

DETECCIÓN DE MARCADORES LIGADOS AL SEXO EN *Gambusia* A PARTIR DE RAD-seq

Estudiant: Silvia Cobelo Gómez

Grau en Biotecnologia

Correu electrònic: silviacobelog@gmail.com

Tutor: Jose Luis García Marín

Cotutor*:

Empresa / institució: Universitat de Girona

Vistiplau tutor (i cotutor*):

Nom del tutor: Jose Luis García Marín

Nom del cotutor*:

Empresa / institució: Universitat de Girona

Correu(s) electrònic(s): joseluis.garcia@udg.edu

*si hi ha un cotutor assignat

Data de dipòsit de la memòria a secretaria de coordinació:

RESUMEN

La obtención de genes representativos de la determinación sexual en peces continúa siendo hoy en día un aspecto poco conocido, en gran parte por los diversos sistemas cromosómicos dentro de un mismo género e incluso especie. La *Gambusia holbrooki* es un pez de la familia de los poecílidos, ecológicamente interesante como una de las especies más invasoras del mundo. En el presente trabajo se pretende buscar un marcador molecular para así distinguir el sexo de *Gambusia holbrooki*, y con el fin de aportar nuevos conocimientos sobre la determinación sexual en peces. Con este fin se utilizó la técnica RAD-seq (restriction site-associated DNA sequencing), por tal de identificar miles de marcadores genéticos (RAD-tags) con variaciones en nucleótidos únicos (SNPs) entre machos y hembras. A partir de 44271 RAD-tags disponibles en la base de datos STACKS se realizó la búsqueda de loci exclusivos de un sexo y loci con un patrón diferencial de SNPs entre los machos y hembras de *Gambusia holbrooki*. Para la detección de estos loci RAD-tags diferenciadores se emplearon diversos programas bioinformáticos basados en el estadístico F_{ST} , así como un posterior análisis funcional de los posibles genes candidatos. Ambas búsquedas concluyeron que ninguno de los loci RAD-tags del estudio era específico para diferenciar entre machos y hembras de la especie *Gambusia holbrooki*, aunque algunos SNPs si parecen tener diferencias alélicas entre los sexos.

Palabras clave: Determinación sexual, Gambusia holbrooki, RAD-tags, SNPs, F_{ST} outlier.

RESUM

L'obtenció de gens representatius de la determinació sexual en peixos continua essent avui en dia un aspecte poc conegut, en major part degut a la diversitat de sistemes cromosòmics dins d'un mateix gènere o inclús espècie. La *Gambusia holbrooki* és un peix de la família dels poecílids, ecològicament interessant com una de les espècies més invasores del món. En el present treball, es pretén buscar un marcador molecular per a distingir el sexe de *Gambusia holbrooki*, i amb el fi d'aportar nous coneixements sobre la determinació sexual en peixos. Amb aquest fi es va utilitzar la tècnica RAD-seq (restriction site-associated DNA sequencing), per tal d'identificar milers de marcadors genètics (RAD-tags) amb variacions en nucleòtids únics (SNPs) entre mascles i femelles. A partir de 44271 RAD-tags disponibles en la base de dades STACKS es va realitzar la cerca de loci exclusius d'un sexe i loci amb un patró diferencial de SNPs entre els mascles i femelles de *Gambusia holbrooki*. Per a la detecció d'aquests loci RAD-tags diferenciadors es van emprar diversos programes bioinformàtics basats en l'estadístic F_{ST} , així com un posterior anàlisi funcional dels possibles gens candidats. Ambdues cerques van concloure que cap dels loci RAD-tags de l'estudi era específic per a diferenciar entre mascles i femelles de l'espècie *Gambusia holbrooki*, tot i que alguns SNPs si que semblen presentar diferències al·lèliques entre sexes.

Paraules clau: Determinació sexual, Gambusia holbrooki, RAD-tags, SNPs, F_{ST} outlier.

ABSTRACT

The obtaining of representative genes of the sex determination in fish remains nowadays a little-known aspect, to a large extent by the various chromosomal systems within a same genus and even species. *Gambusia holbrooki* is a fish of the poeciliids' family, ecologically interesting as one of the most invasive species in the world. The present work intends to find a molecular marker to distinguish the sex of *Gambusia holbrooki*, and in order to provide new knowledge on the sex determination in fish. To this end, the RAD-seq (restriction site-associated DNA sequencing) technique was used to identify thousands of genetic markers (RAD-tags) with variations in single nucleotides (SNPs) between males and females. From 44271 RAD-tags available in the STACKS database, the search for loci exclusive of a sex and loci with a differential pattern of SNPs was carried out among males and females of *Gambusia holbrooki*. For the detection of these differentiating RAD-tag loci, various bioinformatic programs based on the F_{ST} statistic were used, as well as a subsequent functional analysis of the possible candidate genes. Both searches concluded that none of the RAD-tags loci of the study was specific to differentiate between males and females of the species *Gambusia holbrooki*.

Keywords: Sex determination, Gambusia holbrooki, RAD-tags, SNPs, F_{ST} outlier.

AGRADECIMIENTOS

Gracias a todas las personas que me han ayudado a hacer posible el desarrollo de este trabajo, tanto familiares como amigos, y en especial a mi tutor, por haberme dado la oportunidad de trabajar con él.

ÍNDICE

1. INTRODUCCIÓN	1
2. OBJETIVES	3
3. METODOLOGÍA	3
3.1 Origen del material biológico y obtención de RAD-tags.....	3
3.2 Selección de RAD-tags para el estudio	5
3.3 Búsqueda de RAD-tags candidatos	5
3.3.1 RAD-tags expresados en un solo sexo.....	5
3.3.2 Loci outliers.....	6
3.4 Anotación de los RAD-tags	6
3.4.1 Anotación de los loci candidatos	6
3.4.2 Anotación funcional	7
3.5 Criterios éticos y sostenibles del estudio	8
4. RESULTADOS.....	8
4.1 Selección de loci candidatos.....	8
4.1.1 Loci exclusivos de un sexo	8
4.1.2 Loci outliers.....	9
4.2 Genes implicados en la determinación del sexo	10
4.2.1 Anotación de genes candidatos	10
4.2.2 Funcionalidad	13
5. DISCUSIÓN	15
6. CONCLUSIONS	16
7. BIBLIOGRAFÍA.....	17

1. INTRODUCCIÓN

La *Gambusia holbrooki* es un pez de la familia de los poecílidos en el orden de los ciprinodóntidos, la cual habita en aguas poco profundas y lentas, soportando elevadas temperaturas, así como un elevado índice de contaminación. Nativa de Estados Unidos, fue introducida a principios del siglo XX en otros países como España e Italia para controlar los mosquitos que provocaban la malaria, lo cual ha alterado el equilibrio biológico del agua, al ser uno de los peces más invasores del mundo¹. La *Gambusia* es una especie ovovípara de fecundación interna, que realiza varias puestas durante el período reproductor. Alcanza la madurez sobre las seis semanas de vida, siendo posible diferenciar fenotípicamente el sexo de los ejemplares adultos, ya que los machos poseen gonopodio, una modificación de la aleta anal que sirve como aparato de inseminación, y las hembras una mancha oscura en el vientre².

La determinación del sexo en la mayoría de los peces no siempre se rige por un patrón fenotípico distintivo, especialmente en embriones y alevines, siendo necesaria una determinación genotípica del sexo del individuo. Se conoce que a diferencia de los mamíferos que presentan polimorfismos cromosómicos XX-XY (♀-♂) y las aves que son WZ-ZZ (♀-♂), en los peces puede haber ambos sistemas dentro del mismo género o incluso en la misma especie³. Además, mientras que comúnmente los mamíferos poseen el gen determinante del sexo SRY (Sex-determining Region Y), propio de los machos, el cual codifica para la proteína TDF (Testis-determining Factor) y se localiza en el cromosoma Y⁴, en los peces no se ha encontrado ningún gen que pueda ser exclusivo de uno de los dos sexos. En el año 2002, Matsuda *et al.*⁵ observaron que en el pez *Oryzias latipes* el gen DMY (DM-domain gene on the Y chromosome), estaba implicado en la determinación sexual de los machos y por ello podía ser un gen candidato en peces⁶. Sin embargo, este gen no parece estar ampliamente distribuido en todas las familias de peces, por lo que se cree que diversos genes podrían estar implicados. En la especie *Takifugu rubripes* un polimorfismo de SNP (C/G) en el gen Amhr2 (anti-Mullerian hormone receptor type 2) parece ser el único asociado con el sexo fenotípico, mientras que las hembras son homocigotas C/C los machos son heterocigotas C/G⁷. Recientemente, se ha demostrado que en *Gambusia affinis* el marcador específico Gaf88, relacionado con el gen aminometil transferasa, se

encuentra asociado al fenotipo de las hembras. No obstante, esta asociación no se detectó en *Gambusia holbrooki*, probablemente debido a los distintos sistemas cromosómicos, ZW/ZZ en *G. affinis* y XX/XY (♀-♂) en *G. holbrooki*⁸.

Estudios recientes han examinado el número y la ubicación de genes implicados en la determinación del sexo mediante métodos de genotipado para muchos marcadores, que han permitido detectar variaciones en nucleótidos únicos (SNPs) a lo largo de todo el genoma^{7 8}. Actualmente la técnica de RAD-seq es una de las más utilizadas con el fin de identificar y puntuar miles de marcadores genéticos distribuidos al azar a través del genoma. Se basa en un método de secuenciación de DNA que combina el uso de enzimas de restricción con identificadores moleculares, pudiendo así detectar múltiples SNP dentro de un sólo RAD-tag (secuencias de lectura corta adyacentes a los sitios de corte con enzimas de restricción)⁹. Además, esta técnica permite la comparación con un genoma de referencia, así como la distinción entre homocigotos y heterocigotos; pero posee ciertas limitaciones como por ejemplo un muestreo incompleto del genoma debido a la distribución no uniforme de los RAD-tags y la potencial pérdida de SNP adaptativos¹⁰. Pese a estas limitaciones el método RAD-seq aporta una gran cantidad de SNPs a lo largo del genoma de estudio, siendo de gran utilidad para comparar la diferenciación entre grupos de individuos¹¹.

En el presente estudio, a partir de los RAD-tags de *Gambusia holbrooki* obtenidos previamente por el Laboratorio d'Ictologia Genètica de la UdG¹² y la utilización de diversos softwares especializados, se plantea la identificación de RAD-tags con un patrón de variación diferencial de SNPs entre machos y hembras. El estadístico F_{ST} mide las diferencias genéticas a través de desigualdades en las frecuencias alélicas entre grupos, donde una $F_{ST}=0$ indica que las frecuencias alélicas son idénticas en ambos grupos y, por tanto, no hay diferenciación entre ellas, mientras que una $F_{ST}=1$ implica que las frecuencias alélicas son completamente distintas y en cada grupo están fijados alelos diferentes¹³. Así pues, la detección de loci diferenciadores se basa en la aproximación de aquellos que presenten un valor de F_{ST} *outlier*, es decir, loci cuyo coeficiente F_{ST} es mayor o menor que el identificado como neutral, considerando en nuestro caso únicamente loci con una diferenciación genética mayor que los loci neutrales¹⁴. Mediante estos métodos se pretende buscar una discordancia entre ambos sexos de *Gambusia holbrooki* que permita la obtención de algún gen representativo de

la determinación sexual en esta especie, aportando utilidad al limitado conocimiento sobre este aspecto en peces, así como al control de la invasión, donde se ha postulado que la utilización de métodos recombinantes dirigidos hacia un gen involucrado en la sexualidad puede esterilizar, matar o cambiar el sexo de las hembras logrando el control e incluso la erradicación de la especie invasora¹⁵.

2. OBJETIVES

The main aim of the present study is to identify loci related to sex determination in *Gambusia holbrooki*, a highly invasive species belonging to the family of poeciliids, in which despite the phenotype sex differences, has not yet been identified any gen that could be involved in sex determination.

To achieve this goal, differential variations between males and females across loci will be identified by means of bioinformatic analyses on available RAD-tag markers, and the subsequent functional analysis of the identified genes.

3. METODOLOGÍA

3.1 Origen del material biológico y obtención de RAD-tags

Los RAD-tags empleados en este estudio se recopilaron del catálogo correspondiente a la especie *Gambusia holbrooki*, obtenidos para el trabajo de Vera *et al*¹². Estos se elaboraron a partir de 93 individuos de *Gambusia holbrooki*, pertenecientes a 8 localidades incluyendo territorios donde la especie es nativa y otros donde se ha introducido. Del total de individuos se identificaron 53 hembras y 40 machos los cuales se encontraban repartidos casi por igual en todas las poblaciones de estudio. Sin embargo, de la población Florence no hay machos entre los individuos muestreados, por lo que posiblemente los cambios en esta localidad pueden dar un sesgo hacia las hembras (**Tabla 1**).

Tabla 1. Clasificación territorial y separación por sexo de las muestras empleadas. Sitio de muestreo, código de la población, número total individuos (N), número de hembras (♀) y machos (♂).

Sitio de muestreo	Código	N	♀	♂
Everglades (USA)	EV	11	6	5
Florence (USA)	FLO	11	11	-
Potomac (USA)	PO	11	6	5
Coltano (Italia)	CO	12	6	6
Besalú (España)	BL	12	6	6
Sant Pere Pescador (España)	PP	12	6	6
Vilanova de Muga (España)	VM	12	6	6
Empúria-Brava (España)	EP	12	6	6

De cada uno de los individuos el DNA se extrajo del tejido muscular y se envió a FLORAGENEX, donde se generaron y secuenciaron los RAD-tags. Los adaptadores de secuenciación y los códigos de barras individuales (identificadores moleculares) se ligaron a fragmentos de DNA genómico obtenidos mediante la digestión con el enzima de restricción SbfI que genera extremos cohesivos. Los fragmentos resultantes fueron secuenciados mediante la plataforma Illumina HiSeq2000 gracias a los adaptadores específicos que solamente permiten la unión y amplificación a partir de los sitios de restricción. Las lecturas se separaron por individuo mediante los códigos de barras específicos, los cuales posteriormente se eliminaron, obteniendo como resultado final RAD-tags con una longitud de 90pb. Debido a que el genoma de *Gambusia holbrooki* no está disponible, el individuo VM43 recogido en Vilanova de Muga, el cual resultó ser una hembra, se usó como modelo por la alta calidad de sus secuencias. Así pues, alineando las lecturas de secuencia de este ejemplar modelo con las obtenidas en bruto para el resto de las muestras, se pudieron identificar 44271 RAD-tags. Finalmente, los RAD-tags proporcionados para cada individuo por FLORAGENEX se catalogaron y almacenaron en una base datos mediante el programa STACKS 1.46^{16 17 18}, a los que se ha accedido para la realización de este trabajo.

3.2 Selección de RAD-tags para el estudio

Con el fin de exportar los 44271 RAD-tags contenidos en STACKS para posteriores análisis se utilizó el programa interno de STACKS *Populations*. Se realizó una primera exportación a un fichero CSV que contenía los haplotipos de aquellos loci presentes al menos en un 90% de los individuos de un sexo. Posteriormente se realizó una segunda exportación de los SNPs en los RAD-tags a un fichero GENEPOP^{19 20}, un formato con facilidad de conversión y común en gran parte de los programas de genética de poblaciones. Para esta exportación se tuvieron en cuenta ciertas restricciones, así se consideró que cada locus como mínimo debería estar en uno de los sexos y al menos en un 90% de los individuos y que la profundidad, es decir, el número de lecturas para considerar un SNP válido, fuese como mínimo 10. Finalmente, para cada RAD-tag solo se conservó un único SNP de forma aleatoria en el fichero GENEPOP.

3.3 Búsqueda de RAD-tags candidatos

3.3.1 RAD-tags expresados en un solo sexo

A partir del archivo CSV que contenía los haplotipos para cada RAD-tag, se realizó un filtrado manual escogiendo aquellos loci que se encontraban en menos de 53 individuos ($\approx 90\%$ de las hembras) y más de 36 (90% de los machos). Por tal de saber si esos RAD-tags representaban loci presentes en solo uno de los dos sexos y podían por tanto estar relacionados con la determinación del sexo. Con estos loci se creó una White list para una nueva extracción de SNPs mediante el subprograma *Populations*, en este caso sin restricción al 90%. Este nuevo archivo exportado contenía los SNPs de los loci presentes entre 53 y 36 individuos, y mediante inspección visual se comprobó si en alguno de ellos se mostraba alguna variante nucleotídica únicamente en hembras y no en machos o viceversa.

3.3.2 *Loci outliers*

Teniendo en cuenta el archivo exportado en formato GENEPOP se llevó a cabo el análisis para la determinación de los loci con un patrón de variación diferencial de SNPs entre los machos y hembras de la especie *Gambusia holbrooki*. Se decidió utilizar simultáneamente los programas BayeScan²¹ y Arlequin 3.5.2.2²², ya que ambos permiten el análisis de F_{ST} *outliers* en marcadores codominantes, siendo necesaria la conversión del archivo GENEPOP mediante PDGspider²³ al formato adecuado en cada caso para la realización del análisis. En el caso de BayeScan se emplearon los parámetros preestablecidos y mediante el R software²⁴ y la función plot R, se obtuvo también la representación gráfica de los resultados. Paralelamente se ejecutó el programa Arlequin mediante la opción “Detecting loci under selection from F-statistics” y el paquete R (Rcmd) para la posterior representación gráfica de los resultados. Los loci detectados con un patrón de variación diferencial de SNPs entre los machos y las hembras se comprobaron de forma manual a través de la base de STACKS, por tal de verificar si presentaban una distribución nucleotídica distinta entre ambos sexos o simplemente presentaban una distribución diferencial entre las localidades, donde Florence podría sesgar el resultado.

3.4 *Anotación de los RAD-tags*

3.4.1 *Anotación de los loci candidatos*

Para la detección de secuencias genómicas de peces óseos con regiones homólogas a las secuencias de los loci candidatos detectados en el análisis anterior, se utilizó el programa Blast2GO²⁵ mediante la interfaz BLAST (Basic Local Alignment Search Tool). Esta interfaz accede a la base de datos Nucleotide del NCBI (National Center for Biotechnology Information), donde se encuentra una colección de secuencias de varias fuentes, incluyendo GenBank, RefSeq, TPA y PDB. Las secuencias nucleotídicas de los RAD-tags candidatos necesarias para el BLAST se obtuvieron en formato FASTA a partir del archivo que contenía todos los RAD-tags. Para el BLAST se utilizó un “Word Size” de 90 y un “HSP Length Cutoff” de 75, es decir, se buscaron segmentos con una longitud de 90 pb iguales a la secuencia de consulta y con una

alineación local sin espacios lo más exacta posible para las siguientes 75 pb. Por tal de ampliar la búsqueda de genes candidatos se decidió mantener el parámetro de “HSP Length Cutoff”, pero disminuir el “Word Size”. Este juicio se realizó teniendo en cuenta que la búsqueda de BLAST comienza con encontrar una secuencia que coincida perfectamente con la longitud dada por el “Word Size” y, por tanto, al disminuir este parámetro, aumentará el número de secuencias homólogas detectadas²⁶. Con este fin se realizaron dos nuevos BLAST con un “Word Size” de 45 y 30. Cabe señalar que se decidió parar en un “Word Size” de 30, ya que respecto a los valores superiores no se apreció un aumento notorio en el número de resultados del análisis. Por tanto, teniendo en cuenta el BLAST realizado con un “Word Size” de 45 y 30 se consideraron como posibles candidatos aquellos RAD-tags con una similitud (Sim) y un porcentaje de alineamientos locales de alta puntuación (Hsp/Query) $\geq 90\%$ respecto a su secuencia. Por medio del programa GENEPOP se determinó en machos y hembras la frecuencia de nucleótidos del SNP de los RAD-tags candidatos obtenidos del análisis de BLAST.

3.4.2 Anotación funcional

Mediante el programa David Bioinformatics Resources 6.8^{27 28} se obtuvo la función de los genes de aquellos RAD-tags considerados como candidatos en el análisis anterior, por estar posiblemente relacionados con la determinación del sexo en *Gambusia holbrooki*. Por tal de realizar el análisis se utilizó la anotación HGNC²⁹ (*HUGO (Human Genome Organisation) Gene Nomenclature Conmmittee*) que utiliza símbolos y nombres únicos para loci humanos, seleccionando así pues la especie *Homo sapiens* como especie de referencia. De los resultados obtenidos se analizaron aquellos presentes en todos los niveles para funciones relacionadas con procesos biológicos (BP), componentes celulares (CC) y funciones moleculares (MF).

3.5 Criterios éticos y sostenibles del estudio

En el estudio de los marcadores ligados al sexo de *Gambusia holbrooki* a partir de RAD-seq, aunque los análisis realizados fueron bioinformáticos, por tal de conseguir el material de trabajo, fue necesario que previamente se sacrificaran 93 individuos de esta especie. A pesar de no ser una especie protegida sino catalogada como invasora, lo cual aporta menos restricciones a la hora de trabajar, se debería considerar la realización de biopsias como método alternativo por tal de mejorar la ética y sostenibilidad de estudios basados en DNA sobre esta especie.

4. RESULTADOS

4.1 Selección de loci candidatos

4.1.1 Loci exclusivos de un sexo

En el estudio de la determinación del sexo en *Gambusia holbrooki* se partió de un total de 44271 loci presentes en la base de datos STACKS mediante el subprograma *Populations*. Sin embargo, el archivo con los haplotipos para cada RAD-tag obtenido de la primera exportación contenía un total de 38518 loci, de los cuales 1879 se encontraban en el rango de 36-53 individuos. De estos 1879 loci se determinó que 233 no se encontraban en el 90% de los machos y por el contrario 1646 no estaban presentes en el 90% de las hembras. La comprobación del nuevo archivo exportado para los 1879 loci dentro del rango, demostró que ninguna variante era exclusiva de un único sexo y por tanto ningún RAD-tag parece estar relacionado con la determinación del sexo en *Gambusia holbrooki*. Probablemente se trata de loci RAD-tags con bajo rendimiento y que se muestran ausentes en diversos ejemplares.

4.1.2 Loci outliers

A partir de la segunda exportación se obtuvieron 32458 loci, con un único SNP de forma aleatoria por locus, que cumplían las restricciones preestablecidas. En el caso del cálculo de F_{ST} outliers con el programa BayeScan, todos los loci presentaban un valor de F_{ST} entre 0,22 y 0,25 fueron detectados, no habiendo ninguno que tuviera un valor claramente outlier.

El programa Arlequin trabaja exclusivamente con loci que tengan datos para todos los grupos comparados, por eso eliminó 1685 RAD-tags presentes en menos del 90% de ejemplares de uno de los dos sexos y que no habían sido exportados para ese sexo por el subprograma *Populations* de STACKS. Por consiguiente, el análisis se hizo sobre un total de 30773 loci, detectando 904 loci que presentaban una elevada F_{ST} con un p-value inferior a 0,05, indicativo de una posible diferenciación entre los sexos (**Figura 1**). De estos 904 loci, 809 no se consideraron implicados en la determinación del sexo ya que o bien eran prácticamente monomórficos en las poblaciones europeas (733) o presentaban polimorfismos en poblaciones concretas indistintamente del sexo del individuo (76). Sin embargo, los otros 95 loci, 6 con un valor de p-value < 0,01 y 89 con valor < 0,05, podrían ser significativos ya que se encontraban distribuidos en casi todas las poblaciones de *Gambusia holbrooki* y presentaban diferencias significativas entre los sexos.

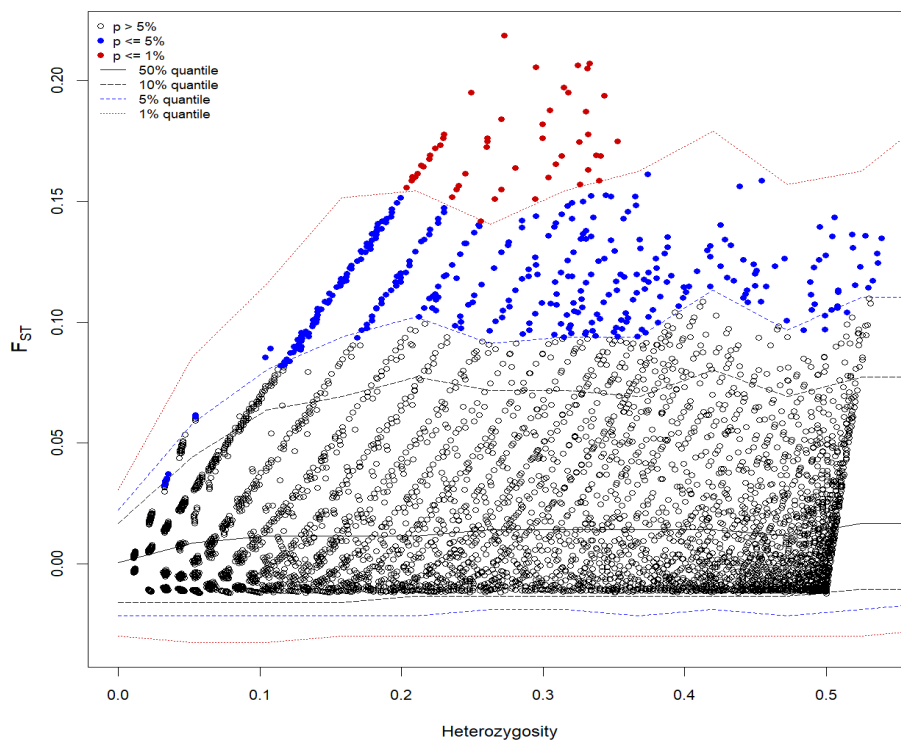


Figura 1. Detección de loci bajo selección a partir de escaneos de genoma basados en F_{ST} mediante el software Arlequin.

4.2 Genes implicados en la determinación del sexo

4.2.1 Anotación de genes candidatos

Los resultados del BLAST para los anteriores 95 RAD-tags candidatos muestran al menos una coincidencia únicamente para 27 de ellos. Todos obtuvieron resultados para un “Word Size” de 30, entre los cuales 16 también lo obtuvieron para 45 y solo uno para un “Word Size” de 90. De estos 27 loci RAD-tags se consideraron relevantes solo los 18 loci que presentaban una similitud (Sim) y un porcentaje de alineamientos locales de alta puntuación ($Hsp/Query$) $\geq 90\%$. En los 18 casos la coincidencia fue con un gen descrito en otra especie de poecílido (**Tabla 2**).

Tabla 2. Resultados del programa Blast2GO. Word Size, Código RAD-tag, Similitud (Sim), Hsp/Query, Accession Number (Acc), Proteína, Código HGNC y especie.

Word Size	RAD-tag	Sim (%)	Hsp/Query (%)	Acc	Protein	HGNC	Especie
30,45	72	98,8	88,9				
30,45	741	97,8	100	XM_014468643	DexH-box helicase 29	DHX29	<i>Xiphophorus maculatus</i>
30	924	97,3	83,3				
30	1815	90,8	96,7	XM_014474682	Activator of 90Dka heat shock protein ATPase homolog 1-like	AHSA1	<i>Xiphophorus maculatus</i>
30,45,90	2461	100	100	XM_016666273	Vasopressin V2 receptor-like	AVPR2	<i>Poecilia formosa</i>
30,45	2643	96,7	100	XM_008410273	Acidic mammalian chitinase-like	CHIA	<i>Poecilia reticulata</i>
30	2889	92,2	100	XM_015031132	Neuron navigator 1	NAV1	<i>Poecilia latipinna</i>
30,45	4883	100	81,1				
30,45	8778	96,7	100	XM_008417353	Multiple epidermal growth factor-like domains protein 9	MEGF9	<i>Poecilia reticulata</i>
30,45	9289	96,7	100	XM_017303343	Gastrula zinc finger protein XICGF52.1-like		<i>Poecilia reticulata</i>
30	9925	100	33,3				
30	10664	100	34,4				
30,45	12802	97,8	100	XM_007541188	Fas activated serine/threonine kinase	FASTK	<i>Poecilia formosa</i>
30,45	17003	95,5	98,9	XM_023340579	Microfibril-associated glycoprotein 4-like	MFAP4	<i>Xiphophorus maculatus</i>
30	18438	97,7	47,8				
30,45	18701	100	66,7				
30	20362	97,5	90,0	XR_001427844	Uncharacterized		<i>Poecilia latipinna</i>
30	22688	96,7	100	XM_023333668	Ubiquitin conjugating enzyme E2 O	UBE2O	<i>Xiphophorus maculatus</i>
30,45	24181	96,7	100	XM_008399219	Protein phosphatase 2 regulatory subunit B''alpha	PPP2R3A	<i>Poecilia reticulata</i>
30,45	27081	96,5	95,6	XM_023335389	Interleukin-20 receptor subunit beta-like	IL20RB	<i>Xiphophorus maculatus</i>
30,45	27526	97,6	92,2	XM_023349363	MBT domain-containing protein 1-like	MBTD1	<i>Xiphophorus maculatus</i>
30,45	28371	96,6	97,8	XM_023349345	Ras-related protein Rab-26-like	RAB26	<i>Xiphophorus maculatus</i>
30	29597	100	33,3				
30,45	30504	98,9	100	XM_023340315	Uncharacterized		<i>Xiphophorus maculatus</i>
30,45	30553	100	96,7	XM_005812326	Clathrin heavy chain 1	CLHC1	<i>Xiphophorus maculatus</i>
30	36975	97,8	100	XM_023328147	FAT atypical cadherin 2	FAT2	<i>Xiphophorus maculatus</i>
30	37575	80,7	97,8				

Los cambios nucleotídicos en el SNP de los 18 loci considerados como significativos no incluyen ninguna fijación alélica diferencial entre los sexos, ni siquiera en el RAD-tag 24181 con un p-value < 0,01 (**Tabla 3**). En 6 RAD-tags (741, 2889, 20362, 24181, 27081, 36975) uno de los nucleótidos se encuentra fijado en los machos y las hembras presentan polimorfismo, mientras que en 4 RAD-tags (1815, 12803, 17003, 30504) se da el caso contrario. Las sustituciones nucleotídicas entre los sexos de estos 18 RAD-tags muestran una proporción semejante entre transiciones (10) y transversiones (8).

Tabla 3. Frecuencias representativas de los cambios nucleotídicos. Código RAD-tag, número SNPs, posición SNP analizado, nucleótidos, mutaciones, frecuencias en machos (♂) y hembras (♀).

RAD-tag	SNPs	Posición	Nucleótido	Mutación	♂	♀
741	3	59 (2° SNP)	A	Transición	0,000	0,113
			G		1,000	0,887
1815	3	49 (3° SNP)	G	Transversión	0,962	1,000
			T		0,038	0,000
2461	2	86 (2° SNP)	A	Transición	0,179	0,433
			G		0,821	0,567
2643	3	58 (1° SNP)	C	Transición	0,013	0,154
			T		0,988	0,846
2889	8	44 (4° SNP)	A	Transversión	1,000	0,877
			T		0,000	0,123
8778	2	21 (1° SNP)	A	Transversión	0,975	0,788
			C		0,025	0,212
9289	2	20 (1° SNP)	A	Transición	0,062	0,236
			G		0,938	0,764
12802	5	40 (4° SNP)	A	Transición	0,037	0,000
			G		0,963	1,000
17003	5	67 (2° SNP)	A	Transversión	0,962	1,000
			T		0,038	0,000
20362	2	17 (2° SNP)	A	Transición	0,000	0,123
			G		1,000	0,877
22688	4	21 (3° SNP)	C	Transversión	0,355	0,635
			G		0,645	0,365
24181	7	31 (4° SNP)	C	Transversión	1,000	0,802
			G		0,000	0,198

RAD-tag	SNPs	Posición	Nucleótido	Mutación	♂	♀
27081	8	60 (7° SNP)	A	Transversión	1,000	0,885
			T		0,000	0,115
27526	2	45 (2° SNP)	C	Transición	0,475	0,745
			T		0,525	0,255
28371	3	23 (1° SNP)	C	Transición	0,562	0,808
			T		0,438	0,192
30504	2	52 (2° SNP)	C	Transición	0,963	1,000
			T		0,037	0,000
30553	2	63 (2° SNP)	C	Transición	0,013	0,144
			T		0,988	0,856
36975	5	46 (4° SNP)	A	Transversión	0,000	0,104
			C		1,000	0,896

4.2.2 Funcionalidad

Para el análisis de la función de los genes de los 18 posibles loci relacionados con la determinación del sexo en *Gambusia holbrooki*, se omitieron las proteínas de los RAD-tags 20362 y 30504 al no estar caracterizadas, así como la Gastrula zinc finger protein XICGF52.1-like del RAD-tag 9289 al no encontrar una anotación adecuada para humanos (**Tabla 4**). Los resultados obtenidos en el análisis mostraron una coincidencia en 9 de los 15 genes para las funciones “single-multicellular organism process” y “multicellular organismal process”, relacionadas en ambos casos con cualquier proceso biológico (BP) que ocurre a nivel de un organismo multicelular. Sin embargo, las funciones relacionadas con los componentes celulares (CC) y la función molecular (MF) no mostraron ninguna concordación para un grupo mayor de 5 genes candidatos.

Tabla 4. Anotación funcional de los genes. Procesos biológicos (BP), componentes celulares (CC) y funciones moleculares (MF).

Genes	Biological process (BP)								Cellular components (CC)					Molecular function (MF)
	Cell adhesion	Biological adhesion	Single-multicellular organism process	Cell-cell adhesion	Interferon-gamma production	Cytokine production	Multicellular organismal process	Regulation of secretion	Cell-cell adherens junction	Cell-cell junction	Adherens junction	Extracellular matrix component	Anchoring junction	Carbohydrate derivative binding
DHX29	✓	✓		✓					✓	✓	✓		✓	✓
AHSA1	✓	✓		✓					✓	✓	✓		✓	
AVPR2			✓		✓	✓	✓	✓						
CHIA			✓			✓	✓	✓						✓
NAV1			✓				✓							
MEGF9												✓		
FASTK														✓
MFAP4	✓	✓	✓				✓					✓		
UBE20														✓
PPP2R3A			✓				✓							
IL20RB	✓	✓	✓	✓	✓	✓	✓							
MBTD1			✓				✓							
RAB26			✓				✓	✓						✓
CLHC1														
FAT2	✓	✓	✓	✓			✓		✓	✓	✓		✓	

5. DISCUSIÓN

En el presente trabajo a partir de marcadores RAD-seq se realizaron diferentes análisis bioinformáticos por tal de determinar aquellos loci implicados en la determinación del sexo en la especie *Gambusia holbrooki*. Los datos obtenidos permitieron considerar 95 loci como posibles candidatos implicados en la determinación del sexo en dicha especie, donde la población PO presentaba una gran similitud con las poblaciones europeas al tener ambas una variabilidad menor respecto a las poblaciones americanas EV y FLO, como ya se afirmaba en el trabajo de Vera *et al.* (2016)¹² del cual se obtuvieron las muestras del análisis. Teniendo en cuenta que la población FLO estaba formada exclusivamente por hembras, la detección de ciertos loci como candidatos a la determinación del sexo en *Gambusia holbrooki* pudo verse sesgada hacia las hembras, cuando en realidad se trataba de un patrón característico de esta población. Este hecho pudo comprobarse tanto en la búsqueda de RAD-tags expresados en un único sexo como en la búsqueda de loci *outliers*, al presentarse una ausencia o variación nucleotídica principalmente en Florence y no en todas las hembras de estudio indistintamente de la población a la que pertenecían.

Los análisis posteriores concluyeron que ninguno de los RAD-tags considerados como candidatos estaba directamente relacionados con la determinación del sexo en *Gambusia holbrooki*, aunque si bien es cierto algunas funciones pueden estar indirectamente implicadas. Uno de los posibles motivos puede ser la utilización del individuo femenino VM43 como modelo de genoma, ya que todos aquellos genes relacionados con el cromosoma Y, tales como el gen DMY hallado en el pez *Oryzias latipes*⁵, no estarían presentes. El uso de marcadores RAD-seq puede haber influido en los resultados obtenidos por su distribución no uniforme de los RAD-tags y la potencial pérdida de SNP adaptativos, dejando ciertas partes del genoma sin muestrear¹⁰. Además, los programas bioinformáticos empleados para la búsqueda de loci *outliers* puede ser otro de los motivos de los resultados concluidos en el presente trabajo. En el estudio de 4 conjuntos de datos con 100 loci cada uno (95 neutrales y 5 con selección divergente) para 10 poblaciones, BayeScan presentaba una sensibilidad más elevada que Arlequin así como un mayor control de falsos positivos³⁰, lo cual

puede ser la razón de que no se encontrase ningún valor claramente outlier mediante BayeScan pero si con Arlequin.

El tamaño del genoma de *Gambusia holbrooki* no se conoce con certeza, pero sí el de algunas especies pertenecientes al género de los poecílidos. Entre ellas, el de *Xiphophorus maculatus* estimado entre 750-950 Mb³¹ y el de *Gambusia affinis* y *Poecilia formosa* constituido por unas 950 Mb y 920 Mb respectivamente³². Por tanto, en el futuro sería de gran interés la utilización de métodos que permitan analizar una mayor cantidad del genoma, ya que en el presente estudio se cubrieron aproximadamente 4 Mb del genoma total, al usar 44271 RAD-tags con una longitud de 90pb. El análisis transcriptómico mediante la técnica RNA-seq en muestras gonadales de hembras y machos realizado en trabajos anteriores como el de Lamatsch *et al.* (2015)⁸ permitió identificar un marcador específico del sexo para *G. affinis*, por lo que esta técnica podría aportar nuevos resultados al estudio de marcadores ligados al sexo en *G. holbrooki*. Además, sería interesante la realización de este mismo estudio sin la población americana FLORENCE, por tal de evitar que los resultados se vean influidos hacia el sexo femenino de la especie *Gambusia holbrooki*.

6. CONCLUSIONS

Based on the objectives set at the beginning of the work and according to the results of the analyses carried out, it has been concluded that:

1. None of the revised 44271 RAD-tag loci in *Gambusia holbrooki* have an exclusive single-sex SNP.
2. However, 95 RAD-tags showed a differential variation for SNPs between males and females, although none of them reveals a function clearly involved in sex determination.

7. BIBLIOGRAFÍA

- ¹Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente. (2013). Catálogo Español de Especies Exóticas Invasoras. Peces: *Gambusia holbrooki*. Recuperado de http://www.mapama.gob.es/es/biodiversidad/temas/conservacion-de-especies/especies-exoticas-invasoras/ce_eei_peces.aspx
- ²Pyke, G. H. (2005). A review of the biology of *Gambusia holbrooki* and *G. Affinis*. *Reviews in Fish Biology and Fisheries*, 15, 339-365.
- ³Lacadena, J. R. (1996). Cromosomas y función genética III: Cromosomas sexuales. *Citogenética* (p. 359-434). Editorial Complutense.
- ⁴Harley, V. R., Goodfellow, P. N. (1994). The biochemical role of SRY in sex determination. *Molecular Reproduction and Development*, 39(2), 184-193.
- ⁵Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C.E., Shibata, N., Asakawa, S., Shimizu, N., Hori, H., Hamaguchi, S., Sakaizumi, M. (2002). *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*, 417, 559-563.
- ⁶Chakraborty, T. Zhou, L. Y., Chaudhari, A., Iguchi, T., Nagahama, Y. (2016). *Dmy* initiates masculinity by altering *Gsdf/Sox9a2/Rspo1* expression in medaka (*Oryzias latipes*). *Scientific Reports*, 6(19480).
- ⁷Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hoyosa, S., Tohari, S., Brenner, S., Miyadai, T., Venkatesh, B., Suzuki, Y., Kikuchi, K. (2012). A Trans-Species Missense SNP in *Amhr2* Is Associated with Sex Determination in the Tiger Pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genetics*, 8(7).
- ⁸Lamatsch, D. K., Adolfsson, S., Senior, A. M., Christiansen, G., Pichler, M., Ozaki, Y., Smeds, L., Scharl, M., Nakagawa, S. (2015). A Transcriptome Derived Female-Specific Marker from the Invasive Western Mosquitofish (*Gambusia affinis*). *PLoS ONE*, 10(2).

- ⁹Davey, J. W., Blaxter, M. L. (2010). RADseq: Next-generation population genetics. *Briefings in Functional Genomics*, 9(5–6), 416–423.
- ¹⁰Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., Storfer, A. (2016). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology*, 17(2), 142-152.
- ¹¹Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., Cresko, W. A. (2010). Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genet*, 6(2).
- ¹²Vera, M., Díez-del-Molino, D., García-Marín, J. L. (2016). Genomic survey provides insights into the evolutionary changes that occurred during European expansion of the invasive mosquitofish (*Gambusia holbrooki*). *Molecular Ecology*, 25(5), 1089-1105.
- ¹³Wright, S. (1951). The genetical structure of populations. *Annals of eugenics*, 15(4), 323-354.
- ¹⁴Foll, M., Gaggiotti, O. (2008). A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *GENETICS*, 180(2), 977-993.
- ¹⁵Thresher, R. E. (2007). Genetic options for the control of invasive vertebrate pests: prospects and constraints. En Witmer, G. W., Pitt, W. C., Fagerstone, K. A. (ed.), *Managing Vertebrate Invasive Species: Proceedings of an International Symposium* (p. 318-331). Fort Collins: USDA/APHIS/WS, National Wildlife Research Center.
- ¹⁶Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., Cresko, W. A. (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124-3140.
- ¹⁷Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, 1(3), 171-182.

- ¹⁸Rochette, N. C., Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, 12(12), 2640–2659.
- ¹⁹Raymond, M., Rousset, F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86(3), 248-249
- ²⁰Rousset, F., (2008). Genepop'007: a complete reimplement of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, 8(1), 103-106.
- ²¹Fischer, M. C., Foll, M., Excoffier, L., Heckel, G. (2011). Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular Ecology*, 20(7), 1450-1462.
- ²²Excoffier, L., Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564-567.
- ²³Lischer, H.E.L., Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2), 298-299.
- ²⁴R Development Core Team (2008). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*.
- ²⁵Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, 36(10), 3420-3435.
- ²⁶Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403-410.
- ²⁷Huang, D. W., Sherman, B. T., Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protocols*, 4(1), 44-57.

- ²⁸Huang, D. W., Sherman, B. T., Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, 37(1), 1-13.
- ²⁹Yates, B., Braschi, B., Gray, K., Seal, R., Tweedie, S., Bruford, E. (2017). Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic Acids Research*, 45(1), 619-625.
- ³⁰Narum, S. N., Hess, J. E. (2011). Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources*, 11(s1), 184-194.
- ³¹Schartl, M., Walter, R. B., Shen, Y., Garcia, T., Catchen, J., Amores, A., Braasch, I., Chalopin, D., Volff, J., Lesch, K., Bissazza, A., Minx, P., Hillier, L., Wilson, R. K., Fuerstenberg, S., Boore, J., Searle, S., Postlethwait, J. H., Warren, W.C. (2013). The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nature Genetics*, 45, 567-572.
- ³²Lamatsch, D. K., Steinlein, C., Schmid, M., Schartl, M. (2000). Noninvasive determination of genome size and ploidy level in fishes by flow cytometry: detection of triploid *Poecilia formosa*. *Cytometry*, 39(2), 91-95.