 3) or sensing (taste receptor type 1 member 1).

**Design of supermultiplex for parentage assignment.** To design high variable PCR multiplex assays (named as supermultiplex) suitable for pedigree reconstruction in breeding programs, a subset of 40 out of 106 markers was selected according to their allelic range and genetic variability markers and they were rearranged in four supermultiplex assays (referred from SMA, SMB, SMC and SMD) ranging from 8- to 11-plex. Allelic allelic ranges are depicted in Fig. 2. Genetic characteristics are shown in Supplementary Table S5. As average, PIC information in the four supermultiplex ranged between 0.79–0.82 and 73% of markers had a PIC value higher than 0.8 and 89% higher than 0.7 (Supplementary Table S5). In total, motifs of 9 markers were dinucleotide, 29 tetranucleotide and 2 pentanucleotide. According to the synteny analysis these markers were positioned in 17 out of 21 chromosomes.

In order to validate the usefulness of the four supermultiplex for parentage assignment in sole, they were tested using different set of parents and offspring. In the case of SMA, an offspring set of 100 individuals and 92 putative parents from 4 different broodstocks (48 females and 44 males) were 100% assigned using to a single parent pair without observing null allele mismatches. For SMB, SMC and SMD, a broodstock tank of 15 parents was characterized and 5 offspring were 100% assigned to a single pair without mismatches. Ranking markers using PIC resulted in accumulative success rate higher than 99% with 7, 5, 4 and 3 markers in SMA, SMB, SMC and SMD, respectively (Fig. 3).

**Construction of an integrated genetic map and synteny analysis.** To construct the integrated genetic map, 121 out of 129 SSRs reported by Molina-Luzon, et al.[14] were succesfully mapped onto the LR-hybrid female genome (Supplementary Table S2 tab "PrimerMappingLuzon"). Overall, a total of 229 SSRs (108 of this study + 121 previously published) were located in genome scaffolds anchored to the 21 linkage groups (SseLGs) of a recenlty high-density SNP genetic linkage map built in the lab that matches with the expected number of chromosomes *S. senegalensis*. The number of markers per LG ranged from 4 located in SseLG13 to 19 in SseLG07 (Table 3; Fig. 4a,b; Supplementary Table S2 tab "Physical_genetic_map"). Eight markers were located in unplaced scaffolds. Interestingly, marker distribution in the SseLGs was highly conincident with LGs of Molina-Luzon, et al.[14]. Only those markers from LG1 were split into the SseLG6 and SseLG19 probably due to a misarrangement in the previous map since these markers moved as two blocks between SseLGs.

Macrosynteny analysis bewteen *S. senegalensis* and *C. semilaevis* chromosomes demonstrated that 17 SseLGs of *S. senegalensis* matched perfectly with different chromosomes of *C. semilaevis* (Table 3). Only four chromosomes in *S. senegalensis* appeared as chromosomal rearrangements of C. *semilaevis* and the sequences of

| **MultiplexA** | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Locus | L | Motif | N | k | Range | Ho | He | PIC | NE-PP | F(N) | HW | NA[&] |
| SSeneg4374 | B | Tetra | 61 | 10 | 96–162 | 0.53 | 0.72 | 0.69 | 0.28 | 0.16 | ns | ns |
| SSeneg5202 | B | Tetra | 63 | 16 | 210–270 | 0.89 | 0.88 | 0.86 | 0.09 | −0.01 | ns | ns |
| SSeneg16258 | G | Tetra | 63 | 4 | 88–104 | 0.48 | 0.51 | 0.47 | 0.54 | 0.02 | ns | ns |
| SSeneg12137 | G | Tetra | 63 | 7 | 141–159 | 0.62 | 0.64 | 0.61 | 0.37 | 0.00 | ns | ns |
| SSeneg6381 | G | Tetra | 63 | 18 | 200–266 | 0.71 | 0.87 | 0.85 | 0.10 | 0.10 | (*) | Yes |
| SSeneg16050 | Y | Tetra | 63 | 10 | 142–184 | 0.49 | 0.62 | 0.59 | 0.38 | 0.12 | ns | Yes |
| SSeneg11269 | Y | Di | 63 | 33 | 183–263 | 0.95 | 0.96 | 0.95 | 0.02 | 0.00 | ns | ns |
| SSeneg162554 | R | Tetra | 63 | 9 | 86–118 | 0.89 | 0.80 | 0.76 | 0.21 | −0.06 | ns | ns |
| SSeneg12054 | R | Tetra | 63 | 6 | 159–179 | 0.78 | 0.75 | 0.70 | 0.29 | −0.03 | ns | ns |
| SSeneg3041 | R | Tetra | 63 | 14 | 207–287 | 0.52 | 0.78 | 0.75 | 0.22 | 0.19 | * | Yes |
| **MultiplexB** | | | | | | | | | | | | |
| SSeneg5772 | B | Tetra | 51 | 11 | 80–130 | 0.77 | 0.81 | 0.78 | 0.19 | 0.03 | (*) | ns |
| SSeneg12300 | B | Tetra | 51 | 5 | 177–193 | 0.67 | 0.61 | 0.53 | 0.51 | −0.06 | ns | ns |
| SSeneg6326 | B | Tetra | 51 | 9 | 231–267 | 0.82 | 0.80 | 0.77 | 0.19 | −0.02 | ns | ns |
| SSeneg6982 | G | Penta | 51 | 2 | 94–100 | 0.28 | 0.27 | 0.23 | 0.81 | −0.02 | ns | ns |
| SSeneg827 | Y | Tetra | 51 | 6 | 91–111 | 0.55 | 0.63 | 0.58 | 0.41 | 0.07 | ns | ns |
| SSeneg395 | Y | Penta | 51 | 8 | 241–276 | 0.80 | 0.82 | 0.79 | 0.18 | 0.01 | ns | ns |
| SSeneg14931 | R | Tetra | 51 | 7 | 89–113 | 0.69 | 0.77 | 0.73 | 0.26 | 0.06 | ns | ns |
| SSeneg2894 | R | Tetra | 51 | 7 | 178–268 | 0.43 | 0.70 | 0.64 | 0.37 | 0.24 | (*) | Yes |
| **MultiplexC** | | | | | | | | | | | | |
| SSeneg12678 | B | Tetra | 46 | 29 | 121–377 | 0.37 | 0.96 | 0.95 | 0.02 | 0.44 | * | Yes |
| SSeneg11209 | G | Tetra | 54 | 10 | 94–134 | 0.78 | 0.79 | 0.76 | 0.20 | 0.01 | (*) | ns |
| SSeneg433 | Y | Tetra | 54 | 6 | 101–174 | 0.56 | 0.62 | 0.57 | 0.44 | 0.05 | (*) | ns |
| SSeneg7919 | Y | Tetra | 53 | 9 | 174–210 | 0.53 | 0.79 | 0.75 | 0.23 | 0.2 | (*) | Yes |
| SSeneg1973 | Y | Tetra | 53 | 23 | 249–329 | 0.94 | 0.93 | 0.92 | 0.04 | −0.01 | (*) | ns |
| SSeneg17673 | R | Tetra | 54 | 5 | 116–177 | 0.44 | 0.73 | 0.67 | 0.35 | 0.24 | * | Yes |
| SSeneg10308 | R | Tetra | 54 | 13 | 161–239 | 0.87 | 0.90 | 0.88 | 0.08 | 0.01 | ns | ns |
| **MultiplexD** | | | | | | | | | | | | |
| SSeneg1505 | B | Tetra | 57 | 7 | 112–136 | 0.75 | 0.78 | 0.75 | 0.23 | 0.01 | ns | ns |
| SSeneg4306 | B | Tetra | 57 | 2 | 204–208 | 0.51 | 0.50 | 0.37 | 0.72 | −0.01 | ns | ns |
| SSeneg10667 | B | Tetra | 54 | 7 | 277–301 | 0.56 | 0.64 | 0.61 | 0.36 | 0.06 | ns | ns |
| SSeneg2307 | G | Tetra | 57 | 6 | 134–166 | 0.53 | 0.55 | 0.50 | 0.51 | 0.02 | (ns) | ns |
| SSeneg13116 | G | Tetra | 57 | 7 | 199–235 | 0.63 | 0.72 | 0.69 | 0.28 | 0.06 | ns | ns |
| SSeneg1201 | Y | Penta | 57 | 13 | 115–180 | 0.40 | 0.83 | 0.81 | 0.16 | 0.35 | * | Yes |
| SSeneg4572 | Y | Tetra | 57 | 2 | 207–215 | 0.26 | 0.48 | 0.36 | 0.73 | 0.29 | * | Yes |
| SSeneg4065 | R | Tetra | 57 | 10 | 117–161 | 0.86 | 0.84 | 0.82 | 0.14 | −0.01 | ns | ns |
| SSeneg8782 | R | Tetra | 57 | 10 | 200–242 | 0.88 | 0.83 | 0.81 | 0.15 | −0.03 | ns | ns |
| **MultiplexE** | | | | | | | | | | | | |
| SSeneg5850 | B | Tetra | 50 | 4 | 74–92 | 0.62 | 0.62 | 0.55 | 0.49 | 0 | ns | ns |
| SSeneg2473 | B | Tetra | 50 | 4 | 204–216 | 0.70 | 0.56 | 0.46 | 0.61 | −0.12 | ns | ns |
| SSeneg544 | B | Tetra | 50 | 4 | 282–290 | 0.70 | 0.61 | 0.53 | 0.52 | −0.08 | ns | ns |
| SSeneg87 | G | Tetra | 50 | 12 | 106–166 | 0.64 | 0.67 | 0.62 | 0.37 | 0.02 | ns | ns |
| SSeneg5828 | G | Tetra | 49 | 7 | 192–224 | 0.55 | 0.48 | 0.46 | 0.52 | −0.11 | ns | ns |
| SSeneg3415 | Y | Tetra | 50 | 8 | 94–132 | 0.56 | 0.64 | 0.60 | 0.39 | 0.06 | (ns) | ns |
| SSeneg5919 | Y | Di | 50 | 8 | 204–224 | 0.66 | 0.74 | 0.69 | 0.31 | 0.05 | ns | ns |
| SSeneg585 | R | Tetra | 50 | 7 | 103–127 | 0.62 | 0.75 | 0.71 | 0.27 | 0.07 | (*) | Yes |
| SSeneg14542 | R | Tetra | 49 | 8 | 202–244 | 0.67 | 0.76 | 0.71 | 0.29 | 0.06 | ns | ns |
| **MultiplexF** | | | | | | | | | | | | |
| SSeneg1411 | B | Tetra | 64 | 3 | 120–128 | 0.22 | 0.25 | 0.23 | 0.78 | 0.05 | ns | ns |
| SSeneg3069 | B | Tetra | 63 | 13 | 183–245 | 0.73 | 0.87 | 0.85 | 0.11 | 0.08 | (*) | ns |
| SSeneg9009 | B | Tetra | 64 | 19 | 286–368 | 0.89 | 0.93 | 0.92 | 0.04 | 0.02 | ns | ns |
| SSeneg437 | G | Tetra | 65 | 9 | 219–249 | 0.52 | 0.81 | 0.78 | 0.19 | 0.22 | ns | Yes |
| SSeneg247 | Y | Tetra | 61 | 7 | 85–122 | 0.71 | 0.72 | 0.68 | 0.30 | −0.02 | ns | ns |
| SSeneg73 | Y | Di | 65 | 16 | 199–255 | 0.85 | 0.86 | 0.84 | 0.12 | 0.01 | ns | ns |
| SSeneg12624 | Y | Penta | 64 | 11 | 311–359 | 0.84 | 0.82 | 0.79 | 0.17 | −0.02 | ns | ns |
| Continued | | | | | | | | | | | | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SSeneg12095 | R | Tetra | 65 | 4 | 148–160 | 0.51 | 0.48 | 0.42 | 0.61 | −0.03 | ns | ns |
| SSeneg582 | R | Di | 62 | 18 | 224–308 | 0.94 | 0.87 | 0.85 | 0.11 | −0.05 | ns | ns |
| **MultiplexG** | | | | | | | | | | | | |
| SSeneg3683 | B | Tetra | 69 | 11 | 125–167 | 0.75 | 0.82 | 0.80 | 0.15 | 0.04 | ns | ns |
| SSeneg5713 | B | Di | 65 | 21 | 227–311 | 0.86 | 0.88 | 0.87 | 0.08 | 0.01 | ns | ns |
| SSeneg1667 | G | Di | 69 | 24 | 225–319 | 0.80 | 0.89 | 0.88 | 0.07 | 0.05 | (ns) | Yes |
| SSeneg2891 | Y | Tetra | 68 | 9 | 150–190 | 0.65 | 0.82 | 0.79 | 0.19 | 0.11 | * | Yes |
| SSeneg45 | Y | Tetra | 69 | 5 | 242–258 | 0.59 | 0.65 | 0.59 | 0.44 | 0.04 | (*) | ns |
| SSeneg12417 | R | Di | 69 | 9 | 199–225 | 0.86 | 0.78 | 0.74 | 0.25 | −0.06 | (ns) | ns |
| SSeneg10524 | R | Tetra | 69 | 7 | 266–286 | 0.75 | 0.71 | 0.66 | 0.34 | −0.05 | ns | ns |
| **MultiplexH** | | | | | | | | | | | | |
| SSeneg4608 | B | Tetra | 71 | 4 | 82–104 | 0.13 | 0.16 | 0.15 | 0.86 | 0.14 | ns | ns |
| SSeneg2868 | B | Tetra | 71 | 9 | 112–172 | 0.78 | 0.83 | 0.80 | 0.17 | 0.03 | (*) | ns |
| SSeneg11316 | B | Tetra | 71 | 10 | 214–292 | 0.72 | 0.71 | 0.67 | 0.33 | 0.00 | (*) | ns |
| SSeneg287 | G | Tetra | 71 | 7 | 68–114 | 0.55 | 0.52 | 0.47 | 0.55 | −0.05 | * | ns |
| SSeneg90 | G | Tetra | 71 | 13 | 133–175 | 0.93 | 0.85 | 0.84 | 0.13 | −0.05 | ns | ns |
| SSeneg2596 | Y | Tetra | 71 | 5 | 78–104 | 0.38 | 0.40 | 0.35 | 0.68 | 0.00 | * | ns |
| SSeneg8412 | Y | Tetra | 71 | 8 | 138–172 | 0.41 | 0.47 | 0.43 | 0.57 | 0.05 | (*) | ns |
| SSeneg6827 | R | Tetra | 71 | 4 | 88–100 | 0.32 | 0.33 | 0.30 | 0.71 | −0.02 | ns | ns |
| SSeneg5412 | R | Tetra | 71 | 7 | 148–216 | 0.41 | 0.46 | 0.43 | 0.57 | 0.05 | * | ns |
| **MultiplexI** | | | | | | | | | | | | |
| SSeneg854 | B | Di | 69 | 6 | 85–95 | 0.52 | 0.60 | 0.53 | 0.50 | 0.07 | ns | ns |
| SSeneg5899 | B | Tetra | 69 | 5 | 164–216 | 0.41 | 0.47 | 0.43 | 0.58 | 0.06 | ns | ns |
| SSeneg5346 | B | Di | 68 | 43 | 184–542 | 0.87 | 0.95 | 0.94 | 0.02 | 0.04 | (*) | ns |
| SSeneg1669 | G | Tetra | 69 | 16 | 94–168 | 0.75 | 0.82 | 0.80 | 0.15 | 0.05 | ns | ns |
| SSeneg7074 | G | Tetra | 69 | 6 | 144–182 | 0.64 | 0.76 | 0.71 | 0.30 | 0.08 | ns | Yes |
| SSeneg4382 | Y | Tetra | 64 | 5 | 92–108 | 0.22 | 0.42 | 0.37 | 0.65 | 0.31 | * | Yes |
| SSeneg53551 | Y | Tetra | 67 | 8 | 142–184 | 0.72 | 0.72 | 0.68 | 0.30 | 0.00 | ns | ns |
| SSeneg3978 | R | Tetra | 69 | 7 | 84–108 | 0.67 | 0.68 | 0.64 | 0.34 | 0.00 | ns | ns |
| SSeneg15332 | R | Tetra | 68 | 19 | 168–250 | 0.91 | 0.89 | 0.87 | 0.08 | −0.02 | ns | ns |
| **MultiplexJ** | | | | | | | | | | | | |
| SSeneg17159 | B | Tetra | 58 | 5 | 75–93 | 0.48 | 0.47 | 0.42 | 0.59 | −0.03 | ns | ns |
| SSeneg9042 | B | Tetra | 56 | 19 | 174–260 | 0.77 | 0.86 | 0.83 | 0.12 | 0.06 | ns | ns |
| SSeneg1723 | G | Tetra | 58 | 7 | 97–127 | 0.55 | 0.51 | 0.47 | 0.53 | −0.08 | ns | ns |
| SSeneg348796 | Y | Tetra | 58 | 6 | 81–101 | 0.78 | 0.67 | 0.61 | 0.41 | −0.09 | (ns) | ns |
| SSeneg7987 | Y | Di | 58 | 32 | 238–354 | 0.88 | 0.94 | 0.93 | 0.03 | 0.03 | (*) | Yes |
| SSeneg3077 | R | Tetra | 58 | 4 | 94–110 | 0.40 | 0.36 | 0.33 | 0.68 | −0.07 | (ns) | ns |
| SSeneg10804 | R | Tetra | 54 | 23 | 261–525 | 0.93 | 0.87 | 0.85 | 0.10 | −0.04 | ns | ns |
| **MultiplexK** | | | | | | | | | | | | |
| SSeneg2083 | B | Tetra | 62 | 9 | 92–124 | 0.69 | 0.65 | 0.61 | 0.36 | −0.05 | ns | ns |
| SSeneg4083 | B | Tetra | 63 | 6 | 220–242 | 0.56 | 0.58 | 0.54 | 0.43 | 0.00 | (ns) | ns |
| SSeneg171 | G | Tetra | 63 | 7 | 136–172 | 0.78 | 0.75 | 0.71 | 0.28 | −0.03 | ns | ns |
| SSeneg2487 | G | Tetra | 50 | 26 | 188–328 | 0.98 | 0.95 | 0.93 | 0.03 | −0.02 | * | ns |
| SSeneg566 | Y | Tetra | 63 | 7 | 114–136 | 0.84 | 0.77 | 0.73 | 0.27 | −0.05 | ns | ns |
| SSeneg6876 | R | Tetra | 63 | 21 | 108–198 | 0.94 | 0.91 | 0.90 | 0.06 | −0.02 | ns | ns |
| SSeneg4081 | R | Tetra | 61 | 19 | 268–374 | 0.90 | 0.88 | 0.87 | 0.08 | −0.02 | ns | ns |
| **MultiplexL** | | | | | | | | | | | | |
| SSeneg7666 | B | Di | 46 | 21 | 162–224 | 0.89 | 0.92 | 0.90 | 0.05 | 0.01 | ns | ns |
| SSeneg4003 | B | Di | 46 | 21 | 244–332 | 0.89 | 0.93 | 0.91 | 0.05 | 0.01 | ns | ns |
| SSeneg5891 | G | Tetra | 46 | 12 | 97–159 | 0.76 | 0.78 | 0.75 | 0.22 | 0.01 | (ns) | ns |
| SSeneg774 | G | Tetra | 46 | 4 | 172–178 | 0.17 | 0.27 | 0.26 | 0.75 | 0.26 | * | Yes |
| SSeneg6689 | Y | Tetra | 44 | 5 | 111–131 | 0.11 | 0.41 | 0.38 | 0.61 | 0.55 | (*) | Yes |
| SSeneg1147 | Y | Tetra | 46 | 14 | 204–252 | 0.80 | 0.91 | 0.89 | 0.06 | 0.06 | ns | Yes |
| SSeneg14333 | R | Tetra | 46 | 8 | 132–172 | 0.37 | 0.83 | 0.79 | 0.18 | 0.38 | * | Yes |
| SSeneg2996 | R | Tetra | 45 | 14 | 229–291 | 0.64 | 0.90 | 0.88 | 0.07 | 0.16 | (*) | Yes |
| **MultiplexM** | | | | | | | | | | | | |
| SSeneg506 | B | Tetra | 63 | 6 | 88–114 | 0.22 | 0.66 | 0.60 | 0.43 | 0.49 | * | Yes |
| Continued | | | | | | | | | | | | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SSeneg387243 | B | Tetra | 62 | 17 | 250–316 | 0.86 | 0.87 | 0.85 | 0.11 | 0.01 | ns | ns |
| SSeneg10877 | G | Tetra | 63 | 12 | 177–223 | 0.71 | 0.80 | 0.77 | 0.19 | 0.04 | * | ns |
| SSeneg14597 | G | Tetra | 62 | 13 | 250–356 | 0.76 | 0.90 | 0.89 | 0.07 | 0.08 | * | Yes |
| SSeneg4328 | Y | Tetra | 63 | 16 | 96–168 | 0.92 | 0.92 | 0.90 | 0.06 | −0.01 | (ns) | ns |
| SSeneg4039 | Y | Di | 60 | 26 | 248–322 | 0.43 | 0.91 | 0.90 | 0.05 | 0.36 | * | Yes |
| SSeneg1988 | R | Tetra | 62 | 2 | 91–95 | 0.02 | 0.02 | 0.02 | 0.98 | 0.00 | ns | na |

**Table 2.** Genetic diversity estimates of 106 by multiplex PCRs (A-M). Fluorescent labelling (B, blue; G, green; Y, yellow; R, red), repeat motif (Di, tetra or pentanucleoide), Number of samples (N), number of alleles (k), Allelic range, observed heterozygosity (Ho) and expected heterozygosity (He), polymorphic information content (PIC), non-exclusion probability of pair parent (NE-PP); null allele frequency (F(N)). Hardy–Weinberg equilibrium (HW; *significant after bonferroni correction; ns, non-significant) and Null alleles as determined by micro-checker (yes, significant after bonferroni correction; ns, non-significant).

Z chromosome were dispersed through the SseLG3, SseLG4 and SseLG5. The SseLG1 appeared as a fusion of chromosomes 3 and 20 of *C. semilaevis*. Moreover, some rearrangements were observed for SseLG2 that included the chromosome 16 and part of 14, the SseLG3 that grouped regions of chromosomes 1, 8 and Z and the SseLG4 that combined the chromosome 11 and regions of Z.

## Discussion

The SSRs are highly abundant in the genome of vertebrates although their use has been limited by the knowledge of flanking regions suitable for primer design. Some authors considered as alternative the cross-species amplification of highly conserved SSRs[14,46,47]. Recently, a study in Senegalese sole based on the 1.1% of the genome information estimated a high density of SSRs (675 per Mb) with dinucleotide SSRs representing overall 59.7%[15]. In this study, we took advantage of a 85 k genome draft[7] and a de novo female hybrid genome based on Nanopore and Illumina reads to overpass the deficit of markers in Senegalese sole. Total size of this new genome was 608 Mb very close to the 600.3 Mb reported for the 85 k Illumina assembly[7] suggesting that Senegalese sole genome is a slightly bigger than other flatfish (up to 584 Mb)[2–4,48] . This assembly had a high-quality gene representativity (completeness was 96.2% similar to previous flatfish assemblies)[3] with the marker density of 886.7 SSRs per megabase (Supplementary Table S1 tab "SSR_genome"). Previous cytogenetic analyses demonstrated that most of di- and tetranucleotides appear widely distributed in subtelomeric position of metacentric, submetacentric and acrocentric chromosomes[15] and hence both of them were considered suitable for primer design and multiplex amplification in this study.

Whole-genome mapping requires high-throughput strategies to save consumables, labour costs and reduce the processing and analysis times. PCR multiplex assays have been successfully developed in seabream[25,49] and grapevine[20] for QTLs identification and pedigree reconstruction. In this study, thirteen PCR multiplex assays comprising 106 markers widespread in the genome were optimized. Although previous studies in sole have reported microsatellite markers derived from EST or SSR-enriched libraries[46,50,51] only three of them considered SSR multiplexing (from 4 to 8-pex)[47,52,53]. These new multiplex PCRs and their integration with the 121 markers previously published[14] represent key genomic tools for QTL detection in sole. The new genome information provided also facilitates the integration with SNP markers and the redesign of some SSR primers in the map to construct new multiplexes that improve the genome coverage.

Tetra- and pentanucleotides predicted motifs were initially selected for multiplex PCRs although finally some of them (12%) followed a dinucleotide allelic series. It has been demonstrated that SSRs with dinucleotide motifs have a higher variability but more prone to genotyping errors than those with larger motifs[28,54]. In this study, the average number of alleles per locus was 10.9 ranging from 2 to 43 in accordance with previous SSR markers in Senegalese sole[14,46,50,51]. As expected, the dinucleotide markers showed a higher variability (average PIC 0.84) than tetra- (0.65) and pentanucleotides (0.66). Moreover, scoring accuracy was estimated using a standardized methodology to identify potential errors in the electropherograms[21] indicating only a small set of markers (17) with stuttering, allele dropout or intermediate alleles, ~ 16% of total markers. In seabream, the percentage of *loci* with some of these errors was similar although with higher rates of intermediate alleles[21]. It should be indicated that stutter peaks have a low effect to assign *loci* size in tetranucleotides as observed by a double validation across two independent labs reaching similar values in genetic diversity parameters.

The use of genetic tools to infer genealogies is a demand for genetic breeding programs in mass-spawning species such as Senegalese sole. Due to the economic value of these species, the optimization of genotyping tools for parental assignment in a feasible, accurate and cost-effective way is a requirement. Moreover, the loss in variability that occurs in subsequent selection cycles makes necessary a minimal number of markers to sustain the program through some generations. Both the number of *loci* and their heterozygosity level may influence the power of markers for parentage exclusion approaches[55]. In this study, a total of 40 high variable and genome widespread markers were selected according to PIC and combined in four supermultiplex (7 to 11-pex). Assignment simulations indicated that a subset of 7, 5, 4 and 3 markers were able to assign 99% offspring with SMA (11-pex), SMB (11-pex), SMC (8-pex) or SMD (10-pex), respectively. Moreover, a real testing using SMA to genotype 92 parents accurately allocated all 100 parent–offspring relationships. All these data indicate that these supermultiplex can be transferred to the industry as standards for pedigree reconstruction to support a long-term use for genetic breeding selection.

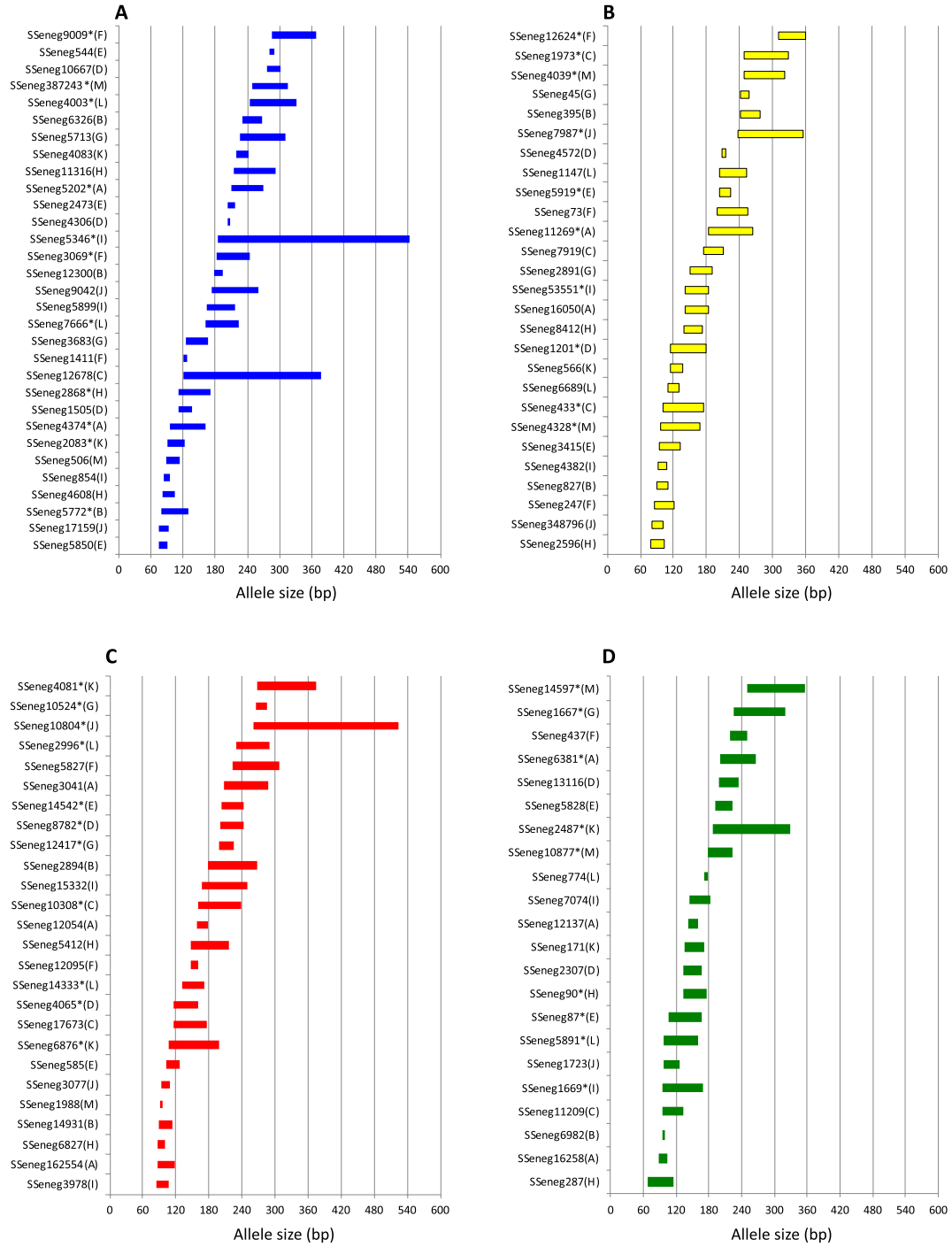


**Figure 1.** Allelic ranges of the 106 SSRs analysed in this study by fluorescence labelling (**A**–**D**). The name of the multiplex PCRs in which each marker is included is indicated between brackets. The asterisk indicates that the marker was selected to be included in the supermultiplex PCRs.

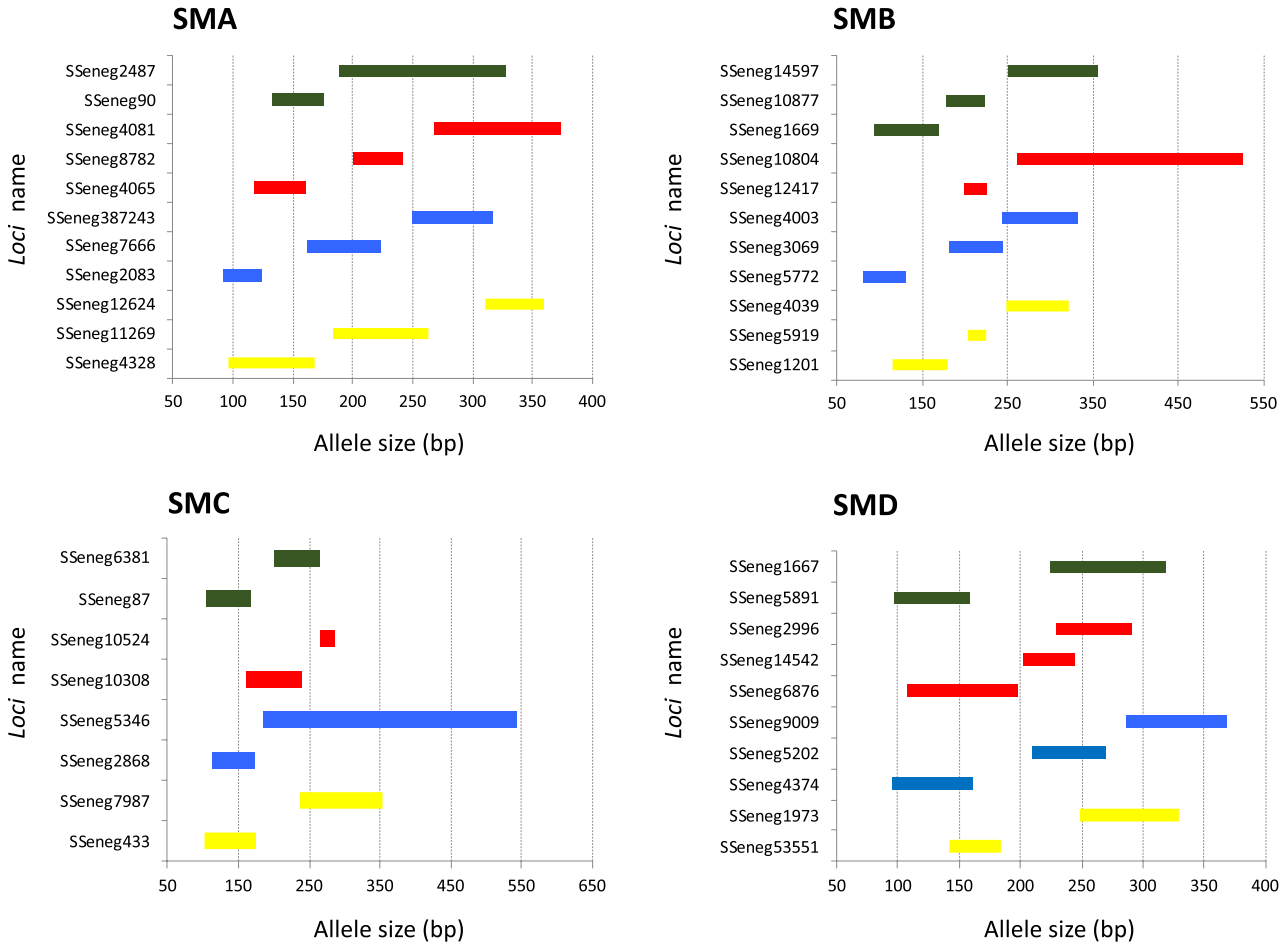


**Figure 2.** Allelic ranges of the 40 SSRs selected for the supermultiplex (SM) PCRs. The markers are shown by SM (**A**–**D**).

An integrated genetic map with 229 SSR markers was generated that improve the current low density genetic linkage map available in Senegalese sole[14] (Fig. 4). Using a high-density SNP genetic map as reference, the whole set of SSR markers was distributed in 21 LGs that fit with the haploid complement in this flatfish species (3 metacentric pairs, 2 submetacentric pairs, 4 subtelocentric pairs and 12 acrocentric pairs)[56]. Our anaysis confirmed that the LGs from the previous genetic map[14] clustered perfectly within the SseLGs after anchoring the LR-hybrid female genome and the high density genetic map (Fig. 4 and Table 3). Only LG1 was split into two SseLGs that might be due to an error in the consensus between gynogenetic families.

Flatfish genome comparisons have demonstrated a high degree of conservation at macrosynteny level[5,57,58]. Our data confirmed that most of chromosomes matched one-by-one with different chromosomes of *C. semilaevis* supporting this high conservation observed in other flatfish. Moreover, chromosome fusions and translocations have occured frequently during flatfish evolution shaping the number of chromosomes from n = 24 pairs in Japanese flounder to n = 20 autosome pairs and one sexual chromosome pair in C. *semilaevis*. In *S. senegalensis*, it has been hypothesized that the largest metacentric chromosome arose from a robertsonian fusion of two acrocentric chromosomes followed by pericentric inversions[16,59]. Our data also support this fusion and chromosome rearrangements between chromosomes 3 and 20 of *C. semilaevis* (Table 3). It should be noted that Senegalese sole has two additional metacentric pairs and 2 submetacentric pairs unlike *C. semilaevis* with all chromosomes telocentric[60]. Three LGs (SseLG02, SseLG03 and SseLG04) were also associated with more than one chromosome of *C. semilaevis* and a fourth LG (SseLG05) was syntenic with the large sexual chromosome Z (Table 3). Some robertsonian translocations (fissions and fusions) could be the origin of these non-acrocentric chromosomes in *S. senegalensis* as previously observed in turbot[5]. Most interestingly, the high remodelling of sexual ZW chromosomes that was also previously assessed by a scaffold mapping strategy[8] suggests that a shift in the sex determining system might have occurred in Senegalese sole. In fact, a sex determination XX-XY system

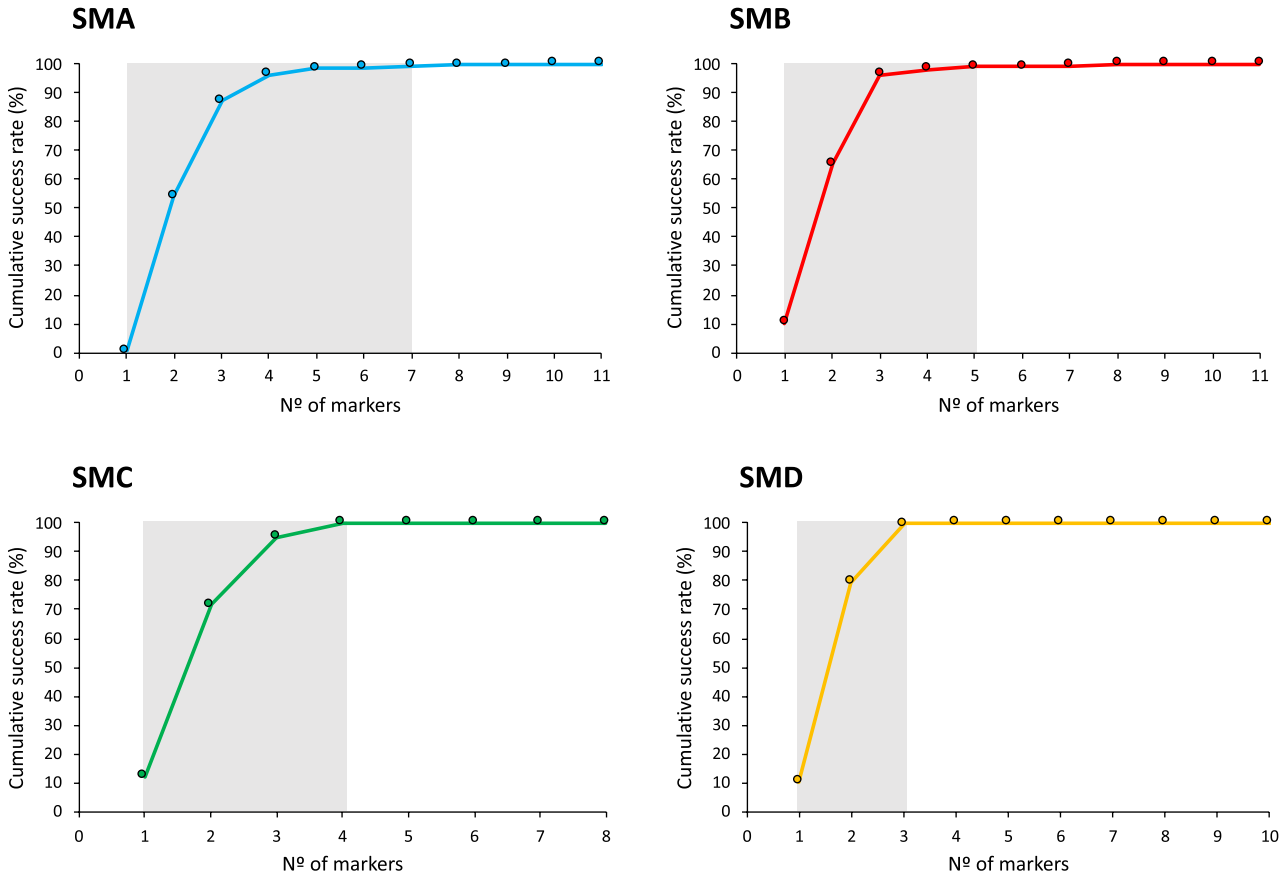


**Figure 3.** Cumulative success rate for parentage assignment based on exclusion with markers ranked on PIC value. The grey area indicates the *loci* required to reach more than 99% probability of assigning a correct parent–offspring relationship. SMA n = 92 parents; SMB, n = 15 parents; SMC, n = 15; SMD, n = 15.

was proposed in this species with the female as homogametic sex[8,61]. Although the SseLG01 has been proposed as a sex proto-chromosome due to the location of some key sex-determining genes and repetitive sequences[12,16], the spreading of Z/W sequences through the genome indicates that a further experimental validation is required to identify a putative major *loci* for sex determination.

In conclusion, this study uses two genome assemblies of Senegalese sole for the identification of SSR markers, sequence validation and cross-species synteny comparison analysis. A total of 106 selected SSR markers were structured in thirteen multiplex PCR assays available for whole-genome mapping. Moreover, forty high-polymorphic markers were used to optimize four high-variable supermultiplex PCRs suitable for pedigree analysis and genetic breeding programs. All SSR markers were positioned in the genome and integrated with previous published SSR markers to generate a new integrated genetic map containing 21 LGs. A macrosynteny comparison with *C. semilaevis* indicated the largest metacentric and submetacentric chromosomes of *S. senegalensis* could be explained by fusions and rearrangements of telocentric chromosomes in C. semilaevis. This integrated genetic map and the new multiplex PCRs provide a valuable resource for association studies, selection breeding and flatfish comparative genomics.

| High density SNP map | SSR markers | | | *Cynoglossus* Chromosomes | LD genetic map |
|---|---|---|---|---|---|
| | This study | LD genetic map | Total | | |
| SseLG01 | 11 | 3 | 14 | chr3,chr20 | LG21,LG27 |
| SseLG02 | 6 | 6 | 12 | chr14,chr16 | LG17,LG18,LG25 |
| SseLG03 | 5 | 5 | 10 | chr1, chr8, chrZ | LG7 |
| SseLG04 | 4 | 8 | 12 | chr11, chrZ | LG2 |
| SseLG05 | 8 | 7 | 15 | chrZ | LG4 |
| SseLG06 | 7 | 10 | 17 | chr9 | LG1 |
| SseLG07 | 4 | 15 | 19 | chr5 | LG3,LG26 |
| SseLG08 | 5 | 4 | 9 | chr4 | LG22,LG24 |
| SseLG09 | 5 | 5 | 10 | chr13 | LG16,LG20 |
| SseLG10 | 4 | 6 | 10 | chr6 | LG6 |
| SseLG11 | 3 | 5 | 8 | chr10 | LG10 |
| SseLG12 | 6 | 8 | 14 | chr15 | LG13,LG23 |
| SseLG13 | 4 | 0 | 4 | chr19 | – |
| SseLG14 | 4 | 8 | 12 | chr2 | LG8 |
| SseLG15 | 5 | 4 | 9 | chr12 | LG12 |
| SseLG16 | 4 | 5 | 9 | chr1 | LG15 |
| SseLG17 | 4 | 6 | 10 | chr7 | LG11 |
| SseLG18 | 5 | 2 | 7 | chr8 | LG19 |
| SseLG19 | 4 | 7 | 11 | chr17 | LG1,LG14 |
| SseLG20 | 4 | 3 | 7 | chr18 | LG5 |
| SseLG21 | 3 | 4 | 7 | chr14 | LG9 |
| Unplaced | 3 | 5 | 8 | – | |
| Total | 108 | 126 | 234 | | |

**Table 3.** SSR distribution. Markers are groups by the 21 linkage groups (SseLG) of the high-density SNP genetic map. The number of SSRs of this study and those from Low-density (LD) genetic linkage map (Molina-Luzon et al., 2015) are indicated. The location of markers in *C. semilaevis* genome by blasting the scaffold containing the SSR marker and the LG in the LD genetic map are indicated.

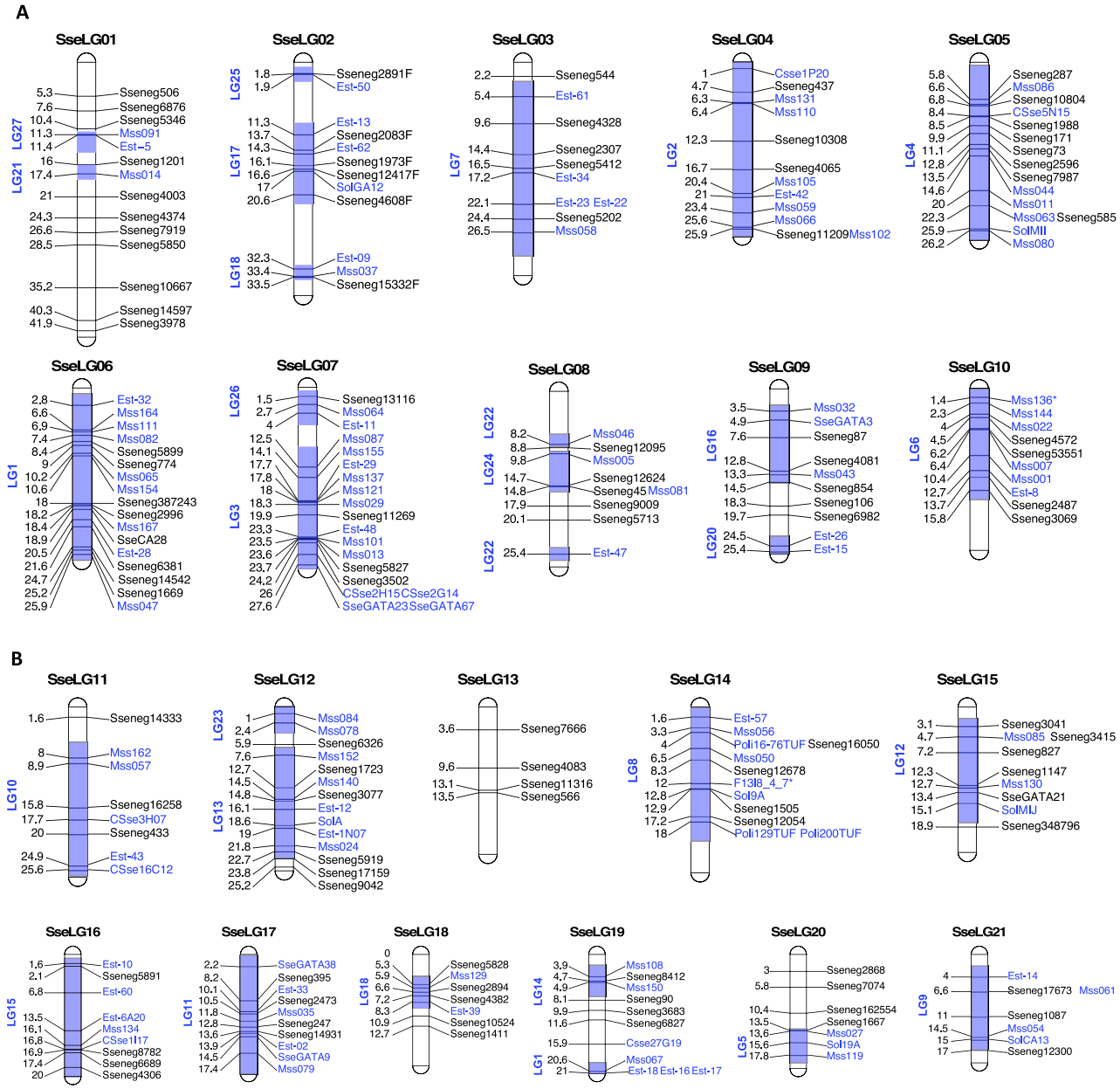


**Figure 4.** Integrated SSR genetic map of Senegalese sole (*S. senegalensis*). SseLG refer to the linkage groups according the high-density SNP genetic map. Genetic distance is indicated on the left. SSRs of this study are indicate in black and those from Molina-Luzón et al., 2015 in blue. The LGs previously assigned to these markers are shaded and indicated on the left. (**a**) SseLG1- SseLG10; (**b**) SseLG11- SseLG21.

Received: 29 July 2020; Accepted: 20 November 2020
Published online: 14 December 2020

## References

1. Cerda, J. & Manchado, M. Advances in genomics for flatfish aquaculture. *Genes Nutr.* **8**, 5–17. https://doi.org/10.1007/s12263-012-0312-8 (2013).
2. Chen, S. *et al.* Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat. Genet.* **46**, 253–260. https://doi.org/10.1038/ng.2890 (2014).
3. Xu, X. W. *et al.* Draft genomes of female and male turbot *Scophthalmus maximus*. *Sci. Data* **7**, 90. https://doi.org/10.1038/s41597-020-0426-6 (2020).
4. Shao, C. *et al.* The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nat. Genet.* **49**, 119–124. https://doi.org/10.1038/ng.3732 (2017).
5. Maroso, F. *et al.* Highly dense linkage maps from 31 full-sibling families of turbot (*Scophthalmus maximus*) provide insights into recombination patterns and chromosome rearrangements throughout a newly refined genome assembly. *DNA Res.* **25**, 439–450. https://doi.org/10.1093/dnares/dsy015 (2018).

6. Robledo, D., Rubiolo, J. A., Cabaleiro, S., Martinez, P. & Bouza, C. Differential gene expression and SNP association between fast- and slow-growing turbot (*Scophthalmus maximus*). *Sci. Rep.* **7**, 12105. https://doi.org/10.1038/s41598-017-12459-4 (2017).
7. Manchado, M., Planas, J. V., Cousin, X., Rebordinos, L. & Claros, M. G. In *Genomics in Aquaculture* (eds Mackenzie, S. & Jentoft, S.) 195–221 (Elsevier, Amsterdam, 2016).
8. Manchado, M., Planas, J. V., Cousin, X., Rebordinos, L. & Claros, M. G. In *The Biology of Sole* (eds Muñoz-Cueto, J. *et al.*) 375–394 (CDC Press, New York, 2019).
9. Souissi, A., Bonhomme, F., Manchado, M., Bahri-Sfar, L. & Gagnaire, P. A. Genomic and geographic footprints of differential introgression between two divergent fish species (*Solea* spp.). *Heredity* **121**, 579–593. https://doi.org/10.1038/s41437-018-0079-9 (2018).
10. Roman-Padilla, J., Rodriguez-Rua, A., Claros, M. G., Hachero-Cruzado, I. & Manchado, M. Genomic characterization and expression analysis of four apolipoprotein A-IV paralogs in Senegalese sole (*Solea senegalensis* Kaup). *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **191**, 84–98. https://doi.org/10.1016/j.cbpb.2015.09.010 (2016).
11. Carballo, C. *et al.* Genomic and phylogenetic analysis of choriolysins, and biological activity of hatching liquid in the flatfish Senegalese sole. *PLoS ONE* **14**, e0225666. https://doi.org/10.1371/journal.pone.0225666 (2019).
12. Portela-Bens, S. *et al.* Integrated gene mapping and synteny studies give insights into the evolution of a sex proto-chromosome in *Solea senegalensis*. *Chromosoma* **126**, 261–277. https://doi.org/10.1007/s00412-016-0589-2 (2017).
13. Cordoba, J. M., Chavarro, C., Schlueter, J. A., Jackson, S. A. & Blair, M. W. Integration of physical and genetic maps of common bean through BAC-derived microsatellite markers. *BMC Genom.* **11**, 436. https://doi.org/10.1186/1471-2164-11-436 (2010).
14. Molina-Luzon, M. J. *et al.* First haploid genetic map based on microsatellite markers in Senegalese sole (*Solea senegalensis*, Kaup 1858). *Mar. Biotechnol. (NY)* **17**, 8–22. https://doi.org/10.1007/s10126-014-9589-5 (2015).
15. Garcia, E. *et al.* Integrative genetic map of repetitive DNA in the sole *Solea senegalensis* genome shows a Rex transposon located in a proto-sex chromosome. *Sci. Rep.* **9**, 17146. https://doi.org/10.1038/s41598-019-53673-6 (2019).
16. Rodriguez, M. E. *et al.* Evolution of the proto sex-chromosome in *Solea senegalensis*. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms20205111 (2019).
17. Wang, W. *et al.* High-density genetic linkage mapping in turbot (*Scophthalmus maximus* L.) based on SNP markers and major sex- and growth-related regions detection. *PLoS ONE* **10**, e0120410. https://doi.org/10.1371/journal.pone.0120410 (2015).
18. Lu, Q. *et al.* Genome-wide identification of microsatellite markers from cultivated peanut (*Arachis hypogaea* L.). *BMC Genom.* **20**, 799. https://doi.org/10.1186/s12864-019-6148-5 (2019).
19. Sundaray, J. K. *et al.* Simple sequence repeats (SSRs) markers in fish genomic research and their acceleration via next-generation sequencing and computational approaches. *Aquac. Int.* **24**, 1089–1102. https://doi.org/10.1007/s10499-016-9973-4 (2016).
20. Zarouri, B. *et al.* Whole-genome genotyping of grape using a panel of microsatellite multiplex PCRs. *Tree Genet. Genomes* https://doi.org/10.1007/s11295-015-0843-4 (2015).
21 Lee-Montero, I. *et al.* Development of the first standardised panel of two new microsatellite multiplex PCRs for gilthead seabream (*Sparus aurata* L.). *Anim. Genet.* **44**, 533–546. https://doi.org/10.1111/age.12037 (2013).
22. Carballo, C. *et al.* Heritability estimates and genetic correlation for gowth traits and LCDV susceptibility in gilthead sea bream (*Sparus aurata*). *Fishes* **5**, 2. https://doi.org/10.3390/fishes5010002 (2020).
23. Garcia-Celdran, M. *et al.* Estimates of heritabilities and genetic correlations of growth and external skeletal deformities at different ages in a reared gilthead sea bream (*Sparus aurata* L.) population sourced from three broodstocks along the Spanish coasts. *Aquaculture* **445**, 33–41. https://doi.org/10.1016/j.aquaculture.2015.04.006 (2015).
24. Lee-Montero, I. *et al.* Genetic parameters and genotype-environment interactions for skeleton deformities and growth traits at different ages on gilthead seabream (*Sparus aurata* L.) in four Spanish regions. *Anim. Genet.* **46**, 164–174. https://doi.org/10.1111/age.12258 (2015).
25. Negrin-Baez, D., Negrin-Baez, D., Rodriguez-Ramilo, S. T., Afonso, J. M. & Zamorano, M. J. Identification of quantitative trait loci associated with the skeletal deformity LSK complex in gilthead seabream (*Sparus aurata* L.). *Mar. Biotechnol.* **18**, 98–106. https://doi.org/10.1007/s10126-015-9671-7 (2016).
26. Benzekri, H. *et al.* De novo assembly, characterization and functional annotation of Senegalese sole (*Solea senegalensis*) and common sole (*Solea solea*) transcriptomes: Integration in a database and design of a microarray. *BMC Genom.* **15**, 952. https://doi.org/10.1186/1471-2164-15-952 (2014).
27. Flores-Renteria, L. & Krohn, A. Scoring microsatellite loci. *Methods Mol. Biol.* **1006**, 319–336. https://doi.org/10.1007/978-1-62703-389-3_21 (2013).
28. Nater, A., Kopps, A. M. & Krutzen, M. New polymorphic tetranucleotide microsatellites improve scoring accuracy in the bottlenose dolphin *Tursiops aduncus*. *Mol. Ecol. Resour.* **9**, 531–534. https://doi.org/10.1111/j.1755-0998.2008.02246.x (2009).
29. Wood, D. E. & Salzberg, S. L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46. https://doi.org/10.1186/gb-2014-15-3-r46 (2014).
30. Marco-Sola, S., Sammeth, M., Guigo, R. & Ribeca, P. The GEM mapper: Fast, accurate and versatile alignment by filtration. *Nat. Methods* **9**, 1185–1188. https://doi.org/10.1038/nmeth.2221 (2012).
31. Vaser, R., Sovic, I., Nagarajan, N. & Sikic, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746. https://doi.org/10.1101/gr.214270.116 (2017).
32. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191 (2018).
33. Zimin, A. V. *et al.* Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res.* **27**, 787–792. https://doi.org/10.1101/gr.213405.116 (2017).
34. Zimin, A. V. *et al.* The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669–2677. https://doi.org/10.1093/bioinformatics/btt476 (2013).
35. Miller, J. R. *et al.* Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* **24**, 2818–2824. https://doi.org/10.1093/bioinformatics/btn548 (2008).
36. Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351 (2015).
37. Waterhouse, R. M. *et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548. https://doi.org/10.1093/molbev/msx319 (2018).
38. Claros, M. G., Seoane, P. & Manchado, M. Sequences and annotations of a provisional genome draft of a *Senegalese sole* female. figshare https://doi.org/10.6084/m9.figshare.12472100.v1 (2020).
39. Beier, S., Thiel, T., Munch, T., Scholz, U. & Mascher, M. MISA-web: a web server for microsatellite prediction. *Bioinformatics* **33**, 2583–2585. https://doi.org/10.1093/bioinformatics/btx198 (2017).
40. Sanchez, J. J. *et al.* Multiplex PCR and minisequencing of SNPs–a model with 35 Y chromosome SNPs. *Forensic. Sci. Int.* **137**, 74–84. https://doi.org/10.1016/s0379-0738(03)00299-8 (2003).
41. Kalinowski, S. T., Taper, M. L. & Marshall, T. C. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* **16**, 1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x (2007).
42. Peakall, R. & Smouse, P. E. GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537–2539. https://doi.org/10.1093/bioinformatics/bts460 (2012).

43. Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. Micro-checker: Software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**, 535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x (2004).
44. Sekino, M. & Kakehi, S. PARFEX v1.0: An EXCEL-based software package for parentage allocation. *Conserv. Genet. Resour.* **4**, 275–278. https://doi.org/10.1007/s12686-011-9523-3 (2012).
45. Ouellette, L. A., Reid, R. W., Blanchard, S. G. & Brouwer, C. R. LinkageMapView-rendering high-resolution linkage and QTL maps. *Bioinformatics* **34**, 306–307. https://doi.org/10.1093/bioinformatics/btx576 (2018).
46. Funes, V., Zuasti, E., Catanese, G., Infante, C. & Manchado, M. Isolation and characterization of ten microsatellite loci for Senegal sole (*Solea senegalensis* Kaup). *Mol. Ecol. Notes* **4**, 339–341. https://doi.org/10.1111/j.1471-8286.2004.00690.x (2004).
47. Castro, J. *et al.* A microsatellite marker tool for parentage analysis in Senegal sole (*Solea senegalensis*): Genotyping errors, null alleles and conformance to theoretical assumptions. *Aquaculture* https://doi.org/10.1016/j.aquaculture.2006.09.001 (2006).
48. Figueras, A. *et al.* Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): A fish adapted to demersal life. *DNA Res.* **23**, 181–192. https://doi.org/10.1093/dnares/dsw007 (2016).
49. Negrin-Baez, D. *et al.* A set of 13 multiplex PCRs of specific microsatellite markers as a tool for QTL detection in gilthead seabream (*Sparus aurata* L.). *Aquac Res.* **46**, 45–58 (2015).
50. Molina-Luzon, M. J. *et al.* Validation and comparison of microsatellite markers derived from Senegalese sole (*Solea senegalensis*, Kaup) genomic and expressed sequence tags libraries. *Mol. Ecol. Resour.* **12**, 956–966. https://doi.org/10.1111/j.1755-0998.2012.03163.x (2012).
51. Chen, S.-L., Shao, C.-W., Xu, G.-B., Liao, X.-L. & Tian, Y.-S. Development of 15 novel dinucleotide microsatellite markers in the Senegalese sole *Solea senegalensis*. *Fish. Sci.* **74**, 1357–1359. https://doi.org/10.1111/j.1444-2906.2008.01668.x (2008).
52. Porta, J., Porta, J. M., Martínez-Rodríguez, G. & Álvarez, M. C. Development of a microsatellite multiplex PCR for Senegalese sole (*Solea senegalensis*) and its application to broodstock management. *Aquaculture* **256**, 159–166. https://doi.org/10.1016/j.aquaculture.2006.02.022 (2006).
53. De La Herran, R. *et al.* A highly accurate, single PCR reaction for parentage assignment in Senegal sole based on eight informative microsatellite loci. *Aquacult. Res.* **39**, 1169–1174. https://doi.org/10.1111/j.1365-2109.2008.01979.x (2008).
54. Zalapa, J. E. *et al.* Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am. J. Bot.* **99**, 193–208. https://doi.org/10.3732/ajb.1100394 (2012).
55. Labuschagne, C., Nupen, L., Kotze, A., Grobler, P. J. & Dalton, D. L. Assessment of microsatellite and SNP markers for parentage assignment in ex situ African Penguin (*Spheniscus demersus*) populations. *Ecol. Evol.* **5**, 4389–4399. https://doi.org/10.1002/ece3.1600 (2015).
56. Merlo, M. A. *et al.* Analysis of the histone cluster in Senegalese sole (*Solea senegalensis*): Evidence for a divergent evolution of two canonical histone clusters. *Genome* **60**, 441–453. https://doi.org/10.1139/gen-2016-0143 (2017).
57. Bouza, C. *et al.* An Expressed sequence Tag (EST)-enriched genetic map of turbot (*Scophthalmus maximus*): A useful framework for comparative genomics across model and farmed teleosts. *BMC Genet.* **13**, 54. https://doi.org/10.1186/1471-2156-13-54 (2012).
58. Garcia-Angulo, A. *et al.* Genome and phylogenetic analysis of genes involved in the immune system of *Solea senegalensis*—Potential applications in aquaculture. *Front. Genet.* **10**, 529. https://doi.org/10.3389/fgene.2019.00529 (2019).
59. Garcia-Angulo, A. *et al.* Evidence for a Robertsonian fusion in *Solea senegalensis* (Kaup, 1858) revealed by zoo-FISH and comparative genome analysis. *BMC Genom.* **19**, 818. https://doi.org/10.1186/s12864-018-5216-6 (2018).
60. Zhang, S., Wang, C. & Chu, J. C-banding pattern and nucleolar organizer regions of amphioxus *Branchiostoma belcheri* tsingtauense Tchang et Koo, 1936. *Genetica* **121**, 101–105. https://doi.org/10.1023/b:gene.0000019939.38428.e5 (2004).
61. Viñas, J., Asensio, E. & Piferrer, F. Gonadal sex differentiation in the Senegalese sole (*Solea senegalensis*) and first data on the experimental manipulation of its sex ratios. *Aquaculture* https://doi.org/10.1016/j.aquaculture.2012.12.012 (2012).


## Acknowledgements

This study was funded by project RTA2017-00054-C03-01, RTA2017-00054-C03-02 and RTA2017-00054-C03-03 funded from MCIU/AEI/FEDER, UE. IGC is funded by a predoctoral fellowship from INIA. We thank Manuel Aparicio for helping in the optimization of SSR Multiplex PCRs and Tyler Alioto for his useful comments about the sequencing strategy and genome assembly. This work would not have been possible without the computer resources and the technical support provided by the Plataforma Andaluza de Bioinformática of the University of Málaga and CNAG.

## Author contributions

I.G.C.: investigation, data curation, writing—original draft preparation. Multiplex. In silico analysis. C.P.: validation. Multiplex SSR. H.B.: Investigation. In silico analysis. J.J.S.: Methodology. Multiplex design. P.S.: Methodology. In silico analysis. F.C.: Investigation. Genome assembly. M.G.: Investigation. Genome sequencing. M.J.Z.: conceptualization, funding acquisition, validation. M.G.C.: supervision, conceptualization, funding acquisition, writing—review & Editing. M.M.: resources, conceptualization, funding acquisition, writing—original draft preparation, writing—review & editing.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-020-78397-w.

**Correspondence** and requests for materials should be addressed to M.M.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Aquaculture 539 (2021) 736665

Contents lists available at ScienceDirect

Aquaculture

journal homepage: www.elsevier.com/locate/aquaculture

# Genetic parameter estimates and identification of SNPs associated with growth traits in Senegalese sole

Israel Guerrero-Cózar [a], Eduardo Jimenez-Fernandez [b], Concha Berbel [a], José Córdoba-Caballero [c], M. Gonzalo Claros [c,d,e,f], Ricardo Zerolo [g], Manuel Manchado [a,h,*]

[a] IFAPA Centro El Toruño, Junta de Andalucía, Camino Tiro Pichón s/n, 11500 El Puerto de Santa María, Cádiz, Spain
[b] Free Radical Research Group, University of the Highlands and Islands, Centre for Health Sciences, Inverness IV2 3JH, UK
[c] Universidad de Málaga, Department of Molecular Biology and Biochemistry, Málaga E-29071, Spain
[d] CIBER de Enfermedades Raras (CIBERER), Málaga E-29071, Spain
[e] Institute of Biomedical Research in Málaga (IBIMA), IBIMA-RARE, Málaga E-29010, Spain
[f] Instituto de Hortofruticultura Subtropical y Mediterránea (IHSM-UMA-CSIC), Málaga E-29010, Spain
[g] CUPIMAR, Ctra. Carraca, s/n, Salina San Juan Bautista, San Fernando, Cádiz, Spain
[h] Crecimiento Azul, Centro IFAPA El Toruño, Unidad Asociada al CSIC, Spain

ARTICLE INFO

*Keywords:*
Senegalese sole
Genetic estimates
Association analysis
SNPs

ABSTRACT

Breeding programs are essential for aquaculture. The present study was conducted to investigate the genetic parameters of four growth traits in Senegalese sole. Families were produced by mass spawning ($n$ = 2171 offspring) from an industrial broodstock. Weight, total area, standard length and width (the two latter measured directly on the fish and by image analysis) were determined before the on-growing stage (age ~ 400 d) and at harvest (~800 d) in RAS. Phenotypic data are presented and discussed. Females grew faster than males and a high variation of sex ratios by family was observed. Heritabilities were high for all traits ranging from 0.568 to 0.609 at 400 d and from 0.424 to 0.500 at 800d. The genetic correlations between traits were very high (>0.94) at both ages. Genetic estimates for traits measured *in vivo* and by digital image analysis were similar. To investigate genomic regions associated with these growth traits, a low-density array using a set of 60 single nucleotide polymorphisms (SNPs) distributed in 17 linkage groups was designed. Using samples from wild fish, a total of 49 assays for SNP analysis were validated. The association analysis was carried out using two fast-growth families and two slow-growth families ($n$ = 279). Two significantly-associated SNPs with most of the traits at both ages were detected. They were located in the general transcription factor 3C polypeptide 4 (*gtf3c4*) and the mitochondrial fission process protein 1 (*mtfp1*). Results are of high relevance for genetic breeding programs in this species.

## 1. Introduction

The Senegalese sole (*Solea senegalensis*) is one the most valuable flatfish in Southern Europe aquaculture. Its production has grown exponentially in the last decade due to significant improvements in larval rearing, optimization of dietary requirements and the use of recirculation technologies (RAS) for on-growing (Manchado et al., 2016; Manchado et al., 2019; Morais et al., 2016). However, reproductive dysfunction of breeders reared in captivity and the understanding of the courtship behavior required for a successful spawning still persist as two major bottlenecks for larval production (Fatsini et al., 2016; Fatsini et al., 2017; Fatsini et al., 2020). Hormonal therapies revealed as unsuccessful to release fertilized eggs (Agulleiro et al., 2006) and the low volume of sperm production (<130 μl) makes impractical the use of *in vitro* techniques as a routine procedure in the hatcheries (Chauvigne et al., 2017). However, in the last years, environmental control techniques based on thermocycles applied to wild males and hatchery-produced females have become a useful strategy to circumvent at least partially some of larval production limitations (Martin et al., 2019). This new approach can be used for the design of breeding programs based on mass spawning and the selection of best-ranked females.

A sustainable aquaculture is dependent on genetic breeding programs that select the best breeders to produce high-quality offspring. Currently, production of most economically important marine species in

* Corresponding author at: IFAPA Centro *El Toruño*, Camino Tiro de Pichón s/n, 11500 El Puerto de Santa María, Cádiz, Spain.
*E-mail address:* manuel.manchado@juntadeandalucia.es (M. Manchado).

https://doi.org/10.1016/j.aquaculture.2021.736665
Received 24 October 2020; Received in revised form 17 March 2021; Accepted 18 March 2021
Available online 20 March 2021

Europe are supported by selection schemes for relevant traits related to growth, disease resistant, morphology and flesh quality (Janssen et al., 2017). These programs have been designed taking into account the reproduction characteristics and the industrial production models of each species. A mass-spawning model followed by pedigree reconstruction using microsatellites and BLUP analysis was used in gilthead seabream (*Sparus aurata*) and common sole (*Solea solea*) (Blonk et al., 2010a; Blonk et al., 2010b; Brown et al., 2005; Carballo et al., 2020; Lee-Montero et al., 2013; Lee-Montero et al., 2015; Navarro et al., 2009). A major issue that these programs have to deal with is the skewed familial contributions and the high variance of family sizes. In gilthead seabream, these effects have been minimized by synchronizing egg release by photoperiod followed by mixing egg batches from different broodstock tanks through four consecutive days (Carballo et al., 2020). Following this strategy, heritabilities for weight at different ages (from hatchery to harvest), flesh quality, disease resistance against bacteria and virus and skeletal deformations were estimated (Aslam et al., 2018; Carballo et al., 2020; Garcia-Celdran et al., 2016; Garcia-Celdran et al., 2015a; Garcia-Celdran et al., 2015b; Lee-Montero et al., 2015). Interestingly, a significant genotype×production system interaction was reported in the close species *S. solea* demonstrating that in soles exist families with different growth performance in RAS (artificial environment) or ponds (natural environment) (Mas-Muñoz et al., 2013). In Senegalese sole, there is still no genetic estimates for growth traits. However, recent advances in methodologies for juvenile tagging (Carballo et al., 2018), genotyping using multiplex PCR for parentage assignment (Guerrero-Cózar et al., 2020), and the synchronization of spawns using thermoperiod control (Martin et al., 2019; Martin et al., 2014) make possible to investigate genetic parameters under industrial conditions in RAS.

Breeding programs in aquaculture benefit of cost-effective sequencing technologies to implement marker-assisted selection (MAS) and genomic selection. The identification of genomic regions associated with a trait of interest (Genome Wide Association Studies;GWAS) are becoming popular in aquaculture and markers associated with growth (Kyriakis et al., 2019), pigmentation (Bertolini et al., 2020) disease resistance (Palaiokostas et al., 2016) or sex (Purcell et al., 2018) have been reported. Although several methodological approaches and genotyping platforms have been used in marine cultivated species, recent studies have demonstrated that that low-density single nucleotide polymorphism (SNP) panels are a cost-effective solution for broadening the impact of genomic selection in aquaculture. However, this approach require non-random SNP selection to increase prediction accuracy (Kriaridou et al., 2020). One source of genetic markers is the RNA sequencing (RNA-seq) (Brouard et al., 2019; Espinosa et al., 2020), most of which are linked to specific genes and suitable for quantitative PCR (qPCR). The use of these markers in low-density arrays (OpenArray® technology) has been successfully applied in marker-assisted selection (MAS) for plant breeding (Chagne et al., 2019) and fine-mapping for disease diagnosis (Gutierrez-Camino et al., 2018; Verbeek et al., 2012).

The aim of this study was to estimate heritabilities and genetic correlations for four growth traits both at the beginning of on-growing period in RAS (~400 days) and at harvest (~800 d) using offspring of a commercial Senegalese sole broodstock. Data for two traits were validated by *in situ* measurement and digital image analysis. Moreover, a SNP-based array was designed and validated using wild fish. An association analysis using low-density arrays was carried out using four families with different growth rates. The results obtained will be useful to design breeding program schemes to enhance the Senegalese sole aquaculture.

## 2. Material and methods

### 2.1. Animals

Genetic families were created by mass-spawning using a wild broodstock ($n = 150$ animals) distributed into nine tanks (ranging from 6 to 26 breeders) with a 1:1 sex ratio. Mass spawning was synchronized by thermoperiod manipulation as previously reported (Martin et al., 2014). Full information about food and animal size is depicted in Suppl. file 1. Animals were sample for blood (~0.5 ml) by puncturing in the caudal vein using a heparinized syringe, adding heparin (100 mU) and keeping at −20 °C until use.

Since not all broodstock tanks responded to each thermocycle treatment, seven evaluation batches were created with eggs from at least three tanks from July 2014 to Nov 2015 (Table 1). To maximize family representation, eggs in each batch were proportionally mixed by considering the total volume of eggs by tank. Moreover, eggs from one broodstock were represented in all batches to make easier data normalization and convergence. Each evaluation batch was always managed as an experimental unit from larvae to harvest under commercial procedures and fish were never graded. For evaluation, a subset of 200 and 550 animals proportional to the estimated number of families in each batch (Table 1) was randomly selected and intraperitoneally tagged as previously reported (Carballo et al., 2018).

Animals were sampled before entering the final on-growing period in recirculation aquaculture systems (RAS; ~400 days) and at harvest (~800 d). Since total number of tagged fish in the evaluation batches represented between 4 and 11% of total population, only those tagged fish identified after one pass through the FISH Reader (Zeus, Trovan) ($n = 1843$) were sampled at 400 d. Later at harvest (800 d), all animals ($n = 2171$) were *in situ* sacrificed using slurring ice following commercial techniques and 60 animals of each batch were kept alive as future breeders. Sacrificed animals were taken a piece of caudal fin that was preserved in 99% alcohol and alive fish were sampled for blood as indicated above for DNA isolation. Moreover, sacrificed animals were dissected and alive fish sampled to record the sex and the presence of white nodules compatible with amoebic disease. Body weight (W), standard length (from mouth to beginning of caudal fin;SL) and width (Wi) were *in situ* measured for all fish. Moreover, animals were individually photographed using a Canon EOs1300D camera following the methodology previously established in PROGENSA® (Navarro et al., 2016). Image analysis was carried out using the Fiji 2.0.0-rc-69/1.52p and standard length (SLI), width (WiI) and total area (A) were measured. Further details on fish handling and culture are indicated in Suppl. file 1. All procedures were authorized by the Bioethics and Animal Welfare Committee of IFAPA and given the registration number 10/06/2016/101.

### 2.2. DNA isolation and parentage assignment

Total DNA from caudal fin (30 mg) or blood (20 μL) was isolated using Isolate II Genomic DNA Kit treated with RNase A (Bioline, London, UK) using the manufacturer's instructions. DNA was quantified spectrophotometrically using the Nanodrop ND-8000. Genotyping was carried out using 11 *loci* in a supermultiplex PCR (Guerrero-Cózar et al., 2020) that were run on an ABI3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Genotypes were collected using Genemapperv3.8 (Applied Biosystems, Foster City, CA, USA) and parentage assignment was determined using the exclusion method with Vitassign v8.2.1 (Vandeputte et al., 2006).

### 2.3. Genetic analysis

Before carrying out the genetic analyses, main factors were identified by ANOVA analysis using SPSS v.23 (SPSS, Chicago, IL, USA). Weight and area were transformed using cubic square root and square root, respectively to fit normality. Sex, amoebic disease and evaluation batch factors were found to be significant for all traits, hence, they were included as fixed factors in the model. Genetic parameters were estimated at the beginning of on-growing in RAS (~400 d) and at harvest (~800 d). Genetic estimates for heritability and correlations were

**Table 1**
Growth traits for Senegalese sole for different batches. The birth date (birth), number of tagged fish (n), age at tagging (Aget), weight at tagging (Wt) and % males (M) of each batch are shown. For samplings at 400 and 800 d, number of fish sampled (n), age, weight (W), standard length (SL) and width (Wi) as *in situ* determined, standard length (SLI), width (WiI) and total area (A) from image analysis are indicated. The number of families in each batch (nF) is also shown. Letters indicate statistically significant differences between batches for each batch.

| Batch | Birth | %M | Aget | Wt | n400 | Age | W400 | SL400 | Wi400 | SLI400 | WiI400 | A400 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Jul_14 | 61.7 | 278 | 5.3 ± 3.1 | 137 | 446 | 29.2 ± 16.4$^{c}$ | 12.1 ± 2.1$^{c}$ | 4.8 ± 1.1$^{c}$ | 12.4 ± 2.2$^{c}$ | 4.8 ± 1.1$^{c}$ | 50.4 ± 19.8$^{d}$ |
| 2 | Oct_14 | 69.1 | 239 | 2.9 ± 1.5 | 287 | 415 | 26.4 ± 19.2$^{bc}$ | 11.6 ± 2.3$^{bc}$ | 4.4 ± 1.0$^{b}$ | 11.9 ± 2.4$^{c}$ | 4.5 ± 1.0$^{b}$ | 44.2 ± 21.0$^{c}$ |
| 3 | Mar_15 | 58.0 | 243 | 5.8 ± 4.5 | 273 | 397 | 18.2 ± 14.9$^{a}$ | 10.4 ± 2.3$^{a}$ | 3.8 ± 1.0$^{a}$ | 10.2 ± 2.3$^{a}$ | 3.7 ± 1.0$^{a}$ | 31.2 ± 15.9$^{a}$ |
| 4 | May_15 | 43.9 | 193 | 3.8 ± 2.8 | 422 | 399 | 22.3 ± 14.3$^{b}$ | 11.3 ± 2.1$^{b}$ | 4.3 ± 1.0$^{b}$ | 11.1 ± 2.1$^{b}$ | 4.2 ± 0.9$^{b}$ | 37.3 ± 15.3$^{b}$ |
| 5 | Jun_15 | 49.0 | 171 | 1.8 ± 0.8 | 229 | 398 | 22.9 ± 14.1$^{b}$ | 11.3 ± 2.0$^{b}$ | 4.4 ± 0.9$^{b}$ | 11.1 ± 2.0$^{b}$ | 4.3 ± 0.9$^{b}$ | 38.1 ± 15.0$^{b}$ |
| 6 | Sep_15 | 43.3 | 150 | 2.2 ± 1.0 | 234 | 395 | 53.9 ± 34.6$^{d}$ | 14.3 ± 3.0$^{d}$ | 5.8 ± 1.4$^{d}$ | 13.8 ± 2.9$^{d}$ | 5.6 ± 1.4$^{d}$ | 62.0 ± 26.7$^{e}$ |
| 7 | Nov_15 | 65.1 | 154 | 1.7 ± 0.9 | 261 | 414 | 60.9 ± 44.0$^{d}$ | 14.6 ± 3.5$^{d}$ | 6.0 ± 1.7$^{d}$ | 14.3 ± 3.5$^{d}$ | 5.8 ± 1.6$^{d}$ | 66.1 ± 32.4$^{e}$ |
| Batch | nF | | | | n800 | Age | W800 | SL800 | Wi800 | SLI800 | WiI800 | A800 |
| 1 | 11 | | | | 167 | 861 | 219.9 ± 123.0$^{b}$ | 22.1 ± 3.8 $^{b}$ | 9.5 ± 2.0 $^{c}$ | 22.2 ± 3.9 $^{b}$ | 9.3 ± 2.0 $^{b}$ | 159.7 ± 59.9 $^{b}$ |
| 2 | 23 | | | | 408 | 844 | 201.2 ± 129.8$^{b}$ | 22.4 ± 4.3 $^{b}$ | 8.9 ± 2.0 $^{b}$ | 22.5 ± 4.3 $^{b}$ | 8.8 ± 1.9 $^{b}$ | 153.0 ± 63.6 $^{b}$ |
| 3 | 14 | | | | 333 | 733 | 157.4 ± 103.1$^{a}$ | 20.6 ± 4.5 $^{a}$ | 8.1 ± 2.1 $^{a}$ | 20.3 ± 4.4 $^{a}$ | 7.9 ± 2.0 $^{a}$ | 128.7 ± 57.7 $^{a}$ |
| 4 | 17 | | | | 490 | 789 | 221.9 ± 113.9$^{b}$ | 22.7 ± 3.8 $^{b}$ | 9.2 ± 1.9 $^{b\ c}$ | 22.2 ± 3.8 $^{b}$ | 9.0 ± 1.9 $^{b}$ | 158.0 ± 55.9 $^{b}$ |
| 5 | 12 | | | | 259 | 817 | 327.3 ± 146.1 $^{c}$ | 25.6 ± 3.9 $^{c}$ | 10.7 ± 1.9 $^{d\ c}$ | 25.1 ± 3.9 $^{c}$ | 10.4 ± 1.9 $^{c}$ | 206.5 ± 64.9 $^{c}$ |
| 6 | 13 | | | | 245 | 780 | 425.3 ± 182.8 $^{d}$ | 27.4 ± 3.7 $^{cd}$ | 12.0 ± 2.0 $^{f}$ | 27.9 ± 3.7 $^{d}$ | 11.8 ± 1.9 $^{e}$ | 246.0 ± 71.6 $^{d}$ |
| 7 | 22 | | | | 269 | 806 | 395.6 ± 206.2 $^{d}$ | 26.4 ± 5.0 $^{cd}$ | 11.4 ± 2.4 $^{e}$ | 25.9 ± 4.8 $^{c}$ | 11.0 ± 2.3 $^{d}$ | 221.4 ± 80.9 $^{c}$ |

calculated using trivariate animal models fitted by restricted maximum likelihood (REML) in WOMBAT (Meyer, 2007): $y = X\beta + Zu + e$, where y is the observed trait, β, is the vector for fixed factor (sex, batch and amoebic lesions), u is the vector for animal random factor and e is the error. The age was initially tested as covariable but it was removed from the model since no effect on genetic estimates was observed. The adjusted breeding values were estimated using BLUP and later used for association tests.

### 2.4. OpenArray® design and methodology

For fish genotyping, a set of 60 SNPs was predicted after mapping 30 Illumina RNA-seq libraries (Benzekri et al., 2014; Cordoba-Caballero et al., 2020) onto a genome draft (Manchado et al., 2019). SNPs were identified with *Varscan mpileup2snp* (Koboldt et al., 2012) with a minimal coverage of 65 reads per position and a minimal amount of 15 reads supporting the SNP. As general criteria for selecting putative candidate sequences for primers and probes design, no ambiguities should exist into the 10 nucleotides (nt) upstream or downstream of identified SNP marker and the total number of ambiguities in the 600 nt surrounding the SNP should not be higher than 5. To position the markers in the Senegalese sole genome, they were mapped in the scaffolds of a female genome (Claros et al., 2020; Guerrero-Cózar et al., 2020) and the anchored megascaffolds (SseLG) to a high-density genetic map of *S. senegalensis* (Supp. file 1, tab "Probe_positioning"). Markers were drawn in the genome using the software linkagemapview (Ouellette et al., 2018). As we were also interested in the identification of markers associated with sex, the sequences were mapped onto the flatfish *Cynoglossus semilaevis* genome to identify markers located in sexual chromosomes ZW and 14.

Primers and probes were designed using the on-line Custom TaqMan® Assay Design Tool (www.thermofisher.com). Assays names are indicated in Suppl. file 2 Tab "Primers_Probes". Genotyping was carried out using a custom TaqMan® OpenArray® Genotyping Plate with 60 assays. Samples (2.5 μl of DNA sample normalized at 40 ng $\mu l^{-1}$) were mixed with an equal volume of TaqMan® OpenArray® Genotyping Master Mix and each subarray was loaded into the OpenArray® plates with the OpenArray® AccuFill™ System according to the manufacturer's protocols. Genotyping plates were run in the QuantStudio™ 12 K Flex Real-Time PCR System (Thermo Fisher Scientific). The samples were amplified using the thermal cycling conditions established manufacturer's protocol.

To test if markers were polymorphic, wild breeders from the broodstock population ($n = 164$; female 83 and male 81) were selected. Results were analyzed using the QuantStudio™ 12 K Flex software and Thermo Fisher cloud. As the assays were newly designed, each assay was manually examined by viewing the real-time trace and the endpoint call. Data were exported as a matrix of genotypic calls for each individual sample. Hardy-Weinberg equilibrium (HWE) was tested using the SNPstats (Sole et al., 2006). Assays calls with successful amplification are indicated in Supp. file 2. Seven assays with a call rate lower than 90% were excluded from analysis (AN3267M, ANMF2WF, ANNKWGD, ANPRP2A, AN2XDMN, ANEPWZV, ANRWJK9). Only three markers appeared as monomorphic (ANFVRKU, ANKA9CK, ANKA9CJ) and one marker was not in HWE (ANAAFAZ) in whole population data set and it was excluded from the association analysis.

### 2.5. Association analysis

Four families ($n = 279$) with different adjusted weight at harvest (800 d) were selected for association analysis. Family 1 ($n = 87$) and 2 ($n = 64$) were considered as fast growing (FG-1) families (average adjusted weight 442.8 ± 19.7 g and 382.0 ± 9.6 g, respectively). The family 3 ($n = 50$) and 4 ($n = 70$) were slow-growing (SG) and half-sibs (average weight of 262.8 ± 6.2 g and 229.5 ± 4.3 g, respectively) (Fig. 1). When the subset of animals also sampled at 400 d was compared, only family 1 but not family 2 had a higher weight than families 3 and 4 at this age (Fig. 1).

To identify SNPs associated with sex, wild fish and animals from the four families were analyzed using a logistic mixed model (multi-step) approach as implemented in the R package GENABEL (v1.8–0) (Aulchenko et al., 2007) that was the best fit model for binary traits (Female = 0 and Male = 1). Growth marker-trait associations using adjusted phenotypic traits for the four families were tested using TASSEL software *v*5 (Bradbury et al., 2007). SNPs were filtered using a minimum allele frequency (MAF) of 0.05. In order to control false associations, population structure (Q) and/or relatedness (K) between individuals were taken into account in the general linear model (GLM) and the mixed linear model (MLM). The Q matrix based on principal component analysis and the kinship (K) matrix were calculated using TASSEL. Four statistical models were tested: naïve-model (GLM without any correction for population structure); Q-model (GLM with Q-matrix as correction for population structure); K-model (MLM with K-matrix as correction for kinship relationship structure). QK-model (MLM with Q-matrix and K-matrix as correction for population structure and kinship relationships). Fitness of different GWAS models for all traits was evaluated using quantile (Q-Q) plots of the observed *vs* expected -log10(*p*) values which should follow a uniform distribution under the null

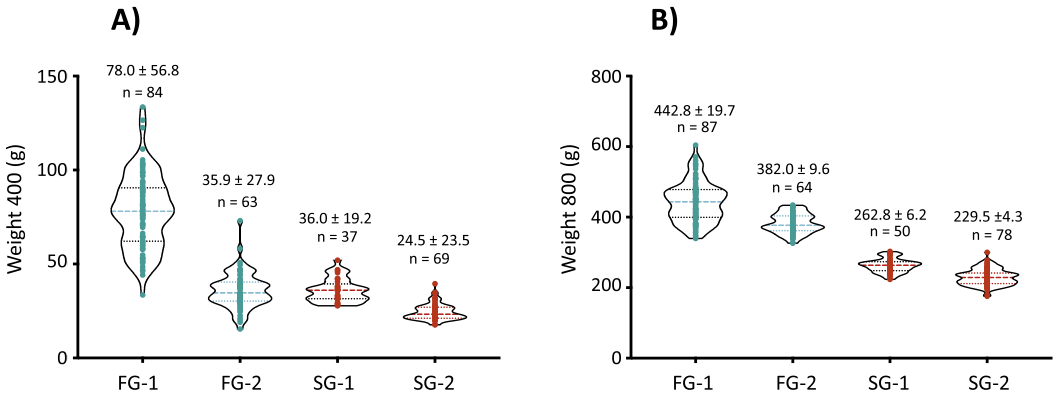


**Fig. 1.** Selected families for association study. Families were selected by their differences in adjusted weight at harvest: two fast-growing (FG-1 and FG-2; blue) and two slow-growing (SG-1 and SG-2; red) families. Weight at 400 (A) and 800 d (B). Average weight ± SD at 400 (A) and 800 d (B). and number of individuals analyzed for in each family are indicated. The number of individuals in each family at 400 d is lower than 800 d since not all animals were sampled to reduce handling effects. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

hypothesis. Association tests and Q-Q plots were further validated using GWASpoly R package (Rosyara et al., 2016) under the QK-model in which K was constructed on DAPC technique. Bonferroni's correction (with genome-wide $\alpha = 0.05$) was used for establishing a *P*-value detection threshold for statistical significance.

## 3. Results

### 3.1. Phenotype data

Seven fish batches (Table 1) from overall 150 breeders (distributed in nine tanks, sex ratio 1:1) were evaluated after synchronizing mass spawning. To reduce environmental variability, larvae and juveniles of each batch were always managed as a single unit and cultivated under industrial conditions from larval mouth opening to harvest. For genetic evaluation, a small subset of specimens (Table 1) was randomly selected and intraperitoneally tagged as previously reported (Carballo et al., 2018). Fish weight at tagging ranged between 1.7 ± 0.9 and 5.8 ± 4.5 g (Table 1). All batches were evaluated at the beginning of on-growing stage in RAS (~400 d) and at harvest (~800 d). As tagged animals were cultivated under commercial conditions in large tanks (4–11% of whole batch population) not all tagged fish was sampled at 400 d to minimize handling effects and maximize survival. The total number of fish evaluated in each batch at 400 and 800 d is indicated in Table 1. Significant differences in growth traits between batches at both ages were observed (Fig. 2, Table 1). The average weight of each batch at 400 d ranged between 18.2 ± 14.9 g (batch 3; age 397 dph) and 60.9 ± 44.0 g (batch 7; age 414 dph). SL oscillated between 10.4 ± 2.3 and 14.6 ± 3.5 cm and Wi between 3.8 ± 1.0 and 6.0 ± 1.7 cm. At 800 d, weight oscillated between 157.4 ± 103.1 and 425.3 ± 182.8, SL between 20.6 ± 4.5 and 27.4 ± 3.7 and Wi between 8.1 ± 2.1 and 12.0 ± 2.0 for batch 3 and 6, respectively.

In addition to direct biometric measures, all sampled animals were photographed for digital image analysis to estimate standard length (SLI), width (WiI) and total area (A). Data are presented in Table 1. SLI and WiI values were similar to the corresponding traits directly measured on fish. Total area oscillated between 44.2 ± 21.0 and 66.1 ± 32.4 $cm^2$ at 400 d and between 128.7 ± 57.7 and 246.0 ± 71.6 $cm^2$ at 800 d.

As sex and disease symptoms are two important factors that modulate growth in sole, they were recorded for each animal at harvest. ANOVA analysis showed that males were statistically smaller than females for all growth traits ($P < 0.05$). Average weight of males was 8.2% smaller than females at 400 d (34.4 *vs* 30.7 g). Later at 800 d, these differences between males and females were more pronounced ($P < 0.05$) for the six growth traits. Males were as average 22.7% smaller than females (240.6 *vs* 295.3 g).

Overall, males were more abundant than females (55.3% of sampled animals) in the tested population. In batches 1, 2, 3 and 7, male percentages ranged between 58.0 and 69.1% while in batches 4–6 between 43.3 and 49.0%. When sex ratios were analyzed by family a high variation was observed with male proportions oscillating between 16% and 90% (Fig. 3). A comparison of sex ratios in four maternal half-sib families indicated that male percentages were always higher than females

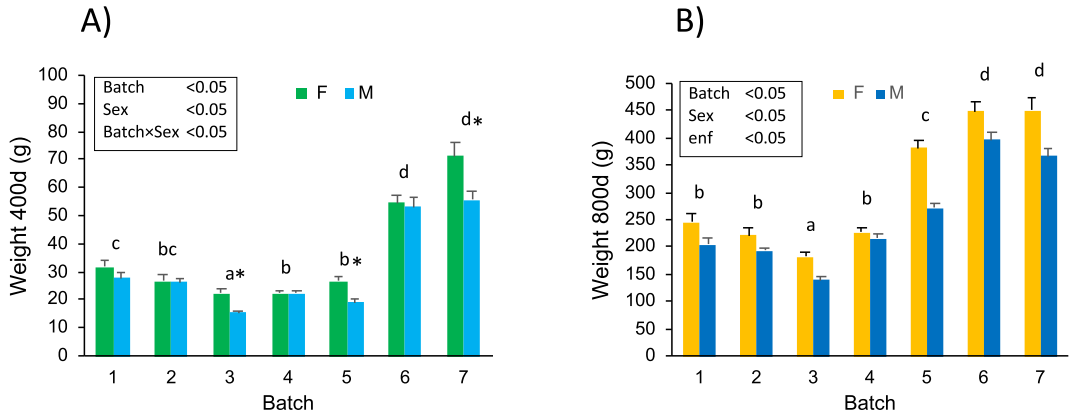


**Fig. 2.** Weight of soles in each evaluation batch by sex at 400 (A) and 800 d (B). ANOVA results are indicated in the square. Letters denote significant differences between batches and asterisks significant differences between sexes (M: male; F: female) for each batch when exist interaction.

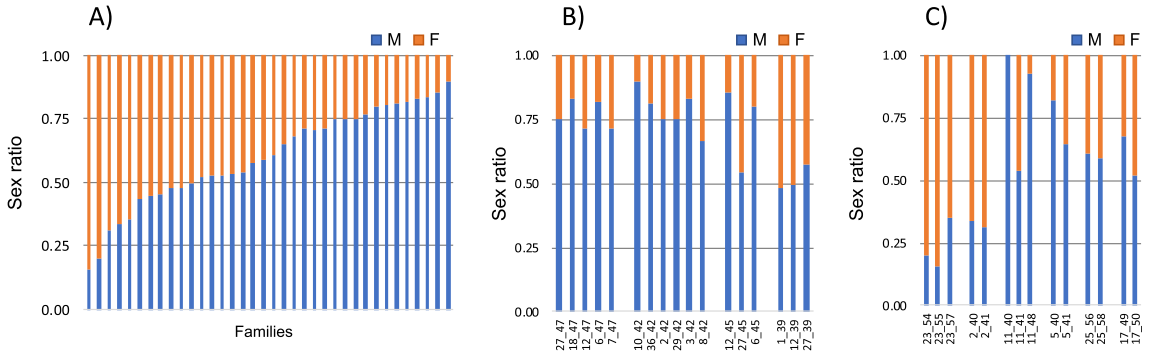


**Fig. 3.** Male (M: blue) and female (F: orange) ratios. A) Sex ratios in families with $n > 10$ individuals. B) Sex ratios in maternal-half sib families (mothers 39, 45, 42 and 47). C) Sex ratios in paternal-half-sib families (fathers 2, 5, 22, 17, 23, 25). The numbers indicate the father code followed by mother code. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(52–78% of males) (Fig. 3B). In contrast, paternal half-sib families produced offspring with a low (24–32%) or a high (60–82%) proportion of males even after crossing with the same females (codes 40 and 41, Fig. 3) in the same batch. These families with an inverted male:female ratio were more represented in batches 4–6.

In some fish, some nodule lesions were observed in liver and/or intestine. These lesions were associated with an amoebic infection of genus Endolimax, very common in RAS systems. The incidence of animals with these lesions ranged between 0.6% in batch 3 and 21% in batch 4. Animal with internal lesions were as average 54.6% lower than healthy fish (179.7 *vs* 277.8 g).

### 3.2. Parentage assignment

Offspring was genotyped using 11 *loci* in a supermultiplex PCR. Parentage assignment was carried out using the exclusion method with a maximal tolerance of two errors. The 98.3% of specimens were successfully assigned to single parent pair. The number of breeders that contributed offspring was 68 (38 males and 30 females). A high bias in offspring contribution was detected since 9 fathers and 8 mothers contributed more than 100 individuals each one that overall represented 62.1 and 71.1% of total population. The total number of families was 70 with an average of 31.0 descendants per family (ranging between 1 and 163). The number of families per batch ranged between 11 (from 8 males and 9 females) and 23 (from 14 males and 10 females) for batch 1 and 2, respectively. One breeder was represented in the seven batches, 4 breeders in 6 batches, 1 breeder in 5 batches, 8 breeders in 4 batches and the remaining 54 breeders in three or less batches.

### 3.3. Genetic estimates

Heritabilities and correlations for growth traits at 400 and 800 days are depicted in Table 2. Heritabilities were higher at 400 than 800 d for the all traits. Wi had the highest heritability (0.643 and 0.500 at 400 d and 800 d, respectively) followed by weight (0.609 and 0.463), area (0.596 and 0.456) and SL (0.593 and 0.425). Heritabilities for SLI and WiI as determined by digital image analysis were similar to those directly measured on the animals.

Genetic correlations between growth traits at 400 d were higher than 0.98. At 800, genetic correlations were a bit lower but still very high (>0.94). Phenotypic correlations were higher than 0.99 at 400 d and higher than 0.96 at 800 d. As a subset of $n = 1843$ was measured at both 400 and 800d, correlations between both ages were also determined. Genetic correlation between 400 and 800 ranged between 0.824 and 0.875 and phenotypic correlations between 0.766 and 0.807 (Table 3).

**Table 2**
Heritabilities(diagonal in bold), phenotypic correlations (below the diagonal) and genetic correlations (above the diagonal), with ± standard error) for growth traits at 400 and 800 d. weight (W), standard length (L) and width (Wi) as *in situ* determined, standard length (SLI), width (WiI) and total area (A) from image analysis are indicated.

| | W400 | SL400 | Wi400 | SLI400 | WiI400 | A400 |
|---|---|---|---|---|---|---|
| W400 | **0.609 ± 0.108** | 0.995 ± 0.002 | 0.995 ± 0.002 | 0.993 ± 0.003 | 0.995 ± 0.002 | 0.997 ± 0.001 |
| SL400 | 0.993 ± 0.001 | **0.593 ± 0.107** | 0.988 ± 0.004 | 1.000 ± 0.000 | 0.988 ± 0.004 | 0.996 ± 0.001 |
| Wi400 | 0.990 ± 0.001 | 0.981 ± 0.002 | **0.643 ± 0.110** | 0.986 ± 0.005 | 0.999 ± 0.000 | 0.996 ± 0.002 |
| SLI400 | 0.990 ± 0.001 | 0.997 ± 0.000 | 0.978 ± 0.002 | **0.568 ± 0.105** | 0.986 ± 0.005 | 0.995 ± 0.002 |
| WiI400 | 0.991 ± 0.001 | 0.982 ± 0.002 | 0.993 ± 0.001 | 0.982 ± 0.002 | **0.631 ± 0.109** | 0.997 ± 0.001 |
| A400 | 0.992 ± 0.001 | 0.989 ± 0.001 | 0.986 ± 0.001 | 0.990 ± 0.001 | 0.989 ± 0.001 | **0.596 ± 0.107** |
| | W800 | SL800 | Wi800 | SLI800 | WiI800 | A800 |
| W800 | **0.463 ± 0.096** | 0.982 ± 0.007 | 0.977 ± 0.008 | 0.982 ± 0.007 | 0.977 ± 0.008 | 0.995 ± 0.002 |
| SL800 | 0.982 ± 0.002 | **0.425 ± 0.091** | 0.944 ± 0.020 | 1.000 ± 0.000 | 0.944 ± 0.02 | 0.983 ± 0.006 |
| Wi800 | 0.982 ± 0.002 | 0.958 ± 0.004 | **0.500 ± 0.099** | 0.944 ± 0.020 | 1.000 ± 0.000 | 0.987 ± 0.005 |
| SLI800 | 0.982 ± 0.002 | 0.996 ± 0.000 | 0.959 ± 0.004 | **0.424 ± 0.091** | 0.943 ± 0.02 | 0.984 ± 0.006 |
| WiI800 | 0.985 ± 0.002 | 0.960 ± 0.004 | 0.994 ± 0.000 | 0.961 ± 0.004 | **0.499 ± 0.099** | 0.985 ± 0.006 |
| A800 | 0.994 ± 0.001 | 0.988 ± 0.001 | 0.985 ± 0.001 | 0.988 ± 0.001 | 0.988 ± 0.001 | **0.456 ± 0.094** |

**Table 3**
Genetic (top) and phenotypic correlations (down) (± standard error) between growth traits at 400 and 800 d. Weight (W); standard length (SL); width (Wi); Total area (A).

| | W800 | SL800 | Wi800 | A800 |
|---|---|---|---|---|
| W400 | 0.850 ± 0.053 | 0.832 ± 0.050 | 0.856 ± 0.049 | 0.861 ± 0.049 |
| SL400 | 0.824 ± 0.060 | 0.835 ± 0.057 | 0.828 ± 0.058 | 0.837 ± 0.056 |
| Wi400 | 0.852 ± 0.052 | 0.844 ± 0.056 | 0.875 ± 0.044 | 0.867 ± 0.047 |
| A400 | 0.848 ± 0.053 | 0.847 ± 0.054 | 0.857 ± 0.049 | 0.861 ± 0.048 |
| W400 | 0.786 ± 0.018 | 0.781 ± 0.019 | 0.786 ± 0.019 | 0.784 ± 0.018 |
| SL400 | 0.785 ± 0.025 | 0.783 ± 0.018 | 0.766 ± 0.019 | 0.774 ± 0.019 |
| Wi400 | 0.796 ± 0.018 | 0.787 ± 0.019 | 0.807 ± 0.017 | 0.797 ± 0.018 |
| A400 | 0.796 ± 0.017 | 0.797 ± 0.017 | 0.801 ± 0.017 | 0.796 ± 0.017 |

### 3.4. OpenArray validation

To identify SNP markers associated with growth traits, an openarray chip containing 60 SNPs markers was designed after mapping RNA-seq information using as reference a female genome draft (Claros et al., 2020; Manchado et al., 2019). The markers were positioned into a high-density SNP genetic linkage map indicating that a high number of markers were located in SseLG02, SseLG05 and SseLG21 that overall covered 17 out of twenty-one SseLGs (Fig. 4). A comparison between species indicated most of the selected SNPs were putatively located in chromosomes Z, W and 14 of *C. semilaevis* (Supp. file 3). Those markers located in the chromosomes ZW of *C. semilaevis* appeared dispersed in several SseLG (SseLG01, SseLG04, SseLG08SseLG05, SseLG07, SseLG09, SseLG16, SseLG17, SseLG19SseLG20) and those SNPs located in chromosome 14 in two SseLGs (SseLG02 and SseLG21).

The SNP array was validated using 164 wild soles (81 males and 83 females). A total of 7 markers (11.7%) failed to amplify and three markers were monomorphic. Moreover, one marker was not in HWE both in males and females and it was excluded from the association study. Finally, a total of 49 markers were selected with a call rate for the individuals higher than 95%, calculated as the proportion of SNPs giving a successful genotypic call for each individual (Suppl. file 2 and 4).

### 3.5. Marker association with growth traits

For the association study, four families with 50 or more individuals were selected by their weight at harvest. Two families were referred to as fast-growing (FG-1 and FG-2) with an average adjusted weight of 442.8 and 382.0 g (Fig. 1). The other two families were named as slow-growing (SG-1 and SG-2) and the weight was 262.8 and 229.5 g. Weights of the individuals did not overlap between FG and SG families. When weight of those specimens also sampled at 400 d was compared, such differences were only observable for FG-1 (Fig. 1).

A total of 47 polymorphic assays were detected with overall had genotyping call rates higher than 94%. PCA analysis identified two full-sib and two half-sib families (Fig. 5). Association analyses were performed using four models: GLM naïve, GLM with Q matrices as covariate, MLM with K matrices as covariate and MLM (Q + K). The quantile-quantile plots (Suppl. file 3) indicated that MLM (K) and MLM (Q + K) models were significantly better than the GLM naïve and GLM (Q) models. The MLM (Q + K) model was selected for the analysis as implemented in Tassel and GWASPoly programs.

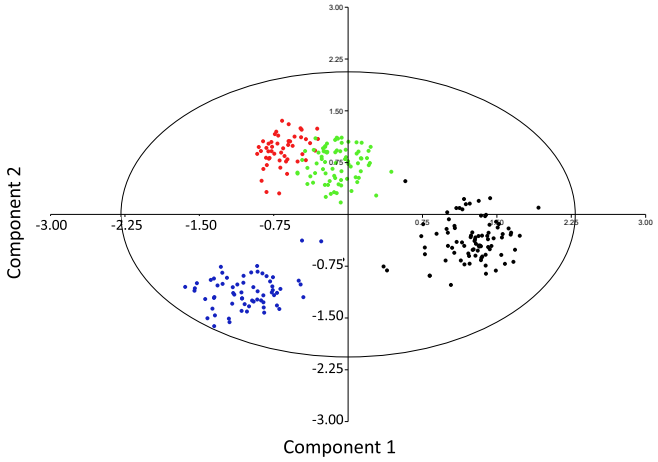


**Fig. 5.** Family structure using a principal component analysis (PC1 and PC2) based on 47 SNPs from 279 animals. The first 2 components explain 37.4 and 30.7% of variance, respectively. Individuals of FG families are indicated in black and blue and those of SG families in red and green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Two significant SNPs (Sosen_s1980 and Sosen_s0233) associated with growth traits after Bonferroni's correction ($P = 0.0011$; $\alpha = 0.05$) were detected at 400 and 800 d (Fig. 6, Table 4). Sosen_s1980 was located in the general transcription factor 3C polypeptide 4 (*gtf3c4*) and it was significant for the four traits at both ages but association scores were stronger at 800 than 400 d. At 800 d, this marker explained from 11.89 to 14.77% of the variation. At 400 d, these values ranged between 5.16 and 7.98%. The scatter violin plots (Fig. 6) show the effects of SNP alleles on weight at 400 and 800 d. Animals with TT and TG were weightier (54 and 130% at 400 d; 23 and 61% at 800 d) than GG (Fig. 6). The marker Sosen_s0233 was located within the mitochondrial fission process protein 1 (*mtfp1*) and it was significant for all traits except width at 800 d and only explained between 3.89 and 5.85% of variation. Animals with CC were weightier than TT and TC (104 and 131% at 400 d; 25 and 61% at 800 d, respectively) (Fig. 7).

No significant association with sex (male or female) using wild fish

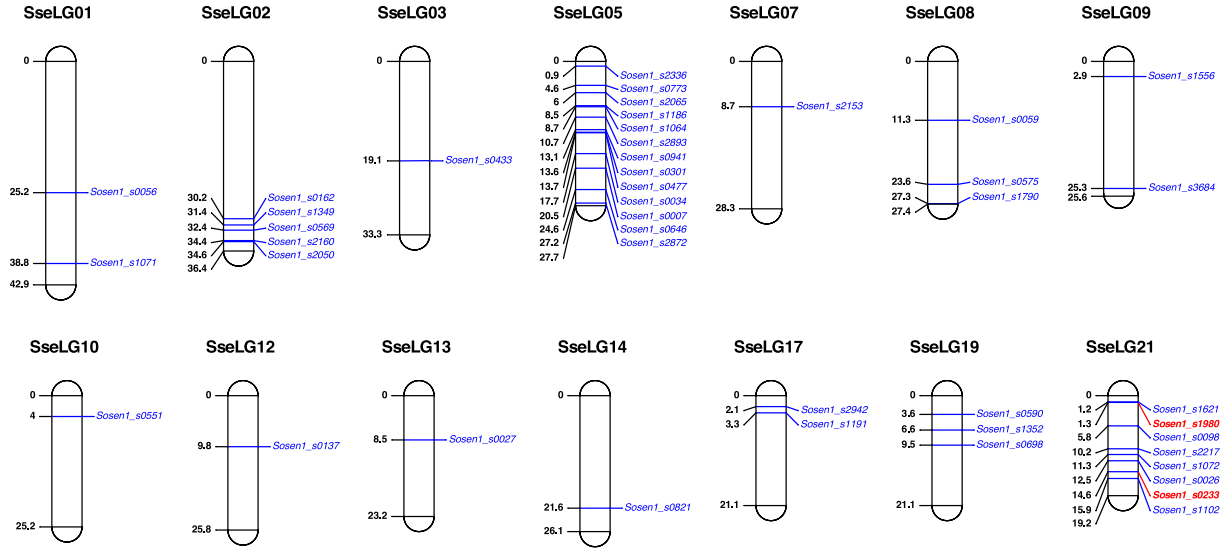


**Fig. 4.** Distribution of SNP markers in the genome of *S. senegalensis*. Markers (in blue) are sorted by linkage groups (SseLG). Numbers on the left denote distance in cM. Significant markers in the association analysis are indicated in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

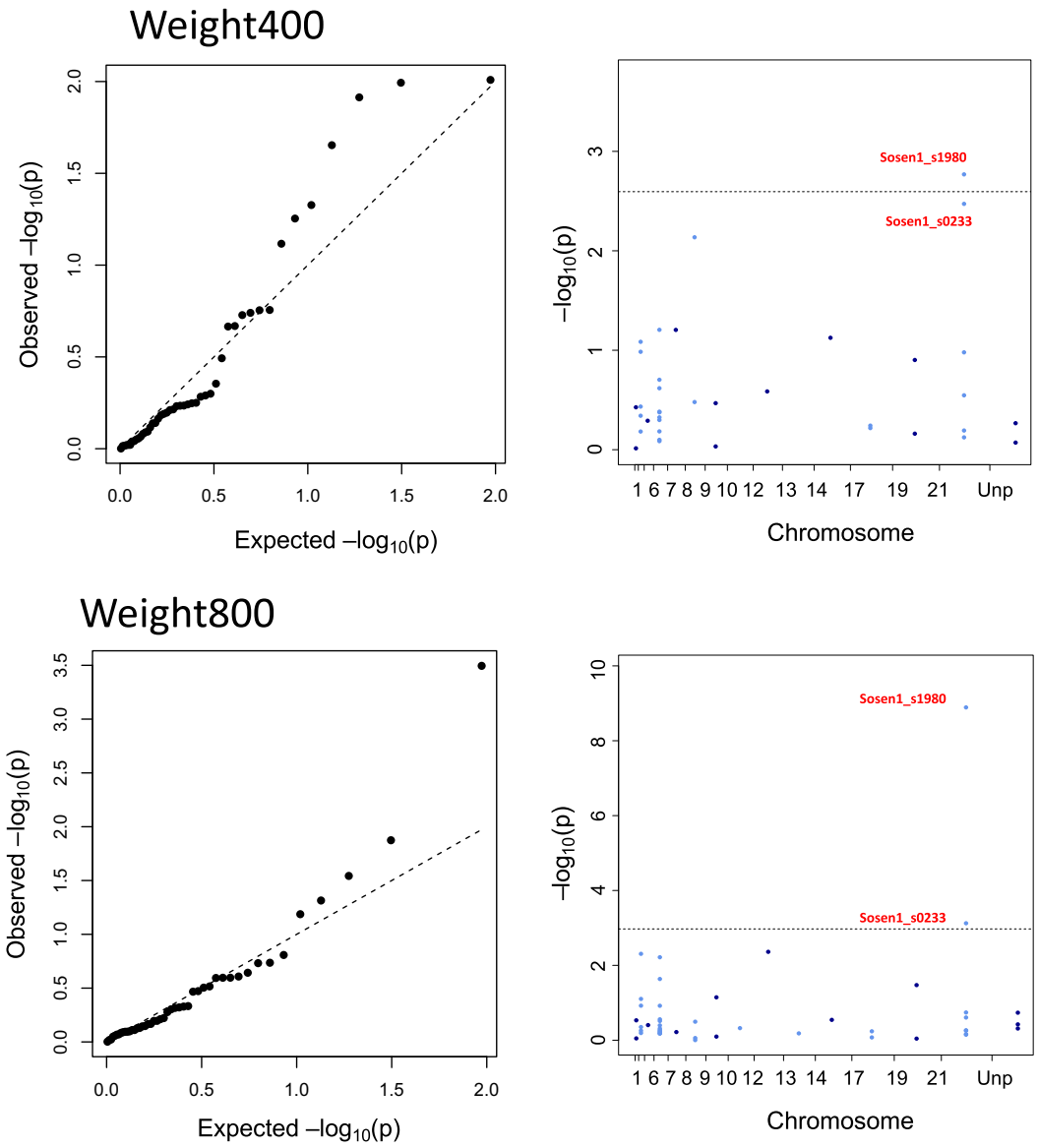


**Fig. 6.** Quantile-quantile (QQ) and Manhattan plots of association study for weight at 400 d and 800 d using MLM (Q + K) model. The two significant SNPs at 800 d are indicated.

($n = 164$) or the four families was observed (data not shown).

## 4. Discussion

Genetic research in Senegalese sole has been hampered until now by the lack of full-control on reproduction success. Nevertheless, the recent advances in thermoperiod manipulation for spawning synchronization have made feasible the design of genetic programs (Martin et al., 2019). In this study, we have genetically evaluated a commercial broodstock of Senegalese sole using a mass spawning methodology followed by molecular pedigree reconstruction as previously reported for other marine species (Garcia-Celdran et al., 2015b; Lee-Montero et al., 2015; Navarro et al., 2009). The evaluation of growth traits was carried out before entering in RAS (~400 d) and at harvest (~800 d). This period appears as critical in sole production since RAS is a technology very different of natural ponds in which soles inhabit and a genotype×environment interaction was previously demonstrated in the close species *S. solea* (Mas-Muñoz et al., 2013). Results from this work generated at industrial scale are high valuable as reference for breeding programs in intensive growing of *S. senegalensis* juveniles in RAS.

In Senegalese sole, females grow faster than males and significant differences in weight are observable from young juveniles (females 13.6% heavier than males) (Carballo et al., 2018) to harvest size (19–32% heavier). These differences are even more evident at high stocking densities (Sanchez et al., 2010). In this study, females were significantly 8.2% heavier than males at the beginning of RAS and 22.7% at harvest. In spite of the interest of cultivating female-enriched populations, sole populations cultured under standard production conditions at 20 °C are normally skewed toward males that usually represent around 60–67% of whole populations (Blanco-Vives et al., 2011; Carballo et al., 2018; Sanchez et al., 2010; Viñas and Asensio, 2012). This skewed abundance of males in sole has been associated with epigenetic effects mediated by environmental temperature that induces masculinization (Blanco-Vives et al., 2011; Viñas and Asensio, 2012). In this study, in which a high number of families is represented, the average percentage of males in the population was 55.3% (ranging from 43.4 to 69.1% between batches). Larval rearing was carried out under commercial conditions using constant temperature (~20 °C) that could explain this slightly higher proportion of males especially in some batches (1, 2, 3 and 7). However, it should be noted a high variation in

**Table 4**
Significance levels for markers Sosen_s1980 and Sosen_s0233 from MLM(Q + K) analyses for growth traits at 400 and 800 d. Additive (add) and dominant (dom) *P*-values as determined by TASSEL (left) and GWASpoly (right) programs are shown. $R^2$ indicates proportion of phenotypic variation as determined by TASSEL. ns, not-significant. *P*-values are indicated in scientific notation.

| Marker | Trait | add *P*-value | dom *P*-value | $R^2$ (%) |
|---|---|---|---|---|
| Sosen_s1980 | W400 | ns/7.8E-06 | ns/ns | 8.0 |
| | SL400 | ns/8.8E-05 | ns/ns | 5.7 |
| | Wi400 | 1.4E-04/ns | ns/3.9E-05 | 5.4 |
| | A400 | ns/1.6E-04 | ns/8.1E-04 | 5.2 |
| | W800 | 2.6E-08/3.2E-04 | 7.3E-06/6.9E-10 | 12.5 |
| | SL800 | 4.2E-08/2.9E-04 | 1.5E-05/7.6E-10 | 11.9 |
| | Wi800 | 3.2E-08/6.5E-04 | 1.5E-07/1.7E-11 | 14.8 |
| | A800 | 3.7E-08/3.1E-04 | 6.2E-06/2.8E-10 | 12.6 |
| Sosen_s0233 | W400 | ns/ns | 4.7E-04/ns | 4.3 |
| | SL400 | ns/ns | 1.6E-04/1.1E-03 | 5.1 |
| | Wi400 | ns/ns | 6.1E-05/4.3E-04 | 5.9 |
| | A400 | ns/ns | 1.7E-04/1.3E-03 | 5.0 |
| | W800 | ns/ns | 3.9E-04/ns | 4.1 |
| | SL800 | ns/ns | 5.2E-04/ns | 4.0 |
| | Wi800 | ns/ns | ns/ns | ns |
| | A800 | ns/ns | 6.7E-04/ns | 3.9 |

sex ratios associated with genetic families (% males ranging from 16 up to 90%; Fig. 3). It was striking that when maternal half-sibs were compared, all of them contributed a higher proportion of males in similar percentages. In contrast, when paternal half-sibs were compared (mainly from batches 4–6), the families could be enriched in females or males. These data indicate that epigenetic effects act differently on fathers than on mothers and that males could be the heterogametic sex able to skew population toward neomales or neofemales. These data agree with XX-XY sex determining system proposed in *S. senegalensis* using gynogenetic families although modulated by other genetic or environmental factors, such autosomal genes or temperature (Molina-Luzon et al., 2015).

Multiplex assays assigned 98.1% offspring (2171 specimens) to a single parent pair supporting the high assignation rates previously reported (Guerrero-Cózar et al., 2020). Heritabilities for all growth traits were higher at 400 d than 800 d and values ranged between 0.568 and 0.643 at 400 d and between 0.424 and 0.500 at 800 d (Table 2). These heritabilities were higher than those observed in *S. solea* cultivated in RAS at harvest (0.23–0.25) (Blonk et al., 2010a; Blonk et al., 2010b) and in gilthead seabream (0.34–0.40 at 509-689d) (Lee-Montero et al., 2015; Navarro et al., 2009). The fast growth rates of *S. senegalensis* and the adaptation to high stocking densities and handling under different production systems (flow-throw or RAS) explain these high genetic estimates. The slight reduction in heritabilities observed from 400 to 800 d could be due to sexual maturation effects that mask the effects associated with growth potential.

Genetic correlations between all growth traits were very high and positive. High correlations between weight, length and width are routinely reported (Blonk et al., 2010b; Lee-Montero et al., 2015; Navarro et al., 2009; Vandeputte et al., 2008). It should be noted that in this study the body width was the trait with the highest heritability at both ages. This is important since this trait has very high genetic correlations with other growth traits, it is important to control shape quality, and it can be measured during fieldwork conditions or from images, hence, this could be a good candidate to be used for genetic selection as alternative to weight in Senegalese sole breeding programs. In the case of gilthead seabream, length was also suggested as more adequate than weight due to higher heritabilities and lower coefficients of variation (Lee-Montero et al., 2015; Navarro et al., 2009). It should be note that genetic correlations between ages at 400 and 800 d were still high. Normally, genetic

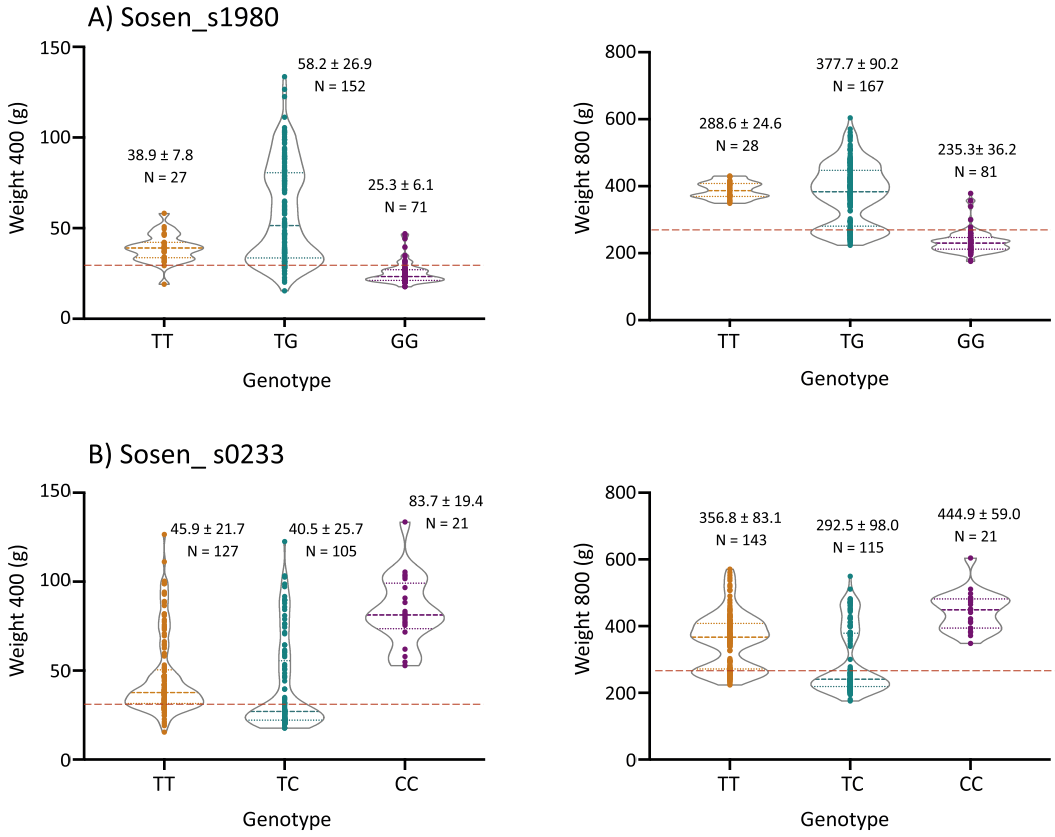


**Fig. 7.** Validation of Sosen_s1980 and Sosen_s0233 markers for weight at 400 and 800. The adjusted weight average ± SD for each genotype and number of individuals analyzed are indicated. The red dashline represents the average population mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

correlations between growth traits at different ages are usually low (0.3–0.5) when long time periods are considered (hatchery ~120–150 d *vs* harvest ~500–800 d) (Navarro et al., 2009; Vandeputte et al., 2008) and they increase to ~0.8 or more when the time gap is smaller as occurs in this study (Lee-Montero et al., 2015; Navarro et al., 2009; Vandeputte et al., 2008). Interestingly, genetic estimates and standard errors for SL, width and total area measured directly on fish and after image analysis were almost identical for equivalent traits confirming that the image analysis was a feasible approach to measure these growth traits reducing the times during sampling and fish stress (Blonk et al., 2010b; Navarro et al., 2016). These data also support the use of this methodology for future studies related to flatfish morphology.

The study of genetic estimates for growth traits was completed with the identification of genetic variants significantly associated with these traits. We exploited the information from several genomic resources available in Senegalese sole for the design of a low-density chip and the study in 4 families with highly different breeding values at harvest. Although the arrays were originally intended to identify sex-related markers and most of the markers were located in sexual chromosomes ZW and chromosome 14 of *C. semilaevis*, we failed to find any association with sex in Senegalese sole. Moreover, the distribution of markers in a genetic linkage map of *S. senegalensis* indicated that sequences ZW were spread through the genome (Fig. 4; Suppl. file 3) as previously indicated (Guerrero-Cózar et al., 2020). In contrast, two markers were significantly associated with adjusted growth traits at both 400 and 800 d. Growth is a polygenic trait controlled by many genes spread through the chromosomes involved in cell growth, cell proliferation, cell cycle, lipid metabolism, proteolytic activities, chromatin modification, and developmental processes (Ali et al., 2020). The Sosen1_s1980 is located in the general transcription factor 3C polypeptide 4 (*gtf3c4*, also known as TFIIIC90). This gene is responsible for the recruitment of RNA polymerase III and initiating of tRNA transcription in eukaryotes and it has also been reported to possess histone acetyltransferase activity (HAT) *in vitro* (Hsieh et al., 1999; Kundu et al., 1999; Trisciuoglio et al., 2018). This HAT activity is important for chromatin relaxation and activation of gene transcription. In rat the *gtf3c4* is located within a region where localize a QTL rat associated with body weight (Casiro et al., 2017). Several SNPs associated with body weight gain in rainbow trout were located in genes related to chromatin modification, and developmental processes (Ali et al., 2020). Moreover, HAT genes were potentially involved in cotton growth and development, fiber-related traits, and plant response to the environment (Imran et al., 2019). The other significant marker is the Sosen1_s0233 encodes for mitochondrial fission process protein 1 (MTFP1, also known as MTP18) that plays an essential role for maintaining mitochondrial integrity and it is highly expressed in organs enriched with mitochondria such as heart and skeletal muscles and whose regulation of mitochondrial physiology is essential for maintenance of muscle mass and function (Aung et al., 2017; Tondera et al., 2004).

In conclusion, this study provides for first time genetic estimates for growth traits in Senegalese sole. Animals were evaluated before on-growing in RAS and at harvest observing high heritabilities for all traits and high genetic correlations between growth traits and between ages. The comparison of *in situ* measured variables with those obtained from image-based analysis showed similar results supporting the latter analysis as a feasible methodology to be used in sole. The analysis of sex ratios identified an important skewness across genetic families indicating that environmental factors act differentially on mothers, that tend toward masculinization, and on fathers, that can skew populations toward males or females. Finally, the use of low-density chip on two fast and two slow-growing families identified two SNPs significantly associated with growth traits at 400 and 800 d. All these data are of high relevance to design genetic breeding programs that boost the Senegalese sole aquaculture, one of the most promising in Southern Europe.

*

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This study was funded by projects RTA2017-00054-C03-01 and RTA2017-00054-C03-03 from MCIU/AEI/FEDER, UE. Moreover, the study has received funding from EU H2020 research and innovation program under grant agreement 817992 ERANET-BLUEBIO COFUND project PCI2020-111994. IGC is funded by a predoctoral fellowship from INIA.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aquaculture.2021.736665.

## References

Agulleiro, M.J., Anguis, V., Cañavate, J.P., Martínez-Rodríguez, G., Mylonas, C.C., Cerdà, J., 2006. Induction of spawning of captive-reared Senegal sole (*Solea senegalensis*) using different administration methods for gonadotropin releasing hormone agonist. Aquaculture. 257, 511–524. https://doi.org/10.1016/j.aquaculture.2006.02.001.

Ali, A., Al-Tobasei, R., Lourenco, D., Leeds, T., Kenney, B., Salem, M., 2020. Genome-wide identification of loci associated with growth in rainbow trout. BMC Genomics 21, 209. https://doi.org/10.1186/s12864-020-6617-x.

Aslam, M.L., Carraro, R., Bestin, A., Cariou, S., Sonesson, A.K., Bruant, J.S., Haffray, P., Bargelloni, L., Meuwissen, T.H.E., 2018. Genetics of resistance to photobacteriosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. BMC Genet. 19, 43. https://doi.org/10.1186/s12863-018-0631-x.

Aulchenko, Y.S., Ripke, S., Isaacs, A., van Duijn, C.M., 2007. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 23, 1294–1296. https://doi.org/10.1093/bioinformatics/btm108.

Aung, L.H.H., Li, R., Prabhakar, B.S., Li, P., 2017. Knockdown of Mtfp1 can minimize doxorubicin cardiotoxicity by inhibiting Dnm1l-mediated mitochondrial fission. J. Cell. Mol. Med. 21, 3394–3404. https://doi.org/10.1111/jcmm.13250.

Benzekri, H., Armesto, P., Cousin, X., Rovira, M., Crespo, D., Merlo, M.A., Mazurais, D., Bautista, R., Guerrero-Fernandez, D., Fernandez-Pozo, N., Ponce, M., Infante, C., Zambonino, J.L., Nidelet, S., Gut, M., Rebordinos, L., Planas, J.V., Begout, M.L., Claros, M.G., Manchado, M., 2014. De novo assembly, characterization and functional annotation of Senegalese sole (*Solea senegalensis*) and common sole (*Solea solea*) transcriptomes: integration in a database and design of a microarray. BMC Genomics 15 (952). https://doi.org/10.1186/1471-2164-15-952.

Bertolini, F., Ribani, A., Capoccioni, F., Buttazzoni, L., Utzeri, V.J., Bovo, S., Schiavo, G., Caggiano, M., Fontanesi, L., Rothschild, M.F., 2020. Identification of a major locus determining a pigmentation defect in cultivated gilthead seabream (Sparus aurata). Anim. Genet. 51, 319–323. https://doi.org/10.1111/age.12890.

Blanco-Vives, B., Vera, L.M., Ramos, J., Bayarri, M.J., Mananos, E., Sanchez-Vazquez, F. J., 2011. Exposure of larvae to daily thermocycles affects gonad development, sex ratio, and sexual steroids in *Solea senegalensis*, kaup. J Exp Zool A Ecol Genet Physiol. 315, 162–169. https://doi.org/10.1002/jez.664.

Blonk, R.J., Komen, H., Kamstra, A., van Arendonk, J.A., 2010a. Estimating breeding values with molecular relatedness and reconstructed pedigrees in natural mating populations of common sole, *Solea solea*. Genetics 184, 213–219. https://doi.org/10.1534/genetics.109.110536.

Blonk, R.J., Komen, H., Tenghe, A., Kamstra, A., van Arendonk, J.A.M., 2010b. Heritability of shape in common sole, *Solea solea*, estimated from image analysis data. Aquaculture 307, 6–11. https://doi.org/10.1016/j.aquaculture.2010.06.025.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 23, 2633–2635. https://doi.org/10.1093/bioinformatics/btm308.

Brouard, J.S., Schenkel, F., Marete, A., Bissonnette, N., 2019. The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments. J Anim Sci Biotechnol. 10, 44. https://doi.org/10.1186/s40104-019-0359-0.

Brown, R.C., Woolliams, J.A., McAndrew, B.J., 2005. Factors influencing effective population size in commercial populations of gilthead seabream, *Sparus aurata*. Aquaculture 247, 219–225. https://doi.org/10.1016/j.aquaculture.2005.02.002.

Carballo, C., Berbel, C., Guerrero-Cozar, I., Jimenez-Fernandez, E., Cousin, X., Bégout, M. L., Manchado, M., 2018. Evaluation of different tags on survival, growth and stress response in the flatfish Senegalese sole. Aquaculture 494, 10–18. https://doi.org/10.1016/j.aquaculture.2018.05.009.

Carballo, C., Shin, H.S., Berbel, C., Zamorano, M.J., Borrego, J.J., Armero, E., Afonso, J. M., Manchado, M., 2020. Heritability estimates and genetic correlation for gowth

traits and LCDV susceptibility in gilthead sea bream (*Sparus aurata*). Fishes. 5, 2. https://doi.org/10.3390/fishes5010002.
Casiro, S., Velez-Irizarry, D., Ernst, C.W., Raney, N.E., Bates, R.O., Charles, M.G., Steibel, J.P., 2017. Genome-wide association study in an F2 Duroc x Pietrain resource population for economically important meat quality and carcass traits. J. Anim. Sci. 95, 545–558. https://doi.org/10.2527/jas.2016.1003.
Chagne, D., Vanderzande, S., Kirk, C., Profitt, N., Weskett, R., Gardiner, S.E., Peace, C.P., Volz, R.K., Bassil, N.V., 2019. Validation of SNP markers for fruit quality and disease resistance loci in apple (Malus x domestica Borkh.) using the OpenArray(R) platform. Hortic Res. 6 (30) https://doi.org/10.1038/s41438-018-0114-2.
Chauvigne, F., Olle, J., Gonzalez, W., Duncan, N., Gimenez, I., Cerda, J., 2017. Toward developing recombinant gonadotropin-based hormone therapies for increasing fertility in the flatfish Senegalese sole. PLoS One 12, e0174387. https://doi.org/10.1371/journal.pone.0174387.
Claros, M.G., Seoane, P., Manchado, M., 2020. Sequences and annotations of a provisional genome draft of a Senegalese sole female. Figshare. https://doi.org/10.6084/m9.figshare.12472100.v1.
Cordoba-Caballero, J., Seoane, P., Jabato, F.M., Perkins, J.R., Manchado, M., Claros, M. G., 2020. An improved de novo assembling and polishing of *Solea senegalensis* transcriptome shed light on retinoic acid signalling in larvae. Sci. Rep. 10, 20654. https://doi.org/10.1038/s41598-020-77201-z.
Espinosa, E., Arroyo, M., Larrosa, R., Manchado, M., Claros, M.G., Bautista, R., 2020. Micro-variations from RNA-seq experiments for non-model organisms. In: Rojas, I., Valenzuela, O., Rojas, F., Herrera, L., Ortuño, F. (Eds.), Bioinformatics and Biomedical Engineering. IWBBIO 2020. Springer, Cham.
Fatsini, E., Bautista, R., Manchado, M., Duncan, N.J., 2016. Transcriptomic profiles of the upper olfactory rosette in cultured and wild Senegalese sole (Solea senegalensis) males. Comp Biochem Physiol Part D Genomics Proteomics. 20, 125–135. https://doi.org/10.1016/j.cbd.2016.09.001.
Fatsini, E., Carazo, I., Chauvigne, F., Manchado, M., Cerda, J., Hubbard, P.C., Duncan, N. J., 2017. Olfactory sensitivity of the marine flatfish Solea senegalensis to conspecific body fluids. J. Exp. Biol. 220, 2057–2065. https://doi.org/10.1242/jeb.150318.
Fatsini, E., Gonzalez, W., Ibarra-Zatarain, Z., Napuchi, J., Duncan, N., 2020. The presence of wild Senegalese sole breeders improves courtship and reproductive success in cultured conspecifics. Aquaculture. 519, 734922. https://doi.org/10.1016/j.aquaculture.2020.734922.
Garcia-Celdran, M., Ramis, G., Manchado, M., Estévez, A., Afonso, J.M., Armero, E., 2015a. Estimates of heritabilities and genetic correlations of carcass quality traits in a reared gilthead sea bream (*Sparus aurata* L.) population sourced from three broodstocks along the Spanish coasts. Aquaculture. 446, 175–180. https://doi.org/10.1016/j.aquaculture.2015.04.028.
Garcia-Celdran, M., Ramis, G., Manchado, M., Estévez, A., Navarro, A., Armero, E., 2015b. Estimates of heritabilities and genetic correlations of raw flesh quality traits in a reared gilthead sea bream (*Sparus aurata* L.) population sourced from broodstocks along the Spanish coasts. Aquaculture. 446, 181–186. https://doi.org/10.1016/j.aquaculture.2015.04.030.
Garcia-Celdran, M., Cutáková, Z., Ramis, G., Estévez, A., Manchado, M., Navarro, A., Maria-Dolores, E., Peñalver, J., Sanchez, J.A., Armero, E., 2016. Estimates of heritabilities and genetic correlations of skeletal deformities and uninflated swimbladder in a reared gilthead sea bream (*Sparus aurata* L.) juvenile population sourced from three broodstocks along the Spanish coasts. Aquaculture. 464, 601–608. https://doi.org/10.1016/j.aquaculture.2016.08.004.
Guerrero-Cózar, I., Perez-Garcia, C., Benzekri, H., Sánchez, J., Seoane, P., Cruz, F., Gut, M., Zamorano, M.J., Claros, M.G., Manchado, M., 2020. Development of whole-genome multiplex assays and construction of an integrated genetic map using SSR markers in Senegalese sole. Sci. Reports. 10, 21905.
Gutierrez-Camino, A., Martin-Guerrero, I., Dolzan, V., Jazbec, J., Carbone-Baneres, A., Garcia de Andoin, N., Sastre, A., Astigarraga, I., Navajas, A., Garcia-Orad, A., 2018. Involvement of SNPs in miR-3117 and miR-3689d2 in childhood acute lymphoblastic leukemia risk. Oncotarget 9, 22907–22914. https://doi.org/10.18632/oncotarget.25144.
Hsieh, Y.J., Kundu, T.K., Wang, Z., Kovelman, R., Roeder, R.G., 1999. The TFIIIC90 subunit of TFIIIC interacts with multiple components of the RNA polymerase III machinery and contains a histone-specific acetyltransferase activity. Mol. Cell. Biol. 19, 7697–7704. https://doi.org/10.1128/mcb.19.11.7697.
Imran, M., Shafiq, S., Farooq, M.A., Naeem, M.K., Widemann, E., Bakhsh, A., Jensen, K. B., Wang, R.R., 2019. Comparative genome-wide analysis and expression profiling of histone acetyltransferase (HAT) gene family in response to hormonal applications, metal and abiotic stresses in cotton. Int. J. Mol. Sci. 20 https://doi.org/10.3390/ijms20215311.
Janssen, K., Chavanne, H., Berentsen, P., Komen, H., 2017. Impact of selective breeding on European aquaculture. Aquaculture. 472, 8–16. https://doi.org/10.1016/j.aquaculture.2016.03.012.
Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., Wilson, R.K., 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22, 568–576. https://doi.org/10.1101/gr.129684.111.
Kriaridou, C., Tsairidou, S., Houston, R.D., Robledo, D., 2020. Genomic prediction using low density marker panels in aquaculture: performance across species, traits, and genotyping platforms. Front. Genet. 11, 124. https://doi.org/10.3389/fgene.2020.00124.
Kundu, T.K., Wang, Z., Roeder, R.G., 1999. Human TFIIIC relieves chromatin-mediated repression of RNA polymerase III transcription and contains an intrinsic histone acetyltransferase activity. Mol. Cell. Biol. 19, 1605–1615. https://doi.org/10.1128/mcb.19.2.1605.
Kyriakis, D., Kanterakis, A., Manousaki, T., Tsakogiannis, A., Tsagris, M., Tsamardinos, I., Papaharisis, L., Chatziplis, D., Potamias, G., Tsigenopoulos, C.S., 2019. Scanning of genetic variants and genetic mapping of phenotypic traits in Gilthead Sea bream through ddRAD sequencing. Front. Genet. 10, 675. https://doi.org/10.3389/fgene.2019.00675.
Lee-Montero, I., Navarro, A., Borrell, Y., Garcia-Celdran, M., Martin, N., Negrin-Baez, D., Blanco, G., Armero, E., Berbel, C., Zamorano, M.J., Sanchez, J.J., Estevez, A., Ramis, G., Manchado, M., Afonso, J.M., 2013. Development of the first standardised panel of two new microsatellite multiplex PCRs for gilthead seabream (*Sparus aurata* L.). Anim. Genet. 44, 533–546. https://doi.org/10.1111/age.12037.
Lee-Montero, I., Navarro, A., Negrin-Baez, D., Zamorano, M.J., Borrell Pichs, Y.J., Berbel, C., Sanchez, J.A., Garcia-Celdran, M., Manchado, M., Estevez, A., Armero, E., Afonso, J.M., 2015. Genetic parameters and genotype-environment interactions for skeleton deformities and growth traits at different ages on gilthead seabream (*Sparus aurata* L.) in four Spanish regions. Anim. Genet. 46, 164–174. https://doi.org/10.1111/age.12258.
Manchado, M., Planas, J.V., Cousin, X., Rebordinos, L., Claros, M.G., 2016. Current status in other finfish species: Description of current genomic resources for the gilthead seabream (*Sparus aurata*) and soles (*Solea senegalensis* and *Solea solea*). In: Mackenzie, S., Jentoft, S. (Eds.), Genomics in Aquaculture. Elsevier, Amsterdam, pp. 195–221.
Manchado, M., Planas, J.V., Cousin, X., Rebordinos, L., Claros, M.G., 2019. Genetic and genomic characterization of soles. In: Muñoz-Cueto, J., Mañanós-Sánchez, E., Sánchez-Vázquez, F. (Eds.), The Biology of Sole. CDC Press, Boca Raton, pp. 375–394.
Martin, I., Rasines, I., Gomez, M., Rodriguez, C., Martinez, P., Chereguini, O., 2014. Evolution of egg production and parental contribution in Senegalese sole, *Solea senegalensis*, during four consecutive spawning seasons. Aquaculture. 424-425, 45–52. https://doi.org/10.1016/j.aquaculture.2013.12.042.
Martin, I., Carazo, I., Rasines, I., Rodriguez, C., Fernandez, R., Martinez, P., Norambuena, P., Chereguini, O., Duncan, N., 2019. Reproductive performance of captive Senegalese sole, *Solea senegalensis*, according to the origin (wild or cultured) and gender. Span. J. Agric. Res. 17, 2171–9292. https://doi.org/10.5424/sjar/2019174-14953.
Mas-Muñoz, J., Blonk, R.J., Schrama, J.W., van Arendonk, J., Komen, H., 2013. Genotype by environment interaction for growth of sole (*Solea solea*) reared in an intensive aquaculture system and in a semi-natural environment. Aquaculture 410–411, 230–235. https://doi.org/10.1016/j.aquaculture.2013.06.012.
Meyer, K., 2007. WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). J Zhejiang Univ Sci B 8, 815–821. https://doi.org/10.1631/jzus.2007.B0815.
Molina-Luzon, M.J., Lopez, J.R., Robles, F., Navajas-Perez, R., Ruiz-Rejon, C., De la Herran, R., Navas, J.I., 2015. Chromosomal manipulation in Senegalese sole (*Solea senegalensis* Kaup, 1858): induction of triploidy and gynogenesis. J. Appl. Genet. 56, 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., 2016. New developments and biological insights into the farming of *Solea senegalensis* reinforcing its aquaculture potential. Rev Aquacult. 6, 1–37. https://doi.org/10.1111/raq.12091.
Navarro, A., Zamorano, M.J., Hildebrandt, S., Ginés, R., Aguilera, C., Afonso, J.M., 2009. Estimates of heritabilities and genetic correlations for growth and carcass traits in gilthead seabream (*Sparus auratus* L.), under industrial conditions. Aquaculture. 289, 225–230. https://doi.org/10.1016/j.aquaculture.2008.12.024.
Navarro, A., Lee-Montero, I., Santana, D., Henríquez, P., Ferrer, M.A., Morales, A., Soula, M., Badilla, R., Negrin-Baez, D., Zamorano, M.J., Afonso, J.M., 2016. IMAFISH_ML: a fully-automated image analysis software for assessing fish morphometric traits on gilthead seabream (*Sparus aurata* L.), meagre (*Argyrosomus regius*) and red porgy (*Pagrus pagrus*). Comput Electron Agr. 121, 66–73. https://doi.org/10.1016/j.compag.2015.11.015.
Ouellette, L.A., Reid, R.W., Blanchard, S.G., Brouwer, C.R., 2018. LinkageMapView-rendering high-resolution linkage and QTL maps. Bioinformatics 34, 306–307. https://doi.org/10.1093/bioinformatics/btx576.
Palaiokostas, C., Ferraresso, S., Franch, R., Houston, R.D., Bargelloni, L., 2016. Genomic prediction of resistance to Pasteurellosis in Gilthead Sea bream (Sparus aurata) using 2b-RAD sequencing. G3 (Bethesda) 6, 3693–3700. https://doi.org/10.1534/g3.116.035220.
Purcell, C.M., Seetharam, A.S., Snodgrass, O., Ortega-Garcia, S., Hyde, J.R., Severin, A.J., 2018. Insights into teleost sex determination from the *Seriola dorsalis* genome assembly. BMC Genomics 19 (31). https://doi.org/10.1186/s12864-017-4403-1.
Rosyara, U.R., De Jong, W.S., Douches, D.S., Endelman, J.B., 2016. Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome. 9 https://doi.org/10.3835/plantgenome2015.08.0073.
Sanchez, P., Ambrosio, P.P., Flos, R., 2010. Stocking density and sex influence individual growth of Senegalese sole (*Solea senegalensis*). Aquaculture. 300, 93–101. https://doi.org/10.1016/j.aquaculture.2009.12.013.
Sole, X., Guino, E., Valls, J., Iniesta, R., Moreno, V., 2006. SNPStats: a web tool for the analysis of association studies. Bioinformatics. 22, 1928–1929. https://doi.org/10.1093/bioinformatics/btl268.
Tondera, D., Santel, A., Schwarzer, R., Dames, S., Giese, K., Klippel, A., Kaufmann, J., 2004. Knockdown of MTP18, a novel phosphatidylinositol 3-kinase-dependent protein, affects mitochondrial morphology and induces apoptosis. J. Biol. Chem. 279, 31544–31555. https://doi.org/10.1074/jbc.M404704200.
Trisciuoglio, D., Di Martile, M., Del Bufalo, D., 2018. Emerging role of histone Acetyltransferase in stem cells and cancer. Stem Cells Int. 2018, 8908751. https://doi.org/10.1155/2018/8908751.

Vandeputte, M., Mauger, S., Dupont-Nivet, M., 2006. An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. Mol. Ecol. Notes 6, 265–267. https://doi.org/10.1111/j.1471-8286.2005.01167.x.

Vandeputte, M., Kocour, M., Mauger, S., Rodina, M., Launay, A., Gela, D., Dupont-Nivet, M., Hulak, M., Linhart, O., 2008. Genetic variation for growth at one and two summers of age in the common carp (*Cyprinus carpio* L.): heritability estimates and response to selection. Aquaculture. 277, 7–13. https://doi.org/10.1016/j.aquaculture.2008.02.009.

Verbeek, E.C., Bakker, I.M., Bevova, M.R., Bochdanovits, Z., Rizzu, P., Sondervan, D., Willemsen, G., de Geus, E.J., Smit, J.H., Penninx, B.W., Boomsma, D.I., Hoogendijk, W.J., Heutink, P., 2012. A fine-mapping study of 7 top scoring genes from a GWAS for major depressive disorder. PLoS One 7, e37384. https://doi.org/10.1371/journal.pone.0037384.

Viñas, J., Asensio, E., Piferrer, F., 2012. Gonadal sex differentiation in the Senegalese sole (*Solea senegalensis*) and first data on the experimental manipulation of its sex ratios. Aquaculture 384–386. https://doi.org/10.1016/j.aquaculture.2012.12.012.

*animals*

MDPI

*Article*

# Genetic Estimates for Growth and Shape-Related Traits in the Flatfish Senegalese Sole

**Israel Guerrero-Cozar [1], Eduardo Jimenez-Fernandez [2], Concha Berbel [1], Elena Espinosa [3], Manuel Gonzalo Claros [3,4,5,6], Ricardo Zerolo [7] and Manuel Manchado [1,8,*]**

1 IFAPA Centro El Toruño, Junta de Andalucía, Camino Tiro Pichón s/n, 11500 El Puerto de Santa María Cádiz, Spain; israel.guerrero@juntadeandalucia.es (I.G.-C.); mariac.berbel@juntadeandalucia.es (C.B.)
2 Free Radical Research Group, Centre for Health Sciences, University of the Highlands and Islands, Inverness IV2 3JH, UK; eduardojf85@icloud.com
3 Molecular Biology and Biochemistry Department, University of Málaga, 29071 Málaga, Spain; elenamariaesga@gmail.com (E.E.); claros@uma.es (M.G.C.)
4 CIBER de Enfermedades Raras (CIBERER), 29071 Málaga, Spain
5 Institute of Biomedical Research in Málaga (IBIMA), IBIMA-RARE, 29010 Málaga, Spain
6 Instituto de Hortofruticultura Subtropical y Mediterránea (IHSM-UMA-CSIC), 29010 Málaga, Spain
7 CUPIMAR, Ctra. Carraca, n° 2 Salina San Juan Bautista, San Fernando, 11100 Cádiz, Spain; ricardo.zerolo@cupimar.com
8 "Crecimiento Azul", Centro IFAPA El Toruño, Unidad Asociada al CSIC, 11500 El Puerto de Sta María, Spain
* Correspondence: manuel.manchado@juntadeandalucia.es

**Citation:** Guerrero-Cozar, I.; Jimenez-Fernandez, E.; Berbel, C.; Espinosa, E.; Claros, M. G.; Zerolo, R.; Manchado, M. Genetic Estimates for Growth and Shape-Related Traits in the Flatfish Senegalese Sole. *Animals* **2021**, *11*, 1206. https://doi.org/10.3390/ani11051206

Academic Editor: Montse Pérez

Received: 18 March 2021
Accepted: 19 April 2021
Published: 22 April 2021



**Simple Summary:** To increase competitiveness, the aquaculture flatfish industry demands animals with optimal growth rates and a high shape quality. Genetic breeding is an essential tool to achieve these goals but it requires the estimation of the genetic components of these traits under industrial conditions. The current study provides phenotypic data and genetic parameters of eight traits related to growth and shape quality. The high heritabilities and correlations obtained support that genetic breeding programs can be successfully implemented in Senegalese sole to optimize production.

**Abstract:** Shape quality is very important in flatfish aquaculture due to the impact on commercialization. The Senegalese sole (*Solea senegalensis*) is a valuable flatfish with a highly elliptic body that slightly changes with age and size, and it is prone to accumulating malformations during the production cycle. The present study aims to investigate the genetic parameters of two growth traits (weight and standard length) and six shape quality predictors (ellipticity, three body heights (body height at the pectoral fin base [BHP], body maximum height [BMH] and caudal peduncle height [CPH]) and two ratios (BMH/BHP and BMH/CPH)). These traits were measured before the on-growing stage (age ~400 days (d)) and at harvest (~800 d). Phenotypic data, heritabilities and genetic and phenotypic correlations between the traits are presented and discussed. High or very high heritabilities (0.433–0.774) were found for growth traits, body heights and ellipticity and they were higher at 400 than 800 d. In contrast, the ratios of BMH/BHP and BMH/CPH were less heritable (0.144–0.306). Positive and very high (>0.95) correlations between growth traits and the three heights were found and decreased with age. In contrast, ellipticity had negative and medium-high genetic correlations with growth traits and heights, indicating fish selected for bigger size would also become rounder. The ratio of BMH/CPH showed low genetic correlations with all traits and provided complementary information to ellipticity for a better fitting to the expected lanceolate body morphology of sole. The genetic correlations for all traits at both ages were very high, indicating that selection before entering the growth-out stage in recirculation aquaculture systems is recommended to accelerate genetic gains.

**Keywords:** Senegalese sole; genetic estimates; shape; growth; breeding

## 1. Introduction

Flatfish is a general name for a diverse group of highly appreciated species worldwide, both in fisheries and aquaculture. They are morphologically unique among fishes due to their body asymmetry, which is acquired after the migration of one eye to the opposite side and the cranium remodeling in early larval stages. This process, known as metamorphosis, also entails a drastic reorganization of the abdominal cavity, skin pigmentation patterns and the development of sensory structures for the adaptation to a bottom-dwelling mode of life. As a consequence, the new flattened bodies acquire species-specific shapes for swimming, and camouflage capabilities as adaptive mechanisms to specific ecological niches [1]. In a general way, flatfish species from the families Bothidae, Cynoglossidae, Poecilopsettidae and Soleidae are characterized by oblong bodies with shorter jaws and longer dorsal and anal fins than the families Citharidae, Paralichthyidae, Pleuronectidae or Scophthalmidae, among others [2]. Although flatfish shape is slightly modified with the age and size, the species-specific morphological features are well-identified by consumers and are usually important criteria in commercial decisions and the price of fresh marketed products. Due to the high relevance of external morphology on commercialization, the production of high-quality shaped fish is highly important in aquaculture to enhance consumers' awareness and support their perception of fish aquaculture products [3].

Senegalese sole (*Solea senegalensis*) is a marine flatfish of high economic value whose aquaculture is rapidly growing in Southern Europe. The shape of this right-eyed flatfish is well-recognized by the lanceolate bodies, short jaws and long fins. However, this species exhibits a high plasticity of the skeletal components, such as the vertebral number, which oscillates between 44 and 48 (mode = 45) with 8–9 in the abdominal region, 34–35 in the caudal region and 3–4 in the caudal complex [4,5]. Moreover, this species is highly prone to vertebral abnormalities and other skeletal malformations that can reach even more than 70% of individuals in cultured populations, most of them corresponding to vertebral fusions in the caudal region and deformities in the caudal complex [4,6–9]. Most of these malformations are usually externally unnoticed or they have a moderate effect on gross phenotypic morphology (approximately 46% of animals with vertebral deformities were categorized as normal) [4]. However, this plasticity and high incidence of malformations can shift the body ellipticity with an impact on the quality of the marketable product; hence, it is very important to identify the phenotypic and genetic determination of the main morphological traits and the association with other productive parameters.

Nutritional factors and environmental conditions have been identified as two major modulators of morphological features and malformations in Senegalese sole. High levels of vitamin A increase the mean number of vertebrae and the malformation rates in the vertebrae and caudal fin [10]. Moreover, a high stocking density (29.8 kg $m^{-2}$) shifts the relative body proportions toward a wider head and a shortened caudal region with an enlarged peduncle [11]. A high temperature (>18 °C) during larval rearing also increased vertebral anomalies in the caudal region and caudal complex, although the effects on external morphology were not evaluated [7]. In the closely related species *Solea solea*, the body ellipticity measured using image analysis was proposed as an optimal trait to assess the quality of external sole shape [12]. This trait showed a moderate heritability (0.34 ± 0.11) and a moderate and negative genetic correlation with body weight, highlighting the importance of controlling for this trait to maintain high-quality shaped fish in genetic breeding programs [12]. This study aimed at estimating the genetic and phenotypic parameters for growth and shape-related traits at two important stages in the production cycle of Senegalese sole, before entering growth-out in recirculation aquaculture systems (RAS) (~400 days (d)) and at harvest (~800 d). Weight, standard length, three body heights

(at the pectoral fin, maximal and in the peduncle), their relative ratios and body ellipticity were evaluated as quality indicators of sole shape. Heritability estimates and genetic and phenotypic correlations at both ages are provided. The data provided are highly relevant in genetic breeding programs.

## 2. Materials and Methods

### *2.1. Animals*

Broodstock used to produce families comprised 150 wild specimens approx. 8 years old caught in salt marshes from the Gulf of Cadiz (Spain). They were fed with frozen feed including mussels, small squids and polychaeta worms (Seabait Ltd., Ashington, UK) on alternative days. Mass spawning strategy to create the families was previously described [13]. Briefly, spawning was synchronized by thermoperiod control [14]. Due to the courtship behavior of sole [15], it is not easy to achieve all the breeder tanks ($n$ = 9) responding simultaneously in the same thermocycle. Hence, with the objective to increase the number of families in the population upon evaluation, seven evaluation batches (EBs) obtained after different thermocycles were created by mixing proportionally the volume of eggs from each tank that contributed offspring in each thermal treatment. To facilitate the data comparison and convergence, the offspring of a breeder tank ($n$ = 6) were always included in all EBs. Larval rearing and weaning protocols for each EB were those previously described [16,17] and each EB was always managed as a unit until harvest without any grading.

For genetic evaluation, fish (ranging from 200 to 550 specimens per EB) were intraperitoneally tagged, with ages ranging between 150 and 278 days post-hatch (dph) as previously reported [13,18]. Later, fish were phenotypically evaluated in vivo at ~400 d (ranging from 395 to 446 dph) before entering the growth-out period in RAS and at harvest age ~800 d (ranging between 733 and 861 dph). No intermediate samplings were carried out to follow standard production practices and minimize animal handling and stress. Cumulative mortality between ages was lower than 5% and a total of 1840 fish (EB1 = 136; EB2 = 289; EB3 = 273; EB4 = 420; EB5 = 229; EB6 = 234; EB7 = 259) sampled at both ages were considered in this study. Information about the full dataset and culture conditions was previously reported [13]. Fish were individually weighted (W) using Gram FC-200 and a photograph was taken using a Canon EOs1300D camera following the methodology previously established in PROGENSA® [19]. Image analysis was carried out using the Fiji 2.0.0-rc-69/1.52p and standard length (SL), body height at the insertion of the pectoral fin (BHP), body maximum height (BMH) and caudal peduncle height (CPH) were measured (Figure 1). The two ratios between heights (BMH to BHP and BMH to CPH) and ellipticity ((SL-BMH)/(SL + BMH)) [12] were calculated. At harvest, fish were sacrificed using slurring ice following commercial techniques and 60 specimens of each batch were kept alive as future breeders. From sacrificed fish were taken a piece of caudal fin that was preserved in 99% alcohol, and alive fish were sampled for blood by puncturing in the caudal vein using a heparinized syringe, adding heparin (100 mU) and keeping at −20 °C until use. All fish were sexed and the presence of white nodules compatible with amoebic disease (AD) were recorded.

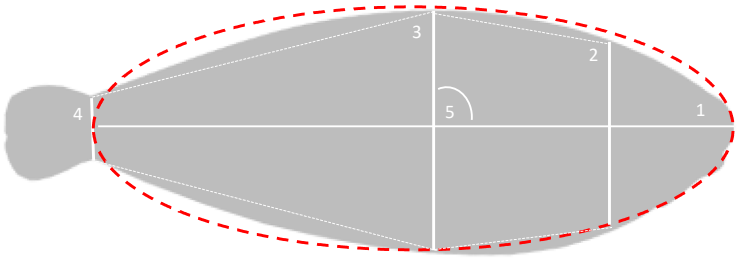


**Figure 1.** Shape measurements: 1: standard length (SL); 2: body height at the insertion of the pectoral fin (BHP); 3: body maximum height (BMH); 4: caudal peduncle height (CPH); and 5: ellipticity ((SL − BMH)/(SL + BMH)). A theoretical ellipse fitting the horizontal axis from the mouth tip to the peduncle center and the vertical axis to BMH is indicated by red dashed line.

### *2.2. DNA Isolation and Parentage Assignment*

DNA isolation from blood (broodstock and non-sacrificed offspring) or caudal fin (slaughtered F1; 30 mg) was carried out using the Isolate II genomic DNA kit (Bioline, London, UK) following the manufacturer's instructions. DNA was quantified using a Nanodrop ND-8000 and quality was evaluated by agarose gene electrophoresis. Genotyping of breeders and offspring was carried out using an 11-loci supermultiplex PCR [20] on an ABI3130 sequencer (Applied Biosystems, Foster City, CA, USA) and genotypes were collected using Genemapperv3.8 (Applied Biosystems, Foster City, CA, USA). Finally, parentage assignment was performed with Vitassign v8.2.1 [21] following the allelic exclusion method. Assignment rates to a single parent pair was 100%. A total number of 71 families from 37 males and 30 females were evaluated. The number of families per batch ranged from 11 (EB1 and EB5) to 23 (EB7). Offspring of seven males and six females were represented in four or more EB.

### *2.3. Statistical Analysis and Genetic Parameters*

All data were tested for normality and homogeneity of variance using SPSS v.23 (SPSS, Chicago, IL, USA). Weight at 400 and 800 d were cube square root and square root transformed, respectively, to fit normality. ANOVA analysis using the General Linear. Models (GLM) procedure was carried out using the gender, EB and AD as fixed factors. To test the effect of age (evaluation of traits between 400 and 800 d), a repeated measures ANOVA was carried out for each trait using the same fixed factors. Regression analysis and slope significance testing were carried out with Prism 9.0 (Graphpad Software Inc., San Diego, CA, USA). Genetic estimates of heritability and correlations were calculated using restricted maximum likelihood adjusted linear mixed models (REML) in WOMBAT [22]: $y = X\beta + Zu + e$, where y is the observed trait, $\beta$ is the fixed factor vector (gender, EB and AD), u is the animal random factor vector and e is the error.

## 3. Results

### *3.1. Phenotypic Data for Growth Traits*

The phenotypic mean ± Standard error (SE) of growth traits (weight and SL) at 400 and 800 d are depicted in Table 1 and in Figure S1. Mean weight and SL were 32.4 ± 29.2 g and 12.00 ± 2.87 cm at 400 d and 264.9 ± 171.9 g and 23.35 ± 4.79 cm at 800 d. Statistical ANOVA analysis showed statistically significant differences associated with the gender, EB and AD (Figures 2 and S1) for both traits. Estimated marginal means indicated that the females appeared on average 16.1% heavier and 2.8% longer than males at 400 d, and 12.2% heavier and 2.5% longer than males at 800 d (Figures 2 and S1). A significant gender × EB interaction was observed at both ages. In addition, significant differences associated with the EB ($p < 0.05$) were found that ranged between 18.9 and 63.3 g at 400 d for EB3 and

EB7, respectively, and between 126.8 and 376.7 g at 800 d for EB3 and EB6, respectively (Figures 2 and S1). A total of 15.3% of evaluated fish at harvest had nodules compatible with amoebic lesions in the gut and/or liver. Fish without hepatic or intestinal amoebic lesions at 800 d were significantly heavier (44.9%) than infected fish (Figure 2). A repeated measures ANOVA analysis revealed significant age × EB and age × AD interactions for weight and length gain and age × gender for weight gain. Tendencies for the three different fixed factors and levels are depicted in Figure 2. A regression weight-length analysis for gender at both ages showed that the coefficients of determination ($R^2$) were ≥ 0.95 with slopes between 3.32 and 3.34 (not statistically significant) (Figure S2).

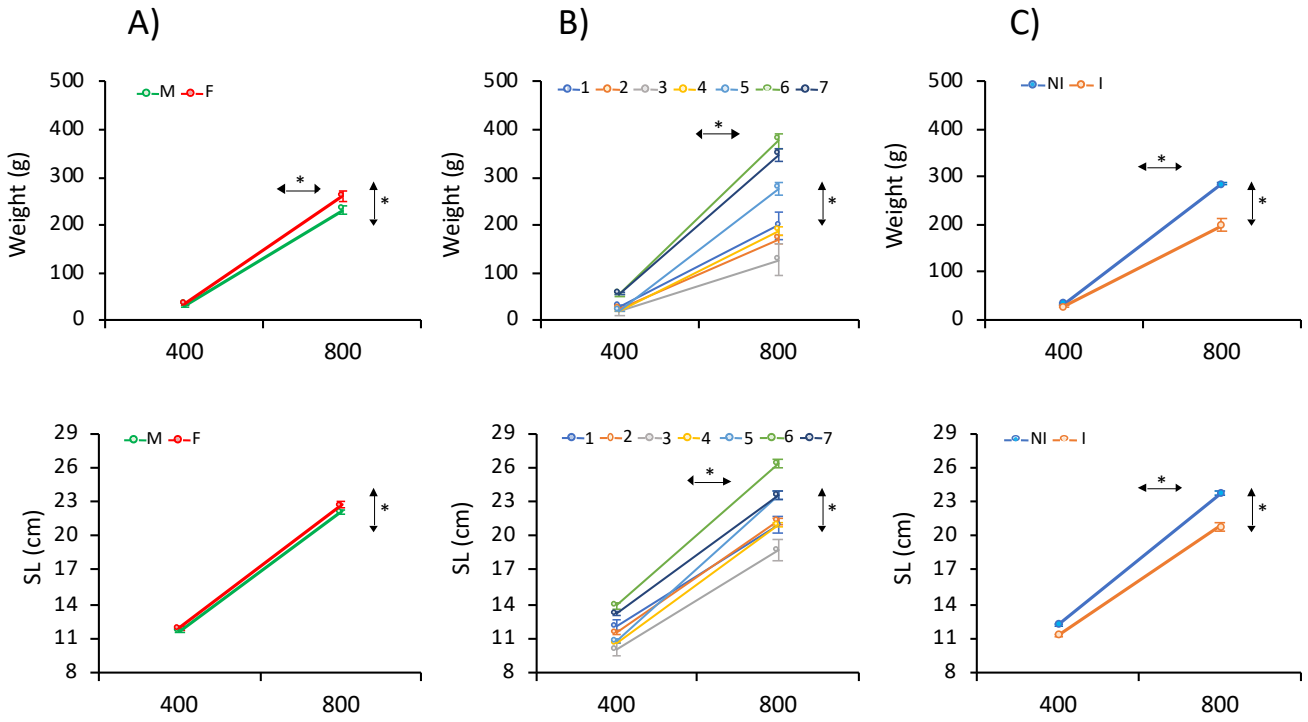


**Figure 2.** Estimated marginal means for weight and standard length (SL) as determined by repeated-measures ANOVA at 400 and 800 days (d) for (**A**) gender (male: M; female: F), (**B**) Evaluation Batch (EB) (1–7) and (**C**) Amoebic disease (AD) (infected: I; non-infected: NI). The asterisks (*) on the horizontal or vertical arrows denote if within- or between-subject effects were significant.

**Table 1.** Phenotypic data for growth traits (weight and standard length (SL)), heights (body height at the insertion of the pectoral fin (BHP), body maximum height (BMH) and caudal peduncle height (CPH)), height ratios (BMH/BHP and BMH/CPH) and ellipticity at 400 and 800 days (d). Overall mean ± standard error by gender is shown. The number (*n*) of soles evaluated at each age is also indicated.

| **400 days (*n* = 1840)** | **Male (*n* = 1007)** | **Female (*n* = 833)** | **Mean** |
|---|---|---|---|
| Weight | 30.7 ± 28.0 | 34.4 ± 30.4 | 32.4 ± 29.2 |
| SL | 11.85 ± 2.87 | 12.18 ± 2.86 | 12.00 ± 2.87 |
| BHP | 3.84 ± 1.00 | 4.00 ± 1.02 | 3.91 ± 1.01 |
| BMH | 4.54 ± 1.33 | 4.71 ± 1.34 | 4.62 ± 1.34 |
| CPH | 1.13 ± 0.34 | 1.10 ± 0.34 | 1.11 ± 0.34 |
| BMH/BHP | 0.45 ± 0.03 | 0.45 ± 0.03 | 0.45 ± 0.03 |
| BMH/CPH | 1.17 ± 0.06 | 1.17 ± 0.06 | 1.17 ± 0.05 |
| Ellipticity | 4.165 ± 0.358 | 4.183 ± 0.335 | 4.173 ± 0.347 |
| **800 days (*n* = 1840)** | **Male (*n* = 1007)** | **Female (*n* = 833)** | **Mean** |
| Weight | 244.0 ± 153.0 | 290.3 ± 189.3 | 264.9 ± 171.9 |
| SL | 22.91 ± 4.64 | 23.88 ± 4.91 | 23.35 ± 4.79 |

| | | | |
|---|---|---|---|
| BHP | 7.51 ± 1.66 | 8.01 ± 1.85 | 7.74 ± 1.77 |
| BMH | 9.24 ± 2.22 | 9.87 ± 2.46 | 9.53 ± 2.35 |
| CPH | 2.55 ± 0.66 | 2.68 ± 0.67 | 2.61 ± 0.67 |
| BMH/BHP | 0.426 ± 0.027 | 0.419 ± 0.028 | 0.424 ± 0.028 |
| BMH/CPH | 1.225 ± 0.054 | 1.227 ± 0.050 | 1.226 ± 0.052 |
| Ellipticity | 3.650 ± 0.325 | 3.699 ± 0.320 | 3.673 ± 0.323 |

### *3.2. Phenotypic Data for Height Traits*

Due to the flattened morphology of sole, the body height at the insertion of the pectoral fin (BHP), body maximum height (BMH) and caudal peduncle height (CPH) and the two ratios of BMH/BHP and BMH/CPH were determined both at 400 and 800 d (Table 1; Figure 1). Mean BHPs were 3.91 ± 1.01 and 7.74 ± 1.77 cm at 400 and 800 d, respectively; the BMHs were 4.62 ± 1.34 and 9.53 ± 2.35 cm, respectively; and the CPHs were 1.11 ± 0.34 and 2.61 ± 0.67 cm, respectively. The three height traits showed statistically significant differences associated with the gender and EB at both ages, and AD at 800 d (Figure 3). Females and non-infected soles had higher heights than males and infected fish. On average, heights in females were 4.6, 4.5 and 4.3% higher than in males and the non-infected fish, and 10.2, 11.2 and 11.9% higher than in infected fish, for BHP, BMH and CPH, respectively. Moreover, EB6 and EB3 showed the largest and lowest heights, respectively. A longitudinal analysis to determine the height gain from 400 to 800 d using repeated measures ANOVA demonstrated significant interactions of age × gender, age × EB and age × AD (Figure 3).

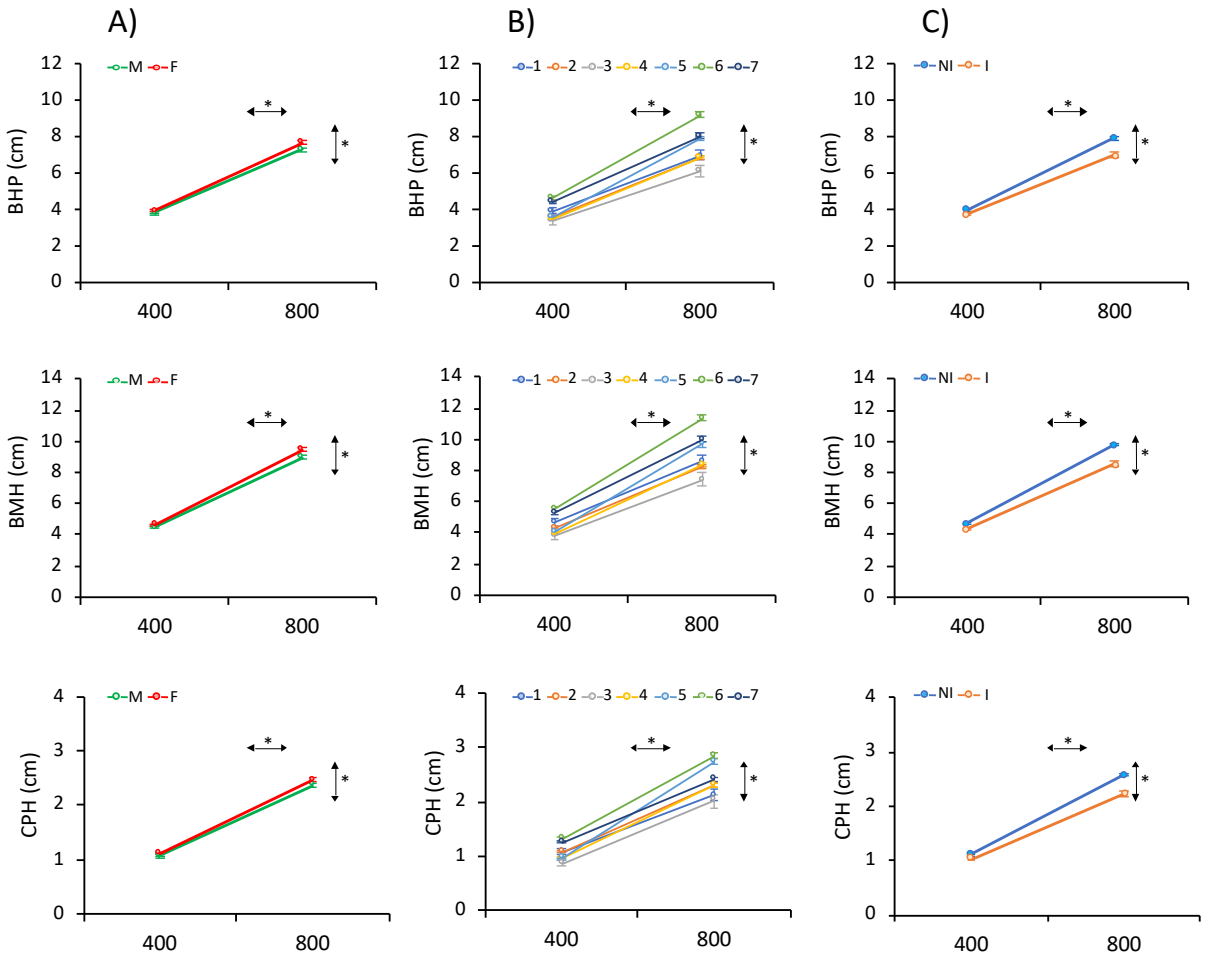

**Figure 3.** Estimated marginal means for height at the pectoral fin base (body height at the insertion of the pectoral fin (BHP), body maximum height (BMH) and caudal peduncle height (CPH) as determined by repeated-measures ANOVA at 400 and 800 days (d) for (**A**) gender (male: M; female: F), (**B**) Evaluation Batch (EB) (1–7) and (**C**) Amoebic disease (AD) (infected: I; non-infected: NI). The asterisks (*) on the horizontal or vertical arrows denote if within- or between-subject effects were significant.

A regression analysis of CPH and BHP on BMH indicated a stronger association between BHP and BMH ($R^2 > 0.97$) than CPH and BMH ($R^2 > 0.86$). Moreover, slopes for males were statistically significantly smaller than females at 800 d at both ages (Figure S3).

With respect to the BMH/CPH and BMH/BHP ratios, BMH/CPH significantly reduced and BMH/BHP increased with age, from 400 to 800 d (Figure 4). A significant effect of the EB on both ratios at 400 and 800 d was detected (Figure S4). Nevertheless, the gender effect was only significant for BMH/BHP at 400 d. The longitudinal analysis only identified a significant interaction of age × EB (Figure S4). In the repeated-measures ANOVA a significant between-subject effect of AD for BMH/BHP was also found.

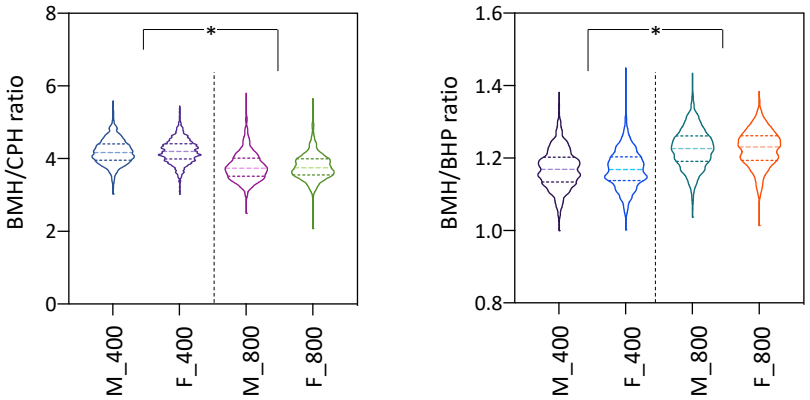


**Figure 4.** Violin plots for body maximum height (BMH) to caudal peduncle height (CPH) and BMH to body height at the insertion of the pectoral fin (BHP) ratios. Data for males (M) and females (F) at both 400 and 800 days (d) are indicated. The asterisk (*) denotes statistically significant differences between ages.

### *3.3. Phenotypic Data for Ellipticity*

Mean ellipticity was 0.449 ± 0.025 at 400 d and 0.422 ± 0.029 at 800 d (Table 1). The distribution of ellipticity at both shapes is shown in Figure 5. Values ranged from 0.32 to 0.52 at 400 d and between 0.24 and 0.51 at 800 d. ANOVA analysis indicated statistically significant differences associated with the gender and EB at both ages and with AD at 800 d (Figure 6). Males and infected fish were more elliptic than females (1.0 and 2.3% higher at 400 and 800 d, respectively) and non-infected fish (1.4% higher) (Table 1; Figure 6). The longitudinal analysis demonstrated a significant interaction of age × gender and age × EB during the cultivation period in RAS (Figure 6).

As ellipticity was significantly and negatively correlated with weight ($R^2$ ranging from 0.362 to 0.443), an ANCOVA analysis using the weight as a covariate was carried out and significant differences associated with the EB and gender at both ages were still observable. An analysis of ellipticity by weight class indicated that females were statistically less elliptic than males in class 0–10 g at 400 d and classes 0–100, 300–400, 400–500 and > 600 g at 800 d.

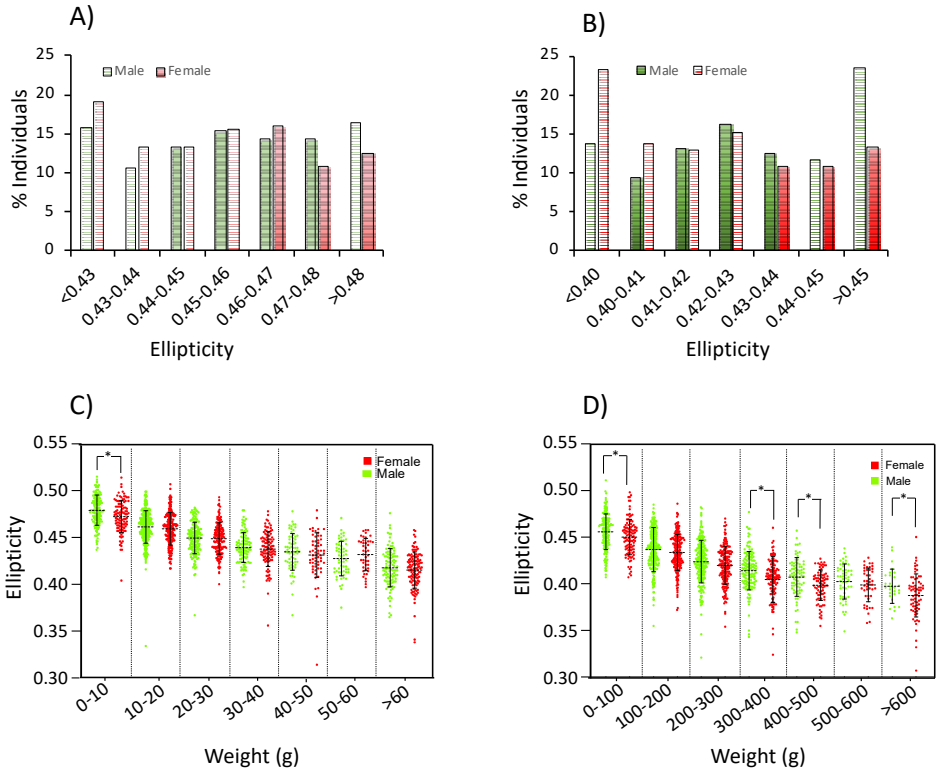


**Figure 5.** Distribution of ellipticity classes by weight. Panels (**A**) (400 days (d)) and (**B**) (800 d) show the frequency of males (green) and females (red) by ellipticity classes. Panels (**C**) (400 d) and (**D**) (800 d) show the ellipticity scatterplot by weight class and gender. The asterisks (*) denote statistically significant differences between gender in a weight class.

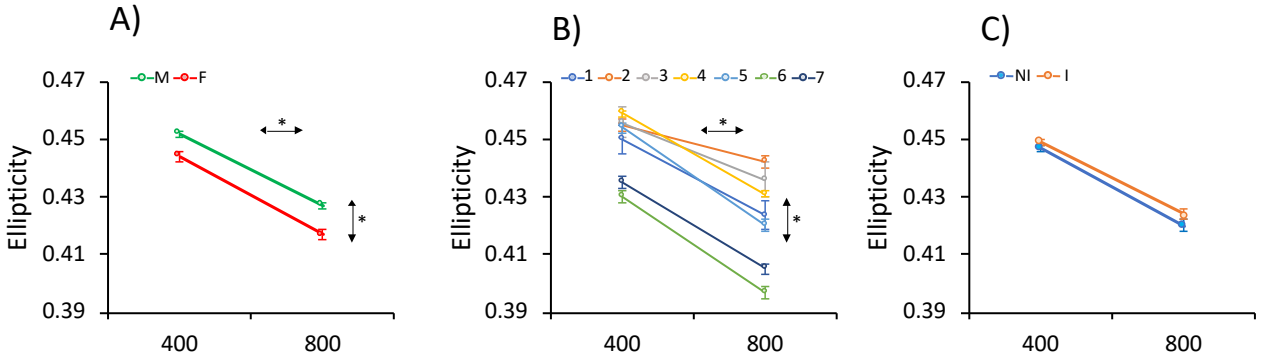


**Figure 6.** Estimated marginal means for ellipticity as determined by repeated-measures ANOVA at 400 and 800 days (d) for the (**A**) gender (male: M; female: F), (**B**) Evaluation Batch (EB) (1–7) and (**C**) Amoebic disease (AD) (infected: I; non-infected: NI). The asterisks (*) on the horizontal or vertical arrows denote if within- or between-subject effects were significant.

### *3.4. Genetic Estimates*

#### 3.4.1. Heritability

Heritabilities and correlations for growth and shape-related traits at 400 and 800 d are depicted in Table 2. Heritabilities were higher for all the traits (except BMH/CPH) at 400 than 800 d. Heritability estimates for weight, SL, the three heights and ellipticity were high or very high at both ages. They ranged between 0.567 and 0.774 at 400 d and between

0.433 and 0.735 at 800 d for ellipticity and SL, respectively. The ratios of BMH/BHP and BMH/CPH had low or moderate heritability values (0.270–0.303 at 400 d and 0.144–0.306 at 800 d).

**Table 2.** Heritabilities (diagonal in bold), phenotypic correlations (below the diagonal) and genetic correlations (above the diagonal) for growth traits (weight (W) and standard length (SL)), heights (body height at the insertion of the pectoral fin (BHP), body maximum height (BMH) and caudal peduncle height (CPH)), height ratios (BMH/BHP and BMH/CPH) and ellipticity (E) at 400 days (d) (top) and 800 d (bottom).

| **400 d** | **W** | **SL** | **BHP** | **BMH** | **CPH** | **BMH/BHP** | **BMH/CPH** | **E** |
|---|---|---|---|---|---|---|---|---|
| W | **0.625 ± 0.109** | 0.991 ± 0.004 | 0.988 ± 0.004 | 0.992 ± 0.003 | 0.990 ± 0.005 | 0.874 ± 0.057 | 0.161 ± 0.189 | −0.768 ± 0.073 |
| SL | 0.983 ± 0.002 | **0.567 ± 0.104** | 0.981 ± 0.007 | 0.986 ± 0.005 | 0.984 ± 0.007 | 0.881 ± 0.054 | 0.167 ± 0.189 | −0.724 ± 0.085 |
| BHP | 0.981 ± 0.002 | 0.976 ± 0.002 | **0.623 ± 0.110** | 0.991 ± 0.001 | 0.948 ± 0.005 | 0.858 ± 0.064 | 0.284 ± 0.180 | −0.738 ± 0.025 |
| BMH | 0.988 ± 0.001 | 0.982 ± 0.002 | 0.999 ± 0.001 | **0.621 ± 0.109** | 0.955 ± 0.004 | 0.878 ± 0.055 | 0.247 ± 0.183 | −0.828 ± 0.056 |
| CPH | 0.952 ± 0.004 | 0.953 ± 0.004 | 0.974 ± 0.010 | 0.979 ± 0.008 | **0.576 ± 0.105** | 0.881 ± 0.055 | 0.044 ± 0.193 | −0.749 ± 0.079 |
| BMH/BHP | 0.591 ± 0.024 | 0.61 ± 0.023 | 0.528 ± 0.028 | 0.622 ± 0.023 | 0.587 ± 0.024 | **0.270 ± 0.069** | 0.102 ± 0.035 | −0.521 ± 0.034 |
| BMH/CPH | 0.062 ± 0.045 | 0.047 ± 0.043 | 0.09 ± 0.044 | 0.094 ± 0.044 | −0.191 ± 0.044 | 0.076 ± 0.204 | **0.303 ± 0.076** | −0.254 ± 0.045 |
| E | −0.673 ± 0.031 | −0.284 ± 0.033 | −0.838 ± 0.054 | −0.750 ± 0.025 | −0.677 ± 0.031 | −0.662 ± 0.115 | −0.487 ± 0.153 | **0.774 ± 0.117** |
| **800d** | **W** | **SL** | **BHP** | **BMH** | **CPH** | **BMH/BHP** | **BMH/CPH** | **E** |
| W | **0.486 ± 0.099** | 0.983 ± 0.007 | 0.974 ± 0.01 | 0.978 ± 0.008 | 0.983 ± 0.008 | 0.546 ± 0.162 | 0.016 ± 0.198 | −0.608 ± 0.115 |
| SL | 0.975 ± 0.002 | **0.433 ± 0.094** | 0.961 ± 0.015 | 0.957 ± 0.016 | 0.964 ± 0.014 | 0.412 ± 0.183 | 0.011 ± 0.198 | −0.509 ± 0.137 |
| BHP | 0.967 ± 0.003 | 0.948 ± 0.005 | **0.549 ± 0.105** | 0.996 ± 0.002 | 0.953 ± 0.018 | 0.506 ± 0.177 | 0.181 ± 0.193 | −0.703 ± 0.092 |
| BMH | 0.982 ± 0.002 | 0.962 ± 0.004 | 0.983 ± 0.001 | **0.515 ± 0.102** | 0.961 ± 0.016 | 0.586 ± 0.155 | 0.182 ± 0.192 | −0.733 ± 0.085 |
| CPH | 0.928 ± 0.005 | 0.918 ± 0.006 | 0.911 ± 0.007 | 0.926 ± 0.006 | **0.463 ± 0.097** | 0.612 ± 0.151 | −0.073 ± 0.195 | −0.586 ± 0.123 |
| BMH/BHP | 0.479 ± 0.027 | 0.486 ± 0.027 | 0.345 ± 0.031 | 0.508 ± 0.026 | 0.474 ± 0.026 | **0.144 ± 0.046** | 0.178 ± 0.211 | −0.389 ± 0.031 |
| BMH/CPH | −0.003 ± 0.042 | −0.015 ± 0.04 | 0.046 ± 0.043 | 0.050 ± 0.043 | −0.314 ± 0.040 | 0.048 ± 0.029 | **0.306 ± 0.075** | −0.217 ± 0.044 |
| E | −0.557 ± 0.037 | −0.447 ± 0.042 | −0.644 ± 0.033 | −0.662 ± 0.030 | −0.548 ± 0.038 | −0.719 ± 0.117 | −0.534 ± 0.144 | **0.735 ± 0.115** |

3.4.2. Genetic Correlations

Genetic correlations between growth and height traits were very high both at 400 and 800 d (>0.95). The ratio of BMH/BHP had moderate-high genetic correlations with growth and height traits that were higher at 400 (0.858–0.881) than 800 d (0.412–0.612). The genetic correlations of BMH/CPH were low (<0.28). The ellipticity had negative and high genetic correlations with growth and height traits ranging from −0.724 to −0.828 at 400 d and from −0.509 to −0.733 at 800 d, and a negative and low correlation with height ratios (Table 2).

Genetic and phenotypic correlations between both ages are depicted in Table 3. Ellipticity (0.912) had the highest genetic correlation when the same traits were compared at 400 and 800 d, followed by growth and height traits (average 0.825 and 0.874). The lowest values were between height ratios (0.663–0.687).

**Table 3.** Genetic (top) and phenotypic correlations between 400 (left) and 800 days (d) (right) for growth traits (weight (W) and standard length (SL)), heights (body height at the insertion of the pectoral fin (BHP), body maximum height (BMH) and caudal peduncle height (CPH)), height ratios (BMH/BHP and BMH/CPH) and ellipticity (E).

| | **Genetic** | **800 d** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **W** | **SL** | **BHP** | **BMH** | **CPH** | **BMH/BHP** | **BMH/CPH** | **E** |
| 400 d | W | 0.843 ± 0.054 | 0.831 ± 0.060 | 0.832 ± 0.057 | 0.838 ± 0.055 | 0.849 ± 0.054 | 0.509 ± 0.169 | 0.041 ± 0.194 | −0.554 ± 0.122 |
| | SL | 0.828 ± 0.058 | 0.837 ± 0.057 | 0.813 ± 0.062 | 0.817 ± 0.061 | 0.826 ± 0.057 | 0.509 ± 0.161 | 0.049 ± 0.194 | −0.511 ± 0.134 |
| | BHP | 0.868 ± 0.047 | 0.859 ± 0.052 | 0.874 ± 0.045 | 0.876 ± 0.044 | 0.853 ± 0.053 | 0.524 ± 0.166 | 0.213 ± 0.186 | −0.637 ± 0.105 |
| | BMH | 0.853 ± 0.051 | 0.856 ± 0.050 | 0.862 ± 0.048 | 0.870 ± 0.046 | 0.846 ± 0.055 | 0.542 ± 0.162 | 0.180 ± 0.188 | −0.618 ± 0.109 |
| | CPH | 0.814 ± 0.063 | 0.821 ± 0.062 | 0.786 ± 0.070 | 0.811 ± 0.063 | 0.825 ± 0.062 | 0.523 ± 0.167 | 0.025 ± 0.195 | −0.533 ± 0.126 |
| | BMH/BHP | 0.633 ± 0.123 | 0.604 ± 0.133 | 0.575 ± 0.141 | 0.653 ± 0.121 | 0.668 ± 0.120 | 0.687 ± 0.139 | 0.038 ± 0.206 | −0.541 ± 0.141 |
| | BMH/CPH | 0.296 ± 0.184 | 0.275 ± 0.186 | 0.408 ± 0.167 | 0.372 ± 0.172 | 0.194 ± 0.193 | −0.009 ± 0.229 | 0.663 ± 0.140 | −0.442 ± 0.161 |
| | E | −0.762 ± 0.079 | −0.712 ± 0.095 | −0.849 ± 0.054 | −0.858 ± 0.051 | −0.733 ± 0.088 | −0.601 ± 0.151 | −0.492 ± 0.152 | 0.912 ± 0.032 |
| | **Phenotypic** | **W** | **SL** | **BHP** | **BMH** | **CPH** | **BMH/BHP** | **BMH/CPH** | **E** |
| 40 | W | 0.786 ± 0.018 | 0.765 ± 0.019 | 0.778 ± 0.021 | 0.783 ± 0.019 | 0.740 ± 0.021 | 0.337 ± 0.031 | 0.015 ± 0.044 | −0.474 ± 0.046 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SL | 0.791 ± 0.018 | 0.790 ± 0.017 | 0.781 ± 0.020 | 0.787 ± 0.019 | 0.746 ± 0.020 | 0.358 ± 0.028 | 0.013 ± 0.039 | −0.430 ± 0.046 |
| BHP | 0.798 ± 0.017 | 0.777 ± 0.019 | 0.809 ± 0.017 | 0.809 ± 0.016 | 0.752 ± 0.020 | 0.328 ± 0.032 | 0.059 ± 0.044 | −0.536 ± 0.041 |
| BMH | 0.798 ± 0.018 | 0.796 ± 0.017 | 0.803 ± 0.018 | 0.810 ± 0.017 | 0.751 ± 0.020 | 0.354 ± 0.031 | 0.056 ± 0.044 | −0.533 ± 0.042 |
| CPH | 0.761 ± 0.020 | 0.765 ± 0.020 | 0.748 ± 0.022 | 0.764 ± 0.020 | 0.739 ± 0.021 | 0.336 ± 0.031 | −0.036 ± 0.042 | −0.464 ± 0.045 |
| BMH/BHP | 0.478 ± 0.030 | 0.474 ± 0.027 | 0.313 ± 0.031 | 0.496 ± 0.029 | 0.458 ± 0.030 | 0.385 ± 0.026 | 0.032 ± 0.035 | −0.362 ± 0.038 |
| BMH/CPH | 0.081 ± 0.043 | 0.072 ± 0.041 | 0.130 ± 0.044 | 0.118 ± 0.043 | −0.001 ± 0.042 | 0.005 ± 0.031 | 0.306 ± 0.031 | −0.197 ± 0.045 |
| E | −0.615 ± 0.035 | −0.521 ± 0.037 | −0.672 ± 0.029 | 0.001 ± 0.028 | −0.574 ± 0.035 | −0.274 ± 0.036 | −0.175 ± 0.047 | 0.797 ± 0.021 |

## 4. Discussion

Genetic breeding programs for growth performance and shape quality are essential for flatfish aquaculture industry competitiveness. This highly plastic taxonomic group transforms during development from a bilateral symmetry to an asymmetric, highly specialized flattened body. Evolutionary studies have demonstrated that different ecological traits act as a driver of body shape in flatfish, acquiring a wide range of body depths, jaw lengths and fin lengths [2]. Hence, flatfish families can be identified by specific shapes and morphological features that should be carefully preserved in aquaculture to maintain consumer acceptance and commercial value. In the case of Senegalese sole, body shape is expected to be highly elliptic and lanceolate with short jaws and long dorsal and anal fins that contrast with the shape of most pleuronectids or scophthalmids with deeper bodies, longer jaws and short dorsal and anal fins. However, several reports that dealt with morphological traits in Senegalese sole in aquaculture reported high rates of malformations that in most cases do not have a severe impact on external gross morphology [4,6–8]. In this study, we investigate for first time in Senegalese sole the phenotypic and genetic variation associated with shape-related traits under industrial conditions in RAS. These results are highly valuable to design genetic breeding programs and integrate the shape quality within the selection schemes.

The reproduction of Senegalese sole is extremely complex due to three singularities: the courtship behavior, the dominance and fidelity of highly successful spawners and the low production of sperm [15,23]. Taking into account these reproductive limitations, this study produced families by mass-spawning using a wild broodstock distributed in nine tanks after spawning synchronization by thermoperiod control [14,23]. Although tagging, stocking density, water temperature, type of tank or feed were common to all EBs, this factor had an important effect on growth and shape-related traits after the RAS growth-out phase. The longitudinal analysis (since the same subset of tagged soles was analyzed at both ages) showed different tendencies in RAS even between EBs with a very similar genetic structure and age at sampling, suggesting that some additional factors such as social interactions or differences in the actual flow-through dynamics could also play a key role in the evaluated traits.

In addition to the EB, gender also had an important effect on growth and shape-related traits. The females appeared 12.2–16.1% heavier and 2.1–2.8% longer than males in the evaluation period. These data are in agreement with the differences previously found in sole juveniles and at harvest [13,18,24]. Moreover, in this study we demonstrate that females are less elliptic than males even after correcting by weight. These differences were more evident at harvest, probably due the ovary maturation increasing the abdominal cavity, which in turn reduces ellipticity. A regression analysis between heights also evidenced a small change in the slopes by gender that was not clearly observable when height ratios were analyzed, indicating that compensatory mechanisms could modify the relative body proportions. In addition to gender effects, the presence of amoebic nodules in liver or intestine at harvest also influenced growth and shape-related traits. This parasite accumulates mainly in the intestinal mucosa and later spreads to some different tissues [25]. Although mortality is scarce, this study demonstrated that non-infected fish were 44.9% heavier than infected fish. Moreover, the infected fish were slightly less elliptic at harvest even after correcting by weight, and changed the ratio of BMH/BHP due to the excess of nodules that in some cases distorted the size of the abdominal cavity.

The ellipticity of the sagittal plan was proposed as the best trait to measure shape quality in sole since this trait could be easily derived from direct measures on fish (using body height and body length) or by image analysis, fitting theoretical ellipses with a similar precision [12]. This theoretical assumption is based on the expected elongated body shape of soleids that differentiates them from pleuronectids or scophthalmids. Although this trait is highly influenced by body size, and bigger fish tend to be rounder, the ellipse fitting still remains as a good predictor of shape for soles. Our ellipticity data confirmed a major effect of weight on ellipticity distribution, with bigger values at 400 than 800 d and a progressive reduction with bigger weight class sizes (Figure 5). Moreover, as indicated above, females were rounder than males even after correcting by weight due to the increase in abdominal size for sexual maturation. In yellowtail flounder, females had relatively deeper abdomens and larger heads than males [26]. However, these differences associated with gender were not observed in *S. solea*, although these authors did not follow a longitudinal approach or provide information about gonad development that could explain such differences. The significant effects of EB conditions, as indicated above, on the ellipticity trajectory also denote the importance of culture conditions on shape, and the relevance of controlling this important feature to maintain high shape-quality standards for fish commercialization.

Heritabilities for growth traits in flatfish are highly influenced by the age and production system. Previous studies in Japanese flounder reported high or very high heritabilities (>0.6) for growth traits (weight and body length) in juvenile stages (<300 dph) that gradually decreased in older flatfish [27–30]. In contrast, heritabilities for weight at harvest in the closely related species *S. solea*, cultivated in RAS, were low-moderate (0.23–0.25), although this species has very low growth rates compared to *S. senegalensis* [12,31,32]. A previous study in Senegalese sole with a higher number of fish ($n$ = 2171 offspring) also estimated higher heritabilities for growth traits in juveniles before growth-out than later at harvest [13]. These results indicate that the high densities reached in RAS, the water flow dynamics and social hierarchies could mask genetic effects in spite of this species still maintaining high growth rates [33,34]. RAS is not a natural environment for sole, modulating the genetic effects on growth, as determined in common sole with a clear genotype, by environment interaction [35]. Moreover, most females at harvest showed some degree of gonad maturation that was reported in rainbow trout to have a huge impact on the additive genetic variability of weight [36]. Interestingly, the high genetic correlation between 400 and 800 d for growth traits support that growth parameters estimated in juvenile stages before RAS could be used as a good predictor of growth performance later at harvest.

Shape predictors such as ellipticity had very high heritabilities (>0.74) at both ages. These values were considerably higher than those obtained in *S. solea* (0.34) [12] or *Oreochromis niloticus* (0.12–0.45) [37,38]. Although some nutritional, management and culture conditions were reported as regulators of meristic characters and malformation rates in sole [7,10,11], our results indicate a high additive genetic component of the external shape-related traits evaluated in this study. Although malformations could exist (they were not evaluated in this study), most of them would have a low impact on gross morphology as previously indicated [4], evidencing a high genetic component for ellipticity. Nevertheless, further studies are required to associate the shape traits with the skeletal characteristics in order to understand the main causes behind the ellipticity range. Moreover, the high genetic correlations (0.91) between both ages confirm that those genetic factors controlling shape are already acting in juveniles, and hence selection could also be carried out in juveniles.

It should be noted that ellipticity was dependent on fish size. In *S. solea*, a moderate negative genetic correlation ($r_g$ = −0.44) between ellipticity and weight was reported [12]. Similarly, in this study, a high and negative genetic correlation between both traits at 400 (−0.768) and 800 d (−0.608) was determined, indicating that fish reaching a bigger size were also rounder. These negative correlations should be carefully considered if selection

for increased weight at harvest is carried out, since less elliptic fish will be produced. Since most soles are sold in fresh markets, ellipticity was proposed as a correction factor for weight-targeted selection breeding programs [12]. A combined selection index setting a zero change in shape reduced by 9.9–13.8% the response to harvest weight that is assumed to preserve a high-quality shape standard [12]. However, no correction would be necessary if, finally, the industry moves toward transformed seafood products, which is one of the most promising markets for flatfish.

The three heights showed a positive and very high genetic correlation with growth traits (>0.95). However, the heritabilities for the two height ratios were low-moderate and the BMH/CPH ratio had very low genetic correlations with growth and height traits. This latter ratio is strongly related to the swimming speed and performance [39,40]. A deep caudal peduncle provides the fish with a superior ability to accelerate and greater power for propulsion, allowing it to reach a high swimming speed and efficiency [39]. Soles are usually very sedentary in the tanks and they do not require high water columns since they are passive feeders in the tank bottom. Hence, there is not expected to exist a high selection pressure on swimming efficiency in the RAS, although fish should adapt to water currents in the tanks. However, this ratio seems to be useful to refine a lanceolate shape toward a more theoretical elliptic one that fits better to the sole body structure (Figure S5). The peduncle is usually considered as the caudal reference point for body length since the caudal fin is highly variable in size and morphology and the BMH/CPH ratio increases with age (Figure 4). Highly pronounced ratios (due to higher BMH or lower CPH) are associated with very high lanceolate shapes (Figure S5) that deviate from the symmetrical body ellipse, giving rise to turbot-like morphologies. The low heritability for this trait could be due to the benthic way of life and the sensitivity of the caudal complex to traumatisms and malformations that in turn can remodel the peduncle. The low genetic correlations with other ellipticity and growth traits indicate that this trait can provide new relevant information for a genetic selection index to preserve a sole high-quality shape.

## 5. Conclusions

This study provides phenotypic and genetic estimates for growth and shape-related traits and supports selective breeding as an effective strategy to improve these traits in Senegalese sole. The gender, EB and AD had significant effects on most traits when evaluated using a longitudinal approach, indicating that these factors need to be carefully controlled to achieve accurate estimates. Moreover, the high correlations at both ages support that selection can be carried out before growth-out in RAS, accelerating the breeding cycles. A combination of ellipticity, BMH/CPH and weight could be used in a multi trait selection index to control the roundness associated with weight gain and select animals with an optimal lanceolate morphology and growth rate.

**Supplementary Materials:** The following are available online at www.mdpi.com/2076-2615/11/5/1206/s1, Figure S1: Phenotypic data for growth traits (weight and SL), heights (BHP, BMH and CPH), height ratios (BMH/BHP and BMH/CPH) and ellipticity at 400 (A) and 800 d (B). Data are shown by EB and gender. The number (*n*) of soles evaluated in each batch is indicated below weight 400 d. The statistical significance of the three fixed factors (gender (G), evaluation batch (EB) and amoebic disease (AD)) is boxed in each trait. The letters denote significant differences between EB. The asterisks indicate significant differences between gender when interaction G × EB is significant; Figure S2: Regressions of log standard length (Log SL) on log weight (Log W) for males and females at 400 and 800 d. The slope and the determination coefficient are indicated; Figure S3: Regressions of BHP on BMH (A,B) and CPH on BMH (C,D) for males and females at 400 (A,C) and 800 d (B/D). The regression equations and determination coefficients are indicated; Figure S4: Estimated marginal means for BMH/BHP on BMH/CPH ratios as determined by repeated-measures ANOVA at 400 and 800 d for the (A) gender (male; female), (B) EB (1–7) and (C) AD (infected vs. non-infected). The asterisks on the horizontal or vertical arrows denote if within- or between-subject effects were significant; Figure S5: Comparison of sole shape with a low (A) vs. high (B) BMH/CPH ratio. The weight (W) and Ellipticity (E) (top) and the BMH/CPH (left down) and BMH/BHP (right down)

ratios are indicated. The theoretical ellipse that fits the mouth tip and the middle of peduncle with the BMH is shown in red dashed line.

**Author Contributions:** Conceptualization, M.M., M.G.C. and R.Z.; methodology, M.M.; formal analysis, I.G.-C. and M.M.; investigation, I.G.-C., E.J.-F., E.E. and C.B.; resources, R.Z. and E.J.-F.; data curation, I.G.-C. and M.M.; writing—original draft preparation, M.M.; writing—review and editing, M.M., I.G.-C. M.G.C. and R.Z.; supervision, M.M., M.G.C. and R.Z.; funding acquisition, M.M., M.G.C. and R.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by MCIU/AEI/FEDER, UE, grant numbers RTA2017-00054-C03-01 and RTA2017-00054-C03-03. Moreover, this research was funded by the EU H2020 research and innovation program under grant agreement 817992 ERANET-BLUEBIO COFUND project PCI2020-111994 "BestBrood". IGC is funded by a predoctoral fellowship from INIA.

**Institutional Review Board Statement:** All procedures were authorized by the Bioethics and Animal Welfare Committee of IFAPA and given the registration number 10/06/2016/101 by the National Authorities for Regulation of Animal Care and Experimentation.

**Data Availability Statement:** The dataset supporting this article has been uploaded as part of the electronic Supplementary Materials.

**Acknowledgments:** We thank CUPIMAR for providing all fish for the analysis.

**Conflicts of Interest:** The authors declare no conflict of interest. The founders had no role in the design of the study, in the collection, analyses or interpretation of data, in the writing of the manuscript or in the decision to publish the result.

## References

1. Akkaynak, D.; Siemann, L.A.; Barbosa, A.; Mathger, L.M. Changeable camouflage: How well can flounder resemble the colour and spatial scale of substrates in their natural habitats? *R. Soc. Open Sci.* **2017**, *4*, 160824.
2. Black, C.R.; Berendzen, P.B. Shared ecological traits influence shape of the skeleton in flatfishes (Pleuronectiformes). *Peer.J.* **2020**, *8*, e8919.
3. Reinders, M.J.; Banovi, M.; Guerrero, L.; Krystallis, A. Consumer perceptions of farmed fish: A cross-national segmentation in five European countries. *Br. Food J.* **2016**, *118*, 2581–2597.
4. de Azevedo, A.M.; Losada, A.P.; Barreiro, A.; Barreiro, J.D.; Ferreiro, I.; Riaza, A.; Vazquez, S.; Quiroga, M.I. Skeletal anomalies in reared Senegalese sole *Solea senegalensis* juveniles: A radiographic approach. *Dis. Aquat. Organ.* **2017**, *124*, 117–129.
5. Fernandez, I.; Ortiz-Delgado, J.B.; Darias, M.J.; Hontoria, F.; Andree, K.B.; Manchado, M.; Sarasquete, C.; Gisbert, E. Vitamin A affects flatfish development in a thyroid hormone signaling and metamorphic stage dependent manner. *Front. Physiol.* **2017**, *8*, 458.
6. de Azevedo, A.M.; Losada, A.P.; Barreiro, A.; Vazquez, S.; Quiroga, M.I. Skeletal anomalies in senegalese sole (*Solea senegalensis*), an anosteocytic boned flatfish species. *Vet. Pathol.* **2019**, *56*, 307–316.
7. Dionisio, G.; Campos, C.; Valente, L.M.P.; Conceicao, L.E.C.; Cancela, M.L.; Gavaia, P.J. Effect of egg incubation temperature on the occurrence of skeletal deformities in *Solea senegalensis*. *J. Appl. Ichthyol.* **2012**, *28*, 471–476.
8. Losada, A.P.; de Azevedo, A.M.; Barreiro, A.; Barreiro, J.D.; Ferreiro, I.; Riaza, A.; Quiroga, M.I.; Vazquez, S. Skeletal malformations in Senegalese sole (*Solea senegalensis* Kaup, 1858): Gross morphology and radiographic correlation. *J. Appl. Ichthyol.* **2014**, *30*, 804–808.
9. Gavaia, P.; Domingues, S.; Engrola, S.; Drake, P.; Sarasquete, C.; Dinis, M.T.; Cancela, M.L. Comparing skeletal development of wild and hatchery-reared Senegalese sole (*Solea senegalensis*, Kaup 1858): Evaluation in larval and postlarval stages. *Aquac. Res.* **2009**, *40*, 1585–1593.
10. Fernández, I.; Pimentel, M.S.; Ortiz Delgado, J.B.; Hontoria, F.; Sarasquete, C.; Estévez, A.; Zambonino Infante, J.L.; Gisbert, E. Effect of dietary vitamin A on Senegalese sole (*Solea senegalensis*) skeletogenesis and larval quality. *Aquaculture* **2009**, *295*, 250–265.
11. Ambrosio, P.P.; Costa, C.; Sánchez, P.; Flos, R. Stocking density and its influence on shape of Senegalese sole adults. *Aquac. Int.* **2008**, *16*, 333–343.
12. Blonk, R.J.; Komen, H.; Tenghe, A.; Kamstra, A.; van Arendonk, J.A.M. Heritability of shape in common sole, *Solea solea*, estimated from image analysis data. *Aquaculture* **2010**, *307*, 6–11.
13. Guerrero-Cozar, I.; Jimenez-Fernandez, E.; Berbel, C.; Cordoba-Caballero, J.; Claros, M.G.; Zerolo, R.; Manchado, M. Genetic parameter estimates and identification of SNPs associated with growth traits in Senegalese sole. *Aquaculture* **2021**, *539*, 736665.
14. Martin, I.; Rasines, I.; Gomez, M.; Rodriguez, C.; Martinez, P.; Chereguini, O. Evolution of egg production and parental contribution in Senegalese sole, *Solea senegalensis*, during four consecutive spawning seasons. *Aquaculture* **2014**, *424-425*, 45–52.
15. Fatsini, E.; Gonzalez, W.; Ibarra-Zatarain, Z.; Napuchi, J.; Duncan, N. The presence of wild Senegalese sole breeders improves courtship and reproductive success in cultured conspecifics. *Aquaculture* **2020**, *519*, 734922.

16. Cañavate, J.P.; Fernandez-Diaz, C. Influence of co-feeding larvae with live and inert diets on weaning the sole *Solea senegalensis* onto commercial dry feeds. *Aquaculture* **1999**, *174*, 255–263.
17. Roman-Padilla, J.; Rodríguez-Rua, A.; Ponce, M.; Manchado, M.; Hachero-Cruzado, I. Effects of dietary lipid profile on larval performance and lipid management in Senegalese sole. *Aquaculture* **2017**, *468*, 80–93.
18. Carballo, C.; Berbel, C.; Guerrero-Cozar, I.; Jimenez-Fernandez, E.; Cousin, X.; Bégout, M.L.; Manchado, M. Evaluation of different tags on survival, growth and stress response in the flatfish Senegalese sole. *Aquaculture* **2018**, *494*, 10–18.
19. Navarro, A.; Lee-Montero, I.; Santana, D.; Henríquez, P.; Ferrer, M.A.; Morales, A.; Soula, M.; Badilla, R.; Negrin-Baez, D.; Zamorano, M.J.; et al. IMAFISH_ML: A fully-automated image analysis software for assessing fish morphometric traits on gilthead seabream (*Sparus aurata* L.), meagre (*Argyrosomus regius*) and red porgy (*Pagrus pagrus*). *Comput. Electron. Agr.* **2016**, *121*, 66–73.
20. Guerrero-Cozar, I.; Perez-Garcia, C.; Benzekri, H.; Sánchez, J.; Seoane, P.; Cruz, F.; Gut, M.; Zamorano, M.J.; Claros, M.G.; Manchado, M. Development of whole-genome multiplex assays and construction of an integrated genetic map using SSR markers in Senegalese sole. *Sci. Rep.* **2020**, *10*, 21905.
21. Vandeputte, M.; Mauger, S.; Dupont-Nivet, M. An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. *Mol. Ecol. Notes* **2006**, *6*, 265–267.
22. Meyer, K. WOMBAT: A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J. Zhejiang Univ. Sci. B.* **2007**, *8*, 815–821.
23. Martin, I.; Carazo, I.; Rasines, I.; Rodriguez, C.; Fernandez, R.; Martinez, P.; Norambuena, P.; Chereguini, O.; Duncan, N. Reproductive performance of captive Senegalese sole, *Solea senegalensis*, according to the origin (wild or cultured) and gender. *Span. J. Agric. Res.* **2019**, *17*, 2171–9292.
24. Sanchez, P.; Ambrosio, P.P.; Flos, R. Stocking density and sex influence individual growth of Senegalese sole (*Solea senegalensis*). *Aquaculture* **2010**, *300*, 93–101.
25. Constenla, M.; Padros, F. Histopathological and ultrastructural studies on a novel pathological condition in *Solea senegalensis*. *Dis. Aquat. Organ.* **2010**, *90*, 191–196.
26. Cadrin, S.X.; Silva, V.M. Morphometric variation of yellowtail flounder. *ICES J. Mar. Sci.* **2005**, *62*, 683–694.
27. Shikano, T. Quantitative genetic parameters for growth-related and morphometric traits of hatchery-produced Japanese founder *Paralichthys olivaceus* in the wild. *Aquac. Res.* **2007**, *38*, 1248–1253.
28. Li, Y.; Zhang, B.; Yang, Y.; Chen, S. Estimation of genetic parameters for juvenile growth performance traits in olive flounder (*Paralichthys olivaceus*). *Aquac. Fish* **2019**, *4*, 48–52.
29. Liu, Y.-X.; Wang, G.-X.; Wang, Y.-F.; Si, F.; Sun, Z.-H.; Zhang, X.-Y.; Wang, J.-D.; Liu, H.-J. Estimation of genetic parameters for growth traits of Japanese flounder *Paralichthys olivaceus* using an animal model. *Fish Sci.* **2011**, *77*, 87–93.
30. Liu, F.; Chen, S.L.; Wang, L.; Zhang, Y.P.; Tian, Y.S.; Chen, H.L. Estimation of genetic parameters for growth related traits at different stages of development in *Paralichthys olivaceus* (Temminck & Schlegel, 1846). *Indian J. Fish* **2016**, *63*, 70–75.
31. Blonk, R.J.; Komen, H.; Kamstra, A.; van Arendonk, J.A. Estimating breeding values with molecular relatedness and reconstructed pedigrees in natural mating populations of common sole, *Solea solea*. *Genetics* **2010**, *184*, 213–219.
32. Imsland, A.K.; Foss, A.; Conceiçao, L.E.C.; Dinis, M.T.; Delbare, D.; Schram, E.; Kamstra, A.; Rema, P.; White, P. A review of the culture potential of *Solea solea* and *S. senegalensis*. *Rev. Fish Biol. Fish* **2003**, *13*, 379–407.
33. Salas-Leiton, E.; Anguis, V.; Martin-Antonio, B.; Crespo, D.; Planas, J.V.; Infante, C.; Canavate, J.P.; Manchado, M. Effects of stocking density and feed ration on growth and gene expression in the Senegalese sole (*Solea senegalensis*): Potential effects on the immune response. *Fish Shellfish Immunol.* **2010**, *28*, 296–302.
34. Salas-Leiton, E.; Anguis, V.; Rodriguez-Rua, A.; Cañavate, J.P. Stocking homogeneous size groups does not improve growth performance of Senegalese sole (*Solea senegalensis*, Kaup 1858) juveniles: Individual growth related to fish size. *Aquac. Eng.* **2010**, *43*, 108–113.
35. Mas-Muñoz, J.; Blonk, R.J.; Schrama, J.W.; van Arendonk, J.; Komen, H. Genotype by environment interaction for growth of sole (*Solea solea*) reared in an intensive aquaculture system and in a semi-natural environment. *Aquaculture* **2013**, *410*, 230–235.
36. Dupont-Nivet, M.; Chevassus, B.; Mauger, S.; Haffray, P.; Vandeputte, M. Side effects of sexual maturation on heritability estimates in rainbow trout (*Oncorhynchus mykiss*). *Aquac. Res.* **2010**, *41*, e878–e880.
37. Omasaki, S.K.; Charo-Karisa, H.; Kahi, A.K.; Komen, H. Genotype by environment interaction for harvest weight, growth rateand shape between monosex and mixed sex Nile tilapia (*Oreochromis niloticus*). *Aquaculture* **2016**, *458*, 75–81.
38. Mengistu, S.B.; Mulder, H.A.; Benzie, J.A.H.; Khaw, H.L.; Megens, H.; Trinh, T.Q.; Komen, H. Genotype by environment interaction between aerated and non-aerated ponds and the impact of aeration on genetic parameters in Nile tilapia (*Oreochromis niloticus*). *Aquaculture* **2020**, *529*, 735704.
39. Assumpção, L.; Makrakis, M.C.; Makrakis, S.; Wagner, R.; Silva, P.S.; Lima, A.F.; Kashiwaqui, E.A.L. The use of morphometric analysis to predict the swimming efficiency of two Neotropical long-distance migratory species in fish passage. *Neotrop. Ichthyol.* **2012**, *10*, 797–804.
40. Fisher, R.; Hogan, J.D. Morphological predictors of swimming speed: A case study of pre-settlement juvenile coral reef fishes. *J. Exp. Biol.* **2007**, *210*, 2436–2443.