

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE MEDICINA



TESIS DOCTORAL

**Estudio evolutivo de los intrones del MHC en aves
passeriformes del género Carduelis**

MEMORIA PARA OPTAR AL GRADO DE DOCTORA

PRESENTADA POR

María Cristina Campos Burgos

DIRECTOR

Antonio Arnáiz Villena

Madrid

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE MEDICINA



TESIS DOCTORAL

**ESTUDIO EVOLUTIVO DE LOS INTRONES DEL
MHC EN AVES PASSERIFORMES DEL GÉNERO
CARDUELIS**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

MARÍA CRISTINA CAMPOS BURGOS

DIRECTOR

ANTONIO ARNAIZ VILLENA

Si no conozco una cosa, la investigaré

Louis Pasteur

AGRADECIMIENTOS

En primer lugar al Dr. Antonio Arnaiz Villena, quien ha sido mi tutor durante todo este periodo de investigación, por su labor como mentor, por su inestimable ayuda y por su dirección en los proyectos en los que he trabajado. Agradecerle la pasión que me ha transmitido por el trabajo científico, y por permitirme el lujo de trabajar con su fascinante y completa recolección de muestras de aves que ha ido recopilando durante muchos años.

A mis compañeros Ester y José, por su apoyo y comprensión en la realización de este trabajo desde el primer día que los conocí. Ha sido una suerte contar con ellos todo este tiempo.

Al Departamento de Inmunología de la Facultad de Medicina de la Universidad Complutense de Madrid, en especial a la Dra. Narcisa Quiles, por su ayuda en este trabajo siempre que lo he necesitado.

Y, por último, al personal del Centro de Transfusión de la Comunidad de Madrid, por su amabilidad y sus sonrisas.

DEDICATORIA

Este trabajo va dedicado a todas aquellas personas que han estado ahí para apoyarme durante esta etapa de mi vida que culmina con este maravilloso trabajo. Han sido unos años cargados de esfuerzo y perseverancia que no hubieran sido posible sin vuestro cariño. En especial, quiero dedicárselo a ciertas personas que han sido clave en el éxito de este proyecto.

En primer lugar, a mi marido, Juli, quien se ha convertido en mi compañero de viaje en esta aventura que es la vida. Es la persona más maravillosa que he conocido, que me complementa día a día y que siempre me apoya cuando más lo necesito.

A mi gran amiga Pili, un miembro más de mi familia, que no ha dejado de animarme en este camino de la investigación y a la que nunca le ha faltado una sonrisa, incluso cuando las cosas no iban bien. Sin ella, no hubiera sido capaz de llegar hasta el final.

A mis padres, mi hermana y mi abuelo, que siempre me preguntaban qué tal lo llevaba todo y me mandaban muchos besos desde Sevilla.

Y a mi Jose favorito, que creyó en mí desde el primer día que me conoció y siempre ha estado ahí cuando lo he necesitado.

Os quiero a todos y os estaré eternamente agradecida.

ÍNDICE

RESUMEN	1
SUMMARY	7
INTRODUCCIÓN.....	11
I. EL COMPLEJO PRINCIPAL DE HISTOCOMPATIBILIDAD	13
1. Definición e historia.....	13
2. El complejo principal de histocompatibilidad en humanos (HLA)	14
2.1 Región HLA	14
2.2 Grupos de genes MHC	17
2.3 Moléculas HLA de clase I	19
2.4 Moléculas HLA de clase II.....	23
2.5 Región HLA de clase III.....	26
3. El complejo principal de histocompatibilidad en el pollo	26
3.1 Región MHC-B	28
3.2 Región MHC-Y	31
3.3 Otros genes fuera de la región de histocompatibilidad.....	31
3.4 Susceptibilidad a enfermedades del pollo	32
3.5 Hipótesis del MHC "mínimo esencial"	32
4. El complejo principal de histocompatibilidad en aves no <i>Passeriformes</i>	33
5. El complejo principal de histocompatibilidad en aves <i>Passeriformes</i>	33
II. INTRONES	37
1. Definición e historia.....	37
2. Maduración del ARN mensajero	37
3. Clasificación de intrones.....	39
4. Funciones de los intrones.....	41
5. Orígenes de los intrones.....	44
III. CLASE AVES	47
1. Orden <i>Passeriformes</i>	48
2. Género <i>Carduelis</i>	48
2.1 <i>Carduelis atrata</i>	49
2.2 <i>Carduelis carduelis parva</i>	49
2.3 <i>Carduelis citrinella</i>	49
2.4 <i>Carduelis pinus</i>	50

2.5	<i>Carduelis spinus</i>	50
2.6	<i>Carduelis crassirostris</i>	50
2.7	<i>Carduelis cucullata</i>	51
2.8	<i>Carduelis magellanica</i>	51
2.9	<i>Carduelis notata</i>	51
2.10	<i>Carduelis spinescens</i>	52
2.11	<i>Carduelis olivacea</i>	52
2.12	<i>Carduelis xanthogastra</i>	52
2.13	<i>Carduelis yarrellii</i>	53
2.14	<i>Carduelis barbata</i>	53
2.15	<i>Carduelis lawrencei</i>	53
2.16	<i>Carduelis dominicensis</i>	54
2.17	<i>Carduelis psaltria hesperophila</i>	54
2.18	<i>Carduelis atriceps</i>	54
2.19	<i>Fringilla coelebs</i>	55
IV. FILOGENIA MOLECULAR		56
1.	Métodos filogenéticos	56
2.	Citocromo b	57
OBJETIVOS.....		59
MATERIALES Y METODOS.....		63
1.	Especies de aves estudiadas	65
2.	Obtención de las muestras	66
3.	Extracción de ADN.....	67
4.	Obtención de secuencias de ADN de los genes del MHC de clase I.....	69
4.1	Amplificación de secuencias del MHC	69
4.2	Electroforesis en gel de agarosa de los fragmentos amplificados	71
4.3	Purificación de la banda de la secuencia amplificada	72
4.4	Secuenciación de los fragmentos amplificados	73
4.5	Clonación de los fragmentos amplificados.....	74
4.6	Obtención de las secuencias de ADN de los fragmentos clonados	77
5.	Obtención de secuencias de ADN del gen de citocromo b.....	78
6.	Análisis de las secuencias obtenidas de los genes del MHC	79
6.1	Caracterización de las secuencias de ADN	79
7.	Filogenia basada en el gen citocromo b.....	80

7.1	Métodos filogenéticos	80
7.1.1	Modelos de sustitución nucleotídica	81
7.1.2	Métodos de reconstrucción de árboles filogenéticos.....	82
RESULTADOS	85
1.	Secuencias del intrón 2 del MHC	87
1.1	Posiciones conservadas.....	87
1.2	Frecuencias nucleotídicas	93
1.3	Distancias genéticas.....	94
1.4	Porcentaje de similitud	98
2.	Filogenia mitocondrial	100
DISCUSIÓN	103
1.	Estudio del intrón 2 del complejo principal de histocompatibilidad	105
1.1	Densidad de intrones	105
1.2	Tamaño de los intrones.....	107
1.3	Funciones de los intrones	109
1.4	Composición nucleotídica de los intrones	110
1.5	Similitud entre los intrones.....	111
1.6	Teoría del MHC “mínimo esencial”	112
2.	Filogenia molecular	113
CONCLUSIONES	117
BIBLIOGRAFÍA	121
ANEXOS	135

RESUMEN

El Complejo Principal de Histocompatibilidad (CPH, en inglés MHC, *Major Histocompatibility Complex*) está formado por un conjunto de genes mayoritariamente polimórficos, cuyos productos se expresan en la superficie de gran variedad de células los cuales son los responsables de la respuesta inmune adaptativa. El mejor MHC estudiado en aves es el de *Gallus gallus*, el cual se encuentra ubicado en el microcromosoma 16 y está formado por la región MHC-B y la región MHC-Y. Estas regiones se caracterizan por contener genes muy próximos entre sí y poseer intrones muy cortos, lo que lleva a plantear la hipótesis del MHC “mínimo esencial” que defiende que este sistema contiene sólo el mínimo número de genes absolutamente esenciales para la función del MHC en un sentido evolutivo. El estudio de estos genes de histocompatibilidad en aves pertenecientes al orden *Passeriformes* es importante porque a estas aves se les consideran modelos naturales en los que la evolución del polimorfismo del MHC ha sido adquirida a lo largo de millones de años. Por ello, es interesante comparar este polimorfismo con el de modelos artificiales como el pollo, el ser humano o el ratón en los que poblaciones pequeñas han crecido rápidamente por consanguinidad y se han convertido en poblaciones relativamente grandes. Otro argumento a favor del estudio del MHC de estas aves es que es fácil identificar las especies y es muy cómodo manejar estas muestras. Además, presentan una amplia distribución geográfica y hay un gran número de especies pertenecientes a este orden.

Los intrones son fragmentos de ADN localizados dentro de un gen, que no codifican ningún fragmento de la proteína, ya que son eliminados en el proceso de maduración del ARN. La presencia de intrones en el genoma es una característica propia de los organismos eucariotas, aunque algunos procariotas también presentan intrones en su material genético. A día de hoy, se pueden describir 8 tipos de intrones. Tradicionalmente, se ha defendido que los intrones son fragmentos de ADN carentes de información cuya presencia en los genomas supone una carga para la célula. Sin embargo, en la actualidad, se ha demostrado que los intrones presentan un amplio espectro de funciones y están involucrados en cada paso del procesamiento de ARNm.

La realización de este trabajo tiene el objetivo fundamental de analizar el polimorfismo y la evolución del MHC en especies del género *Carduelis* a través del estudio del intrón 2 de estos genes. El análisis de estos intrones nos puede dar una idea de si éstos cumplen una función esencial en el MHC que ha hecho que se hayan mantenido sin apenas

variación a lo largo de millones de años. Además, el análisis filogenético de estas especies que se ha llevado a cabo en este trabajo nos permite estudiar la evolución molecular del MHC y compararla con las relaciones filogenéticas de estos jilgueros mediante el estudio del gen del citocromo b. Para estudiar la filogenia se emplean dos métodos de reconstrucción de árboles filogenéticos: Máxima verosimilitud o *Maximum likelihood* (ML) e Inferencia Bayesiana (IB).

En este trabajo se han secuenciado los genes del MHC de 18 especies del género *Carduelis* y se ha estudiado el intrón 2 de cada una de ellas. Este intrón ha sido comparado con la secuencia de *Fringilla coelebs*, especie considerada como grupo externo de este estudio. También se ha comparado el intrón 2 de todas estas especies pertenecientes al orden *Passeriformes*, con el de *G. gallus*, que pertenece al orden *Galliformes*. A partir de muestras de sangre se obtuvo ADN, el cual fue amplificado, clonado y secuenciado dos veces, obteniéndose un resultado idéntico de cada réplica para cada individuo. Las secuencias de ADN obtenidas se correspondían con los genes del MHC de clase I que previamente han sido descritos en estos géneros. Los análisis de estas secuencias se llevaron a cabo con los programas MEGA 7 y BioEdit.

El primer resultado que se obtuvo fue que la divergencia entre las secuencias del intrón 2 del MHC de las aves estudiadas del género *Carduelis* es muy pequeña, habiendo especies que incluso comparten el 100% de similaridad. Se analizó el tamaño del intrón 2 del MHC y se vio que es muy similar en todas las aves *Passeriformes* estudiadas con una longitud media de 311 pb en los jilgueros y 293 pb en el pinzón común. El intrón 2 del pollo es mucho más corto y está formado por 229 pb. Estos datos demuestran que el MHC “mínimo esencial” descrito en *G. gallus* no se ajusta al MHC de las aves *Passeriformes* porque sus intrones son significativamente más largos que el intrón 2 del pollo. Cuando se compara el intrón 2 de las 18 especies del género *Carduelis* se observó que un total de 283 posiciones se conservan en todos los jilgueros variando, únicamente, 28 posiciones. Este alto grado de conservación que presenta el intrón 2 de las especies del género *Carduelis* puede deberse a que estas secuencias tengan alguna función en el MHC o bien, presenten elementos reguladores de la expresión génica. Se calculó también el contenido medio en GC del intrón 2 de las aves *Passeriformes* y el resultado fue del 55,2%, un contenido alto en GC que lleva a pensar que sea un indicio de que presente elementos reguladores en su secuencia. Con respecto a las relaciones filogenéticas de los jilgueros estudiados, los resultados obtenidos mediante los métodos ML e IB son prácticamente

idénticos y en ellos se ve cómo todas las especies están agrupados de acuerdo a las distintas radiaciones que ha habido en América. Además, los resultados obtenidos mediante el análisis de las secuencias de citocromo b no coinciden con los resultados que se obtienen al estudiar la filogenia de las secuencias de MHC, ya que estos genes evolucionan de forma diferente dependiendo del ambiente en el que se desarrollen y de los patógenos a los que se enfrenten.

SUMMARY

The Major Histocompatibility Complex (MHC) comprises the most polymorphic loci in vertebrates and its molecules present antigenic peptides to clonotypic T cell receptor in order to start an immune response. The chicken's MHC is the best-studied MHC in birds. It is located in microchromosome 16 and it is formed of MHC-B region and MHC-Y region. These regions are characterized by containing genes that are very close to each other and by having very short introns; that supported the hypothesis that the chicken's MHC represented a 'minimal essential MHC'. The study of these genes in birds that belong to order *Passeriformes* is important because these birds are considered natural models in which evolution of MHC polymorphism has been acquired over millions of years. Therefore, it is interesting to compare this polymorphism with artificial models such as chicken, human or mouse in which small populations have grown rapidly by consanguinity and they have become relatively large populations. Another argument in favor of MHC study in these birds is that it is easy to identify species and it is very convenient to handle these samples. In addition, they have a wide geographical distribution and there is a large number of species belonging to this order.

Introns are DNA fragments located within a gene, which do not encode any part of the protein and they are eliminated in RNA maturation. The presence of introns in genome is an own characteristic of eukaryotic organisms, although some prokaryotes also have introns in their genetic material. Today, eight types of introns are described. Traditionally, it has been argued that introns are DNA fragments lacking information whose presence in genomes is a burden on the cell. However, at present, it has been shown introns present a broad spectrum of functions and they are involved in each step of mRNA processing.

In the present work, we have aimed to analyze MHC polymorphism and evolution in species of genus *Carduelis* through intron 2 study of these genes. Our analysis of these introns can give us an idea of whether they fulfill an essential function in the MHC that has kept them with not much variation over millions of years. As well, phylogenetic analysis of these species allow us to study MHC molecular evolution and compare it with phylogenetic relationships of these goldfinches through study of cytochrome b gene sequence. To study phylogeny, two methods for reconstructing phylogenetic trees were used: Maximum likelihood (ML) and Bayesian Inference (BI).

In this work, MHC genes of 18 species of genus *Carduelis* have been sequenced and intron 2 of each of them has been studied. This intron has been compared to *Fringilla coelebs* sequence, a species considered as an outgroup of this study. Furthermore, intron

2 of all these *Passeriformes* species has also been compared with intron 2 of *G. gallus*, which belongs to order *Galliformes*. It was obtained DNA from blood samples, which was amplified, cloned and sequenced twice, obtaining an identical result of each replica for each individual. DNA sequences obtained corresponded with MHC class I genes that have previously been described in these genus. Analysis of these sequences were carried out with MEGA 7 and BioEdit programs.

Our first result obtained was divergence between MHC intron 2 sequences of the studied birds of genus *Carduelis* was very small, with species that even share 100% similarity. It was analyzed size of intron 2 of the MHC and it was found to be very similar in all *Passeriformes* birds studied with an average length of 311 bp in goldfinches and 293 bp in chaffinch. Intron 2 of chicken was much shorter and it was formed by 229 bp. These results demonstrated that a minimal essential MHC described in *G. gallus* does not coincident with the MHC of *Passeriformes* birds because their introns are significantly longer than intron 2 of chicken. When it was compared intron 2 of 18 species of genus *Carduelis* it was observed that 283 positions were conserved in all goldfinches varying 28 positions. This conservation of intron 2 of species of genus *Carduelis* may be due to the fact that these sequences have some function in the MHC or that they have elements that regulate gene expression. Intron 2 GC content of *Passeriformes* birds was 55.2%, a high GC content that leads to think that these sequences present regulatory elements. Regarding phylogenetic relationships of goldfinches studied, we results obtained using ML and IB methods are practically identical and they show how all species are grouped according to different radiations that have occurred in America. In addition, our results obtained after analyzing cytochrome b sequences do not coincide with our date obtained at our phylogenetic study with MHC sequences, since these genes evolve differently depending on environment in which they are developed.

INTRODUCCIÓN

I. EL COMPLEJO PRINCIPAL DE HISTOCOMPATIBILIDAD

1. Definición e historia

El Complejo Principal de Histocompatibilidad (CPH, en inglés MHC, *Major Histocompatibility Complex*) está formado por un conjunto de genes mayoritariamente polimórficos, cuyos productos se expresan en la superficie de gran variedad de células los cuales son los responsables de la respuesta inmune adaptativa. Estos genes están presentes en todos los vertebrados y se encargan de presentar antígenos a los linfocitos T, células inmunitarias que una vez que han sido activadas son capaces de poner en marcha una respuesta inmunológica eficaz y específica contra el agente causante de su activación (Abbas *et al.* 2019).

Los genes de histocompatibilidad fueron identificados por primera vez en la década de los 40, gracias al estudio del cáncer mediante el trasplante de tejidos tumorales en ratones. Años antes se identificaron varios antígenos eritrocitarios en ratones receptores de trasplante, relacionándose uno de ellos (el antígeno II) con la supervivencia del tumor trasplantado, ya que en las cepas de ratón donde este antígeno no estaba presente se correlacionaba con el rechazo al trasplante, lo que permitió elaborar una teoría inmunológica del trasplante (Gorer 1936). Posteriormente, G. Snell identificó la región génica responsable de este rechazo, siendo denominados genes de histocompatibilidad, y en el caso concreto del ratón, genes H-2, en referencia al antígeno II (Snell 1948).

Los genes de histocompatibilidad en humanos fueron descubiertos en los años 50 por J. Dausset cuando detectó la presencia de anticuerpos capaces de provocar la aglutinación de los leucocitos del donante pero no de los del receptor, al reconocer los antígenos en la superficie de estas células, en personas que habían recibido múltiples transfusiones sanguíneas (Dausset 1958). En la década de los 60 se comprobó que estos antígenos leucocitarios representaban el complejo principal de histocompatibilidad (Dausset *et al.* 1965; Ceppellini *et al.* 1967). Sin embargo, actualmente se conoce que el papel biológico fundamental de las moléculas HLA consiste en la presentación antigénica mediante su interacción con los linfocitos T (Stern *et al.* 1994; Madden 1995).

2. El complejo principal de histocompatibilidad en humanos (HLA)

En humanos, el complejo principal de histocompatibilidad se denomina sistema HLA (del inglés *Human Leukocyte Antigen*). Se trata de un sistema codominante y muy poligénico que conforma uno de los sistemas genéticos más complejos que se conocen en el ser humano. Es muy polimórfico porque existen múltiples alelos para un mismo locus y posee desequilibrio de ligamiento, una propiedad que presentan algunos genes que tienen diferentes alelos en el mismo cromosoma con una frecuencia mayor a la teóricamente esperada en una combinación al azar.

2.1 Región HLA

El sistema HLA se encuentra ubicado en el brazo corto del cromosoma 6 y ocupa una longitud de 4 centimorgan. El primer mapa genético de esta región se completó en 1991 (Trowsdale *et al.* 1991), pero años más tarde se elaboró otro mapa a partir de secuencias parciales obtenidas en individuos diferentes, según el cual la región HLA ocupa 3,6 Mb y está compuesta por 224 genes de los cuales 128 podrían expresarse y aproximadamente el 40% tendrían función inmunológica (The MHC sequencing consortium 1999). Posteriormente, este mapa fue actualizado con una secuencia completa obtenida a partir de una línea celular homocigota (Stewart *et al.* 2004). Además, la secuenciación completa del cromosoma 6 (Mungall *et al.* 2003) reveló la existencia de otros genes relacionados con el MHC más allá de las 3,6 Mb descritas en 1999, por lo que se elaboró un nuevo mapa genético que recibió el nombre de MHC extendido (xMHC) (Horton *et al.* 2004) para diferenciarlo del mapa constituido anteriormente, denominado mapa MHC clásico. El mapa clásico está dividido en 3 regiones denominadas región de clase I (HLA-I), región de clase III (HLA-III) y región de clase II (HLA-II), situadas en este orden de telómero a centrómero en el brazo corto del cromosoma 6. El mapa extendido completa el anterior al añadir dos nuevas regiones: una en el extremo telomérico (región de clase I extendida) y otra en el extremo centromérico (región de clase II extendida), ampliándose así las dimensiones de la región HLA hasta las 7,6 Mb (Figura II). Según el mapa extendido, la región HLA contiene 421 genes, de los cuales el 60% expresan proteínas funcionales, el 33% son pseudogenes (expresan proteínas aberrantes o bien no se expresan) y el 7% restante se transcriben dando lugar a moléculas de ARN transferente (ARNt), que participan en el proceso de traducción del ARN mensajero (ARNm) en el

ribosoma. La mayor concentración de genes se observa en la región de clase III, que contiene 58 genes funcionales (23%) en 0,7 Mb, lo que la convierte además, en la región de mayor densidad génica de todo el genoma humano (Horton *et al.* 2004).

2.2 Grupos de genes MHC

La región MHC extendida está formada por distintos grupos de genes:

- Los genes de las histonas están involucrados en el correcto plegamiento del ADN para la formación del nucleosoma que conforma la cromatina. La mayor concentración de estos genes en el genoma humano se encuentra en la región HLA de clase I extendida, que contiene 55 genes que expresan histonas de las 5 clases existentes (6 H1, 12 H2A, 15 H2B, 10 H3 y 12 H4) y 11 pseudogenes distintos.
- Los genes de ARNt son muy cortos (75-90 nucleótidos) y están implicados en la síntesis de proteínas. Estas moléculas de ARNt reconocen los tripletes del ARNm para suministrar al ribosoma el aminoácido correspondiente en el proceso de traducción. Son 157 genes (6 de ellos no funcionales) que incluyen todo los tipos existentes salvo Asn- y Cys-ARNt. Están distribuidos a lo largo de la región de clase I extendida y representan el mayor agrupamiento de este tipo de genes de todo el genoma humano.
- Los genes transportadores de solutos (SLC, del inglés *Solute Carrier*), que aparecen tanto en la región de clase I extendida como en la región de clase II extendida, dan lugar a proteínas transportadoras de nutrientes.
- Los genes de las butirofilinas (BTN) son 8, están presentes mayoritariamente en la región de clase I extendida y son miembros de la superfamilia de las inmunoglobulinas. Son muy similares a los genes MOG de los mamíferos (situado entre las regiones de clase I y clase III en humanos) y a los genes de la familia *BG* del complejo principal de histocompatibilidad del pollo. Tienen diversas funciones en el sistema inmunitario, destacando la proteína BTN2A2 que es expresada por las células presentadoras de antígenos (CPA) modulando la respuesta de los linfocitos T (Sarter *et al.* 2016).
- Los genes de los receptores vomeronasales (VNR) se encuentran también en la región de clase I extendida y pertenecen a la familia de genes que codifican receptores de feromonas. Los únicos representantes de esta familia en humanos son pseudogenes.
- Los genes de los receptores olfatorios son 34 (al menos 14 funcionales) y se encuentran en la región de clase I extendida. Presentan polimorfismo y se les atribuye una función relacionada con la supervivencia en los procesos del comportamiento.

- Los genes que dan lugar a proteínas con dedos de zinc presentan funciones muy diversas, pudiendo actuar como enzimas, factores de transcripción y proteínas de replicación. En la región de clase I existen 36 de estos genes, divididos en 3 grupos: los que tienen un dominio SCAN (13 genes), los que tienen un dominio triple denominado TRIM (8 genes) y el resto (15 genes). Diez de ellos no se expresan.
- Los genes HLA-I codifican para proteínas muy similares en cuanto a su estructura, pero muy distintas desde el punto de vista funcional. Según esto, se clasifican en genes de clase I clásicos (*HLA-A*, *HLA-B* y *HLA-C*, también denominados HLA-Ia), genes de clase I no clásicos (*HLA-E*, *HLA-F*, *HLA-G* y 12 pseudogenes, también conocidos como HLA-Ib). Existen otros genes pertenecientes al grupo HLA de clase I no considerados genes HLA que son los genes MIC (*MIC-A* y *MIC-B*) y el gen de la hemocromatosis (*HFE*). Excepto éste último, que se encuentra en el extremo telomérico de la región de clase I extendida, todos están localizados en la región de clase I. Los genes MIC se expresan en varios tipos celulares incluyendo células endoteliales y monocitos (Zwirner *et al.* 1999). Estos genes expresan proteínas que son reconocidas por los receptores activadores NKG2D de las células NK y que pueden tener una función en el sistema inmune de la mucosa intestinal. Por otro lado, el gen *HFE* está implicado en el transporte de hierro.
- Los genes de los factores de necrosis tumoral (TNF) se encuentran situados en la región de clase III. Codifican para 3 tipos de citoquinas (TNF, LTA y LTB) pertenecientes a la superfamilia TNF y participan en procesos inflamatorios.
- Los genes del antígeno 6 linfocitario (LY6) codifican proteínas de membrana asociadas a glicosil-fosfatidil-inositol y pueden tener una función inmunológica. Están localizados en la región de clase III.
- Los genes de las proteínas de choque térmico (HSP, del inglés *Heat Shock Proteins*) aumentan su expresión cuando se encuentran en un medio que le provoca cualquier tipo de estrés. Se encuentran en la región de clase III y están involucrados en la señalización inducida por estrés del sistema inmune para la eliminación de células dañadas, infectadas o tumorales.
- Algunos genes del complemento de la vía clásica (*C2*, *C4A*, *C4B*) y de la vía alternativa (Factor B) también están presentes en la región de clase III.
- Los genes HLA-II están ubicados en la región de clase II y codifican para proteínas polimórficas implicadas en la presentación antigénica. Se clasifican en genes de clase

II clásicos (HLA-DP, HLA-DQ y HLA-DR) y genes de clase II no clásicos (HLA-DM y HLA-DO). Se han descrito numerosos genes funcionales y también pseudogenes. Los genes clásicos se expresan en células presentadoras de antígenos. Su función es unir péptidos de origen extracelular (generalmente procedentes de bacterias) y presentarlos a linfocitos T cooperadores (Th, CD4+) para llevar a cabo una respuesta inmunológica humoral mediada por anticuerpos. Los genes no clásicos no se expresan en la superficie celular, pero forman complejos que están involucrados en el intercambio y carga de péptidos en las moléculas de clase II clásicas.

2.3 Moléculas HLA de clase I

Las moléculas HLA de clase I son proteínas que resultan de la expresión de los genes HLA de clase I clásicos (*HLA-A*, *-B*, *-C*) y no clásicos (*HLA-E*, *-F*, *-G*), quedando excluidas las proteínas expresadas por los genes MIC y el gen *HFE* (Figura I2). Se tratan de glicoproteínas de membrana que se expresan en todas las células nucleadas del organismo, aunque su nivel de expresión varía desde un máximo en células pertenecientes al sistema inmune (linfocitos T, B y macrófagos) hasta un mínimo en células musculares, del sistema nervioso y fibroblastos. Las moléculas HLA de clase I son muy polimórficas y están implicadas en el reconocimiento de péptidos antigénicos de origen intracelular (por degradación de proteínas citosólicas derivadas de virus y tumores), que son presentados a los linfocitos T citotóxicos (Tc, CD8+) para poner en marcha una respuesta inmunológica de tipo celular (citotóxica) mediada por linfocitos Tc y células NK (del inglés, *Natural Killer*).

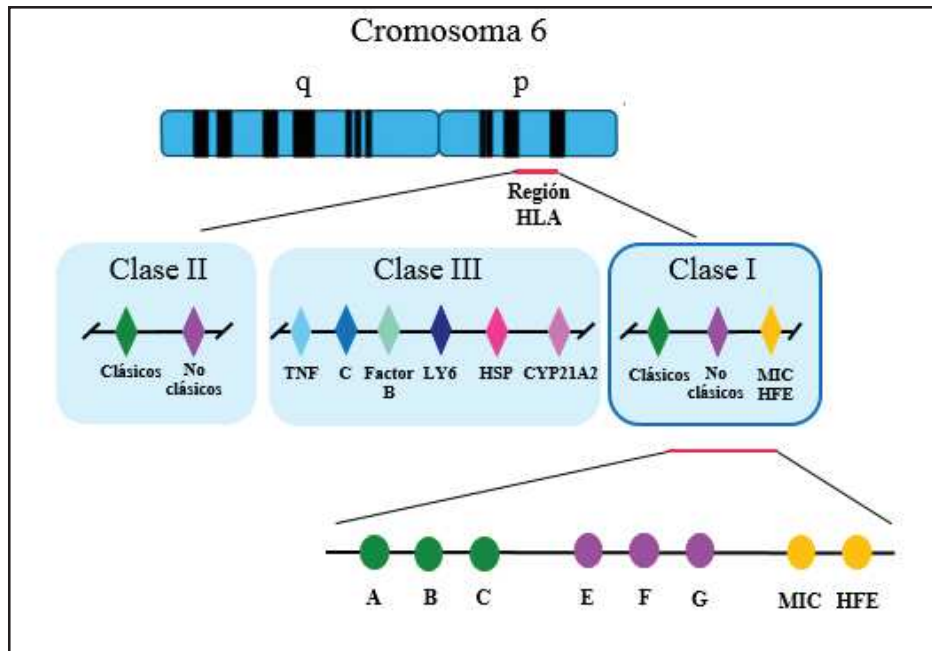


FIGURA I2. Genes de la región HLA de clase I.

Las moléculas HLA de clase I son glicoproteínas de membrana resultantes de la asociación entre un polipéptido denominado cadena alfa y la proteína beta-2-microglobulina (β 2m) (Figura I3). Las cadenas alfa están codificadas en la región de clase I del sistema HLA, mientras que la β 2m está codificada en el cromosoma 15. La cadena

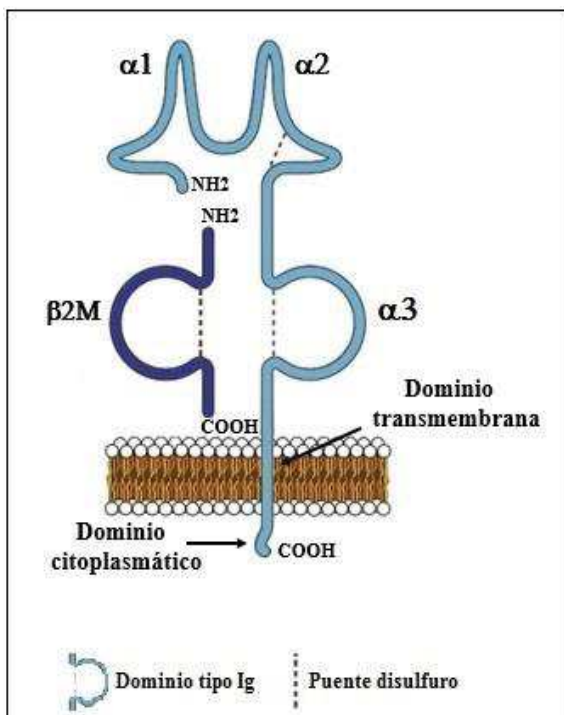


FIGURA I3. Esquema de una molécula de HLA de clase I.

alfa se divide en 3 regiones: una extracelular formada por 3 dominios (alfa 1, alfa 2, alfa 3), otra transmembrana y una tercera citoplásmica. La β 2m está acoplada al dominio alfa 3 mediante una unión no covalente el cual forma una subestructura que recuerda mucho a la de las inmunoglobulinas. Todo el complejo proteico está anclado a la membrana plasmática, de manera que el extremo amino terminal de la cadena alfa queda en el exterior celular y el extremo carboxilo terminal en el citosol (Abbas *et al.* 2019).

Las estructuras de las distintas proteínas HLA de clase I son muy similares entre sí debido a que la estructura de los genes que codifican para sus respectivas cadenas alfa es muy parecida. Como se ve en la Figura I4, los genes HLA de clase I están constituidos por 7 intrones y 8 exones (Malissen *et al.* 1982; Jordan *et al.* 1985). El exón 1 codifica para la región 5'UTR (del inglés, *untranslated region*) y para un péptido líder que contiene un núcleo hidrófobo característico de las proteínas que migran desde el retículo endoplasmático (RE). Los exones 2 y 3 son los más polimórficos y codifican respectivamente para los dominios alfa 1 y alfa 2, mientras que el exón 4 es menos polimórfico y codifica para el dominio alfa 3. El exón 5 codifica para el dominio transmembrana y el dominio citoplasmático está codificado por los exones 6 y 7 más parte del 8 en las moléculas HLA-A, HLA-C y HLA-F. El exón 8 codifica además para la región 3'UTR.

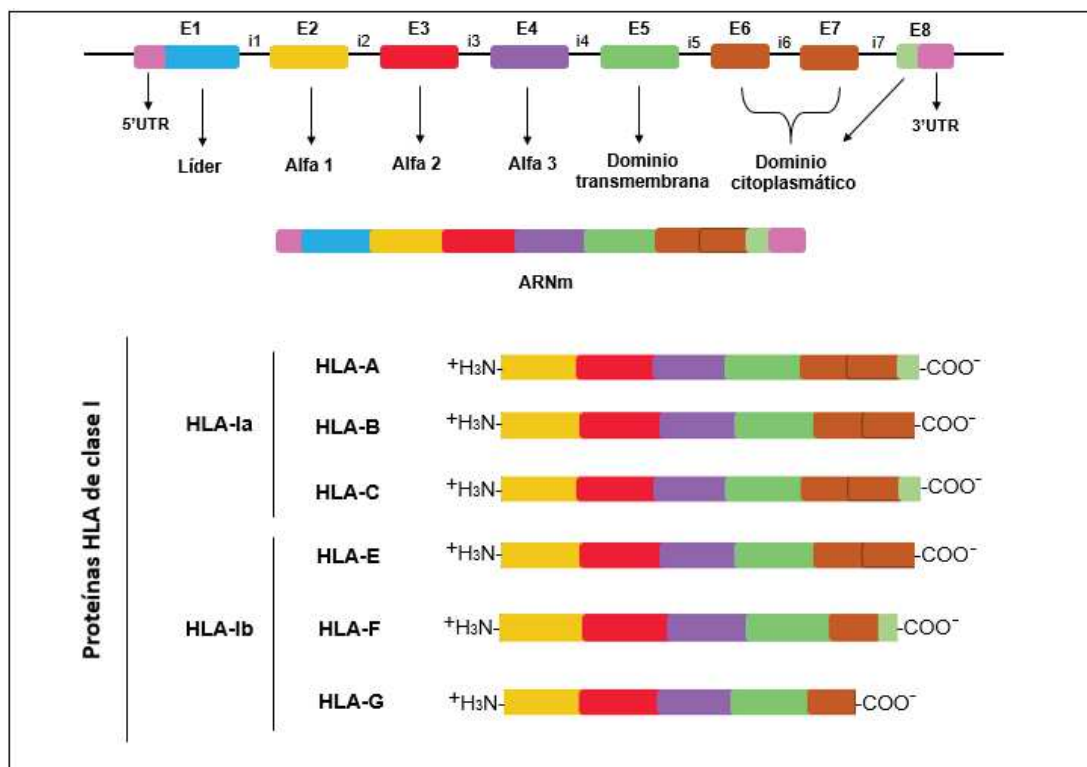


FIGURA I4. Estructura de los genes HLA de clase I. Exón (E). Intrón (i).

Las diferencias más notables que hay entre ellas se encuentran en la cola citoplasmática siendo unas más largas que otras pudiendo tener implicaciones funcionales. La molécula HLA-G prácticamente no tiene cola citoplasmática debido a un codón de terminación al inicio del exón 6 y las moléculas HLA-B y HLA-E presentan una región citoplasmática

más corta debido a la presencia de un codón de terminación al final del exón 7. En las moléculas HLA-Ia, la situación que ocupa el codón de terminación en el dominio citoplasmático es en el exón 8 para HLA-A y HLA-C, y en el exón 7 para HLA-B. Y en HLA-F, este dominio está codificado por el exón 6 y parte del 8, ya que el exón 7 es eliminado en el proceso de maduración del ARNm debido a una mutación puntual.

Los dominios alfa 1 y alfa 2 de las moléculas HLA de clase I forman una especie de concha, comúnmente denominada “valva” (Figura I5), que está formada por 8 láminas beta (4 de cada dominio) y 2 hélices alfa (1 de cada dominio), capaces de alojar un pequeño péptido (8-10 aminoácidos).

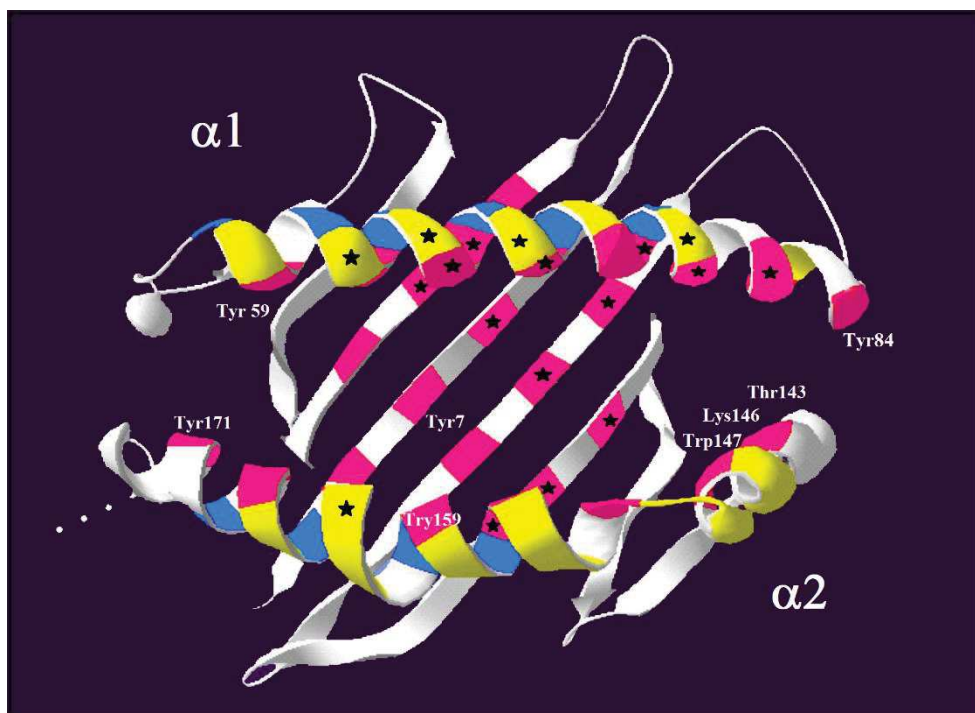


FIGURA I5. Estructura del sitio de unión al péptido de la molécula HLA de clase I formada por los dominios alfa 1 y alfa 2. Los aminoácidos que componen el sitio de reconocimiento antigénico (ARS) aparecen coloreados. Las cadenas laterales están dirigidas hacia el interior de la valva (en rosa, participan en la unión del péptido), hacia arriba (en amarillo, interaccionan con el TCR) o bien hacia el exterior de la valva (en azul, tal vez también interaccionan con el TCR) (Bjorkman *et al.* 1987). Las estrellas señalan las posiciones hipervariables descritas en las moléculas HLA de clase I clásicas (Parham *et al.* 1988). Los aminoácidos destacados no varían en las secuencias humanas y son los responsables del anclaje del péptido a través de sus extremos (Bjorkman *et al.* 1987).

La mayor variabilidad se encuentra en los dominios alfa 1 y alfa 2 y muchas de las posiciones variables forman parte del denominado sitio de reconocimiento antigénico (ARS, *Antigen Recognition Site*) (Parham *et al.* 1988), cuyos residuos interaccionan con los distintos péptidos y/o con el TCR (Bjorkman *et al.* 1987). Los residuos polimórficos relacionados con la unión al antígeno, se encuentran principalmente en el dominio alfa 1 mientras que en el dominio alfa 2 se encuentran mayoritariamente los residuos polimórficos relacionados con la unión al TCR (Parham *et al.* 1988). La variabilidad que presentan las moléculas HLA de clase I se traduce en cambios topológicos en los sitios de unión al péptido.

Las moléculas HLA de clase I no clásicas presentan la misma composición polipeptídica que las moléculas clásicas, pero con un polimorfismo en la región distal que conforma la valva muchísimo menor.

2.4 Moléculas HLA de clase II

Las moléculas HLA de clase II son proteínas que resultan de la expresión de los genes HLA de clase II clásicos (*HLA-DP*, *-DQ*, *-DR*) y no clásicos (*HLA-DM*, *-DO*) (Figura I6).

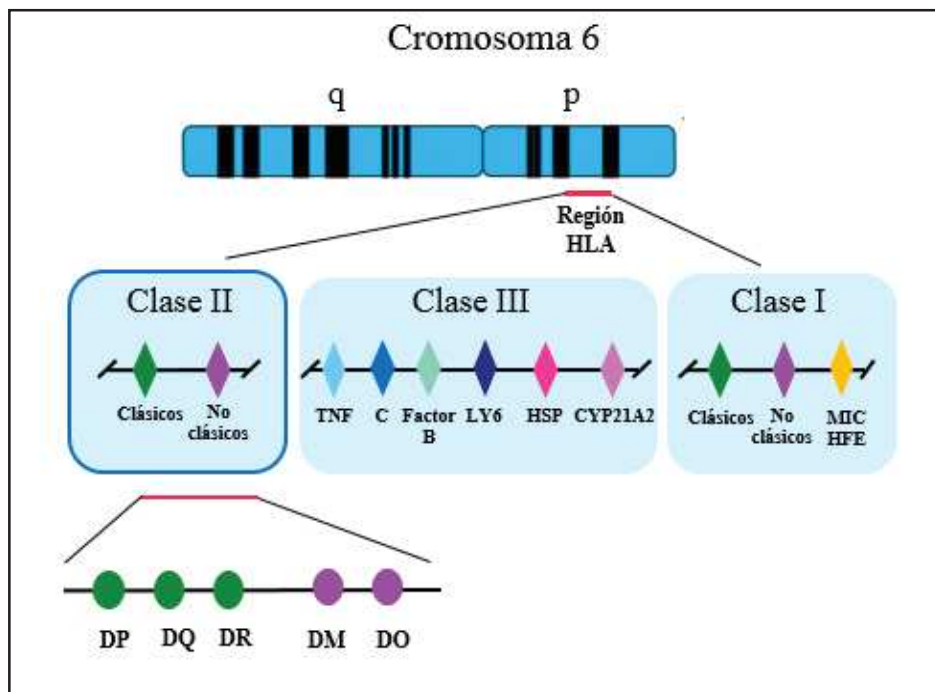
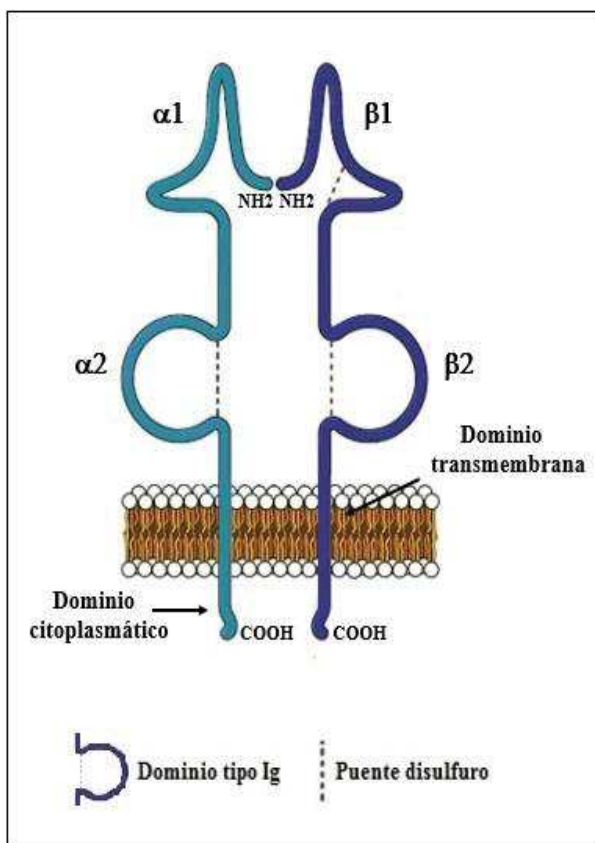


FIGURA I6. Genes de la región MHC de clase II.

Estos genes codifican para proteínas que están implicadas en fenómenos de restricción del reconocimiento antigénico mediado por linfocitos T cooperadores CD4⁺. La

distribución tisular de estas moléculas queda restringida fundamentalmente a células del sistema inmune presentadoras de antígenos: macrófagos, células DC, linfocitos B y, cuando están activados, también linfocitos T y células NK. Igualmente se expresan en alta densidad en otras células que actúan como presentadoras, tales como las células endoteliales.

Las moléculas HLA de clase II son proteínas de membrana constituidas por dos cadenas proteicas (denominadas alfa y beta) ensambladas mediante uniones no covalentes y ancladas a la membrana plasmática (Figura I7). La cadena alfa está formada por dos



dominios extracelulares (alfa 1 y alfa 2) y la cadena beta está formada por otros dos dominios extracelulares (beta 1 y beta 2). Ambas tienen un dominio transmembrana y una región citoplásmica. Los dominios extracelulares alfa 1 y beta 1 contienen los extremos amino terminales, mientras que los extremos carboxilo terminales se encuentran en la porción citoplásmica de cada cadena. Nuevamente, los dominios extracelulares más cercanos a la membrana (alfa 2 y beta 2) forman una subestructura similar a la de las inmunoglobulinas (Abbas *et al.* 2019).

FIGURA I7. Esquema de una molécula de HLA de clase II.

Las moléculas HLA de clase II están formadas por la unión de dos cadenas proteicas, cada una de ellas codificadas por un gen diferente, el gen A y el gen B (Figura I8). Los productos de estos genes se denominan cadena pesada alfa y cadena ligera beta de la molécula HLA de clase II. El gen A está formado por 5 exones. El exón 1 incluye la región 5'UTR y codifica para el péptido líder; los exones 2 y 3 codifican para los dominios alfa 1 y alfa 2 respectivamente; el exón 4 da lugar a la región transmembrana y

al dominio citoplasmático; la región 3'UTR está incluida en el exón 5. El gen B está formado por 6 exones. El exón 1 incluye la región 5'UTR y codifica para el péptido líder; los exones 2 y 3 codifican para los dominios $\beta 1$ y $\beta 2$ respectivamente; el exón 4 codifica para el péptido conector, el dominio transmembrana y parte del dominio citoplasmático; el resto de este dominio está originado por al menos un exón más que puede variar en los diferentes genes de esta región; la región 3'UTR está incluida en el último exón que presente el gen (Kappes *et al.* 1984).

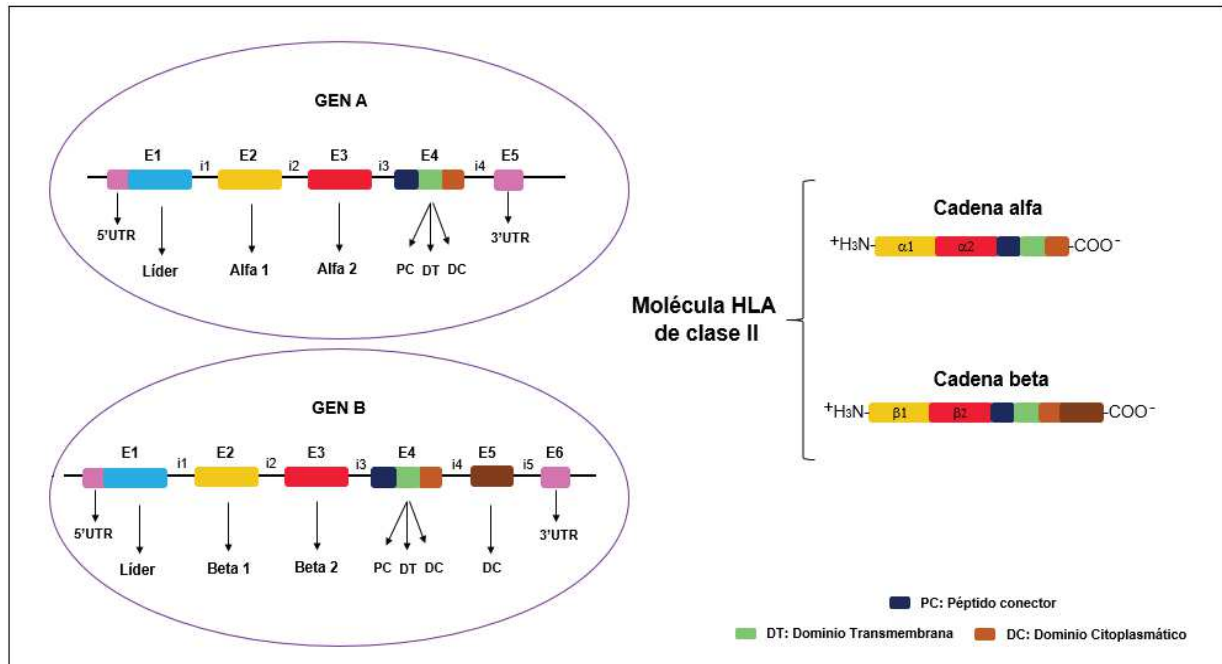


FIGURA I8. Estructura de los genes HLA de clase II. Exón (E). Intrón (i).

Los genes HLA de clase II son altamente polimórficos, pero este polimorfismo se restringe de forma casi exclusiva al exón 2 de los genes de tipo B (*HLA-DRB1*, *HLA-DQB1* y *HLA-DPBI*), mientras que los genes del tipo A apenas presentan polimorfismo.

La estructura tridimensional de las moléculas de clase II es muy similar a la de clase I descrita anteriormente. Los dominios alfa 1 y beta 1 en estas proteínas HLA de clase II conforman una valva muy parecida a la de las moléculas HLA de clase I, y también son capaces de unir péptidos, en este caso de mayor longitud (entre 12 y 24 aminoácidos). El polimorfismo también se concentra en el ARS y afecta fundamentalmente a la cadena beta (dominio beta 1), ya que la cadena alfa suele ser monomórfica en la mayoría de estas proteínas. La variabilidad está relacionada con la función presentadora de antígenos que desempeñan estas moléculas, que una vez unido el péptido (que es de origen extracelular,

generalmente bacteriano) contactan con el receptor de los linfocitos T cooperadores para activar la respuesta inmune.

2.5 Región HLA de clase III

Esta región se sitúa entre las regiones de clase I y de clase II y se trata de la zona con mayor densidad génica del genoma humano, lo que provoca un elevado desequilibrio de ligamiento entre los genes de esta región.

Los genes que se localizan en esta zona son tanto de tipo inmunológico como no inmunológico. Entre los primeros se encuentran aquellos que expresan proteínas con función inmunológica como los genes TNF (*TNF-A* y *-B*), los genes del complemento de la vía clásica (*C4A*, *C4B* y *C2*) y de la vía alternativa (gen del Factor B) y los genes LY6. En cuanto a los genes de tipo no inmunológicos, se localizan los genes HSP, que expresan proteínas que intervienen en procesos de síntesis, plegamiento, ensamblaje, transporte y degradación de proteínas, aumentando su expresión en condiciones de estrés celular; y el gen *CYP21A2* que codifica para la esteroide 21-hidroxilasa, miembro de la superfamilia de enzimas del citocromo P450 que catalizan reacciones involucradas en la síntesis del colesterol y otro lípidos (Abbas *et al.* 2017).

3. El complejo principal de histocompatibilidad en el pollo

El complejo principal de histocompatibilidad del pollo (*Gallus gallus*) es el primer MHC descrito en vertebrados y el mejor estudiado en aves. A finales de los años 40 fue descrito por Briles como parte de un sistema de grupos sanguíneos (Briles *et al.* 1950) y a esta parte altamente polimórfica le denominó B. Más tarde, se vio que este sistema no sólo determinaba aloantígenos eritrocitarios, sino que también controlaba el rechazo a aloinjertos (Schierman & Nordskog 1961), participaba en reacciones linfocitarias (Miggiano *et al.* 1974) y participaban en la cooperación celular (Vainio *et al.* 1984). La existencia de este complejo B fue corroborada en investigaciones posteriores (Guillemot *et al.* 1988) y estudios moleculares demostraron la existencia de otra región de histocompatibilidad a la que se le denominó sistema Rfp-Y, del inglés *Restriction fragment pattern-Y* (Miller *et al.* 1994; Miller *et al.* 1996). Ambas regiones fueron secuenciadas a partir de especies pertenecientes a la raza Leghorn, obteniéndose así el primer mapa completo de la región MHC del pollo (Kaufman *et al.* 1999). La raza

Leghorn es la más utilizada en investigación y alimentación y descende de una población de pollos salvajes que habitaba en las selvas del sudeste asiático (Komiyama *et al.* 2003). El material genético del pollo está distribuido en 39 parejas de cromosomas, 5 de gran tamaño (macro cromosomas) y 34 de pequeño tamaño (micro cromosomas). El MHC del pollo se encuentra ubicado en el micro cromosoma 16 y está formado por la región MHC-B y la región MHC-Y (Figura I9). Además de estas regiones de histocompatibilidad, se encuentran otros genes de gran importancia como son los genes *CD1*, localizados en la parte más telomérica del micro cromosoma, o los genes que codifican para los receptores olfativos y receptores basureros, ubicados justo antes del centrómero. Las regiones MHC-B y MHC-Y se encuentran separadas por una serie de repeticiones en tándem ricas en guanina y citosina denominada región PO41, la cual juega un papel importante en la formación de puntos calientes de recombinación entre los complejos MHC-B y MHC-Y (Solinhaç *et al.* 2010). A continuación de la región MHC-Y, está la región organizadora nucleolar (NOR). A día de hoy, los genes que se encuentran en el brazo corto de este micro cromosoma no han sido mapeados, y por tanto, no se conoce nada de ellos aún (Miller & Taylor Jr. 2016).

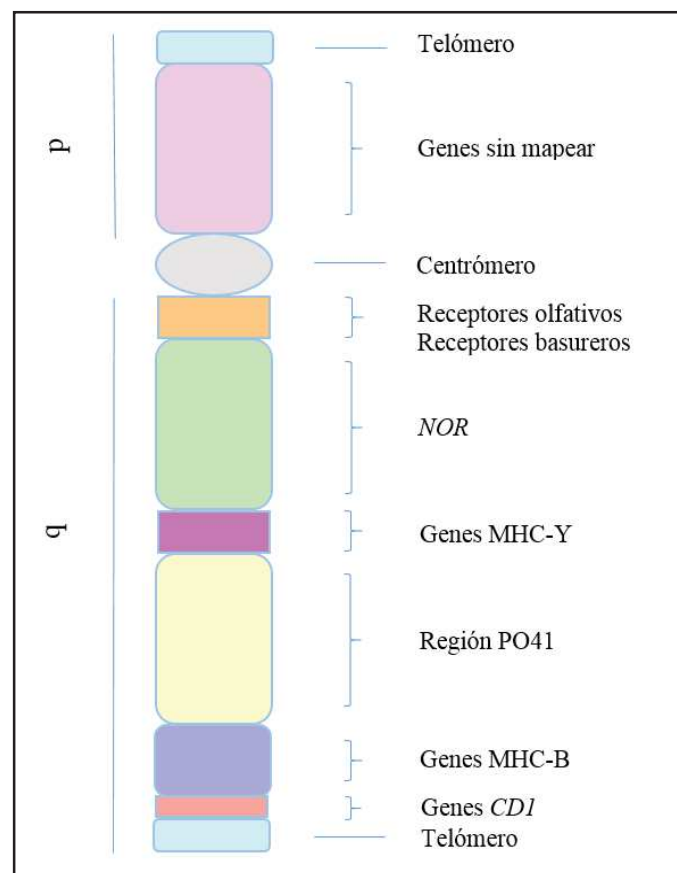


FIGURA I9. Esquema del microcromosoma 16 del pollo.

3.1 Región MHC-B

La región MHC-B ha sido, desde su descubrimiento, la región de histocompatibilidad mejor estudiada debido al papel fundamental que juega en la presentación antigénica y la resistencia a enfermedades. Está dividida en cinco subregiones: la subregión que contiene los genes del MHC de clase I y II, la subregión que contiene la familia de genes BG, la subregión de los genes del complemento, la subregión de los genes TRIM y Blec y la subregión de los genes CD1 (Figura I10).

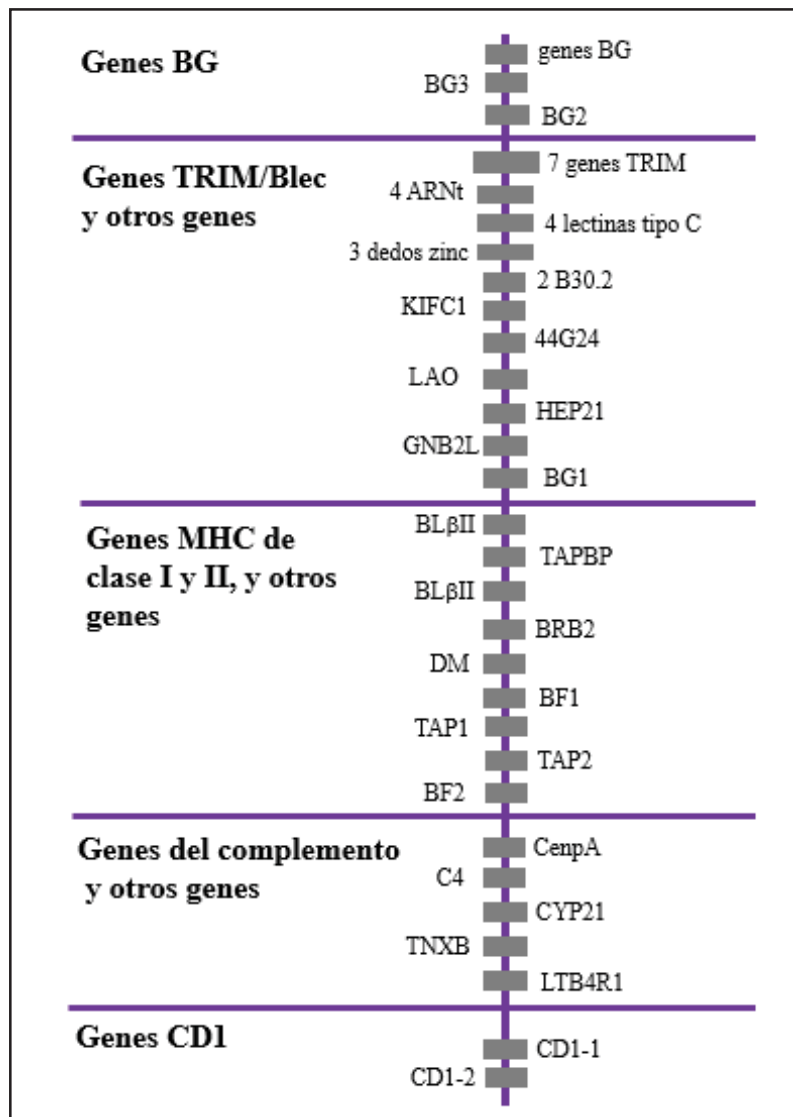


FIGURA I10. Mapa genético de la región MHC-B del pollo.

- **Subregión de genes MHC de clase I y II**

Dentro de esta subregión hay dos genes del MHC-I, dos genes del MHC- β II y genes que se encargan del procesamiento y carga del péptido. Estos genes están muy próximos entre sí.

Los genes del MHC de clase I, llamados *BF1* y *BF2*, codifican para moléculas clásicas del MHC-I. Ambos genes son muy similares, pero *BF2* está más involucrado en la presentación antigénica y *BF1* codifica para moléculas que pueden servir de ligando a las células NK. El grado de expresión de *BF2* es mayor que el de *BF1* (Miller & Taylor Jr. 2016).

Los genes del MHC- β II, llamados *BL β II*, codifican para la cadena β de las moléculas del MHC-II. Se trata de dos genes que son descritos como gen principal y gen menor, siendo el principal el que más se expresa. Con respecto a la cadena α de las moléculas del MHC-II, se ha encontrado una única copia de este gen localizado a 5 centimorgan de la región MHC-B (Kaufman *et al.* 1995).

Los otros genes que se encuentran en esta subregión son el gen de la tapasina (*TAPBP*), que se encuentra entre los genes *BL β II* y se encarga del transporte de péptidos; el gen *BRB2* que codifica para la quinasa RING3; los genes TAP, formados por las subunidades TAP1 y TAP2, son transportadores del retículo endoplasmático que transportan los péptidos desde el citosol al interior del lumen donde los péptidos son cargados en las moléculas de clase I; y los genes DM involucrado en la presentación del péptido de las moléculas de clase II (Kaufman *et al.* 1999).

- **Subregión de la familia de genes *BG***

La familia BG está formada por 14 genes y casi todos estos genes están agrupados en repeticiones en tándem dando lugar a distintos haplotipos. Sólo dos genes se encuentran al margen de este grupo. Uno de ellos es *BG0*, localizado fuera de la región MHC, concretamente en el cromosoma 2; y el otro es *BG1*, localizado cerca de los genes del MHC de clase I y II (Miller & Taylor Jr. 2016). Todos los genes BG son polimórficos, pero se desconoce cómo contribuyen a la inmunidad. A día de hoy se sabe que algunos miembros de la familia BG se expresan en eritrocitos, donde son reconocidos como aloantígenos. Las proteínas BG también se expresan en células epiteliales, particularmente en las del duodeno, ciego e hígado (Miller *et al.* 1990). Y, además, estos genes BG se expresan en muchos tipos de células inmunológicas, como son los linfocitos

T y B, trombocitos y células dendríticas entre otros (Salomonsen *et al.* 1991). El único gen del que se tiene más conocimiento es el *BG1*, que participa en la respuesta inmunitaria y juega un papel importante en la resistencia a la enfermedad de Marek (Miller & Taylor Jr. 2016).

- **Subregión de los genes del complemento**

A esta subregión también se le conoce como región de clase III y contiene 5 genes. Cuatro de ellos presentan un alto grado de conservación y son *C4*, que codifica para el factor del complemento C4; *CenpA*, se trata de una proteína del centrómero; *CYP21*, que codifica para la esteroide 21-hidroxilasa; y *TNXB* que codifica para la tenascina XB. El quinto gen es *LTB4R1* y codifica para un receptor de leucotrieno (Chen *et al.* 2015).

- **Subregión TRIM/Blec**

Dentro de la subregión TRIM/Blec hay 25 genes, además del gen *BG1* ya descrito anteriormente. Parece ser que todos los genes localizados en esta región contribuyen de alguna manera a la defensa inmunitaria. Los genes TRIM participan en la respuesta inmune innata defendiendo al organismo contra infecciones virales. Los genes Blec son cuatro y codifican para lectinas tipo C dependientes del calcio, las cuales tienen una gran variedad de funciones en la respuesta inmunitaria. Entre estos genes destacan *B-NK* y *B-lec*, también conocido como *Blec1*, que son similares a los receptores mamíferos NKR-P1 y LLT1, respectivamente. *B-NK* se expresa en las células NK y *B-lec* se expresa en los linfocitos (Roger *et al.* 2005).

- **Subregión genes CD1**

Los genes CD1 del pollo, *CD1-1* y *CD1-2*, están localizados muy cerca de los genes del MHC de clase I. En mamíferos, los genes CD1 son cinco y están localizados en el cromosoma 1. Se tratan de moléculas presentadoras de antígenos de naturaleza lipídica y glucolipídicas a linfocitos T y contribuyen a la defensa del organismo contra infecciones bacterianas. No se conoce mucho sobre las dos moléculas CD1 del pollo, aunque su estructura sí que ha sido estudiada. Ambas estructuras confirmaron su relación con otros miembros de esta familia de genes y revelaron sitios de unión característicos no presentes en otras moléculas CD1. La estructura del sitio de unión de la molécula CD1-2 muestra una hendidura primitiva muy pequeña que sugiere que el pollo sólo puede reconocer

cadena de lípidos simples de no más de 16 carbonos (Zajonc *et al.* 2008). Por otro lado, la estructura cristalina de la isoforma CD1-1 presenta un sitio de unión mucho más elaborado con una hendidura con capacidad para unir cadenas de lípidos más grandes (Dvir *et al.* 2010). Esta molécula es más similar a la molécula CD1 de los mamíferos con respecto a su capacidad para unir lípidos compuestos.

3.2 Región MHC-Y

La región MHC-Y, originariamente llamada Rfp-Y, fue descubierta cuando fragmentos de restricción de la región MHC-B hibridaron con genes localizados en otra región diferente (Briles *et al.* 1993). Este gran hallazgo supuso que varios genes del MHC y de lectina que se pensaba que estaban en la región MHC-B se reasignaran a la nueva región descubierta.

La región MHC-Y está formada por dos genes del MHC-I, tres genes del MHC-II y dos genes que codifican para receptores de leucina dependientes de calcio. Los dos genes del MHC-I se denominan *YF* y codifican para la molécula BF del MHC-I. Los tres genes del MHC-II se denominan *YLB1*, *YLB2* e *YLB3* y codifican para la molécula BL del MHC-II.

Por último, los genes de receptores de leucina dependientes de calcio se denominan *Ylec* y juegan un papel muy importante en la respuesta inmune innata ya que las células NK lo presentan en su superficie celular (Miller & Taylor Jr. 2016).

3.3 Otros genes fuera de la región de histocompatibilidad

Además de las regiones de histocompatibilidad MHC-B y MHC-Y, en el microcromosoma 16 se encuentran localizados otros genes de gran importancia como son los genes que codifican para los receptores olfativos y los receptores basureros (Miller *et al.* 2014). Estos genes segregan de manera conjunta con la región MHC-Y, ya que ambas regiones están muy cerca entre sí. Entre los genes olfativos destacan los genes *OR14J1*, los cuales también se encuentran en la región extendida del MHC humano. Y con respecto a los genes basureros, son 9, presentan un dominio rico en cisteína y están implicados en la respuesta inmune (Miller & Taylor Jr. 2016).

3.4 Susceptibilidad a enfermedades del pollo

En el MHC del pollo, existe una fuerte vinculación entre la presencia de haplotipos concretos y la capacidad del individuo para hacer frente a enfermedades infecciosas concretas. La mayoría de los haplotipos que determinan resistencia o susceptibilidad a ciertas enfermedades se encuentran en la región MHC-B cuyo papel ha sido estudiado en una amplia variedad de enfermedades causadas por virus, bacterias y parásitos patógenos. Por el contrario, hay muy pocos estudios sobre la contribución del polimorfismo de MHC-Y a la respuesta contra infecciones. Uno de los ejemplos más conocidos de resistencia a patógenos es la enfermedad de Marek la cual es provocada por un herpesvirus que induce una proliferación tumoral de células linfoides en gran número de órganos y tejidos, y que puede causar un grave peligro económico para las granjas de adultos jóvenes. Los individuos que tienen el haplotipos *B21* o el *B19* presentan resistencia a esta enfermedad (Briles *et al.* 1977). Otra de las enfermedades bien estudiadas es el sarcoma de Rous, causado por un retrovirus que provoca tumores en los pollos. En este caso, cuando se cruzaron dos líneas homocigóticas (*B13B13* y *B6B6*) para estudiar su respuesta contra el sarcoma de Rous, se observó que los individuos heterocigotos presentaban una regresión del tumor en un 94% (Schierman *et al.* 1977). También, hay haplotipos del MHC-B que influyen en la resistencia a determinadas bacterias como *Staphylococcus aureus*, *Pasteurella multocida* o especies del género *Salmonella*, y a determinados parásitos como el nematodo *Ascaridia galli* o coccidios del género *Eimeria* (Miller & Taylor Jr. 2016).

3.5 Hipótesis del MHC "mínimo esencial"

El MHC del pollo presenta una serie de características muy particulares que no se dan en el sistema HLA humano ni en otros MHC de mamíferos. Estas peculiaridades son:

- Los genes localizados en esta región se encuentran muy próximos entre sí
- No todos los genes de histocompatibilidad se encuentran contenidos en la región MHC
- Sus genes poseen intrones muy cortos

Estas características llevaron a Kaufman y colaboradores (1999) a proponer la hipótesis del MHC “mínimo esencial” que defiende que este sistema contiene sólo el mínimo número de genes absolutamente esenciales para la función del MHC en un sentido

evolutivo (Kaufman *et al.* 1995; Kaufman *et al.* 1999). Kaufman descubrió este MHC tan singular en el pollo y lo generalizó a todas las aves.

4. El complejo principal de histocompatibilidad en aves no *Passeriformes*

Tras estudiar en profundidad el complejo principal de histocompatibilidad del pollo, muchos investigadores llevaron sus líneas de investigación hacia el estudio de otras aves con el objetivo de dilucidar la evolución del MHC, además de preservar la conservación de las especies y mantener el medio ambiente donde han vivido a lo largo del tiempo. Algunos de los complejos principales de histocompatibilidad mejores estudiados fuera del orden *Passeriformes*, sin contar el del pollo, son el del faisán común (*Phasianus colchicus*), en el que se han identificado dos genes polimórficos ortólogos de los genes *BLβII* de la región MHC-B del pollo (Witzell *et al.* 1999); el del faisán dorado (*Chrysolophus pictus*), cuya región MHC-B contiene 20 genes de histocompatibilidad que presentan una sintenia similar a la del pollo (Ye *et al.* 2012), el de la codorniz japonesa (*Coturnix japonica*), cuya región MHC es de 180 kb englobando un total de 41 genes, en comparación con las 92 kb de la región genómica del pollo que contienen 19 genes (Shiina *et al.* 2004) presentando ambas especies genes del MHC homólogos entre sí (Shiina *et al.* 1999); el del pavo (*Meleagris gallopavo*), que presenta dos regiones homólogas a la región MHC-B y MHC-Y del pollo localizada en el mismo microcromosoma que contiene la región organizadora nucleolar (Chaves *et al.* 2009); el del urogallo (*Tetrao tetrix*), que contiene 19 genes en la región MHC-B con una sintenia prácticamente idéntica a la de *G. gallus*, a excepción de una inversión en el orden de los genes *TAPBP* y *TAP* (Wang *et al.* 2012); y el del pato (*Anas platyrhynchos*), cuyos genes del MHC de clase I contienen cinco tipos de genes de los cuales sólo uno es el que más se expresa (Moon *et al.* 2005).

5. El complejo principal de histocompatibilidad en aves *Passeriformes*

El orden *Passeriformes* engloba a las denominadas aves pequeñas silvestres de canto que representan aproximadamente la mitad de las casi 10.000 especies de aves existentes (Sibley 1996). El estudio del complejo principal de histocompatibilidad en *Passeriformes* es importante porque a estas aves se les consideran modelos naturales en los que la evolución del polimorfismo del MHC ha sido adquirida a lo largo de millones de años

(MA). Por ello, es interesante comparar este polimorfismo con el de modelos artificiales como el pollo, el ser humano o el ratón en los que poblaciones pequeñas han crecido rápidamente por consanguinidad y se han convertido en poblaciones relativamente grandes. Otro argumento a favor del estudio del MHC de estas aves es que es fácil identificar las especies y es muy cómodo manejar estas muestras. Además, presentan una amplia distribución geográfica y hay un gran número de especies pertenecientes a este orden. Todo ello, permite realizar un estudio evolutivo de los genes del MHC y un análisis de los diferentes procesos selectivos-ambientales que inducen su variación.

A día de hoy no hay un MHC de aves *Passeriformes* tan bien estudiado como en el caso del pollo. Sin embargo, se ha estudiado el MHC de diversas aves *Passeriformes* y en todas se ha visto que el número de genes presentes en esta región es mayor que en aves pertenecientes al orden *Galliformes*. Los genes de clase IIB están ampliamente duplicados en *Passeriformes*, aunque todavía queda mucho por descubrir sobre la expresión de estos genes; y el número de genes de clase I expresados en estas especies es mayor que en el pollo. Uno de los *Passeriformes* mejor estudiado es el diamante mandarín (*Taeniopygia guttata*), la primera especie perteneciente a este orden cuyo genoma ha sido secuenciado y su MHC ha sido caracterizado. Con la secuenciación y caracterización del MHC de esta especie, *Taeniopygia guttata* se ha convertido en un modelo para el estudio de otros genomas de aves *Passeriformes*. El MHC del diamante mandarín presenta múltiples genes de clase I y II, incluyendo algunos pseudogenes, que abarcan una región genómica mucho más extensa que la del MHC del pollo. Se ha visto que esta región está distribuida por varios cromosomas, aunque todavía no se sabe la localización exacta de ciertos genes. Los genes *Blecl* y *NKR* están localizados en el cromosoma Z y los genes CD1 se encuentran en el cromosoma 12. Los genes del MHC de clase I parece ser que están en el cromosoma 22, pero los genes de clase II aún no ha sido mapeados. Un hecho interesante es que los genes TAP y los genes del MHC-I están localizados en distintos cromosomas, a pesar de ser genes que funcionalmente están muy relacionados entre sí y de que presentan sintenia en muchos MHC estudiados, incluyendo al pollo y al ser humano (Balakrishnam *et al.* 2010).

Los genes del MHC han sido secuenciados y estudiados en otras aves pertenecientes al orden *Passeriformes* y en todas ellas se ha visto que presentan entre dos y ocho loci del MHC-I, aunque muchos estudios apuntan que puede haber incluso más loci (Karlsson & Westerdahl 2013). Entre estas especies destacan el carricero tordal (*Acrocephalus*

arundinaceus) (Westerdahl *et al.* 1999), el camachuelo carminoso (*Carpodacus erythrinus*) (Promerová *et al.* 2009) y el herrerillo común (*Cyanistes caeruleus*) (Schut *et al.* 2011). Otra especie que hay que resaltar por su importancia en la evolución y la ecología es de gorrión común (*Passer domesticus*), un ave que ha colonizado numerosos ecosistemas del planeta y que probablemente esté sujeta a una presión patógena variable. La caracterización de los genes del MHC de clase I y clase II reveló que el gorrión común tiene menos loci y menos alelos que el carricero tordal. Además, los genes del MHC de clase I están divididos en dos grupos distintos con diferentes niveles de polimorfismo, posiblemente indicando diferentes papeles funcionales para cada familia (Bonneaud *et al.* 2004).

▪ MHC y enfermedad

Los genes del complejo principal de histocompatibilidad tienen un papel crucial en la resistencia a enfermedades y su alto polimorfismo hace que sean unos candidatos excepcionales para el estudio de las variaciones genéticas. La diversidad alélica del MHC se mantiene si los diferentes alelos dan protección al individuo contra los diferentes parásitos que hay en su entorno. Por ello, se han encontrado múltiples asociaciones entre ciertos alelos del MHC y resistencias a enfermedades en numerosas especies como el ser humano, el ratón, el pollo o la oveja (Miller & Taylor Jr. 2016). En aves *Passeriformes* no se conocen tantas asociaciones, pero existen algunas y el caso de la resistencia o susceptibilidad a la malaria es una de ellas muy bien conocida. Los parásitos que pueden infectar de malaria a las aves son especies de los géneros *Plasmodium*, *Haemoproteus* y *Leucocytozoon* entre otros, y dependiendo de las cepas los genes del MHC pueden proteger contra una infección por malaria o bien, pueden conferir protección frente a una malaria letal y direccionarla a cuadros más leves. En el carricero tordal se observó una asociación positiva entre el número de alelos de clase I y la susceptibilidad a desarrollar malaria de la cepa GRW2 (Westerdahl *et al.* 2005). Estudios similares se llevaron a cabo en dos poblaciones de gorrión común y se vio que la cepa SGS1 era la que más prevalecía en los individuos analizados con un porcentaje del 54% (Bonneaud *et al.* 2006). Se ha visto que la diversidad del MHC puede afectar a la supervivencia, ya que los individuos con un gran número de alelos diferentes pueden reconocer y combatir un amplio rango de patógenos y es más probable que puedan sobrevivir durante la cría. La presencia del alelo 1105 en el gorrión común mostró una probabilidad muy alta de supervivencia en

polluelos de esta especie (Karlsson *et al.* 2015). Otros estudios llevados a cabo han sido analizar la asociación que hay entre los alelos MHC de clase I y ciertas infecciones parasitarias en el herrerillo común. En ellos se observaron que los alelos UA104, UA108 y UA117 estaban asociados al grado de intensidad de la infección producida por *Leucocytozoon* y además, estaban relacionados con la edad de los individuos. Los individuos adultos que presentaban el alelo UA104 presentaban infecciones más leves que las presentadas por los individuos jóvenes (Rivero-de-Aguilar *et al.* 2016).

II. INTRONES

1. Definición e historia

Los intrones son fragmentos de ADN localizados dentro de un gen, que no codifican ningún fragmento de la proteína, ya que son eliminados en el proceso de maduración del ARN. Los genes eucariotas están constituidos por intrones y exones y el descubrimiento de que estos genes sean discontinuos se lo debemos a Richard J. Roberts y Phillip A. Sharp en 1977. A comienzos de los años setenta Richard J. Roberts trabajó en el laboratorio de Cold Spring Harbor de Nueva York, donde coincidió con Phillip A. Sharp. Ambos investigadores trabajaban en la transcripción de los adenovirus y probaron por separado que los genes que codifican polipéptidos en los eucariotas se encuentran interrumpidos por secuencias no codificantes, ya que observaron que había segmentos de los genes virales que no se encontraban presentes en sus correspondientes ARN mensajeros (Berget *et al.* 1977; Chow *et al.* 1977). El descubrimiento de los intrones supuso una gran revolución en la genética molecular porque hasta la fecha los genes se concebían como bloques contiguos de secuencias codificadoras de proteínas; por ello, el 11 de octubre de 1993 ambos investigadores fueron galardonados con el Premio Nobel de Medicina por sus hallazgos sobre la estructura de los genes.

2. Maduración del ARN mensajero

El ARN mensajero contiene la información genética procedente del ADN para la síntesis de proteínas. Es transcrito por la ARN polimerasa II en el núcleo y las moléculas de pre-ARNm obtenidas deben sufrir diferentes procesos de maduración antes de salir al citoplasma. Entre estos procesos postranscripcionales se encuentra la eliminación de los intrones y la unión de los exones mediante un mecanismo de corte y empalme, comúnmente denominado como *splicing*. Los intrones que se eliminan mediante este mecanismo se conocen con el término de intrones espliceosomales y los más frecuentes son los intrones GU-AG, que se denominan así porque los primeros dos nucleótidos de la secuencia del intrón son 5'-GU-3' y los dos últimos 5'-AG-3'. Los motivos GU-AG son partes de secuencias conservadas que abarcan los sitios de corte y empalme (Figura I11). Estas secuencias varían en diferentes tipos de eucariotas y en los vertebrados se las puede describir como:

Sitio de corte y empalme 5': 5'-AG↓GUAAGU-3'

Sitio de corte y empalme 3': 5'-PyPyPyPyPyNCAG↓-3'

En estas designaciones, "Py" es uno de los dos nucleótidos pirimidínicos (U o C), "N" es cualquier nucleótido y la flecha indica el límite exón-intrón.

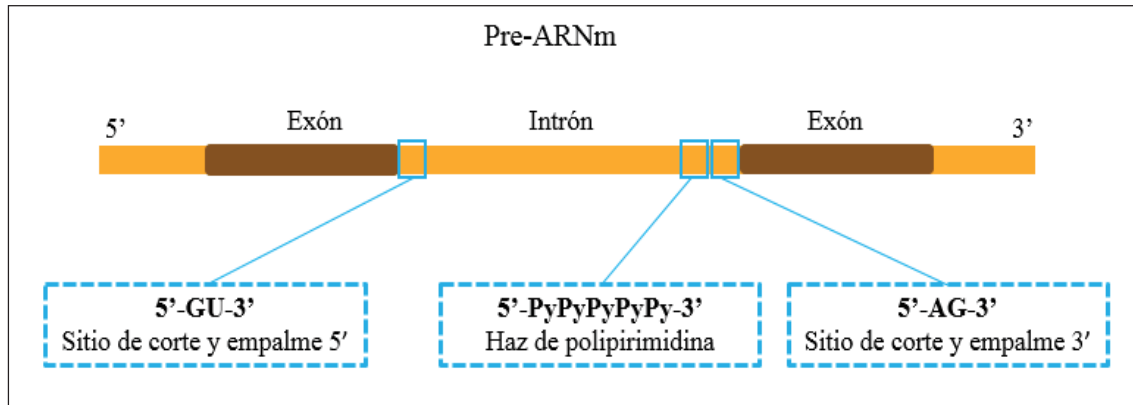


FIGURA I11. Secuencias conservadas en intrones de vertebrados

Como se observa en la Figura I12, la vía de corte y empalme se puede dividir en dos pasos:

- La escisión del sitio de corte y empalme 5' se produce por una reacción de transesterificación promovida por los grupos oxhidrilos unidos al carbono 2' de un nucleótido de adenosina localizado dentro de la secuencia del intrón. El resultado del ataque del oxhidrilo es la degradación de un enlace fosfodiéster en el sitio de corte y empalme 5', acompañada por la formación de un nuevo enlace fosfodiéster 5', que es el que une el primer nucleótido del intrón con la adenosina interna. Esto significa que el intrón ahora ha formado un bucle hacia atrás sobre sí mismo para crear una estructura en lazo.
- La escisión del sitio de corte y empalme 3', y la unión de los exones obedece a una segunda reacción de transesterificación, promovida por el grupo 3'-OH unido al extremo del exón corriente arriba. Este grupo ataca el enlace fosfodiéster en el sitio de corte y empalme 3', lo degrada y libera así al intrón como la estructura en lazo, que después vuelve a convertirse en ARN lineal y se degrada. Al mismo tiempo, el extremo 3' del exón corriente arriba se une al extremo 5' recién formado del exón corriente abajo, lo que completa el proceso de corte y empalme.

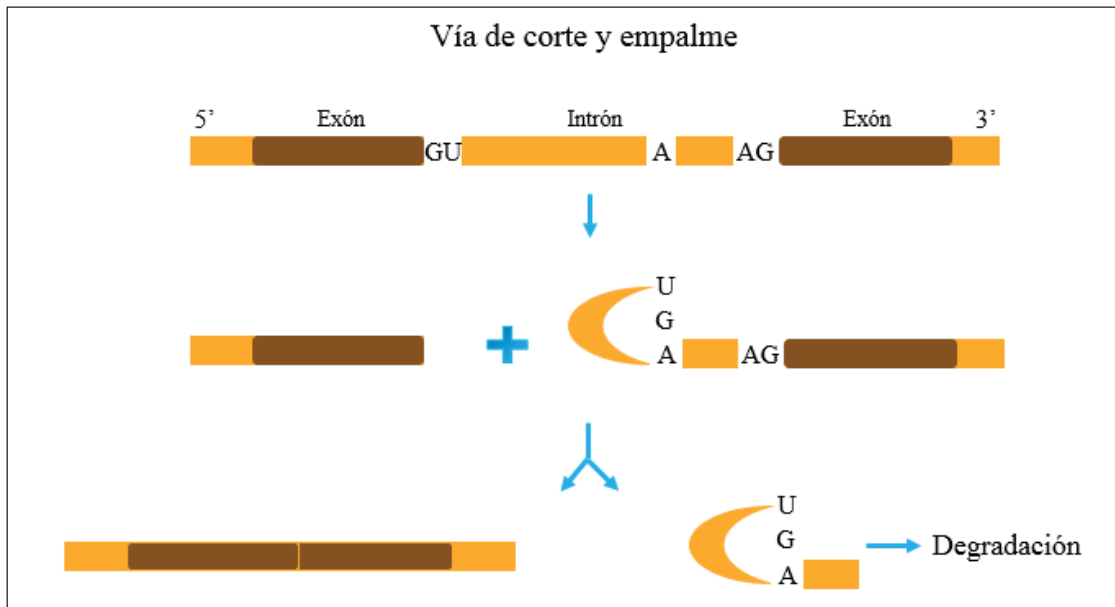


FIGURA I12. Esquema de la vía de corte y empalme.

Este proceso de corte y empalme no representa un gran desafío para la célula ya que se trata de una doble reacción de transesterificación, no más complicada que muchas otras reacciones bioquímicas que se dan en el medio celular. Sin embargo, se ha desarrollado una maquinaria compleja que se ocupa de este proceso para evitar ciertos problemas. El primero de ellos es que los sitios de corte y empalme pueden encontrarse a una distancia sustancial y es necesario que se aproximen para su correcta eliminación. Y el segundo de ellos concierne a la selección del sitio de corte y empalme correcto. Todos los sitios de *splicing* son similares, de manera que si un pre-ARNm contiene dos o más intrones, existe la posibilidad de que se puedan unir sitios de corte y empalme incorrectos, lo que determina la pérdida de un exón del ARNm maduro (Brown 2017).

3. Clasificación de intrones

La presencia de intrones en el genoma es una característica propia de los organismos eucariotas, sin embargo algunos procariotas también presentan intrones en su material genético. En la actualidad, se conocen 8 tipos de intrones y, a continuación, se describen brevemente cada uno de ellos (Brown 2017).

- Los **intrones de los pre-ARNm nucleares de la clase GU-AG** se encuentran en genes que codifican proteínas de eucariotas y el proceso de eliminación se explica detalladamente en el apartado 2 del bloque Intrones.

- Los **intrones de los pre-ARNm nucleares de la clase AU-AC** presentan una vía de corte y empalme muy similar a la de los intrones GU-AG, pero a diferencia de ellos, involucra un grupo diferente de factores de corte y empalme. Este tipo de intrones se han identificado en alrededor de 20 genes de organismos tan diversos como seres humanos, planta e insectos como *Drosophila*.
- Los **intrones de los pre-ARNt** son relativamente comunes en los eucariotas inferiores, pero menos frecuentes en los vertebrados. El corte y empalme de estos intrones no implica transesterificaciones, sino que en este caso los dos sitios de corte y empalme son cortados por una endonucleasa. Esta enzima contiene cuatro subunidades no idénticas, una de las cuales emplea la estructura del intrón apareada por sus bases como guía para identificar las posiciones correctas en las que se debe cortar el ARN. Después, dos de las otras subunidades de la enzima practican los cortes corriente arriba y corriente abajo. La escisión deja una estructura de fosfato cíclico unida al extremo 3' del exón corriente arriba y un grupo oxhidrilo en el extremo 5' de exón corriente abajo. El fosfato cíclico es convertido en un 3'-OH por una fosfodiesterasa y el extremo 5'-OH es convertido en 5'-P por una cinasa. Estos dos extremos se mantienen cerca por el apareamiento de bases natural adoptado por la secuencia de ARNt y son ligados por una ARN ligasa. Una sola proteína aporta las actividades de fosfodiesterasa, cinasa y ligasa.
- Los **intrones del grupo I** se encuentran en pre-ARNm y pre-ARNr (ARN ribosómico) de genomas de mitocondrias y cloroplastos, en algunos pre-ARNr eucariotas microbianos y en ciertos ARN de bacterias. La vía de corte y empalme de este tipo de intrones es similar a la de los intrones del pre-ARNm, ya que se producen dos transesterificaciones. La primera es inducida, no por un nucleótido dentro del intrón, sino por un nucleósido o un nucleótido libre, alguno de guanosina o guanosina mono, di o trifosfato. El 3'OH de este cofactor ataca el enlace fosfodiéster del sitio de corte y empalme 5', y lo degrada, con transferencia de la G al extremo 5' del intrón. La segunda transesterificación involucra el 3'-OH del extremo del exón, que ataca al enlace fosfodiéster del sitio de corte y empalme 3', lo que causa su degradación, la unión de los dos exones y la liberación del intrón. La principal característica de esta vía de corte y empalme es que tiene lugar en ausencia de proteínas, y por lo tanto, es autocatalítica: el propio ARN tiene actividad enzimática.

- Los **intrones del grupo II** se hallan en los orgánulos de hongos y plantas, tanto en los pre-ARNm como en los pre-ARNr, y en algunos casos también en procariotas. Adoptan una estructura secundaria característica y pueden autocortarse y empalmarse en el tubo de ensayo, pero son distintos a los intrones del grupo I. La estructura secundaria es diferente, y el mecanismo de corte y empalme está más estrechamente relacionado con el de los intrones del pre-ARNm: la transesterificación inicial es promovida por el grupo oxhidrilo de un nucleótido de adenosina interno y el intrón se convierte en una estructura en lazo. Estas similitudes han llevado a sugerir que los intrones del grupo II y los del pre-ARNm pueden tener un origen evolutivo común.
- Los **intrones del grupo III** también se encuentran en genomas de orgánulos, y se autocortan y empalman mediante un mecanismo muy similar al de los intrones del grupo II, pero son más pequeños y tienen su propia estructura secundaria distintiva. La similitud con los intrones del grupo II vuelve a sugerir una relación evolutiva.
- Los **intrones gemelos** son estructuras compuestas formadas por dos o más intrones del grupo II o del grupo III. Los intrones gemelos más simples consisten en un intrón sepultado en otro, pero los más complejos contienen múltiples intrones sepultados. Por lo general, cada intrón de un intrón gemelo es cortado y empalmado en una secuencia definida.
- Los **intrones arqueobacterianos** están presentes en los genes de ARNt y ARNr. Son escindidos por una ribonucleasa similar a la que participa en el corte y empalme del pre-ARNt eucariota.

4. Funciones de los intrones

Tradicionalmente, se ha defendido que los intrones son fragmentos de ADN carentes de información cuya presencia en los genomas supone una carga para la célula. El primer argumento que se da es que para eliminar los intrones espliceosomales (GU-AG y AU-AC) se requiere una maquinaria molecular muy compleja formada por 5 ARNnp (ARN nuclear pequeño) y más de 150 proteínas. En segundo lugar, la transcripción de los intrones es costosa en términos de tiempo y energía, ya que el ratio de elongación de una ARNasa es de 60 bases por segundo, lo que significa que la transcripción de algunos intrones puede durar hasta muchas horas. Tercero, el reconocimiento de las uniones de corte y empalme por el espliceosoma es llevado a cabo directamente por elementos

reguladores en *cis* y esto hace que los organismos sean vulnerables a mutaciones que puedan tener efectos aparentemente imperceptibles. Y finalmente, un mal funcionamiento de los ARNnp y de las proteínas que son necesarias para llevar a cabo el proceso de corte y empalme puede originar un efecto perjudicial para la célula.

En la actualidad, se ha demostrado que los intrones presentan un amplio espectro de funciones y están involucrados en cada paso del procesamiento de ARNm. Esta gran diversidad de funciones intrónicas apoya la idea de que los intrones fueron elementos egoístas en eucariotas tempranos, pero luego han ido ganando numerosas funciones en diferentes linajes eucarióticos (Chorev & Carmel 2012). Las funciones de los intrones se pueden dividir en cinco categorías:

- **Funciones asociadas al intrón genómico**

Los intrones modifican el nivel de expresión de sus genes hospedadores de diferentes maneras y lo pueden hacer a través del espliceosoma o de algunas de sus proteínas satélites, o a través del propio intrón. El proceso de corte y empalme está fuertemente acoplado a la formación del extremo 3' y a la regulación del procesamiento de este extremo. Un ejemplo bien conocido es el segundo intrón del gen la β -globina. Si este intrón es eliminado o sustituido por otros intrones se reduce la eficiencia con la que se forma el extremo 3' (Antoniou *et al.* 1998). Por otro lado, los intrones son más largos en genes específicos de tejidos porque alojan elementos regulatorios que sirven de soporte para asegurar el correcto ensamblaje de los nucleosomas (Vinogradov 2006). Además, los nucleosomas preferentemente ocupan los exones y esta preferencia es independiente de si el exón es constitutivo o alternativo, de su nivel de expresión o de su contenido en GC. Se desconoce cómo se lleva a cabo este marcaje de nucleosomas, pero se ha sugerido que las secuencias cercanas al final de los intrones funciona como elemento disuasorio para los nucleosomas, favoreciendo su acoplamiento hacia los exones (Schwartz *et al.* 2009).

- **Funciones asociadas a los intrones transcritos**

Algunos intrones que forman parte del ARNm tienen ciertas funciones en las células asociadas únicamente al hecho de hayan sido transcritos, independientemente del contenido de su secuencia, de su posición o del hecho de que después sean eliminados. Muchos intrones necesitan minutos, horas e incluso días para ser transcritos y este

incremento de tiempo favorece la posibilidad de que los intrones puedan servir como herramienta para orquestar el tiempo de demora entre la activación de un gen y la aparición de sus proteínas. Un caso conocido es el del gen *Hes7* que es cíclicamente expresado en el mesodermo presomítico y regula la segmentación de los somitas. Los intrones que se encuentran dentro del gen *Hes7* de ratón provocan un retraso de 19 minutos en la transcripción y sin este retraso las oscilaciones desaparecerían y *Hes7* se expresaría continuamente provocando defectos en la segmentación (Takashima *et al.* 2011). La longitud de estos intrones es altamente conservada a lo largo de los dominios eucariotas.

- **Funciones asociadas a los intrones cortados y empalmados**

El proceso de corte y empalme de muchos intrones ocurre al mismo tiempo que la transcripción debido a que ambos procesos están conectados. Algunos factores que forman parte del espliceosoma modulan las tres fases de transcripción. Un ejemplo de ello es el ARNnp U1, un componente del espliceosoma que está asociado al factor de iniciación de la transcripción TFIIF y estimula la formación del primer enlace fosfodiéster por la ARN polimerasa II. Otros elementos que componen la maquinaria de corte y empalme favorecen la fase de elongación y terminación del proceso de transcripción.

Probablemente, algunos sitios de corte y empalme pueden perderse y no ser reconocidos por el espliceosoma dando lugar al llamado “*splicing* alternativo”.

- **Funciones asociadas a los intrones eliminados**

Los ARN no codificantes (ARNnc) son moléculas funcionales de ARN que no codifican proteínas, pero que tienen funciones importantes en la célula. Los tipos de ARNnc más relevantes son microARN (ARNmi), ARNnp, ARN interferente pequeño (ARNip), ARN asociados a Piwi (ARNpi) y ARN largos no codificantes (ARNlnc), y muchos de ellos se encuentran dentro de los intrones que han sido eliminados en el proceso de corte y empalme.

- **Funciones asociadas al complejo EJC**

El EJC (complejo de unión de exones) es una estructura multiproteica que se une al ARNm una vez finalizado el proceso de *splicing*. La unión está mediada por el

espliceosoma y se localiza aguas arriba del sitio de corte y empalme, a una distancia de 20-24 nucleótidos. Este complejo está implicado en varias funciones celulares siendo una de ellas la de ayudar en la exportación del ARNm ya maduro desde el núcleo hasta el citoplasma. La presencia del EJC también favorece la formación de polirribosomas, siendo más rápida la síntesis proteica, ya que se pueden sintetizar varias cadenas peptídicas al mismo tiempo. Y, además, hay ARNm que tienen que ser traducidos en una zona concreta del citosol y EJC juega un papel importante en reclutar las proteínas lanzaderas necesarias para esta localización citoplasmática.

Aunque EJC no participe en el proceso de eliminación de intrones, es de vital importancia que los intrones sigan estando presentes en los genomas eucariotas, ya que sin ellos, EJC no se uniría al ARNm y no podría asegurar que la traducción a proteínas se hiciera correctamente.

5. Orígenes de los intrones

Desde que se descubrieron los intrones hasta el día de hoy, se ha debatido con gran interés sobre el origen de estos fragmentos de ADN. Hay pocas controversias acerca de los tipos de intrones de los grupos I, II y III porque se suele aceptar que todos ellos con capacidad de autocorte y empalme evolucionaron en el mundo del ARN y han sobrevivido desde entonces sin sufrir demasiadas modificaciones. Sin embargo, hay problemas respecto al origen de los intrones espliceosomales, que se encuentran en grandes cantidades en los genomas nucleares eucariotas.

A día de hoy, se barajan dos posibles hipótesis sobre la aparición de los intrones, la hipótesis de los **intrones tempranos** y la hipótesis de los **intrones tardíos**. Ambas teorías han sido muy debatidas y tienen argumentos a favor y en contra, por lo que el origen de los intrones sigue siendo aún desconocido. La hipótesis de los intrones tempranos defiende que el ancestro común a eucariotas y procariotas poseía intrones, pero estos últimos los perdieron durante el proceso evolutivo; mientras que la hipótesis de los intrones tardíos postula que los intrones aparecieron durante la formación de los eucariotas.

Hay varios modelos diferentes para cada hipótesis. En el caso de los intrones tempranos, el modelo más convincente es el denominado “teoría exónica de los genes”, que sostiene que los intrones se formaron cuando se construyeron los primeros genomas de ADN, poco después del fin del mundo de ARN. Estos genomas contendrían muchos genes cortos,

cada uno de ellos derivado de una sola molécula de ARN, la cual codificaría para un péptido pequeño con un solo dominio estructural. Es probable que estos polipéptidos tuvieran que asociarse entre sí para formar proteínas más grandes que dieran lugar a enzimas con mecanismos catalíticos más eficientes. Y para ayudar a la síntesis de una enzima de múltiples dominios, es probable que estos pequeños genes se agruparan para tener posiciones más cercanas. En otras palabras, estos pequeños genes se convirtieron en exones y las secuencias de ADN localizadas entre ellos, en intrones. Según los modelos que apoyan la hipótesis de los intrones tempranos, todos los genomas (procariotas y eucariotas) tenían intrones, pero actualmente sabemos que no hay intrones espliceosomales en los genomas bacterianos. Por tanto, si esta hipótesis es correcta, debemos presuponer que, por alguna razón, el genoma ancestral bacteriano perdió, en algún momento de su evolución, todos estos intrones. Este hecho es difícil de concebir porque es prácticamente imposible perder un gran número de intrones de un genoma sin riesgo de alterar funciones génicas.

La hipótesis de los intrones tardíos evita esta incongruencia proponiendo que, al principio de los tiempos, ningún gen contenía intrones. Posteriormente, el genoma nuclear eucariota temprano fue invadido por estas estructuras que con el tiempo, fueron proliferando hasta el día de hoy. Las similitudes entre las vías de corte y empalme de los intrones GU-AG y del grupo II sugieren que los invasores que dieron lugar a los primeros, podrían haber sido secuencias del grupo II que escaparon de genomas de orgánulos. Sin embargo, la similitud entre los intrones GU-AG y del grupo II no demuestra la hipótesis de los intrones tardíos porque es igual de posible crear un modelo de intrones tempranos, diferente a la teoría exónica de los genes, en el cual secuencias del grupo II den origen a los intrones GU-AG, pero en un estadio muy precoz de la evolución de los genomas, poco antes de que aparecieran los eucariotas.

Una de las razones por las cuales el debate respecto al origen de los intrones espliceosomales ha continuado durante más de 25 años es que ha sido difícil hallar evidencias que avalen una u otra hipótesis. Una predicción de la hipótesis de los intrones tempranos es que debería haber una estrecha similitud entre las posiciones de los intrones de genes homólogos de organismos no relacionados porque todos estos genes descienden de un gen ancestral que contiene intrones. Se obtuvo un aval inicial para los intrones tempranos cuando se demostró que esto era así en el caso de cuatro intrones del gen de la triosafosfato isomerasa de animales y vegetales. Sin embargo, al examinar una mayor

cantidad de especies, se volvió más difícil interpretar las posiciones de los intrones de este gen, ya que parecía que se habían perdido intrones en algunos linajes, pero se habían ganado en otros. Este panorama se adapta tanto a los intrones tempranos como a los intrones tardíos, ya que ambas hipótesis tienen en cuenta la pérdida, la ganancia o el reposicionamiento de intrones por eventos de recombinación en linajes individuales. Un enfoque alternativo ha sido intentar correlacionar los exones con dominios estructurales proteicos. La hipótesis de los intrones tempranos predice que sería evidente un razonamiento como este, ya que los genes primitivos pequeños fueron ensamblados entre sí para formar los genes que hoy conocemos. Un estudio de las globinas de los vertebrados concluyó que cada una de estas proteínas está formada por cuatro dominios estructurales, el primero correspondiente al exón 1 del gen de globina; el segundo y el tercero, al exón 2; y el cuarto, al exón 3. La predicción de que debe haber genes de globina con otro intrón que divida el segundo y el tercer dominio resultó correcta cuando se mostró que el gen de la leghemoglobina de la soja tenía un intrón en la posición exacta prevista. Lamentablemente, a medida que se secuenciaron más genes de globina se descubrieron más intrones, cuyas posiciones no corresponden, en la mayoría de los casos, a uniones entre dominios (Brown 2017).

III. CLASE AVES

Las aves constituyen una de las cinco clases de vertebrados pertenecientes al filo Cordados. Se han descrito alrededor de 10.000 especies de aves que comprenden desde los pequeños colibríes a los grandes avestruces. Son animales homeotermos que conservan su temperatura corporal y no dependen del medio ambiente tanto como los animales de sangre fría; por ello, pueden vivir en medios muy inhóspitos permitiéndole la posibilidad de vivir en una gran variedad de hábitats. Las aves tienen el cuerpo cubierto de plumas y las extremidades anteriores modificadas en alas adaptadas generalmente para el vuelo, y gracias a importantes adaptaciones como éstas, las aves han podido colonizar muy exitosamente el medio aéreo permitiéndoles vivir en prácticamente cualquier parte del planeta.

Hace unos 60 MA, en la frontera entre los periodos Cretácico y Terciario, un asteroide chocó contra la Tierra provocando uno de los peores episodios de extinción de todos los tiempos. Muchas especies que poblaban entonces el planeta desaparecieron por completo, pero otras, como los dinosaurios, no murieron en su totalidad y sus descendientes dieron lugar a las aves, que han sido los únicos colonizadores del medio aéreo con una diversidad que duplica la alcanzada por los mamíferos, que también aparecieron en el Triásico y empezaron a dominar el medio terrestre tras la desaparición de los dinosaurios (Sanz 1999).

Archaeopteryx lithographica es el ave más antigua conocida y fue descubierta en 1861 a partir de un fósil encontrado en las calizas de Solnhofen (Alemania) del periodo Jurásico. Éste ave, con una altura de unos 25 cm y un peso de no más de 1 kg, ya presentaba plumas y alas que le permitían volar. El registro fósil de aves es mucho menor con respecto a otros grupos de vertebrados como los peces y los reptiles, pero aun así, hay evidencias suficientes sobre la aparición de los grandes grupos de aves. Así pues, en el Terciario aparecieron las grandes aves carnívoras que ocuparon los nichos de los dinosaurios bípedos. Hasta la última edad de hielo estas aves con forma de buitre dominaban los cielos y llegaban a tener una envergadura de entre 4 y 8 m. Las aves acuáticas como los colimbos y las gaviotas aparecieron en el Eoceno hace entre 37 y 54 MA. Y en el Mioceno, tuvo lugar la radiación de las aves pequeñas silvestres de canto como consecuencia de la rápida evolución de plantas con flores e insectos que generó una diversidad ecológica con nuevos nichos para que estas aves se alimentaran de insectos, frutas o néctar (Regal 1977; Arnaiz-Villena *et al.* 1998).

1. Orden *Passeriformes*

El orden *Passeriformes* abarca a más de la mitad de especies de aves que existen en el planeta. El nombre del orden viene del género *Passer*, donde se incluye el gorrión, y significa "los que tienen forma de gorrión". Las aves pertenecientes al orden *Passeriformes* son generalmente pequeñas, aunque hay excepciones como las que se incluyen en la familia *Corvidae* que pueden llegar a medir hasta 67 cm de largo. Los *Passeriformes* son morfológicamente similares entre ellos, diferenciándose fundamentalmente en la estructura del pico, las patas y la siringe. Esta última se trata de un órgano localizado en una cavidad en la unión de la tráquea con los dos bronquios primarios, y sirve para abrir y cerrar la glotis manteniendo así el agua y la comida fuera del tracto respiratorio. El sonido se produce cuando el aire que recorre los pasillos de la siringe vibra a su paso por ésta. El orden *Passeriformes* presenta dos subórdenes que se establecen por la conformación del órgano de canto: *Tyranni* o pájaros clamadores, que presentan una siringe simple, y *Passeri* o pájaros cantores, cuya siringe es mucho más compleja (Greenewalt 1969; Gaunt & Wells 1973).

2. Género *Carduelis*

El género *Carduelis* está formado por 31 especies incluyendo a todos los jilgueros, lúganos y verderones. Está dentro de la familia *Fringillidae* y todas las especies presentan un pico de tamaño medio más o menos robusto según el método que empleen para alimentarse de las semillas. Las aves pertenecientes a este género se distribuyen por toda Europa, América y Asia. En África, únicamente se pueden encontrar algunas especies en el extremo nororiental del continente, y en Australia y Nueva Zelanda han sido introducidas algunas especies que han sabido adaptarse al nuevo hábitat con mucho éxito (Clement *et al.* 1993).

2.1 *Carduelis atrata*

Carduelis atrata, también conocido como negrito de Bolivia o lúgano negro, se distribuye



desde el centro de Perú hasta el noroeste de Argentina. Mide unos 12-13 cm y tienen todo el cuerpo de color negro, excepto unas franjas amarillas que presentan las alas y ambos lados de la cola. Se localiza en las estepas altas de los Andes y en zonas templadas entre los 1800 m y los 4800 m de altitud. Puede encontrarse en zonas rocosas con escasa vegetación y también próximos a

granjas y pueblos (Clement *et al.* 1993).

2.2 *Carduelis carduelis parva*

Carduelis carduelis parva, también conocido como jilguero europeo, se distribuye por



toda Europa, Asia occidental y el norte de África. Mide 12 cm y la cabeza muestra una característica careta roja con franjas laterales blancas y negras. Su cola es negra, con el obispillo y el extremo distal blanco; además, las plumas más externas de la cola pueden tener amplias manchas blancas. Puede localizarse en una gran variedad de hábitats, siempre y cuando

haya cierta cobertura arbórea y el clima sea suave (Clement *et al.* 1993).

2.3 *Carduelis citrinella*

Carduelis citrinella, también conocido como verderón serrano, se distribuye por casi toda



Europa, gran parte de Asia y el norte de África. Mide 12 cm, tiene la cabeza gris y el resto del cuerpo presenta un plumaje de color amarillo verdoso. Exhiben un plumaje moteado, especialmente en pecho y vientre. Se trata de una

especie endémica de las montañas de Europa occidental (Clement *et al.* 1993).

2.4 *Carduelis pinus*

Carduelis pinus, también conocido como jilguero de los pinos, se distribuye por toda Norteamérica. Mide 11-13 cm y las alas y la cola son de color marrón, con una extensión variable de color amarillo en las alas y en la base de la cola. Habita en zonas de coníferas, bosques mixtos, matorrales de aliso, y árboles caducifolios y ornamentales. Se alimenta de diferentes semillas o brotes, sobre todo de piceas, cicuta, cedro blanco, abedul y aliso, que toman de los propios árboles (Clement *et al.* 1993).



2.5 *Carduelis spinus*

Carduelis spinus, también conocido como jilguero euroasiático, se distribuye por toda Europa, parte de Asia y el norte de África. Mide 12 cm y su plumaje es de color amarillo-verdoso y rayado. Presenta una corona negra, buena parte de las alas y de la cola es negra y la rabadilla es amarillo-verdosa. Habita en árboles propios de brezales, bosques, matorrales, huertos y jardines (Clement *et al.* 1993).



2.6 *Carduelis crassirostris*

Carduelis crassirostris, también conocido como lúgano de pico grueso, se distribuye por los Andes del oeste de Bolivia, Chile y el oeste de Argentina. Mide 14 cm y su cabeza y cuello es de color negro. El abdomen es amarillo y grisáceo, y las alas son negras con matices amarillos. El pico es mucho más grueso que el de otras especies del género *Carduelis*. Suele vivir en zonas elevadas de entre 3000 y 4800 m de altura (Clement *et al.* 1993).



2.7 *Carduelis cucullata*

Carduelis cucullata, también conocido como cardenalito de Venezuela, se distribuye por



el norte de la cordillera de Colombia y el norte de Venezuela entre 300 y 1200 m de altura. Mide 11 cm y presenta un cuerpo rojo anaranjado con una cabeza, alas y cola de color negro. Como en todos los casos, las hembras presentan colores más pálidos que los machos (Clement *et al.* 1993).

2.8 *Carduelis magellanica*

Carduelis magellanica, también conocido como lúgano encapuchado, se distribuye por



toda Sudamérica. Mide 11 cm y presenta una cabeza de color negro. El resto del cuerpo es amarillo verdoso, y las alas y la cola son negras con franjas amarillas. Vive en hábitats muy variados desde el nivel del mar hasta los 5000 m. Es común verlo en jardines y parques (Clement *et al.* 1993).

2.9 *Carduelis notata*

Carduelis notata, también conocido como lúgano de cabeza negra, se distribuye por



América Central y México. Mide 11 cm y posee una cabeza y la parte superior del pecho negras en forma de “v”. De este color también son las alas y la cola, aunque en este caso ambas están salpicadas de tonos dorados. El resto del cuerpo es de color amarillo. Su hábitat son las montañas húmedas tropicales o subtropicales (Clement *et al.* 1993).

2.10 *Carduelis spinescens*



Carduelis spinescens, también conocido como lúgano de los Andes, se distribuye por Colombia, Venezuela y Ecuador. Mide 12 cm y posee una frente de color negro y un dorso de tono verde oliva. El pecho y el vientre son amarillos con tonalidades verdosas, y las alas y la cola negras. Esta especie habita en las regiones andinas de los países mencionados anteriormente (Clement *et al.* 1993).

2.11 *Carduelis olivacea*



Carduelis olivacea, también conocido como lúgano oliváceo, se distribuye por Bolivia, Ecuador y Perú. Mide 11 cm y es muy parecido a *Carduelis magellanica* con su cabeza, su cola y sus alas negras. A diferencia de ésta, el resto del cuerpo tiene un color verde aceituna. Habita en regiones húmedas localizadas entre 1200 y 3000 m de altitud (Clement *et al.* 1993).

2.12 *Carduelis xanthogastra*

Carduelis xanthogastra, también conocido como lúgano de vientre amarillo, se distribuye



por Costa Rica, Panamá, Colombia, Venezuela y Ecuador. Mide 11 cm y la cabeza, el cuello, la parte superior del pecho, las alas y la cola son de color negro brillante. La parte inferior del pecho y el vientre presenta un tono amarillo intenso. Su hábitat son los bosques de montaña (Clement *et al.* 1993).

2.13 *Carduelis yarrellii*

Carduelis yarrellii, también conocido como lúgano de cara amarilla, se distribuye por



Venezuela y el norte de Brasil. Mide 11 cm y posee un característico casquete de color negro que llega hasta la nuca pasando por los ojos. La parte dorsal es amarilla verdosa, y la región subcaudal y el obispillo son de un tono amarillo intenso. Habita en regiones de clima templado en zonas no muy elevadas (Clement *et al.* 1993).

2.14 *Carduelis barbata*

Carduelis barbata, también conocido como lúgano de barbilla negra, se distribuye por



Chile y Argentina. Mide 13 cm y tiene la parte superior de la cabeza y la garganta de color negro. La espalda es amarilla verdosa oscura con listas negras. El pecho, vientre y obispillo son amarillo intenso, y las alas y la cola son negras con franjas amarillas. Habita en zonas costeras (Clement *et al.* 1993).

2.15 *Carduelis lawrencei*



Carduelis lawrencei, también conocido como jilguero de Lawrence, se distribuye por California y Baja California. Mide 11 cm y de color negro presenta la parte frontal de la cabeza y parte de las alas. El pecho y algunas zonas de las alas son amarillas y el resto del cuerpo es gris. Habita en zonas secas y suelen anidar en robles y pinos (Clement *et al.* 1993).

2.16 *Carduelis dominicensis*



Carduelis dominicensis, también conocido como lúgano de las Antillas, se distribuye por Haití y República Dominicana. Mide 12 cm y presenta la cabeza y parte de las alas de color negro. El resto del cuerpo es amarillo verdoso. Habita en zonas de montaña con abundantes pinares (Clement *et al.* 1993).

2.17 *Carduelis psaltria hesperophila*

Carduelis psaltria hesperophila, también conocido como jilguero dominico, se distribuye



desde el noroeste de EEUU hasta el noroeste de México. Mide 10 cm y presenta cabeza negra, espalda verdosa y partes inferiores de color amarillo con marcas blancas en el ala. Habita en diversos nichos desde bosques abiertos hasta jardines y campos de cultivo (Clement *et al.* 1993)

2.18 *Carduelis atriceps*



Carduelis atriceps, también conocido como jilguero capirotado, se distribuye por el sur de México y Guatemala. Mide 12 cm y, su cuerpo y parte de la cabeza es amarilla verdosa. La parte superior de cabeza y la garganta son negras y las alas presentan franjas también de este color. Habita en zonas elevadas localizadas entre 2350 y 3050 m de altitud, concretamente en bosques de coníferas y bordes de cultivos (Clement *et al.* 1993).

2.19 *Fringilla coelebs*

Fringilla coelebs, también conocido como pinzón común, se distribuye por toda Europa,



Asia occidental y norte de África. Posee una llamativa mancha blanca en el hombro y dos bandas blancas en las plumas externas de la cola. Presenta una tonalidad salmón en el vientre, el pecho y las mejillas; su píleo y nuca reflejan colores grises-

azulados. Suele habitar tanto en bosques cerrados como en masas arbóreas abiertas (Clement *et al.* 1993).

F. coelebs no pertenece al género *Carduelis* sino al género *Fringilla*. Ambos géneros están dentro de la subfamilia *Fringillinae* y es importante incluir al pinzón común en los estudios de este trabajo porque se utiliza como grupo externo, es decir, como una especie externa y ancestral al grupo de estudio de *Carduelis* que se está analizando. Para saber cuántos millones de años hace que apareció cada especie, el reloj molecular que se usa en este trabajo está basado en el tiempo de divergencia entre el pollo y el faisán, el cual está datado en 20-19 MA. Para ello, se analizaron las diferencias entre las secuencias del citocromo b del pollo y del faisán y se compararon con los estratos geológicos de ambas especies (Helm-Bychowski & Wilson 1986). De este modo, podemos saber que nuestro grupo externo, *F. coelebs*, apareció en la tierra hace aproximadamente 17,5 MA (Arnaiz-Villena *et al.* 1998).

IV. FILOGENIA MOLECULAR

El término “filogenia” proviene del griego (*φύλον*, raza y *genia*, nacimiento, producción) y hace referencia a las relaciones evolutivas que existen entre los organismos a nivel de especies, géneros y familias. El estudio de las relaciones entre las especies ha formado parte de los estudios de taxonomía clásica desde el siglo XVI, tal y como puede verse en los trabajos de evolución de autores como Linneo. No obstante, no fue hasta mediados del siglo XX cuando se formalizó el interés de aclarar el parentesco que existía entre las especies, tras el planteamiento de la Teoría de la Evolución descrita por Darwin muchos años atrás. Posteriormente, se desarrollaron distintos métodos para la reconstrucción de relaciones filogenéticas basadas en datos morfológicos; sin embargo, no fue hasta el desarrollo de la técnica molecular de la PCR, del inglés *Polymerase Chain Reaction*, cuando se pudo incorporar el estudio comparado de secuencias de ADN al estudio de las relaciones evolutivas entre las especies (Graur & Li 2000).

1. Métodos filogenéticos

El objetivo de la filogenia molecular basada en el ADN es plantear las relaciones evolutivas posibles entre distintos taxones a través de la comparación de secuencias de ADN de cada uno de ellos. La comparación de las secuencias se hace a través de distintas metodologías y las relaciones taxonómicas se representan a través de árboles filogenéticos, también llamados dendogramas. Los métodos filogenéticos constan de dos partes. La primera de ellas se lleva a cabo a través de los modelos de sustitución nucleotídica, y la segunda a través de los modelos de reconstrucción de árboles filogenéticos, es decir, primero se calcula una matriz de distancias genéticas resultante de la comparación de todas las secuencias por parejas, y después se construye un árbol filogenético basado en las distancias genéticas anteriores.

Tras obtener las distancias genéticas de las secuencias estudiadas pueden utilizarse tres métodos de reconstrucción de árboles filogenéticos para expresar gráficamente las relaciones evolutivas. El primero de ellos es la **Máxima verosimilitud o *Maximum likelihood* (ML)** cuyo árbol es construido con el programa PAUP4.0b10 (Swofford 2002); el segundo de ellos es la **Inferencia bayesiana (IB)** cuyo dendograma es creado por el programa MrBayes (Huelsenbeck & Ronquist 2001); y el tercero de ellos es el **Método del vecino más próximo o *Neighbor-Joining* (NJ)** (Saitou & Nei 1987) cuyo

árbol es realizado por el programa MEGA 7 (Kumar et al. 2016). Todos estos modelos de reconstrucción de árboles filogenéticos son explicados en el apartado 7.1.2. de Materiales y Métodos.

2. Citocromo b

Cuando se va a realizar un estudio evolutivo basado en ADN, hay que elegir muy bien la secuencia que queremos analizar de cada especie. Los genes más utilizados en los estudios filogenéticos se encuentran en la mitocondria y el más empleado en este tipo de análisis es el citocromo b.

Los citocromos son unas de las proteínas transportadoras de electrones dentro de la cadena respiratoria. El complejo III de la cadena respiratoria está formado por el citocromo b, junto con el citocromo c1 y un centro ferro-sulfato específico. El gen del citocromo b está codificado en la cadena pesada del ADN mitocondrial (ADNmt) y se localiza entre las posiciones 14.747 y 15.887 con una longitud de 1.143 pb (Anderson *et al.* 1981). Es de herencia materna y codifica para un único polipéptido sin intrones. Se trata de un gen pequeño que está en múltiple copias y es muy fácil de purificar. Otro aspecto relevante, propio del ADNmt, es la ausencia de alelos y la ausencia de recombinación de estos genes. Todas estas características propias del citocromo b han revolucionado los estudios evolutivos moleculares permitiendo analizar desde organismos que no están estrechamente emparentados hasta realizar estudios a nivel poblacional. Además, hay que destacar que se trata de una proteína importante en el mecanismo de fosforilación oxidativa y, por tanto, funcional y estructuralmente está muy conservada (Irwin *et al.* 1991).

OBJETIVOS

La realización de este trabajo tiene el objetivo fundamental de analizar el polimorfismo y la evolución del complejo principal de histocompatibilidad en especies del género *Carduelis* a través del estudio del intrón 2 de estos genes. El análisis de estos intrones nos puede dar una idea de si éstos cumplen una función esencial en el MHC que ha hecho que se hayan mantenido sin apenas variación a lo largo de millones de años. Además, el análisis filogenético de estas especies que se ha llevado a cabo en este trabajo nos permite estudiar la evolución molecular del complejo principal de histocompatibilidad y compararla con las relaciones filogenéticas de estos jilgueros. Para llevar a cabo todo ello, nos hemos centrado en el estudio del intrón 2 de los genes de clase I del MHC de 18 especies del género *Carduelis*. Las otras dos especies estudiadas que complementan nuestro análisis son *Fringilla coelebs* y *Gallus gallus*.

Los principales objetivos son:

- 1- Comparar las secuencias de ADN del intrón 2 y determinar la localización de las posiciones variables.
- 2- Calcular la longitud del intrón 2 de todas las especies analizadas.
- 3- Analizar la composición nucleotídica de las secuencias de intrones y estudiar qué diferencias se aprecian.
- 4- Estudiar la variabilidad de estos intrones y analizar la similaridad que hay entre ellos.
- 5- Comprobar si el MHC de estas pequeñas aves de canto es similar al MHC “mínimo esencial” descrito en el pollo.
- 6- Realizar un análisis filogenético de las especies estudiadas del género *Carduelis* utilizando el gen del citocromo b para obtener unos datos más exactos que los de la filogenia fenotípica, que permitan estudiar la evolución molecular del complejo principal de histocompatibilidad.

MATERIALES Y MÉTODOS

1. Especies de aves estudiadas

En este trabajo se utilizan muestras de 18 especies de aves del género *Carduelis* y una especie del género *Fringilla* pertenecientes al orden *Passeriformes*. Para la comparación de las secuencias obtenidas del MHC con la de *Gallus gallus* se utilizó la base de datos GenBank. La tabla M1 recoge el nombre de las distintas especies utilizadas junto a los códigos de acceso de cada una de las secuencias del MHC.

Nombre científico	GenBank	Procedencia
<i>C. atrata</i>	DQ257462	Sucre, Bolivia
<i>C. atriceps</i>	HM041047	Quetzaltenango, Guatemala
<i>C. barbata</i>	DQ257463	Magallanes, Chile
<i>C. carduelis parva</i>	FJ266447	Alcalá de Henares, España
<i>C. citrinella</i>	DQ257482	Alcalá de Henares, España
<i>C. crassirostris</i>	DQ257464	Pisac, Perú
<i>C. cucullata</i>	DQ257465	Mérida, Venezuela
<i>C. dominicensis</i>	HM041048	Constanza, Rep. Dominicana
<i>C. lawrencei</i>	HM041049	San Diego, EEUU
<i>C. magellanica</i>	DQ257467	Lima, Perú
<i>C. notata</i>	DQ257468	Xalapa, México
<i>C. olivacea</i>	DQ257470	Lima, Perú
<i>C. pinus</i>	FJ266376	Jackson, EEUU
<i>C. psaltria hesperophila</i>	HM041050	Sacramento, EEUU
<i>C. spinescens</i>	DQ257472	Mérida, Venezuela
<i>C. spinus</i>	FJ266399	Madrid, España
<i>C. xanthogastra</i>	DQ257473	San José, Costa Rica
<i>C. yarrellii</i>	DQ257475	Recife, Brasil
<i>F. coelebs</i>	DQ257477	Madrid, España
<i>G. gallus</i>	AY489160	Brasil

TABLA M1. Especies utilizadas en este estudio junto a los códigos de acceso del MHC a la base de datos GenBank.

Para estudiar las relaciones filogenéticas de las 19 especies pertenecientes al orden *Passeriformes* e interpretar el proceso evolutivo al que han estado sometidas las secuencias de histocompatibilidad analizadas en este trabajo, se recurrió al gen mitocondrial del citocromo b para analizar su evolución a lo largo de millones de años. En la tabla M2 se recogen todas las especies utilizadas y los códigos de acceso de cada una de las secuencias de citocromo b.

Nombre científico	GenBank	Procedencia
<i>C. atrata</i>	L76385	Sucre, Bolivia
<i>C. atriceps</i>	AF342863	Quetzaltenango, Guatemala
<i>C. barbata</i>	L77868	Magallanes, Chile
<i>C. carduelis parva</i>	L76387	Alcalá de Henares, España
<i>C. citrinella</i>	L77872	Alcalá de Henares, España
<i>C. crassirostris</i>	L77869	Pisac, Perú
<i>C. cucullata</i>	L76299	Mérida, Venezuela
<i>C. dominicensis</i>	AF342864	Constanza, Rep. Dominicana
<i>C. lawrencei</i>	L76392	San Diego, EEUU
<i>C. magellanica</i>	U79016	Lima, Perú
<i>C. notata</i>	U79019	Xalapa, México
<i>C. olivacea</i>	L77871	Lima, Perú
<i>C. pinus</i>	U79020	Jackson, EEUU
<i>C. psaltria hesperophila</i>	L76390	Sacramento, EEUU
<i>C. spinescens</i>	U79017	Mérida, Venezuela
<i>C. spinus</i>	L76391	Madrid, España
<i>C. xanthogastra</i>	L76389	San José, Costa Rica
<i>C. yarrellii</i>	U83200	Recife, Brasil
<i>F. coelebs</i>	L76609	Madrid, España

TABLA M2. Especies utilizadas en el estudio filogenético junto a los códigos de acceso del citocromo b a la base de datos GenBank.

2. Obtención de las muestras

Todas las muestras empleadas para el desarrollo del presente estudio, pertenecen a la colección que el Dr. Arnaiz Villena ha ido reuniendo a lo largo de los últimos 20 años y

corresponden a individuos salvajes capturados en sus entornos naturales. La captura de aves se realizó mediante reclamo y redes muy finas. Dichas redes son el método más utilizado, ya que permiten capturar una gran variedad de aves. La red se extiende entre dos postes formando unas bolsas. Las aves vuelan hacia la red y, generalmente, caen en estas bolsas y se enredan. El tamaño de malla utilizado fue de 35 mm, que es apropiado para la captura de aves de tamaño pequeño. El emplazamiento de las redes fue determinado por miembros de grupos locales de anillamiento, considerando para ello los movimientos usuales de las aves, la estructura y altura de la vegetación, la accesibilidad, proximidad al sitio de procesado, pendiente, y tipo de suelo. Se utilizaron postes de metal para sostener las redes y una vez instaladas, éstas eran revisadas cada veinte minutos.

Las muestras de sangre fueron obtenidas mediante el corte de una uña tras anestésiar la zona con una crema de lidocaína. La sangre fue depositada en tubos estériles de 3 ml que contenían una solución salina formada por cloruro sódico y dimetilsulfóxido (DMSO) en una proporción de 3:1. Las muestras se conservaron a 4 °C hasta ser almacenadas a -20 °C.

3. Extracción de ADN

Los glóbulos rojos de las aves son células nucleadas y gracias a esta característica se puede obtener una buena cantidad de ADN a partir de volúmenes muy pequeños de sangre. La extracción de ADN se llevó a cabo con un extractor automático (*Nucleic Acid Extraction System, QuickGene-810*, FUJIFILM) después de someter las muestras a un tratamiento previo con un kit comercial (*QuickGene Whole Blood Extraction Kit S*, FUJIFILM) indicado para la extracción de ADN a partir de sangre total. El proceso de extracción consiste en:

- Diluir 50 µl de sangre en 150 µl de solución salina (*PBS pH 7,4*, GIBCO) en un tubo de 1,5 ml
- Añadir 30 µl de proteasa (*DNA whole blood Protease EDB-01*, FUJIFILM)
- Añadir 250 µl de tampón de lisis (*DNA whole blood Lysis Buffer LDB-04*, FUJIFILM)
- Incubar en un baño a 56 °C hasta que desaparezcan los posibles coágulos (2-10 minutos)
- Añadir 250 µl de etanol absoluto (*Ethanol Absolute PRS*, PANREAC)

- Pasar la muestra a una columna del kit. Estas columnas poseen una membrana de 80 µm de espesor que atrapan las moléculas de ADN, debido a su carácter hidrofílico
- Introducir la columna en el extractor automático, donde será sometida a 3 ciclos de lavado (*DNA whole blood Wash Buffer WDB-03*, FUJIFILM) y posteriormente será diluida en 200 µl de una solución de elución (*DNA whole blood Elution Buffer CDB-02*, FUJIFILM). Una vez completado el proceso, se obtiene en el tubo colector suministrado por la casa, entre unos 150 a 200 µl de ADN.

En las muestras de aves donde existe una gran cantidad de ADN, la adición de etanol absoluto provoca, con frecuencia, la aglutinación del ADN (formándose la denominada “medusa”, que se observa a simple vista). En estos casos hay que retirar la medusa antes de pasar la muestra a la columna. Para ello, se procesa manualmente sometiéndola a un lavado con etanol absoluto (*Ethanol Absolute PRS*, PANREAC) y otro con etanol 70% (*Ethanol 70% v/v BP*, PANREAC). Finalmente, tras eliminar completamente los restos de etanol, se añaden 200 µl de agua destilada o bien de la solución de elución (*DNA whole blood Elution Buffer CDB-02*, FUJIFILM) suministrada con el kit de extracción.

La concentración y la pureza del ADN resultante fueron medidas en un espectrofotómetro (*Nanodrop ND-1000 Spectrophotometer*), utilizando para ello 1,5 µl del material genético obtenido. Este espectrofotómetro mide la absorbancia de la muestra a 230, 260 y 280 nm. Como es conocido, los ácidos nucleicos absorben luz a una longitud de onda de 260 nm. Gracias a esta tecnología se puede calcular la concentración del ADN mediante la medida de la Densidad Óptica (DO) siguiendo la relación: $1A_{260} = 1 \text{ DO ADN} = 50 \text{ ng}/\mu\text{l}$. También se puede conocer su pureza midiendo longitudes de onda de 260 y 280 nm (A_{260}/A_{280}) cuyos valores deben estar entre 1,6 y 1,8 respectivamente para que la muestra sea válida. De cada una de las muestras se extrajo una alícuota que fue ajustada a una concentración de 100-200 ng/µl mediante la adición de agua destilada cuando requería ser diluida o mediante evaporación en estufa a 60 °C cuando requería ser concentrada. Todas las muestras fueron conservadas a -20 °C.

4. Obtención de secuencias de ADN de los genes del MHC de clase I

Para estudiar el polimorfismo del complejo principal de histocompatibilidad hay que identificar las diferentes secuencias de ADN presentes en cada individuo y para ello, se recurre a técnicas de amplificación, secuenciación y clonación.

4.1 Amplificación de secuencias del MHC

Para el estudio de secuencias de ADN es necesario disponer de un número de copias muy superior al obtenido tras el proceso de extracción. Para ello, se lleva a cabo la técnica de la reacción en cadena de la polimerasa, más conocida como PCR (Figura M1). Se trata de una técnica de biología molecular desarrollada en 1986 por Kary Mullis, cuyo objetivo es obtener un gran número de copias de un fragmento de ADN particular, partiendo de un fragmento original, o molde.

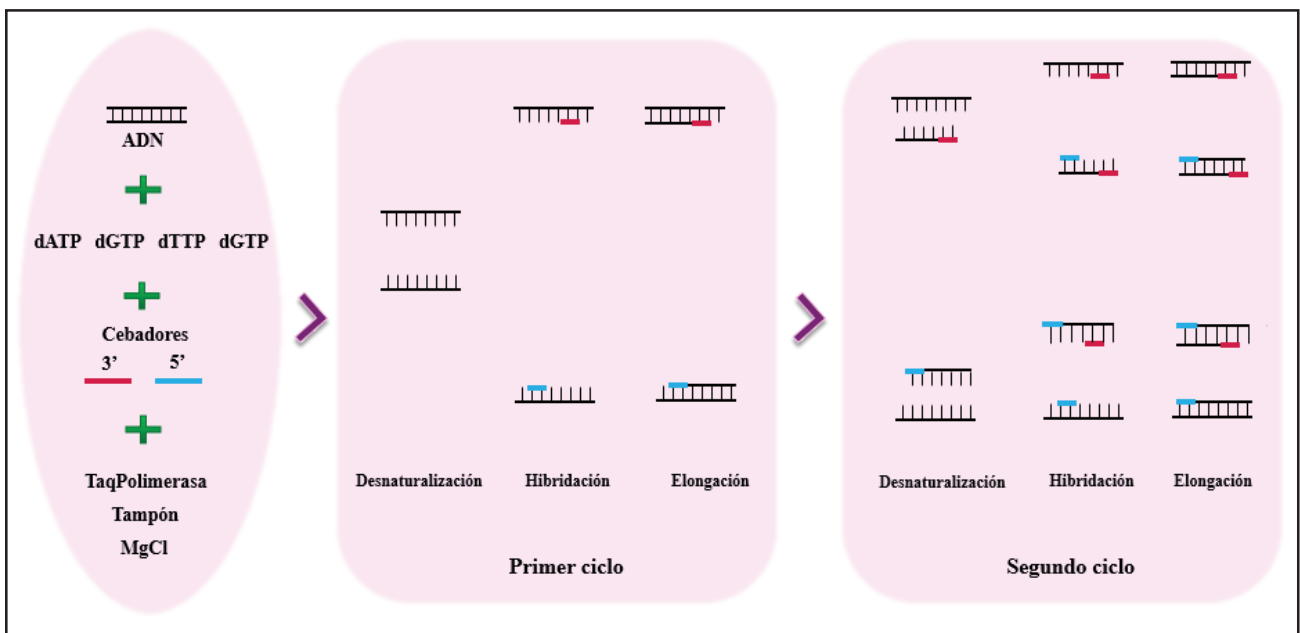


FIGURA M1. Esquema de la PCR.

La PCR está diseñada según el principio natural de replicación del ADN y consta de tres fases: desnaturalización, hibridación y elongación. En primer lugar es necesario que el ADN se desnaturalice, es decir, que las dos cadenas de ADN se separen. Esta primera fase se conoce como desnaturalización y se lleva a cabo elevando la temperatura hasta 94-95 °C. El siguiente paso consiste en un descenso de la temperatura para permitir que los cebadores se unan por complementariedad al ADN molde. Esta segunda fase se

conoce como hibridación y las temperaturas habituales en esta fase oscilan entre 50 y 60 °C dependiendo de la composición nucleotídica y la especificidad de los cebadores empleados. Por último, en la fase de elongación, la enzima polimerasa incorpora nucleótidos complementarios a partir del extremo 3' libre de la región en que han hibridado los cebadores. La temperatura para este paso depende de la ADN polimerasa que usemos. Para la Taq polimerasa, la temperatura que se suele emplear son 72 °C.

Aunque en este trabajo, únicamente, se haya llevado a cabo el estudio del intrón 2 de los genes del MHC de clase I, se han amplificado las secuencias de ADN que contienen el exón 2, el intrón 2 y el exón 3 de los genes del MHC de clase I de las distintas especies estudiadas, como se ve en el esquema de la Figura M2.

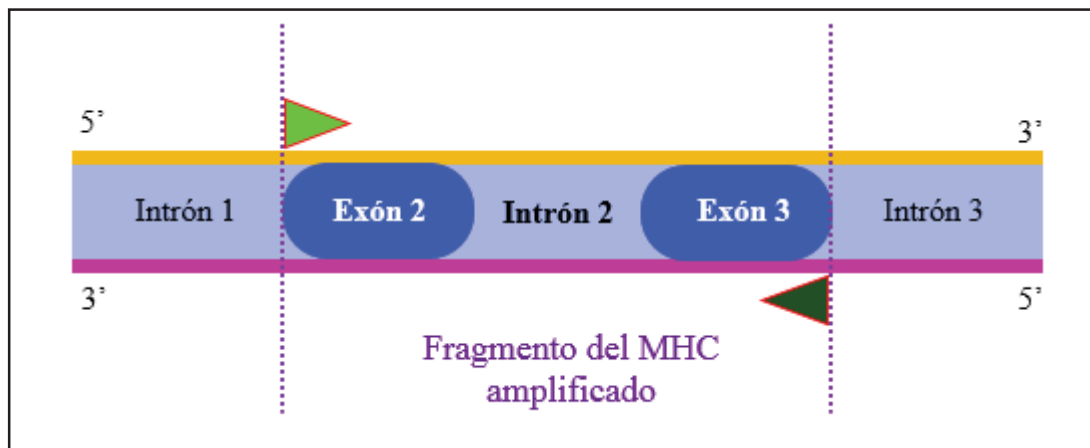


FIGURA M2. Esquema del fragmento del MHC amplificado.

Las reacciones de amplificación que se llevaron a cabo bajo las condiciones descritas en las Tablas M3 y M4, se realizaron en un termociclador (*Mastercycler egradient S*, EPPENDORF). La ADN polimerasa utilizada fue *Taq DNA Polymerase* (INVITROGEN), suministrada junto con el tampón *Taq DNA Polymerase PCR Buffer*, 10X (INVITROGEN), el cloruro de magnesio de 50 mM (INVITROGEN) y los dNTPs (100 mM dNTP *Set*, INVITROGEN). El cebador directo utilizado fue: E2.1: 5'-GTTCTCCACTCCCTGGATTACC-3' y el cebador inverso fue A23b: 5'-GCGCTCCAGCTCCTTCTGCCCA/GTA-3' (Westerdahl *et al.* 1999). Estos cebadores fueron suministrados por la casa *Thermo Fisher Scientific*.

Reactivos	Concentración	Volumen
dNTP	2,5 mM	1,6 µl
Tampón	5x	4 µl
MgCl ₂	25 mM	2 µl
Cebador E2.1	10 µm	0,8 µl
Cebador A23b	10 µm	0,8 µl
Taq Polimerasa	5 U/µl	0,3 µl
ADN	100-200 ng/µl	1 µl
Agua	-	8,5 µl
Total:		20 µl

TABLA M3. Cantidad de cada reactivo, según la concentración dada, para un volumen de reacción de 20 µl por muestra empleada en la técnica de PCR.

Nº Ciclos	Temperatura	Tiempo	Proceso
1	95 °C	5 min	Desnaturalización inicial
42	95 °C	10 s	Desnaturalización
	65 °C	30 s	Hibridación
	72 °C	1 min	Elongación
1	72 °C	10 min	Elongación final

TABLA M4. Programa de amplificación utilizado en la técnica de PCR.

4.2 Electroforesis en gel de agarosa de los fragmentos amplificados

Los procesos de amplificación no siempre son específicos, por eso es necesario realizar una electroforesis para identificar y aislar el fragmento deseado. Esta técnica nos permite separar especies químicas (ácidos nucleicos o proteínas) a lo largo de un campo eléctrico en función de su tamaño y de su carga eléctrica. Los ácidos nucleicos, tienen por naturaleza carga negativa. Al ponerlos sobre un gel poroso y aplicar un campo eléctrico,

se produce la migración diferencial de los fragmentos a través de los poros de la matriz. Su posterior tinción y exposición con luz ultravioleta, permite observar el resultado de ésta migración. El tamaño de los fragmentos de ADN se estima comparándolos con el patrón de bandas que se obtiene de marcadores comerciales de peso molecular conocido. Nuestros productos obtenidos en la amplificación fueron sometidos al proceso de electroforesis en gel de agarosa (*BIOTOOLS HR Agarose*, BIOTOOLS) al 2% en tampón 1X TBE (10X TBE *Buffer*, INVITROGEN). En el proceso se incluyó un marcador de peso molecular (*DNA Molecular Weight Marker XIV - 100 bp ladder* - ROCHE) para identificar el fragmento deseado, en torno a las 850 pb. La presencia de fragmentos de ADN en el gel fue revelado tiñéndolo con un agente intercalante (*GelRed Dropper Bottle*, OLERUP) y observado al ser iluminado con radiación ultravioleta. La banda correspondiente fue recortada del gel con un bisturí estéril desechable y depositada en un tubo de 1,5 ml.

4.3 Purificación de la banda de la secuencia amplificada

El fragmento de ADN fue extraído del gel con un kit comercial de purificación (*MinElute Gel Extraction Kit*, QIAGEN). Este kit consta de unas columnas especiales con una membrana de sílice en la que se queda adherido el material genético mientras que los contaminantes (restos de gel, producto de tinción, etc.) se desechan en un tubo colector tras varios lavados. Los reactivos que contiene este kit son *Solubilization Buffer (Buffer QG)* y *Wash Buffer (Buffer PE)*. El protocolo de purificación es el siguiente:

- Añadir 300 µl de *Buffer QG* en el tubo que contiene la banda de gel recortada anteriormente e introducir en un baño a 56 °C hasta que se disuelva el gel (5-10 min).
- Pasar el contenido del tubo a una columna del kit. Centrifugar 1 minuto a 14000 rpm y desechar el filtrado.
- Añadir 500 µl de *Buffer QG*. Centrifugar 1 minuto a 14000 rpm y desechar el filtrado.
- Añadir 650 µl de *Buffer PE* y dejar actuar 5 minutos.
- Centrifugar 1 minuto a 14000 rpm y desechar el filtrado
- De nuevo, centrifugar 1 minuto a 14000 rpm, desechar el filtrado y tirar el colector.

- Poner la columna en un tubo de 1,5 ml.
- Añadir 50 µl de agua destilada y centrifugar 1 minuto a 14000 rpm.
- Tirar la columna y guardar el ADN purificado a -20 °C

4.4 Secuenciación de los fragmentos amplificados

Generalmente, los métodos de secuenciación que se utilizan en la actualidad están basados en el método de Sanger (Sanger *et al.* 1977). Este método está basado en el empleo de didesoxinucleótidos trifosfatos (ddNTP) que carecen de uno de los grupos hidroxilo, de manera que cuando uno de estos nucleótidos se incorpora a una cadena de ADN en crecimiento, está cadena no puede continuar elongándose ya que la ADN polimerasa necesita un extremo 3' OH para añadir el siguiente nucleótido. En la secuenciación se utiliza un cebador marcado que suministra el extremo 3'OH que necesita la ADN polimerasa. Se preparan cuatro tubos de reacción, cada uno con el ADN molde de hélice sencilla que se desea secuenciar, la ADN polimerasa, el cebador marcado y los cuatro nucleótidos trifosfato. A cada tubo se le añade una pequeña proporción de un ddNTP, un primer tubo con ddCTP, un segundo con ddGTP, un tercero con ddATP y un cuarto con ddTTP. En cada uno de estos tubos se producirán cadenas de ADN de distintas longitudes, terminando todas en el lugar en el que se incorporó el ddNTP correspondiente añadido al tubo. Posteriormente, estas piezas de ADN se separan mediante electroforesis en gel de acrilamida. Las piezas más pequeñas migran más rápidamente que las grandes y la secuencia se puede leer directamente sobre el gel empezando por el extremo más alejado de la zona de carga (extremo positivo), donde se encuentra el fragmento de menor tamaño, y siguiendo un orden de tamaño ascendente. La figura M3 muestra un ejemplo de un posible resultado de la reacción de secuenciación por el método de Sanger.

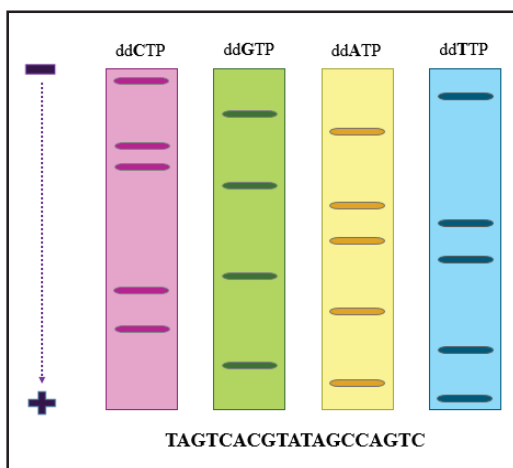


FIGURA M3. Esquema de un posible resultado de las reacciones de secuenciación por el método de Sanger.

Actualmente, las reacciones de secuenciación se llevan a cabo en secuenciadores automáticos de ADN en los que las reacciones se desarrollan en un mismo tubo ya que existe la posibilidad de marcar diferencialmente los cuatro tipos de nucleótidos didesoxinucleótidos. Este marcaje, comúnmente hecho por fluorescencia, es detectado directamente por el secuenciador. En este caso se recurrió al servicio de secuenciación de la empresa SECUGEN S.L. Para desarrollar el proceso requieren el envío de 15 µl del fragmento de ADN amplificado y purificado a una concentración de 10-20 ng/ µl (para fragmentos de más de 500 pb), y 1,5 µl de cebador a una concentración de 5 µM.

4.5 Clonación de los fragmentos amplificados

Los productos de amplificación no pueden ser directamente secuenciados ya que las secuencias de ADN resultantes pueden presentar ambigüedades lo que significa que en determinadas posiciones ofrecen dos (o más) posibles alternativas (Figura M4). Estas ambigüedades se representan en el cromatograma como un doble pico y esto puede indicar que el individuo es heterocigoto para el fragmento amplificado, o incluso que existen dos o más *loci* que presentan secuencias muy parecidas y que han sido amplificados simultáneamente con los cebadores utilizados.

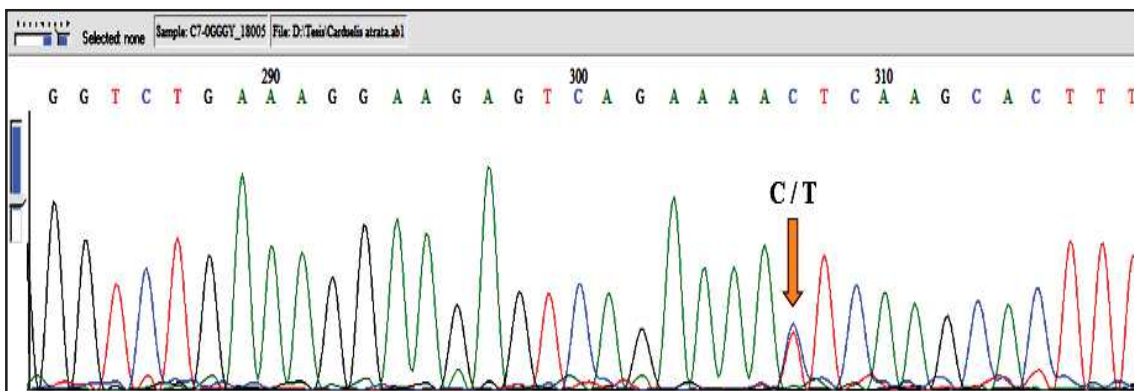


FIGURA M4. Ejemplo de ambigüedades representadas por dobles picos en un cromatograma.

Para resolver dichas ambigüedades, los productos de amplificación fueron sometidos a un proceso de clonación antes de ser definitivamente secuenciados. La clonación es una técnica empleada para obtener múltiples copias de un gen o de su producto génico que a la vez, es una herramienta muy útil para aislar y caracterizar secuencias de ADN de forma individual. Se basa en la capacidad natural que tienen algunas bacterias para incorporar plásmidos presentes en su entorno en un proceso denominado transformación bacteriana.

La estrategia que se emplea consiste en colocar un fragmento de ADN portador del gen en cuestión dentro de un plásmido bacteriano, acompañado de genes que permitan la selección de las moléculas híbridas, en el interior de una bacteria, que puede multiplicarse a gran velocidad y permitir que la maquinaria de la bacteria transcriba y traduzca el segmento de ADN que se le ha introducido. En este trabajo se ha utilizado el sistema de clonación *pGEM-T Easy Vector System II* (PROMEGA), basado en el plásmido que se muestra en la Figura M5.

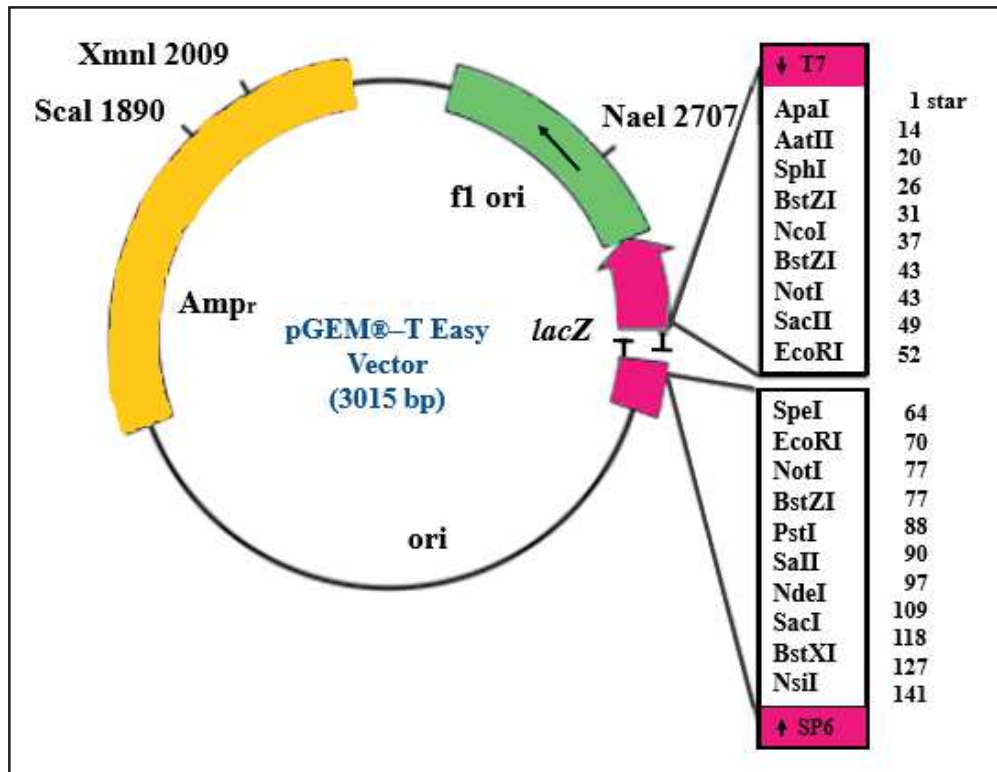


FIGURA M5. Esquema del vector utilizado en el proceso de clonación.

La utilidad de este vector se debe a la presencia de un gen de resistencia a ampicilina (*Amp^r*), que permite seleccionar las bacterias que han incorporado el plásmido, con o sin fragmento; y a que el lugar de inserción se localiza a la mitad del gen *lacZ*. Éste último codifica la enzima β -galactosidasa que capaz de hidrolizar el X-gal disuelto en el medio de agar de una placa de cultivo produciendo un compuesto azul insoluble. Esto indica que las colonias crecidas en la placa que posean un gen *lacZ* funcional serán de coloración azul y las que tengan interrumpido el operón Lac con la secuencia del gen de interés a clonar serán blancas. La figura M6 muestra un esquema del proceso de clonación.

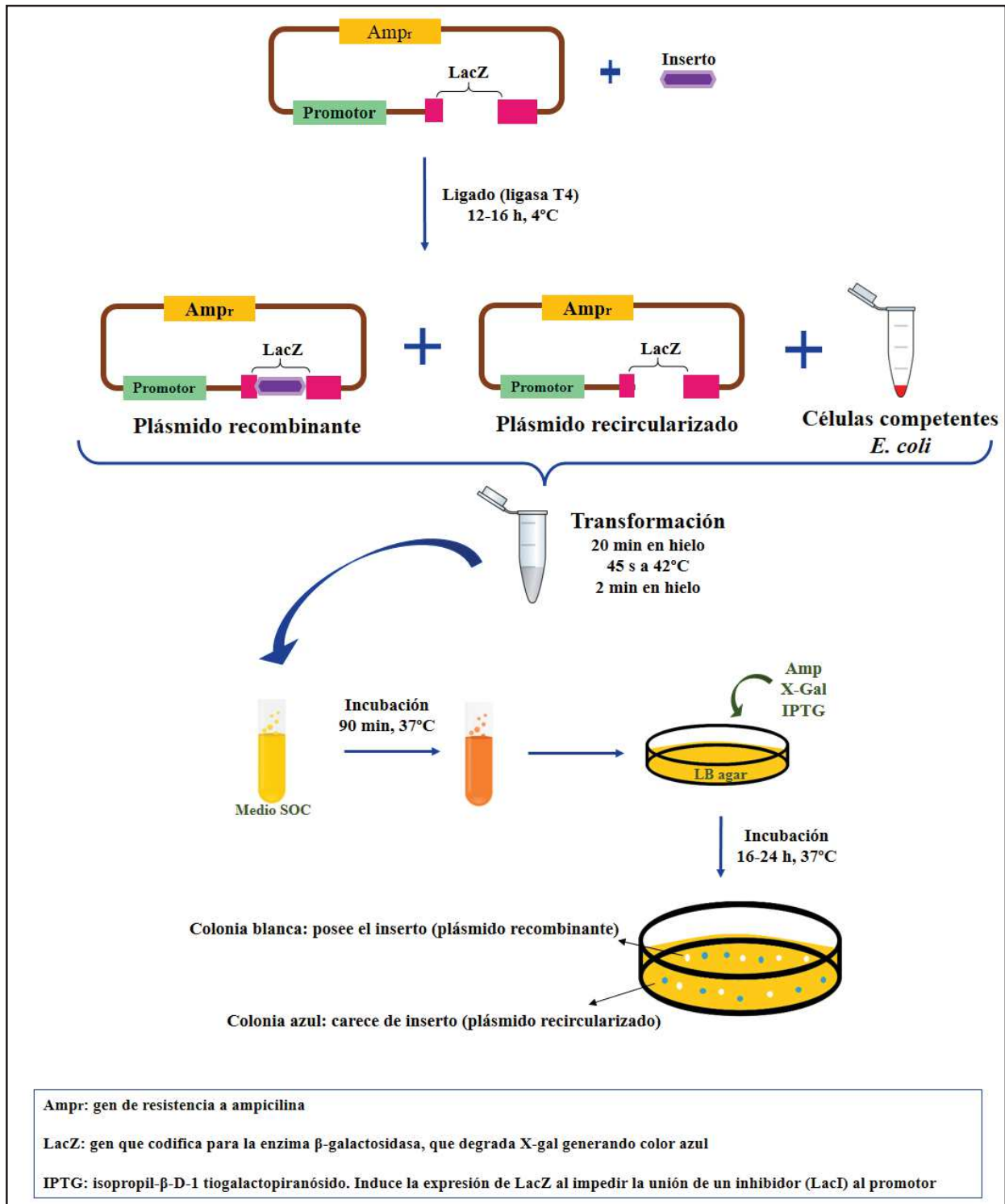


FIGURA M6. Esquema del proceso de clonación.

El protocolo de clonación empleado en este estudio fue el siguiente:

- Preparar la reacción de ligado mezclando 5 µl de tampón (*2x Rapid Ligation Buffer*, PROMEGA), 1 µl de plásmido (*pGEM-T Easy Vector*, PROMEGA), 3 µl del producto de amplificación y 1 µl de ligasa (*T4 DNA Ligase*, PROMEGA).
- Incubar durante 12-16 horas a 4 °C.
- Añadir 50 µl de células competentes (*JM109 Competent Cells, High efficiency*, PROMEGA) a un tubo conteniendo 2 µl de la mezcla de ligado.
- Incubar en hielo durante 20 minutos.
- Incubar en un baño a 42 °C durante 40-45 segundos.
- Incubar en hielo durante 2 minutos.
- Añadir 950 µl de medio SOC, que contiene 30,7 g/l de medio SOB (*SOB media*, PRONADISA) y glucosa 20 mM (dextrosa, PRONADISA).
- Incubar a 37 °C en agitación (150 rpm) durante 90 minutos.
- Sembrar 100 µl en placas con medio LB Agar conteniendo ampicilina, IPTG y X- GAL (2 placas por muestra). Un litro de medio contiene 35 g de LB Agar (LB Agar LENNOX, PRONADISA), 10 ml de ampicilina a 10 mg/ml (ampicillin sodium salt irradiated, GIBCO), 0,5 ml de IPTG 1M (IPTG *Dioxane-Free*, PROMEGA) y 1,6 ml de X-Gal 50 mg/ml (X-GAL, PROMEGA).
- Incubar a 37 °C durante 16-24 horas.
- Opcionalmente, las placas pueden mantenerse a 4 °C durante unas horas después de la incubación para favorecer la aparición de color azul en las colonias “negativas”.

4.6 Obtención de las secuencias de ADN de los fragmentos clonados

Una vez incubadas las placas se seleccionaron 10 colonias blancas de cada una (20 colonias por muestra) para extraer el ADN bacteriano mediante el siguiente procedimiento:

- Resuspender la colonia en 20 µl de agua destilada en un tubo de 0,2 ml.
- Incubar durante 5 minutos a 95 °C en un termociclador.
- Incubar durante 5 minutos en hielo.
- Centrifugar 3 minutos a 14000 rpm.
- Pasar el sobrenadante (unos 15 µl) a un tubo de 1,5 ml. Conservar a -20 °C.

- Posteriormente se amplificaron las secuencias de histocompatibilidad de clase I clonadas utilizando como molde el ADN extraído de las bacterias (que incluye el cromosoma bacteriano y el plásmido con el inserto). La PCR se desarrolló en las mismas condiciones en que se hizo con el ADN genómico (ver apartado 4.1). Los productos de amplificación fueron sometidos a electroforesis para separar el fragmento correspondiente (ver apartado 4.2) y finalmente fueron secuenciado (ver apartado 4.3).

5. Obtención de secuencias de ADN del gen de citocromo b

Es necesario conocer las relaciones evolutivas existentes entre las especies del género *Carduelis* de este trabajo para poder así interpretar el proceso evolutivo al que han estado sometidos los intrones de las secuencias de histocompatibilidad de estas especies. Ya que las secuencias del MHC no están recomendadas para analizar relaciones filogenéticas debido a su peculiar modo de evolución, se recurrió al gen mitocondrial del citocromo b para analizar la filogenia de las especies estudiadas.

Las reacciones de amplificación de las secuencias de citocromo b se llevaron a cabo en un termociclador (*Mastercycler epgradient S*, EPPENDORF). La ADN polimerasa utilizada fue *Taq DNA Polymerase* (INVITROGEN), suministrada junto con el tampón *Taq DNA Polymerase PCR Buffer*, 10X (INVITROGEN), el cloruro de magnesio de 50 mM (INVITROGEN) y los dNTPs (100 mM dNTP *Set*, INVITROGEN). El cebador directo utilizado fue L14841: 5'-AAAAGCTTCCATCCAACATCTCAGCATGATGAAA-3' (Kocher *et al.* 1989), y el inverso utilizado fue H15767: 5'ATGAAGGGATGTTCTACTGGTTG-3' (Edwards *et al.* 1991). Estos cebadores fueron suministrados por la casa *Thermo Fisher Scientific*.

Tras el proceso de amplificación bajo las condiciones descritas en las Tablas M5 y M6, los productos de amplificación fueron sometidos a electroforesis para separar el fragmento deseado siguiendo el mismo protocolo descrito en el apartado 4.2. Después de cortar la banda correspondiente al gen de citocromo b de cada una de las muestras, se realizaron las purificaciones siguiendo las pautas del apartado 4.3 y finalmente fueron secuenciadas (ver apartado 4.4).

Reactivos	Concentración	Volumen
dNTP	2,5 mM	1,6 µl
Tampón	10x	2 µl
MgCl ₂	25 mM	2 µl
Cebador L14841	10 µM	1,2 µl
Cebador H15767	10 µM	1,2 µl
Taq Polimerasa	5 U/µl	0,3 µl
ADN	100-200 ng/µl	1 µl
Agua		10,7 µl
Total:		20 µl

TABLA M5. Cantidad de cada reactivo, según la concentración dada, para un volumen de reacción de 20 µl por muestra empleada en la técnica de PCR.

Nº Ciclos	Temperatura	Tiempo	Proceso
1	95 °C	5 min	Desnaturalización inicial
	96 °C	30 s	Desnaturalización
45	50 °C	35 s	Hibridación
	72 °C	1 min 10 s	Elongación
1	72 °C	20 min	Elongación final

TABLA M6. Programa de amplificación utilizado en la técnica de PCR.

6. Análisis de las secuencias obtenidas de los genes del MHC

6.1 Caracterización de las secuencias de ADN

Para verificar que las secuencias de ADN obtenidas corresponden a los genes del MHC de clase I se llevó a cabo un alineamiento con otras secuencias previamente descritas con las que comparten una gran identidad. Para ello, se utilizó el programa MEGA 7 (Kumar *et al.* 2016). Una vez obtenidas las secuencias de estos genes, se detectó el inicio y el final

del intrón 2, y con este mismo programa se seleccionaron las secuencias intrónicas para su posterior análisis. Los análisis realizados fueron:

- **Composición nucleotídica**

Se realizó un cálculo de las frecuencias nucleotídicas de cada secuencia y se calculó el porcentaje GC de cada intrón. Para ello, se utilizó el programa MEGA 7 (Kumar *et al.* 2016).

- **Distancia genética**

La distancia genética entre dos secuencias de ADN se calcula en base a las diferencias que se observan entre ellas. La manera más simple de calcularlo consiste en contar el número de posiciones variables existentes entre dos secuencias y dividirlo por el número total de nucleótidos de la secuencia; ésta es la denominada “distancia p” (Nei 1987). Para ello utilizamos el programa MEGA 7 (Kumar *et al.* 2016).

- **Porcentaje de similitud**

Para completar el análisis de la conservación y evolución de las secuencias del intrón 2 del MHC de clase I, calculamos el porcentaje de similitud que hay entre ellas con el programa BioEdit (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>).

7. Filogenia basada en el gen citocromo b

Para comprender el proceso evolutivo al que han estado sometidos los genes de histocompatibilidad, es necesario conocer las relaciones evolutivas entre las distintas especies estudiadas en este trabajo. Las secuencias de los genes del MHC no están recomendadas para analizar relaciones filogenéticas debido a su peculiar modo de evolución, y por ello, se utiliza un gen que evolucione a una velocidad más constante como es el caso del citocromo b.

7.1 Métodos filogenéticos

La comparación de las secuencias se hace a través de distintas metodologías y las relaciones taxonómicas se representan a través de árboles filogenéticos. Los métodos filogenéticos constan de dos partes. La primera de ellas se lleva a cabo a través de los modelos de sustitución nucleotídica, y la segunda a través de los modelos de reconstrucción de árboles filogenéticos.

7.1.1 Modelos de sustitución nucleotídica

Para calcular una distancia genética es necesario contar el número de posiciones variables existentes entre dos secuencias y dividirlo por el número total de nucleótidos de la secuencia; ésta es la denominada 'distancia p' (Nei 1987). La distancia p asume que todos los cambios de nucleótido ocurren con la misma frecuencia y en cualquier parte de la molécula. Sin embargo, la experiencia nos dice que el proceso evolutivo no es tan simple como eso. Se sabe que ciertos tipos de mutaciones ocurren con más frecuencia que otros y que estas mutaciones no se distribuyen por igual a lo largo de toda la secuencia. Por todo ello, se han desarrollado los modelos de sustitución nucleotídica que tienen en cuenta ciertas variables antes de realizar el cálculo de las distancias genéticas. Estas variables son:

- **Frecuencias nucleotídicas**

En los programas de análisis filogenéticos se puede asumir que la frecuencia de cada nucleótido es la misma para todos (0,25) o bien que es distinta para cada uno de ellos. Si es diferente dependerán de si la frecuencia de cada nucleótido puede calcularse empíricamente a partir de la matriz de secuencias de ADN, o bien de si se estima como un valor de probabilidad aplicable a cualquier matriz de ese tipo independientemente de las secuencias que la compongan.

- **Número de tipos de sustituciones**

En los programas de análisis filogenéticos se puede asumir que cualquier cambio nucleotídico es igualmente probable ($A/C = A/G = A/T = C/T = C/G = G/T$) o bien que cada tipo de cambio nucleotídico tiene su propia probabilidad. En el caso de que fueran iguales, el número de tipos de sustituciones sería 1, y si fueran diferentes sería 6. De todas las posibilidades intermedias, la más utilizada es la que tiene en cuenta dos tipos de cambios ($nst=2$), denominados transiciones y transversiones.

- **Proporción de sitios invariantes**

En una matriz de secuencias de ADN, el número de posiciones conservadas puede afectar el resultado filogenético. Un número elevado de posiciones invariantes reduce las posibilidades de obtener un resultado filogenético fiable (ya que son escasos los elementos variables de los que se dispone para realizar el análisis) y un número elevado de posiciones variables también puede afectar al resultado porque indica un ritmo

evolutivo elevado que puede estar enmascarando dobles cambios en una posición (por ejemplo: A→G→A), fenómeno denominado saturación, que provoca una subestimación de la distancia genética. Esta variable puede ser tomada en cuenta o no. En caso de considerarse, puede tomar un valor fijo o bien estimado.

7.1.2 Métodos de reconstrucción de árboles filogenéticos

Una vez obtenidas las distancias genéticas de las secuencias de estudio, hay tres métodos de reconstrucción de árboles filogenéticos para expresar gráficamente las relaciones evolutivas.

- **Máxima verosimilitud o *Maximum likelihood* (ML)**

El método de Máxima verosimilitud, también llamado método ML (del inglés, *Maximum likelihood*) evalúa la probabilidad de obtener un árbol final a partir de una matriz de secuencias dada, de acuerdo con los valores que toman las variables que definen el modelo de sustitución nucleotídica. El árbol (o árboles) resultante será el que tenga mayor valor de probabilidad (denominado *likelihood score*). El método ML busca el árbol más probable que hayan generado los datos que hemos observado. En este método partimos de los datos y de un modelo de evolución. A partir de esta base se calcula la probabilidad de que nuestros datos hayan sido generados por los distintos árboles posibles y se devuelve el árbol que presenta una máxima probabilidad. Este método tiene la ventaja frente a los métodos de distancias de utilizar con mayor eficiencia la información filogenética contenida en el alineamiento múltiple. Es decir, dado un mismo alineamiento éste método tiende a generar un resultado más cercano a la realidad (Felsenstein 1981). El modelo escogido fue GTR+I+G (General Time Reversible), que considera 6 tipos de sustitución nucleotídica y tiene en cuenta la proporción de sitios invariantes y la distribución gamma de las tasas de sustitución en los sitios variables. El árbol fue linearizado de acuerdo con el modelo de Thorne (Thorne *et al.* 1998), que admite un ritmo de cambio variable a lo largo del tiempo y distinto para cada linaje, con el fin de conocer la antigüedad relativa de cada grupo. Además se asignaron tiempos de divergencia aproximados en base a análisis anteriores del grupo (Arnaiz-Villena *et al.* 2014).

- **Inferencia bayesiana (IB)**

La Inferencia bayesiana (IB) calcula una probabilidad *a posteriori* de una serie de modelos a partir de unas observaciones y de una probabilidad *a priori* de esos mismos modelos. Es decir, dadas unas observaciones la Inferencia bayesiana actualiza las probabilidades de que los modelos sean correctos. En caso de la filogenia bayesiana, dados los datos que hemos observado, en nuestro caso una serie de secuencias y unas probabilidades *a priori*, se generan unas probabilidades tanto para el conjunto de posibles árboles como para los parámetros del modelo de mutación. Este método es el que más tiempo y recursos computacionales requiere, pero suele considerarse como el método que arroja unos mejores resultados. Un árbol no tiene demasiada utilidad si no evaluamos su significación estadística. Un método filogenético siempre generará un resultado a partir de cualquier conjunto de datos, pero esto no implica que este resultado sea fiable.

En los métodos bayesianos los nodos tienen asociados una fiabilidad o probabilidad posterior que indica la confianza que podemos tener en ellos, pero en el resto de metodologías debemos utilizar algún método para hacer esta evaluación. Un método ideal consistiría en obtener varios conjuntos de datos independientes y generar a partir de ellos distintos árboles. Comparando qué nodos son compartidos por estos árboles y cuáles no podríamos hacernos una idea de qué resulta fiable y qué no. Desgraciadamente este método no resulta práctico por lo que se han desarrollado otros algoritmos que nos ofrecen algo similar a tener varios conjuntos de datos para obtener distintos árboles. Uno de los más utilizados es el bootstrap. Este método se puede aplicar a todos los métodos y consiste en crear réplicas de los alineamientos a partir del original, eliminando cierto número de posiciones al azar en cada réplica. El número final de posiciones se mantiene constante añadiendo duplicaciones de los sitios que han permanecido. Para cada una de estas réplicas aplicaremos el método de reconstrucción filogenética y generaremos un árbol. El paso final será evaluar para cada nodo el porcentaje de árboles en los que aparece. Los nodos con un alto valor de bootstrap tienen, si se cumplen las asunciones del método utilizado, una probabilidad alta de ser correctos mientras que los que tienen un bajo valor de bootstrap podrían haberse generado simplemente por azar. Se ha discutido mucho cuales son los valores límite que indicarían que un nodo es fiable y no se ha llegado a conclusiones demasiado claras. Evidentemente un nodo que aparece en el 95% de los árboles tiene una apariencia de ser sólido mientras que uno que aparezca en un 25% de

los árboles no parece demasiado fiable. Pero en los casos intermedios es difícil llegar a una conclusión demasiado clara (Huelsenbeck & Ronquist 2001).

- **Método del vecino más próximo o *Neighbor-Joining* (NJ)**

El Método del vecino más próximo o más conocido como método NJ (del inglés, *Neighbor-Joining*) analiza todos los árboles resultantes de agrupar dos secuencias, dejando las demás desagrupadas. Una vez construidos todos esos árboles (uno por posible pareja) se calcula la longitud total de cada uno y se selecciona como válido el que muestra la menor longitud total. Esto significa que las secuencias quedan agrupadas de manera que se minimiza la distancia genética existente entre ellas tomadas en su conjunto, lo que no implica que dos secuencias agrupadas en el mismo nodo no puedan tener una tercera más próxima de acuerdo con la matriz de distancias inicial. (Saitou & Nei 1987).

RESULTADOS

1. Secuencias del intrón 2 del MHC

En este trabajo se han secuenciado los genes del MHC de 18 especies del género *Carduelis* y se ha estudiado el intrón 2 de cada una de ellas. Este intrón ha sido comparado con la secuencia de *Fringilla coelebs*, especie considerada como grupo externo de este estudio. Además, también se ha comparado el intrón 2 de todas estas especies pertenecientes al orden *Passeriformes*, con el de *Gallus gallus*, que pertenece al orden *Galliformes*.

A partir de muestras de sangre se obtuvo ADN, el cual fue amplificado, clonado y secuenciado dos veces, obteniéndose un resultado idéntico de cada réplica para cada individuo. Las secuencias de ADN obtenidas se correspondían con los genes del MHC de clase I que previamente han sido descritos en estos géneros. Los análisis de estas secuencias se llevaron a cabo con los programas MEGA 7 y BioEdit y los resultados se muestran en los siguientes apartados.

1.1 Posiciones conservadas

Se realizó un alineamiento múltiple de todas las secuencias de intrones obtenidas para conocer el número de posiciones conservadas que presentaban estas especies. Cuando comparamos las 18 especies del género *Carduelis* se observó que un total de 283 posiciones se conservan en todos los intrones de jilgueros variando, únicamente, 28 posiciones. Los resultados del análisis se muestran en la Figura R1. Cuando incluimos a *F. coelebs* en el alineamiento se observó que el número de posiciones conservadas disminuyó a 244, variando en este caso 67 posiciones. La longitud total de los intrones de nuestras especies de jilgueros fue de 311 pb, mientras que la longitud del intrón del pinzón común fue de 293 pb. También, se quiso alinear el intrón de *G. gallus* con todas las especies *Passeriformes* de este trabajo. En esta comparativa se pudo apreciar que el intrón del pollo es mucho más corto que los del resto de especies, con una longitud de 229 pb. Este intrón, al pertenecer al orden *Galliformes*, es muy diferente del resto con solamente 87 posiciones conservadas en todas las especies.

	1										10										20									
<i>C. atrata</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	T	T	G	G	A	A		
<i>C. atriceps</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	C	T	G	G	A	A		
<i>C. barbata</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	T	T	G	G	A	A		
<i>C. c. parva</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	C	T	G	G	A	A		
<i>C. citrinella</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	C	T	G	G	A	A		
<i>C. crassirostris</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	T	T	G	G	A	A		
<i>C. cucullata</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	T	T	G	G	A	A		
<i>C. dominicensis</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	T	T	G	G	A	A		
<i>C. lawrencei</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	C	T	G	G	A	A		
<i>C. magellanica</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	T	T	G	G	A	A		
<i>C. notata</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	T	T	G	G	A	A		
<i>C. olivacea</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	T	T	G	G	A	A		
<i>C. p. hesperophila</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	T	T	G	G	A	A		
<i>C. pinus</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	C	T	G	G	A	A		
<i>C. spinescens</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	T	T	G	G	A	A		
<i>C. spinus</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	C	T	G	G	A	A		
<i>C. xanthogastra</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	T	T	G	G	A	A		
<i>C. yarrellii</i>	G	T	G	A	G	T	G	G	G	G	G	A	T	-	C	T	G	T	G	C	A	G	T	T	G	G	A	A		
<i>F. coelebs</i>	G	T	G	A	G	T	G	G	G	G	G	G	T	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A		
<i>G. gallus</i>																												G		

	30										40										50									
<i>C. atrata</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	G
<i>C. atriceps</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	A
<i>C. barbata</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	G
<i>C. c. parva</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	G
<i>C. citrinella</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	A
<i>C. crassirostris</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	G
<i>C. cucullata</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	G
<i>C. dominicensis</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	G
<i>C. lawrencei</i>	T	G	A	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	A
<i>C. magellanica</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	G
<i>C. notata</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	G
<i>C. olivacea</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	G
<i>C. p. hesperophila</i>	T	G	G	G	A	A	C	T	G	T	G	G	C	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	G
<i>C. pinus</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	A
<i>C. spinescens</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	G
<i>C. spinus</i>	T	G	A	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	G
<i>C. xanthogastra</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	G
<i>C. yarrellii</i>	T	G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	G	T	G	G	A	G	A	T	C	A	G
<i>F. coelebs</i>	T	G	G	A	A	A	C	T	G	T	G	G	G	G	C	T	T	G	T	G	T	G	G	A	G	A	T	T	G	A
<i>G. gallus</i>	T	G	A	G	C	A	C	G	G	C	C	G	G	G	G	C	C	G	C	G	-	-	-	-	G	C	T	C	C	G

	60										70										80									
<i>C. atrata</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	G	G	A	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. atriceps</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	G	G	T	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. barbata</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	G	G	A	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. c. parva</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	G	G	A	T	C	T	G	T	A	G	G	T	C	T	C	A	G
<i>C. citrinella</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	G	G	A	T	T	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. crassirostris</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	G	G	A	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. cucullata</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	G	G	A	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. dominicensis</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	G	G	A	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. lawrencei</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	G	G	A	T	T	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. magellanica</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	A	G	A	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. notata</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	A	G	A	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. olivacea</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	A	G	A	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. p. hesperophila</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	G	G	A	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. pinus</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	G	G	T	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. spinescens</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	A	G	A	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. spinus</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	G	G	A	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. xanthogastra</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	A	G	A	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>C. yarrellii</i>	T	G	G	G	G	C	T	G	G	G	G	T	G	A	G	A	T	C	T	A	T	A	G	G	T	C	T	C	A	G
<i>F. coelebs</i>	T	G	G	G	G	C	T	T	G	G	A	T	G	G	G	A	T	C	T	C	T	A	G	G	T	C	T	C	G	G
<i>G. gallus</i>	T	G	G	G	T	G	T	G	G	G	A	T	G	G	G	C	T	C	C	A	T	G	G	-	-	C	G	C	A	G

	90										100										110									
<i>C. atrata</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. atriceps</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. barbata</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. c. parva</i>	A	T	-	T	G	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. citrinella</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. crassirostris</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. cucullata</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. dominicensis</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. lawrencei</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. magellanica</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. notata</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. olivacea</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. p. hesperophila</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. pinus</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. spinescens</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. spinus</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. xanthogastra</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>C. yarrellii</i>	A	T	G	C	A	G	A	T	T	C	A	C	A	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	A
<i>F. coelebs</i>	A	T	A	C	G	G	A	T	T	C	A	C	A	G	G	G	C	T	C	T	G	A	G	G	G	G	C	T	G	A
<i>G. gallus</i>	-	T	G	C	C	G	-	C	C	C	A	C	A	-	-	-	C	C	C	C	C	C	A	G	G	C	C	T	G	G

	120										130										140														
<i>C. atrata</i>	A	A	-	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. atriceps</i>	A	A	-	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. barbata</i>	A	A	-	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. c. parva</i>	A	A	-	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. citrinella</i>	A	A	-	T	G	G	G	T	A	-	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. crassirostris</i>	A	A	-	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. cucullata</i>	A	A	A	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. dominicensis</i>	A	A	A	T	G	G	G	-	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. lawrencei</i>	A	A	A	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. magellanica</i>	A	A	A	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. notata</i>	A	A	A	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. olivacea</i>	A	A	A	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. p. hesperophila</i>	A	A	A	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. pinus</i>	A	A	-	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. spinescens</i>	A	A	A	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. spinus</i>	A	A	A	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. xanthogastra</i>	A	A	A	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>C. yarrellii</i>	A	A	-	T	G	G	G	G	A	T	C	C	A	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>F. coelebs</i>	A	A	-	T	G	G	G	G	T	T	C	T	G	T	T	A	G	G	C	T	G	G	G	A	T	G	G	G	A	T	G	G	G	A	T
<i>G. gallus</i>	C	C	C	T	G	-	-	-	C	C	C	G	G	C	G	G	C	A	C	C	G	T	C	C	C	G	G	G	G	C	G	G	G	G	C

	150										160										170																
<i>C. atrata</i>	A	T	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. atriceps</i>	A	T	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. barbata</i>	A	T	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. c. parva</i>	A	T	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. citrinella</i>	A	G	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. crassirostris</i>	A	T	C	T	T	G	T	G	A	T	C	C	C	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. cucullata</i>	A	T	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. dominicensis</i>	A	T	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. lawrencei</i>	A	T	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. magellanica</i>	A	T	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. notata</i>	A	T	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. olivacea</i>	A	T	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. p. hesperophila</i>	A	T	C	T	A	A	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T		
<i>C. pinus</i>	A	T	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. spinescens</i>	A	T	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. spinus</i>	A	T	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. xanthogastra</i>	A	T	C	T	T	G	T	G	A	T	C	C	A	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>C. yarrellii</i>	A	T	C	T	T	G	T	G	A	T	C	C	C	T	A	G	G	T	T	G	G	G	A	T	G	T	G	G	A	T	G	T	G	G	A	T	
<i>F. coelebs</i>	A	T	C	T	T	G	T	G	A	T	C	C	-	-	-	-	-	-	-	-	-	-	-	A	T	G	T	G	G	A	T	G	T	G	G	A	T
<i>G. gallus</i>	T	G	C	C	C	G	T	C	A	C	A	G	C	C	C	C	A	C	C	G	C	G	C	T	-	C	G	G	G	G	C	G	G	G	G	C	

	180	190										200																		
<i>C. atrata</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. atriceps</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. barbata</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. c. parva</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	A	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. citrinella</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. crassirostris</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. cucullata</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. dominicensis</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. lawrencei</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. magellanica</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. notata</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. olivacea</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. p. hesperophila</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. pinus</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. spinescens</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. spinus</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. xanthogastra</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>C. yarrellii</i>	T	G	A	T	G	T	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>F. coelebs</i>	T	G	A	T	G	G	G	G	C	T	C	C	A	A	G	G	G	G	C	T	G	G	G	A	T	G	A	A	G	C
<i>G. gallus</i>	T	G	C	C	G	C	G	T	C	C	C	G	G	G	G	G	G	A	C	C	C	C	A	A	C	C	C	A	T	C

	210	220										230																		
<i>C. atrata</i>	T	C	T	G	A	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	C	G	G	G	G	C	T	C
<i>C. atriceps</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	T	G	G	G	G	C	T	C
<i>C. barbata</i>	T	C	T	G	A	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	C	G	G	G	G	C	T	C
<i>C. c. parva</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	T	G	G	G	G	C	T	C
<i>C. citrinella</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	T	G	G	G	G	C	T	C
<i>C. crassirostris</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	C	G	G	G	G	C	T	C
<i>C. cucullata</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	C	G	G	G	G	C	T	C
<i>C. dominicensis</i>	T	C	A	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	C	G	G	G	G	C	T	C
<i>C. lawrencei</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	T	G	G	G	G	C	T	C
<i>C. magellanica</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	C	G	G	G	G	C	T	C
<i>C. notata</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	C	G	G	G	G	C	T	C
<i>C. olivacea</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	C	G	G	G	G	C	T	C
<i>C. p. hesperophila</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	C	G	G	G	G	C	T	C
<i>C. pinus</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	T	G	G	G	G	C	T	C
<i>C. spinescens</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	C	G	G	G	G	C	T	C
<i>C. spinus</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	C	G	G	G	G	C	T	C
<i>C. xanthogastra</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	C	G	G	G	G	C	T	C
<i>C. yarrellii</i>	T	C	T	G	G	G	G	A	T	C	T	C	C	A	T	C	A	G	A	C	T	C	C	G	G	G	G	C	T	C
<i>F. coelebs</i>	T	C	T	G	G	G	G	G	T	C	-	C	C	A	T	C	A	G	A	C	T	C	-	-	-	-	-	-	-	-
<i>G. gallus</i>	C	C	C	G	-	-	-	-	C	T	G	C	A	G	T	G	G	G	A	G	C	C	C	C	G	G	A	G	C	C

	240				250				260																					
<i>C. atrata</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. atriceps</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. barbata</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. c. parva</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. citrinella</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. crassirostris</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. cucullata</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. dominicensis</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. lawrencei</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. magellanica</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. notata</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. olivacea</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. p. hesperophila</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. pinus</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. spinescens</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. spinus</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. xanthogastra</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>C. yarrellii</i>	C	A	T	A	G	G	G	C	T	G	G	A	A	T	T	T	G	G	C	T	T	T	G	T	G	G	G	A	T	
<i>F. coelebs</i>	C	A	T	-	G	G	G	C	T	G	G	A	A	T	G	G	G	G	C	T	T	T	G	-	A	G	G	A	T	
<i>G. gallus</i>	G	G	A	G	G	G	G	C	C	C	C	-	-	T	C	A	C	C	C	C	C	T	G	C	C	C	G	G	C	T

	270				280				290																					
<i>C. atrata</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	G	T	C	T	C	T	A	T	G	G	G	T	A
<i>C. atriceps</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	A	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. barbata</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	G	T	C	T	C	T	A	T	G	G	G	T	A
<i>C. c. parva</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	A	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. citrinella</i>	G	G	G	A	C	A	C	A	A	T	T	C	T	A	T	G	G	G	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. crassirostris</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	G	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. cucullata</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	G	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. dominicensis</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	G	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. lawrencei</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	A	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. magellanica</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	G	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. notata</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	G	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. olivacea</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	G	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. p. hesperophila</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	G	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. pinus</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	A	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. spinescens</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	G	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. spinus</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	G	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. xanthogastra</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	G	T	C	T	C	T	G	T	G	G	G	T	A
<i>C. yarrellii</i>	G	G	G	A	C	A	C	A	A	T	T	C	C	A	T	G	A	G	T	C	T	C	T	G	T	G	G	G	T	A
<i>F. coelebs</i>	G	G	G	A	T	A	C	A	A	T	T	C	C	A	T	G	G	G	T	C	T	C	T	G	T	G	G	G	T	A
<i>G. gallus</i>	G	T	G	T	T	T	C	A	G																					

	300	310										
<i>C. atrata</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. atriceps</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. barbata</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. c. parva</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. citrinella</i>	C	A	A	T	T	C	C	C	C	C	A	G
<i>C. crassirostris</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. cucullata</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. dominicensis</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. lawrencei</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. magellanica</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. notata</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. olivacea</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. p. hesperophila</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. pinus</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. spinescens</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. spinus</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. xanthogastra</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>C. yarrellii</i>	C	A	A	T	C	C	C	C	C	C	A	G
<i>F. coelebs</i>	A	A	A	T	C	C	C	C	C	C	A	G
<i>G. gallus</i>												

FIGURA R1. Alineamiento de todas las secuencias del intrón 2. Las columnas moradas indican que todas las especies *Passeriformes* tienen el mismo nucleótido en esa posición. Las columnas blancas indican que hay alguna especie *Passeriformes* que no tiene el mismo nucleótido en esa posición, el cual se marca en color negro.

1.2 Frecuencias nucleotídicas

En la Tabla R1 se muestran las frecuencias nucleotídicas de cada una de las secuencias del intrón 2. El porcentaje medio de G en aves *Passeriformes* es del 38,1% y es más alto que el porcentaje de C con un valor del 17,1%. El porcentaje medio de A y T en estas mismas aves es del 19,8% y 25,0%, respectivamente. Los valores son muy similares tanto en las especies del género *Carduelis* como en *F. coelebs*, y el porcentaje medio de GC de estas aves *Passeriformes* es del 55,2%. Si se analizan las frecuencias nucleotídicas de *G. gallus*, se observa que los valores son muy diferentes a los obtenidos en los jilgueros con un porcentaje de T del 11,8%, de C del 42,4%, de A del 9,6% y de G del 36,2%. El contenido en GC del pollo es de 78,6%, mucho más elevado que el que presentan las aves *Passeriformes*.

	T	C	A	G
<i>C. atrata</i>	24,9	17,2	20,1	37,9
<i>C. atriceps</i>	25,2	17,2	19,7	37,9
<i>C. barbata</i>	24,9	17,2	20,1	37,9
<i>C. carduelis parva</i>	25,2	17,2	19,4	38,2
<i>C. citrinella</i>	25,6	16,2	19,5	38,6
<i>C. crassirostris</i>	24,9	17,5	19,1	38,5
<i>C. cucullata</i>	24,8	17,1	19,7	38,4
<i>C. dominicensis</i>	24,6	17,2	20,1	38,2
<i>C. lawrencei</i>	25,2	16,8	20,6	37,4
<i>C. magellanica</i>	24,8	17,1	20,0	38,1
<i>C. notata</i>	24,8	17,1	20,0	38,1
<i>C. olivacea</i>	24,8	17,1	20,0	38,1
<i>C. pinus</i>	25,2	17,2	19,7	37,9
<i>C. psaltria hesperophila</i>	24,5	17,4	20,6	37,4
<i>C. spinescens</i>	24,8	17,1	20,0	38,1
<i>C. spinus</i>	24,5	17,4	20,0	38,1
<i>C. xanthogastra</i>	24,8	17,1	20,0	38,1
<i>C. yarrellii</i>	24,9	17,5	19,4	38,2
<i>F. coelebs</i>	25,7	17,1	18,2	39,0
<i>G. gallus</i>	11,8	42,4	9,6	36,2

TABLA R1. Frecuencias nucleotídicas de las secuencias del intrón 2 de las especies estudiadas.

1.3 Distancias genéticas

Las distancias genéticas fueron calculadas entre las distintas especies estudiadas y se recogen en la Tabla R2. Los valores oscilan entre 0 y 1, indicando 0 que las secuencias son idénticas. Los valores más altos se corresponden con la única especie perteneciente al orden *Galliformes*, *G. gallus*, indicando que es la secuencia que más difiere del resto. De las especies que forman parte del orden *Passeriformes*, *F. coelebs* presenta los valores más elevados con respecto a todos los jilgueros. Hay especies que obtuvieron una distancia genética de 0,000 indicando que la secuencia del intrón 2 era la misma en las dos especies comparadas. Este hecho fue observado en tres casos diferentes. El primer

grupo de especies con la misma secuencia está formado por *C. atrata*, *C. barbata*, *C. crassirostris* y *C. cucullata*. El segundo grupo con la secuencia del intrón 2 idéntica lo forman *C. pinus* y *C. atriceps*. Y el último de ellos está formado por las especies *C. magellanica*, *C. notata*, *C. olivacea*, *C. spinescens*, *C. xanthogastra* y *C. yarrellii*. Dentro de los jilgueros, la distancia genética más alta fue de 0,038 y se corresponde con *C. c. parva* al compararla con *C. citrinella*, *C. lawrencei* y *C. p. hesperophila*.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
1 <i>C. atrata</i>																				
2 <i>C. atriceps</i>	0,010																			
3 <i>C. barbata</i>	0,000	0,010																		
4 <i>C. c. parva</i>	0,024	0,033	0,024																	
5 <i>C. citrinella</i>	0,014	0,014	0,014	0,038																
6 <i>C. crassirostris</i>	0,000	0,010	0,000	0,024	0,014															
7 <i>C. cucullata</i>	0,000	0,010	0,000	0,024	0,014	0,000														
8 <i>C. dominicensis</i>	0,005	0,014	0,005	0,029	0,019	0,005	0,005													
9 <i>C. lawrencei</i>	0,014	0,014	0,014	0,038	0,010	0,014	0,014	0,019												
10 <i>C. magellanica</i>	0,005	0,014	0,005	0,029	0,019	0,005	0,005	0,010	0,019											
11 <i>C. notata</i>	0,005	0,014	0,005	0,029	0,019	0,005	0,005	0,010	0,019	0,000										
12 <i>C. olivacea</i>	0,005	0,014	0,005	0,029	0,019	0,005	0,005	0,010	0,019	0,000	0,000									
13 <i>C. pinus</i>	0,010	0,000	0,010	0,033	0,014	0,010	0,010	0,014	0,014	0,014	0,014	0,014								
14 <i>C. p. hesperophila</i>	0,014	0,024	0,014	0,038	0,029	0,014	0,014	0,019	0,029	0,019	0,019	0,019	0,019	0,024						
15 <i>C. spinescens</i>	0,005	0,014	0,005	0,029	0,019	0,005	0,005	0,010	0,019	0,000	0,000	0,000	0,014	0,019						
16 <i>C. spinus</i>	0,005	0,014	0,005	0,029	0,019	0,005	0,005	0,010	0,010	0,010	0,010	0,010	0,014	0,019	0,010					
17 <i>C. xanthogastra</i>	0,005	0,014	0,005	0,029	0,019	0,005	0,005	0,010	0,019	0,000	0,000	0,000	0,014	0,019	0,000	0,010				
18 <i>C. yarrellii</i>	0,005	0,014	0,005	0,029	0,019	0,005	0,005	0,010	0,019	0,000	0,000	0,000	0,014	0,019	0,000	0,010	0,000			
19 <i>F. coelebs</i>	0,115	0,115	0,115	0,120	0,120	0,115	0,115	0,120	0,120	0,120	0,120	0,120	0,115	0,129	0,120	0,120	0,120	0,120	0,120	
20 <i>G. gallus</i>	0,512	0,517	0,512	0,522	0,517	0,512	0,512	0,512	0,517	0,517	0,517	0,517	0,517	0,522	0,517	0,507	0,517	0,517	0,517	0,531

TABLA R2. Distancias genéticas de las secuencias del intrón 2.

Con las distancias genéticas de cada una de las especies se realizó un dendograma mediante el método *Neighbour Joining* (NJ) para visualizar con mayor claridad la similitud existente entre las secuencias de intrones. El resultado se puede ver en la Figura R2 y se observa que al igual que se ve en la Tabla R2, se diferencian tres grupos de especies con la misma secuencia del intrón 2 entre ellas. El primer grupo está marcado de color rosa y lo engloban las especies *C. magellanica*, *C. notata*, *C. olivacea*, *C. spinescens*, *C. xanthogastra* y *C. yarrellii*. El segundo grupo se indica de color verde y está formado por *C. atrata*, *C. barbata*, *C. crassirostris* y *C. cucullata*. El tercer grupo lo forman *C. pinus* y *C. atriceps* y destaca en color azul. Como se observa en la Tabla R2, la distancia genética más alta dentro del grupo de los jilgueros, se corresponde con *C. c. parva* al compararla con *C. citrinella*, *C. lawrencei* y *C. p. hesperophila*. En el dendograma, se refleja este hecho siendo el jilguero europeo el taxón más alejado del resto de especies del género *Carduelis*. Las especies más distantes son *G. gallus* y *F. coelebs* porque sus secuencias son las más diferentes, como indican sus elevadas distancias genéticas.

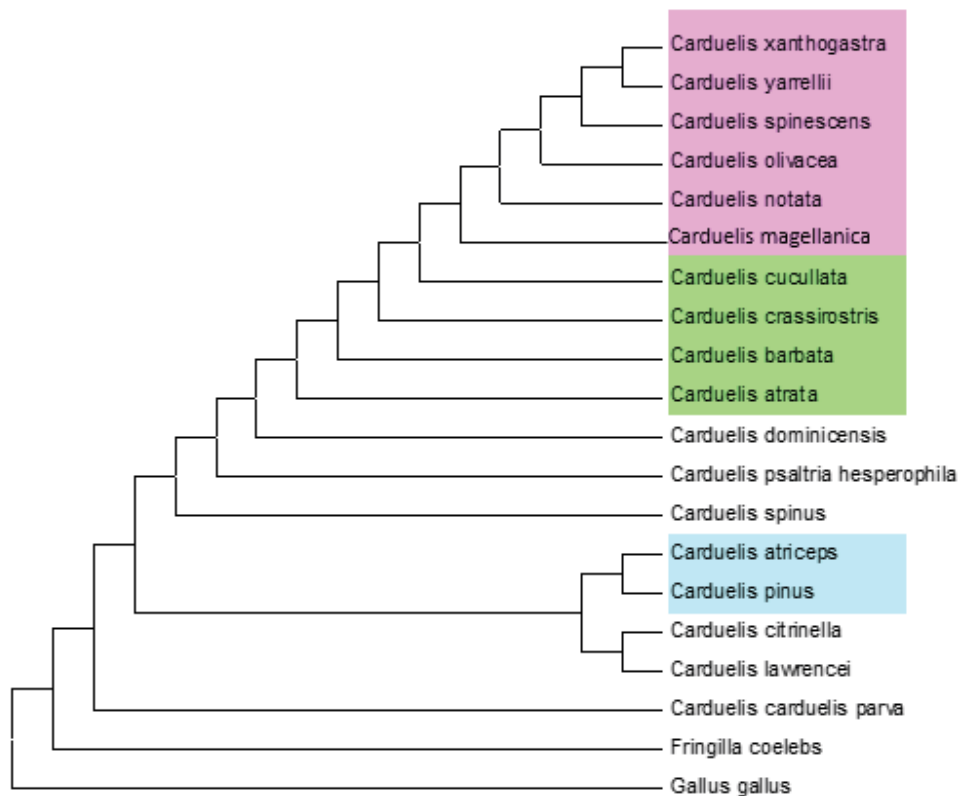


FIGURA R2. Dendograma de las distancias genéticas del intrón 2 realizado mediante el método *Neighbour Joining* (NJ). Los colores engloban las especies que tienen secuencias idénticas.

1.4 Porcentaje de similitud

Para calcular el porcentaje de similitud de dos secuencias se emplea la siguiente fórmula:

$$S = [(L_s \times 2) / (L_a + L_b)] \times 100$$

donde S es el porcentaje de similitud de las secuencias, L_s es el número de residuos alineados con características similares, y L_a , L_b son las longitudes totales de cada secuencia. La similitud de secuencias se refiere al porcentaje de residuos alineados que tienen características fisicoquímicas similares y que pueden ser sustituidos entre sí.

El porcentaje de similitud que existe entre las secuencias de intrones estudiadas se recoge en la Tabla R3. El intrón 2 con mayor diferencia fisicoquímica con respecto al resto de secuencias comparadas es el de *G. gallus*, con porcentajes que no superan el 20% de similitud. *F. coelebs* también presenta porcentajes bajos con valores que oscilan entre el 26 y el 40%. Hay secuencias que presentan un porcentaje de similitud del 100% indicando que sus características fisicoquímicas son idénticas. Se distinguen tres grupos con la máxima similitud. El primero de ellos lo forman *C. atrata* y *C. crassirostris*; en el segundo destacan *C. pinus* y *C. atriceps*; y el tercero está formado por *C. magellanica*, *C. notata*, *C. olivacea*, *C. spinescens* y *C. xanthogastra*. Las especies de jilgueros con características fisicoquímicas similares son *C. atrata*, *C. atriceps*, *C. barbata*, *C.c. parva*, *C. crassirostris*, *C. dominicensis*, *C. pinus* y *C. yarrellii*, ya que todas ellas presentan un porcentaje de similitud mayor del 90%. La especie del género *Carduelis* con la secuencia del intrón 2 más diferente al resto de jilgueros es *C. citrinella*, con unos porcentajes de similitud que no superan el 60%.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
1 <i>C. atrata</i>																				
2 <i>C. atriceps</i>	98,0																			
3 <i>C. barbata</i>	99,0	97,7																		
4 <i>C. c. parva</i>	97,1	97,7	96,8																	
5 <i>C. citrinella</i>	58,3	58,6	57,9	57,3																
6 <i>C. crassirostris</i>	100	98,1	99,0	97,1	58,3															
7 <i>C. cucullata</i>	58,7	57,4	58,1	56,5	55,5	58,7														
8 <i>C. dominicensis</i>	98,7	97,4	98,4	96,4	57,6	98,7	59,0													
9 <i>C. lawrencei</i>	57,4	57,4	56,8	55,8	56,1	57,4	98,1	57,7												
10 <i>C. magellanica</i>	58,4	57,1	57,7	56,1	55,2	58,4	99,7	58,7	97,7											
11 <i>C. notata</i>	58,4	57,1	57,7	56,1	55,2	58,4	99,7	58,7	97,7	100										
12 <i>C. olivacea</i>	58,4	57,1	57,7	56,1	55,2	58,4	99,7	58,7	97,7	100	100									
13 <i>C. pinus</i>	98,1	100	97,7	97,7	58,6	98,1	57,4	97,4	57,4	57,1	57,1	57,1								
14 <i>C. p. hesperophila</i>	58,7	57,4	58,1	56,4	54,5	58,7	98,7	59,0	96,8	98,4	98,4	98,4	57,4							
15 <i>C. spinescens</i>	58,4	57,1	57,7	56,1	55,2	58,4	99,7	58,7	97,7	100	100	100	57,1	98,4						
16 <i>C. spinus</i>	58,1	57,4	57,4	56,4	55,5	58,1	99,4	58,4	98,7	99,0	99,0	99,0	57,4	98,1	99,0					
17 <i>C. xanthogastra</i>	58,4	57,1	57,7	56,1	55,2	58,4	99,7	58,7	97,7	100	100	100	57,1	98,4	100	99,0				
18 <i>C. yarrellii</i>	99,7	97,7	98,7	96,8	57,9	99,7	58,4	98,4	57,1	58,7	58,7	58,7	97,7	58,4	58,7	57,7	58,7			
19 <i>F. coelebs</i>	27,2	26,5	27,5	27,2	29,9	27,2	40,6	27,5	40,3	40,3	40,3	40,3	26,3	39,4	40,3	40,3	40,3	26,9		
20 <i>G. gallus</i>	18,1	18,8	18,4	19,7	19,5	18,1	16,1	18,4	15,8	16,1	16,1	16,1	18,8	15,8	16,1	16,1	16,1	18,1	16,4	

Tabla R3. Porcentajes de similitud de las secuencias del intrón 2.

2. Filogenia mitocondrial

Para estudiar la relación filogenética de las especies del género *Carduelis* y compararla con la evolución del intrón 2, se llevaron a cabo distintos estudios filogenéticos utilizando el gen del citocromo b para la comparación de las secuencias. En la Figura R3 se recoge el árbol filogenético de las secuencias del citocromo b realizado mediante el método *Maximum likelihood* (ML), el cual fue linearizado para determinar la antigüedad relativa de cada grupo, y en la Figura R4 se observa otro dendograma realizado con las mismas secuencias, pero en este caso realizado mediante Inferencia bayesiana (IB).

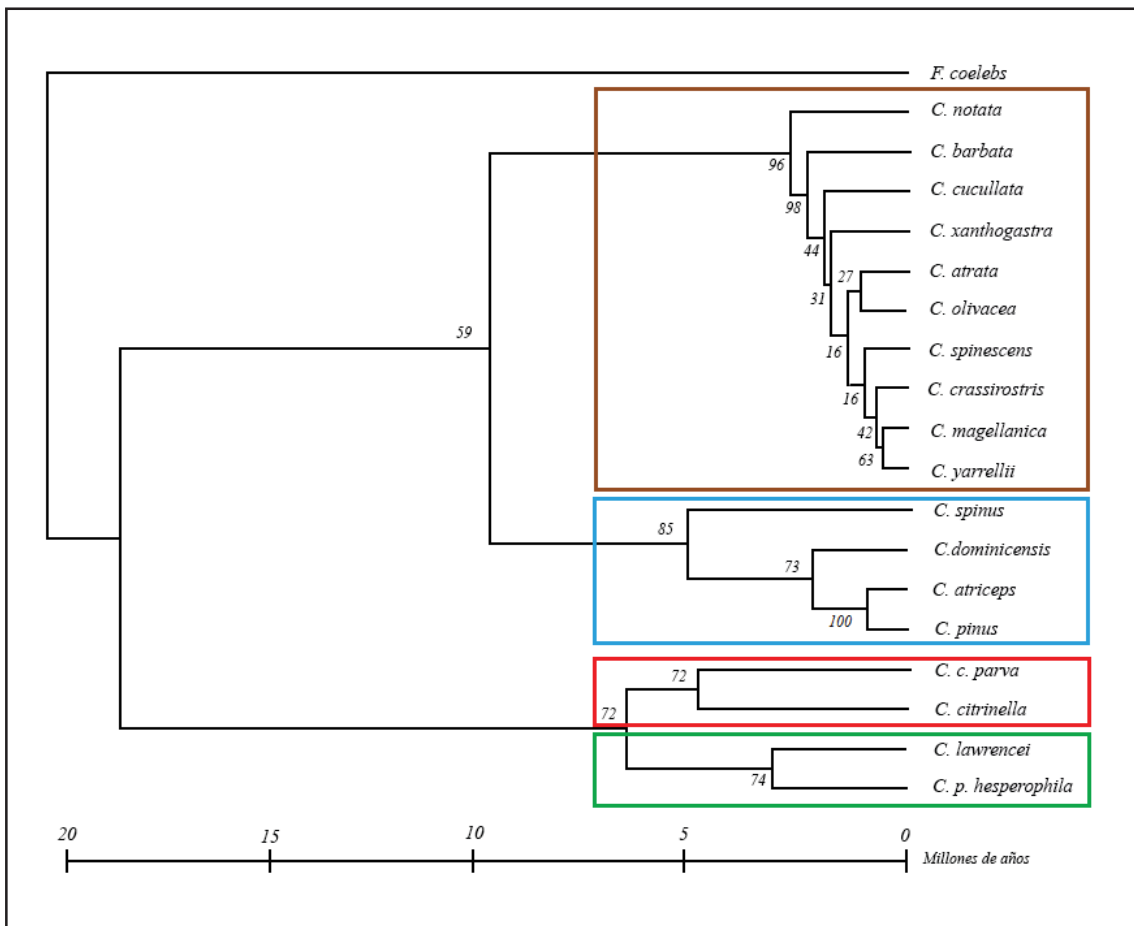


FIGURA R3. Dendograma realizado mediante el método *Maximum likelihood* (ML) basado en secuencias de citocromo b mitocondrial. Los porcentajes de fiabilidad de los nodos (valores de bootstrap) se obtuvieron tras 1.000 replicaciones y están indicados en cada nodo. Las especies señaladas con el marco azul corresponden a las especies de la radiación norteamericana, las marrón a la radiación sudamericana, las rojas a jilgueros europeos y las verdes a la radiación mesoamericana.

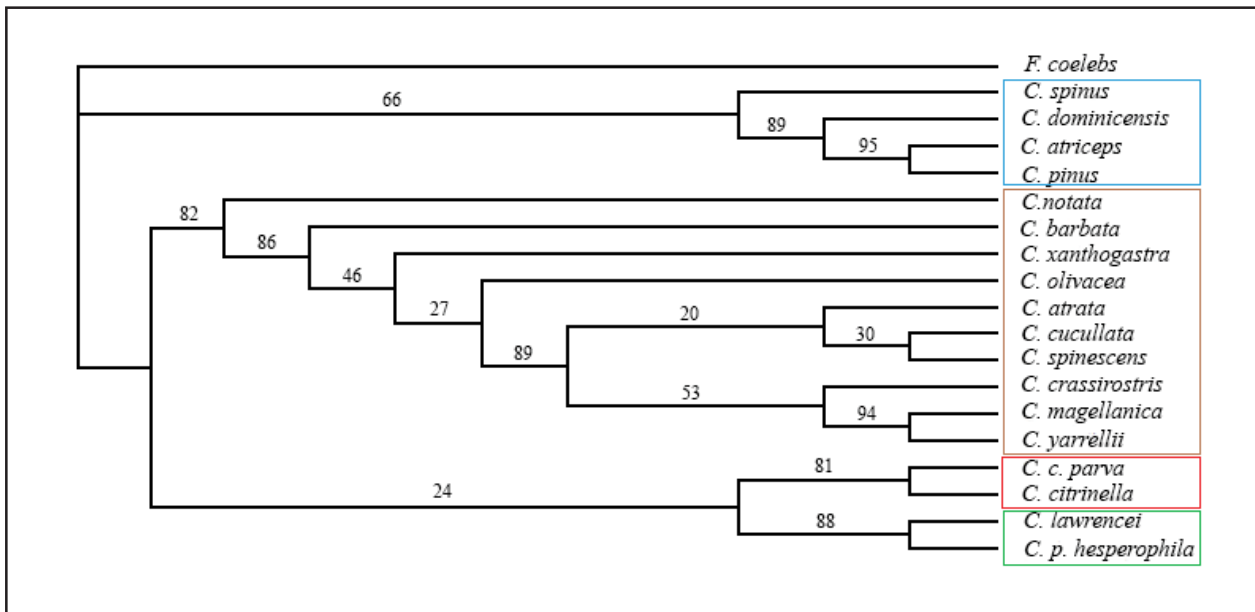


FIGURA R4. Dendrograma realizado mediante Inferencia bayesiana (IB) basado en secuencias de citocromo b mitocondrial. Los porcentajes de fiabilidad de los nodos se obtuvieron tras 1.000 replicaciones y están indicados en cada nodo. Las especies señaladas con el marco azul corresponden a las especies de la radiación norteamericana, las marrón a la radiación sudamericana, las rojas a jilgueros europeos y las verdes a la radiación mesoamericana.

Ambos resultados son prácticamente idénticos y en ellos se ve cómo todas las especies están agrupadas de acuerdo a las distintas radiaciones que ha habido en América. En la Figura R3 se observa, en azul, que *C. spinus* apareció, aproximadamente, hace 5 MA y que es el ancestro común a los jilgueros *C. dominicensis*, *C. atriceps* y *C. pinus*, que forman parte de la radiación norteamericana. El otro grupo a destacar está señalado en color marrón y *C. notata* es el ancestro común a todos ellos. Esta especie apareció hace 3 MA y dio lugar a los lúganos sudamericanos *C. barbata*, *C. xanthogastra*, *C. olivacea*, *C. atrata*, *C. cucullata*, *C. spinescens*, *C. crassirostris*, *C. magellanica* y *C. yarrellii*. Todos ellos forman parte de la radiación sudamericana. El grupo marcado en color rojo incluye a *C. lawrencei* y *C. p. hesperophila*, dos especies que aparecieron durante la radiación mesoamericana. Y el grupo señalado en color verde incluye a *C. c. parva* y *C. citrinella*, dos especies que surgieron en Europa hace 4,9 MA, aproximadamente. En la Figura R4, el dendrograma realizado mediante Inferencia bayesiana, agrupa las especies en los mismos grupos que se han determinado mediante máxima verosimilitud. En ambos

análisis, *F. coelebs* es el grupo externo a todos los jilgueros analizados y como se ve en la Figura R3, el pinzón común apareció hace 17,5 MA, aproximadamente. Los valores altos de fiabilidad apoyan nuestros resultados filogenéticos obtenidos. Sin embargo, hay veces que no se obtienen valores muy elevados y esto se debe a que el número de muestras analizadas es bajo y, por tanto, a los métodos filogenéticos les falta información de los cambios que se producen en las secuencias.

En la Figura R5 se muestra el análisis realizado con las secuencias del intrón 2 de cada especie mediante Inferencia bayesiana y se observa que las relaciones obtenidas no se corresponden con las observadas en las Figuras R3 y R4. Esto demuestra que las secuencias del MHC no son una buena opción a la hora de estudiar relaciones filogenéticas entre especies, ya que estos genes presentan un modo de evolución muy peculiar. En este dendograma no es posible reconocer los cuatro grupos filogenéticos destacados en las Figuras R3 y R4 y los valores de fiabilidad son muy bajos, indicando que los nodos no son muy fiables.

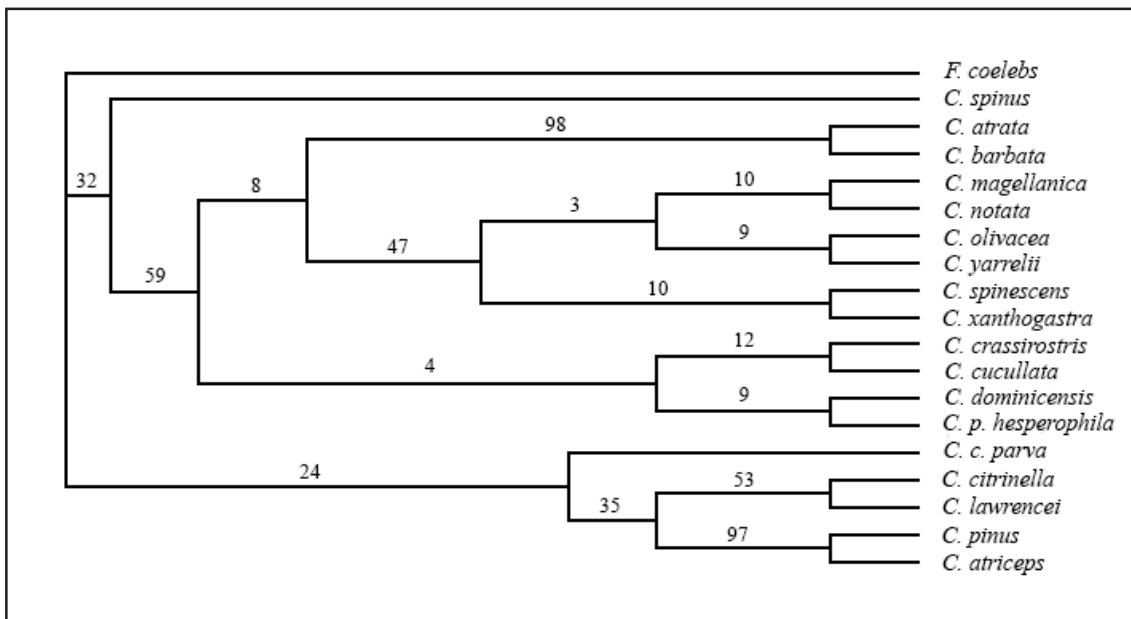


FIGURA R5. Dendograma realizado mediante Inferencia bayesiana (IB) basado en las secuencias del intrón 2 del MHC. Los porcentajes de fiabilidad de los nodos se obtuvieron tras 1.000 repeticiones y están indicados en cada nodo.

DISCUSIÓN

El estudio de los genes de histocompatibilidad de aves *Passeriformes* tiene múltiples aplicaciones y en la realización de este trabajo nos hemos centrado en analizar el polimorfismo y la evolución del complejo principal de histocompatibilidad de 18 especies del género *Carduelis* a través del estudio del intrón 2 de estos genes. Además, lo hemos complementado con el análisis filogenético de estas especies que nos ha permitido estudiar la evolución molecular del MHC y compararla con las relaciones filogenéticas de estos jilgueros.

1. Estudio del intrón 2 del complejo principal de histocompatibilidad

Hasta el día de hoy, no se había hecho un estudio tan completo sobre la evolución de intrones de especies pertenecientes al género *Carduelis* y en él se ha visto que el intrón 2 del MHC de dicho género prácticamente ha conservado su secuencia a lo largo de millones de años. Además, se trata de un intrón muy largo en comparación con otras especies de aves, como es el caso del pollo. Actualmente, apenas hay estudios sobre los intrones del MHC de aves y es difícil llegar a una conclusión con un número tan pequeño de especies estudiada. Sin embargo, los resultados que han obtenido otros investigadores coinciden con lo hallado en este trabajo. Hay un estudio de variabilidad del MHC de dos subespecies del diamante mandarín, *Taeniopygia guttata castanotis* y *Taeniopygia guttata guttata*, en el que se observa que al caracterizar el patrón de variación de todos los intrones del MHC de estas especies, el intrón 2 es el que menos varía (Newhouse & Balakrishnan 2015). Otros autores analizan el intrón 2 del MHC de varias especies pertenecientes al orden *Passeriformes* y observan que casi el 90% de la secuencia del intrón se conserva en todas las aves. (Arnaiz-Villena *et al.* 2010b). Los hallazgos de ambos grupos de investigación coinciden con los resultados obtenidos en nuestro trabajo sobre la conservación del intrón 2 de los jilgueros y del pinzón común y, por tanto, apoyan nuestra hipótesis de que es posible que este intrón tenga alguna función en el complejo principal de histocompatibilidad.

1.1 Densidad de intrones

Cincuenta años después del descubrimiento de los intrones, todavía no hay una teoría aceptada sobre el origen de estas secuencias. Los investigadores no se ponen de acuerdo sobre si los ancestros de los eucariotas y los procariotas ya presentaban intrones en sus

genomas, pero éstos últimos los perdieron durante su proceso evolutivo, o bien, si estos ancestros no poseían intrones y los eucariotas los adquirieron durante la evolución. Lo que sí se sabe con certeza es que todos los eucariotas con genomas secuenciados, incluidos organismos unicelulares con genomas compactos, presentan múltiples intrones y un completo espliceosoma. La densidad de intrones en eucariotas va desde unos pocos por genoma, como es el caso de *Giardia intestinalis* que presenta únicamente 4 intrones, a más de ocho por genes en muchos mamíferos (Jeffares *et al.* 2006). Los eucariotas multicelulares presentan mayor número de intrones que los eucariotas unicelulares (Lynch 2002) y el 25% del genoma humano está constituido por intrones (Sakharkar *et al.* 2004). A lo largo de la evolución, los genomas han ido ganando y perdiendo intrones, y la densidad intrónica que presentan las distintas especies es muy variada. La cantidad de intrones presentes en un genoma está influenciada por la selección natural y la deriva, y Farlow y colaboradores (2011) sugieren que la recombinación homóloga (RH) y la recombinación no homóloga (NHEJ, del inglés *Non-Homologous End Joining*) durante la reparación de los cortes de doble cadena del ADN tienen un papel fundamental en la generación de mutaciones que provocan pérdida o ganancia de intrones. La reparación por el sistema NHEJ puede dar lugar a la pérdida o ganancia de intrones dependiendo de si en la reparación hay una deleción del fragmento del intrón o de si se inserta un fragmento de ADN. El sistema de reparación RH sólo puede originar pérdidas de intrones porque si el molde para la reparación es un ADNc (ADN complementario), éste no lleva intrones y por tanto, se produce una deleción. NHEJ se da con mayor frecuencia en los mamíferos y en *Drosophila*, y las reparaciones por RH son más comunes en *Saccharomyces cerevisiae*, una especie que ha sufrido la pérdida casi completa de sus intrones (Farlow *et al.* 2011). Otros autores como Carmel y colaboradores (2007) proponen un modelo de evolución de intrones que sugiere que la mayor densidad de intrones se alcanzó muy temprano. El último ancestro común a los eucariotas contenía más de 2,5 intrones por kilobase y el último ancestro común a los organismos multicelulares presentaba 3,4 intrones por kilobase, unas densidades mayores que las que se aprecian hoy en la mayoría de los hongos existentes y en algunos animales. Estos autores también afirmaron que en los últimos 1.300 MA, el número de ganancias de intrones en los genomas había disminuido (Carmel *et al.* 2007). Una posible explicación a este hecho sería que los intrones producen un cierto retraso en la aparición de las proteínas y la selección natural no favorece su presencia en genes que requieren un ajuste

rápido al ambiente en el que se encuentra la célula. Jeffares y colaboradores (2008) estudian genes de ratón, levadura y *A. thaliana* y observan que los genes con altos niveles de expresión en respuesta al estrés, presentan una densidad de intrones menor que la que presentan otros genes (Jeffares *et al.* 2008). A día de hoy, no hay ningún estudio realizado sobre la densidad de intrones en el genoma de alguna especie de ave, pero para cuando se realice, los resultados aquí presentes pueden contribuir a ello con la caracterización del intrón 2 del MHC de los jilgueros que se ha hecho en este trabajo.

1.2 Tamaño de los intrones

El tamaño de los intrones puede variar entre las diferentes especies debido a las mutaciones y a la selección natural. Hay autores que proponen que el tamaño de los intrones presentes en las aves es mucho más reducido que el de los mamíferos debido a la presión selectiva impuesta por comportamientos metabólicamente exigentes como el vuelo (Hughes & Hughes 1995) donde los intrones pequeños podrían proporcionar una ligera mejora en la eficacia de la transcripción o en la precisión del proceso de corte y empalme (Lynch 2002). Autores como Zhang y Edwards (2012) apoyan esta hipótesis tras estudiar distintas especies amniotas y verificar que los intrones de los mamíferos son más grandes que los de las aves y los reptiles. Además, ellos aportan que dentro de los mamíferos, los intrones de los murciélagos tienden a ser más cortos que los de sus grupos hermanos no voladores sugiriendo que el vuelo o características relacionadas con ello puedan plantear restricciones selectivas sobre la evolución del tamaño de los intrones (Zhang & Edwards 2012). Esta reducción en el tamaño de los intrones de las especies voladoras no coincide con los datos obtenidos en este trabajo ya que todas las especies de aves estudiadas son voladoras y, únicamente, el intrón 2 del MHC de *G. gallus* es el más reducido de todos con una longitud de 229 pb, con respecto a las demás especies de jilgueros que presentan intrones de 311 pb. Sin embargo, no podemos generalizar este hecho porque deberíamos de estudiar más genes con más intrones tanto en el pollo como en las aves *Passeriformes*, ya que hay autores como Vinogradov y colaboradores (1999) que examinaron 176 intrones de 55 genes homólogos de pollo y humano y no pudieron encontrar diferencias significativas entre los tamaños de los intrones de éstas dos especies (Vinogradov 1999).

Hay intrones que se encuentran en regiones codificantes del ADN llamadas CDSs, del inglés *CoDing Sequence*. Bradnam y Korf (2008) analizan la longitud del primer intrón

de al menos 500 CDSs en 36 especies distintas y observan que éste era significativamente más largo que el resto de intrones de la región codificante, estableciendo que esto podría ser una característica general de la estructura de los genes eucariotas. Además, ellos sugieren que este aumento de longitud en el primer intrón en *A. thaliana* puede deberse en parte a la presencia de un motivo IME (del inglés, *intron mediated enhancement*) que mejora la expresión del gen que contiene este intrón (Bradnam & Korf 2008). Usando perfiles genómicos de varias marcas epigenómicas de la base de datos ENCODE (<http://www.nature.com/encode/#/threads>), Park y colaboradores (2014) muestran que los primeros intrones también están enriquecidos con marcas de cromatina indicativas de regiones reguladoras activas con bloques conservados, que probablemente sean importantes para regular patrones complejos de expresión génica de genes muy grandes (Park *et al.* 2014). En este trabajo, solamente, se ha estudiado el intrón 2 del complejo principal de histocompatibilidad de aves *Passeriformes* y no existen datos de que se haya estudiado con anterioridad el intrón 1 de estas aves; de tal modo que no podemos comparar ambos intrones y comprobar si lo que aseguran otros autores sobre la longitud del primer intrón se cumple en el MHC de jilgueros. Lo que sí podemos afirmar es que al comparar el intrón 1 y 2 del gen HLA-B del ser humano, el intrón 1 tiene una longitud de 473 pb, mientras que el intrón 2 presenta casi la mitad de su longitud con 245 pb. No sabemos si dentro del intrón 1 existen elementos reguladores del gen, pero sí podemos asegurar que los intrones de estos genes tienen un alto grado de conservación entre ellos (Gomez-Casado *et al.* 1999) y probablemente tengan un papel relevante en la funcionalidad del gen.

Hay autores que defienden que la transcripción es un proceso muy costoso debido a que para transcribir aproximadamente 20 nucleótidos en un segundo, la célula necesita al menos dos moléculas de ATP por nucleótido. Castillo-Davis y colaboradores (2002) estudian la expresión de genes que codifican proteínas en *Caenorhabditis elegans* y *Homo sapiens*, y demuestran que los intrones de los genes altamente expresados son sustancialmente más cortos que los de los genes que se expresan a un nivel más bajo. Ellos defienden que la selección natural parece favorecer los intrones cortos en genes muy expresados para minimizar el coste energético de la transcripción y otros procesos moleculares, como el de *splicing* (Castillo-Davis *et al.* 2002). Por el contrario, Huang y Niu (2008) no están de acuerdo con esta hipótesis y defienden que los intrones cortos en genes altamente expresados no deberían atribuirse a la restricción energética. Ellos

consideran que un gen expresado en un órgano grande, como el hígado, se transcribe mucho más que un gen con un nivel de expresión similar en un órgano pequeño, como el hipotálamo y, por tanto, esperan hallar intrones más pequeños en los genes de los órganos grandes. Sin embargo, estos autores no encuentran diferencias significativas en la longitud de los intrones entre unos órganos y otros, y atribuyen la presencia de intrones cortos a la reducción del tiempo en la maduración del ARNm (Huang & Niu 2008). Los genes de histocompatibilidad pueden aumentar o disminuir su expresión dependiendo de múltiples factores, por lo que la hipótesis de Castillo-Davis y sus colaboradores no se cumple con los genes del MHC-I de las especies del género *Carduelis*, ya que estos genes pueden elevar su expresión en presencia de citoquinas producidas durante la respuesta inmune y, por el contrario, presentan un intrón 2 muy largo cuya eliminación supondría un gasto energético muy alto para la célula.

1.3 Funciones de los intrones

No hace mucho, los intrones se definían como fragmentos de ADN carentes de información cuya presencia suponía un gasto de energía para la célula, pero actualmente, sabemos que aunque algunos intrones se comporten como elementos neutrales, muchos otros presentan una amplia diversidad de funciones que en algunas ocasiones son imprescindibles para el funcionamiento del organismo. Como ya se detallaba en la introducción de este trabajo, los intrones están involucrados en numerosos procesos celulares, entre los que destaca la regulación de la iniciación, elongación y terminación de la transcripción del ARNm y del tiempo en el que se lleva a cabo este proceso. Sería muy interesante identificar la funcionalidad de los intrones y comprobar verdaderamente que desempeñan un papel importante en el genoma. Sin embargo, no es posible detectar intrones funcionales usando estrategias tradicionales, como buscar secuencias altamente conservadas, ya que las secuencias intrónicas raramente presentan fragmentos muy conservados. Una razón que impide detectar estas funciones es que no todas dependen de la secuencia. Algunas funciones intrónicas dependen de la longitud del intrón más que de su secuencia. Un ejemplo es el gen *E74* en *D. melanogaster*. El pre-ARNm de este gen tiene 60 kb de longitud, pero el ARNm maduro tiene solo 6 kb de longitud. Thummel y colaboradores (1990) midieron la tasa de elongación de la ARN polimerasa II en este gen y observaron que el tiempo que lleva transcribir los intrones generaba retrasos que son cruciales para la función de este gen durante la metamorfosis. Y otra razón es que hay

funciones intrónicas que dependen del hecho de que el proceso de corte y empalme ocurra durante la transcripción, ya que algunos factores que forman parte del espliceosoma modulan las tres fases de transcripción (Thummel *et al.* 1990).

A pesar de las pocas secuencias conservadas de los intrones, si es cierto que algunos intrones muestran una gran conservación de su posición con respecto a la secuencia codificante, que a veces se mantiene a lo largo de millones de años. En este trabajo se ha estudiado el grado de conservación del intrón 2 del MHC de los jilgueros y se observa que un total de 283 posiciones se conservan en todos los intrones variando, únicamente, 28 posiciones. Además, la posición de este intrón es invariable en todas las especies estudiadas ya que siempre se localiza entre el exón 2 y el exón 3. Este alto grado de conservación que presenta el intrón 2 puede deberse a que estas secuencias tengan alguna función en el sistema principal de histocompatibilidad. En ese caso, habría que examinar las propiedades que caracterizan los genes del MHC y compararlas con las propuestas por Chorev y colaboradores (2017) en su modelo estadístico. Estos autores han desarrollado un modelo que calcula la probabilidad de que cada intrón sea funcional basándose en 13 características propias de los intrones que albergan elementos funcionales (Chorev *et al.* 2017).

1.4 Composición nucleotídica de los intrones

La composición nucleotídica de los genomas es un carácter muy estudiado. Los genomas de vertebrados muestran un contenido GC bastante uniforme, que oscila entre el 40% y el 45%, aproximadamente (Graur & Li 2000). Debido a la importancia que están adquiriendo los intrones, tras ir conociendo el papel fundamental que juegan en el genoma, cada vez hay más estudios que arrojan luz sobre sus propiedades, tales como el contenido GC. Estudios en intrones de *Drosophila melanogaster* y *Drosophila simulans* afirman que existe una correlación negativa entre la longitud de los intrones y el contenido GC, es decir, que cuanto mayor es la longitud de un intrón menor es su contenido GC (Haddrill *et al.* 2005). Otros análisis en intrones humanos y de chimpancés corroboran esta correlación negativa entre la longitud de los intrones y el contenido GC (Gazave *et al.* 2007). En este trabajo, se calculan las frecuencias nucleotídicas de cada una de las secuencias del intrón 2 del MHC y se observa que el porcentaje medio del contenido GC de todas las aves *Passeriformes* estudiadas es del 55,2%, a diferencia de lo obtenido por Gazave y colaboradores (2007) cuyo porcentaje medio es del 43,5%. Parece ser que lo

obtenido en este trabajo no coincide con lo afirmado por estos autores anteriores, ya que el intrón 2 del MHC de aves *Passeriformes* se considera un intrón largo y su contenido en GC es elevado. Bien es cierto, que deberíamos analizar más intrones de estos jilgueros para llegar a una conclusión más fehaciente sobre si se cumple o no esta correlación. Sin embargo, lo afirmado por Gazave y colaboradores (2007) sí se cumple en el intrón 2 del pollo, ya que se trata de un intrón corto con un porcentaje medio de GC bastante elevado del 78,6%. Estos mismos autores analizan la posición de los intrones en los genes y observan que los primeros son más ricos en GC y más largos que el resto de intrones del gen. Ellos defienden que los primeros intrones presentan un alto contenido GC porque contienen elementos reguladores ricos en islas CpG que se asocian con frecuencia a los extremos 5' de los genes y se cree que juegan un papel importante en la regulación de la expresión génica (Gazave *et al.* 2007). Haddrill y colaboradores (2005) observan que cuanto mayor sean los intrones, menos divergencia existe entre ellos y que, probablemente, se deba a la presencia de elementos reguladores en estas secuencias (Haddrill *et al.* 2005). Según estas afirmaciones, es posible pensar que el intrón 2 del MHC presente elementos reguladores ya que su contenido en GC es alto y, además, la divergencia entre las secuencias de jilgueros analizadas es muy pequeña.

1.5 Similitud entre los intrones

En este trabajo se ha analizado la semejanza que hay entre las secuencias del intrón 2 del complejo principal de histocompatibilidad de las aves estudiadas mediante dos parámetros, la distancia genética y el porcentaje de similitud, ambos recogidos en las Tablas R2 y R3, respectivamente.

Con las distancias genéticas obtenidas, se realiza un dendograma por el método NJ (Figura R2) que clasifica las especies de jilgueros que tienen la misma secuencia del intrón 2 en tres grupos. Dos de los grupos, engloban a todos los lúganos sudamericanos, y el tercer grupo con secuencias iguales engloba a dos jilgueros de la radiación norteamericana, *C. pinus* y *C. atriceps*, pero no a *C. dominicensis*. Probablemente, el intrón 2 de esta especie evolucionó de forma diferente a los otros dos jilgueros porque *C. dominicensis* apareció en la tierra casi 2 MA antes que *C. pinus* y *C. atriceps*, y pudo estar sometido a distintos patógenos que a los que estuvieron expuestos el jilguero de los pinos y el jilguero capiroto.

El porcentaje de similitud mide el porcentaje de residuos alineados que tienen características fisicoquímicas similares. Las especies que presentan porcentajes de similitud idénticos son especies que están muy emparentadas: *C. pinus* y *C. atriceps* pertenecen a la radiación norteamericana; *C. atrata* y *C. crassirostris* se distribuyen por los Andes; y *C. magellanica*, *C. notata*, *C. olivacea*, *C. spinescens* y *C. xanthogastra* son lúganos pertenecientes a la radiación sudamericana. Sin embargo, tanto las especies con porcentajes de similitud de más del 90% como las cercanas al 50%, no están tan relacionadas geográficamente. No obstante, hay que tener en cuenta que los porcentajes de similitud obtenidos son bastante elevados indicando un alto grado de conservación de las secuencias, si los comparamos con los de *G. gallus*. Esto sugiere que el intrón 2 ha podido estar sometido a distintas presiones que han hecho que no varíe a lo largo de la evolución, quizás porque tenga un papel importante en el MHC de estas aves.

Debido a la importancia que están adquiriendo los intrones, cada vez más son el objeto de estudio de muchos investigadores. La similaridad entre secuencias intrónicas de un determinado gen en diferentes especies no es el parámetro que más se analiza, pero hay algunos estudios que sí lo hacen. Por ejemplo, cuando se estudia la similaridad de la secuencia del gen *St7* en humano, ratón, vaca y babuino se observa que el intrón 8 está muy conservado en las cuatro especies, sugiriendo que este intrón está sometido a una presión selectiva que impide su variación, y que probablemente sea porque incluya algún elemento regulador en su secuencia (Frazer *et al.* 2003). Otro caso de similitud entre especies es el que presentan *Caenorhabditis briggsae* y *Caenorhabditis elegans* en dos genes. Al analizar los intrones del gen *let-2*, aparecen unas secuencias conservadas entre los exones 9 y 10, y los exones 10 y 11 de ambas especies. Se trata de un motivo poli-GT, también encontrado en el gen *bli-4*, donde se sabe que esta secuencia conservada está involucrada en el *splicing* alternativo de este gen (Kent & Zahler 2000).

1.6 Teoría del MHC “mínimo esencial”

Tras la caracterización del MHC de *G. gallus*, Kaufman y sus colaboradores proponen la hipótesis del MHC “mínimo esencial” que defiende que este sistema contiene sólo el mínimo número de genes absolutamente esenciales para la función del MHC en un sentido evolutivo (Kaufman *et al.* 1995; Kaufman *et al.* 1999). Una de las características que llevan a estos autores a proponer esta hipótesis es que los genes del MHC poseen intrones muy cortos. Como se ve en este trabajo, esta característica no se cumple en el

pinzón común ni en los jilgueros estudiados, ya que el intrón del pollo presenta una longitud de 229 pb frente a los 293 pb de *F. coelebs* y los 311 pb de las especies del género *Carduelis*.

Hay especies pertenecientes al orden *Galliformes* que sí presentan un MHC “mínimo esencial”, como es el caso del pavo (*Meleagris gallopavo*) el cual presenta dos regiones homólogas a la región MHC-B y MHC-Y del pollo localizada en el mismo microcromosoma que contiene la región organizadora nucleolar (Chaves *et al.* 2009); el faisán dorado (*Chrysolophus pictus*), cuya región MHC-B contiene 20 genes de histocompatibilidad que presentan una sintenia similar a la del pollo (Ye *et al.* 2012); el pato (*Anas platyrhynchos*), que a pesar de tener 5 genes del MHC de clase I, esta región de histocompatibilidad se parece mucho funcionalmente al “mínimo esencial” del pollo (Moon *et al.* 2005); y el urogallo (*Tetrao tetrix*), que contiene 19 genes en la región MHC-B con una sintenia prácticamente idéntica a la de *G. gallus*, a excepción de una inversión en el orden de los genes *TAPBP* y *TAP* (Wang *et al.* 2012). Sin embargo, existen otras especies del orden *Galliformes* que no tienen un MHC “mínimo esencial” como es el caso de la codorniz (*Coturnix japonica*), cuya región MHC es de 180 kb englobando un total de 41 genes, en comparación con las 92 kb de la región genómica del pollo que contienen 19 genes (Shiina *et al.* 2004).

2. Filogenia molecular

Para completar el estudio de evolución del complejo principal de histocompatibilidad en especies del género *Carduelis*, es necesario realizar un análisis filogenético que nos permita comparar los datos obtenidos del intrón 2 con las relaciones filogenéticas de los jilgueros. Para llevar a cabo el análisis de filogenia molecular de estas especies, se utilizan las secuencias del citocromo b, obteniéndose como resultados varios dendogramas generados por los distintos métodos de reconstrucción de árboles filogenéticos utilizados en este trabajo.

Las secuencias de citocromo b corresponden al género *Carduelis*, excepto la de *F. coelebs*, que pertenece al género *Fringilla* y que, únicamente, la utilizamos como grupo externo de los jilgueros. Los resultados de estos análisis filogenéticos se muestran en el dendograma realizado mediante el método ML (Figura R3) y en el dendograma realizado mediante Inferencia bayesiana (Figura R4). Ambos resultados son prácticamente idénticos y en ellos se ve cómo todas las especies están agrupadas de acuerdo a las

distintas radiaciones que ha habido en América. Las especies pertenecientes a las tres radiaciones de *Carduelis* que ha habido en América se muestran en la Figura D1.

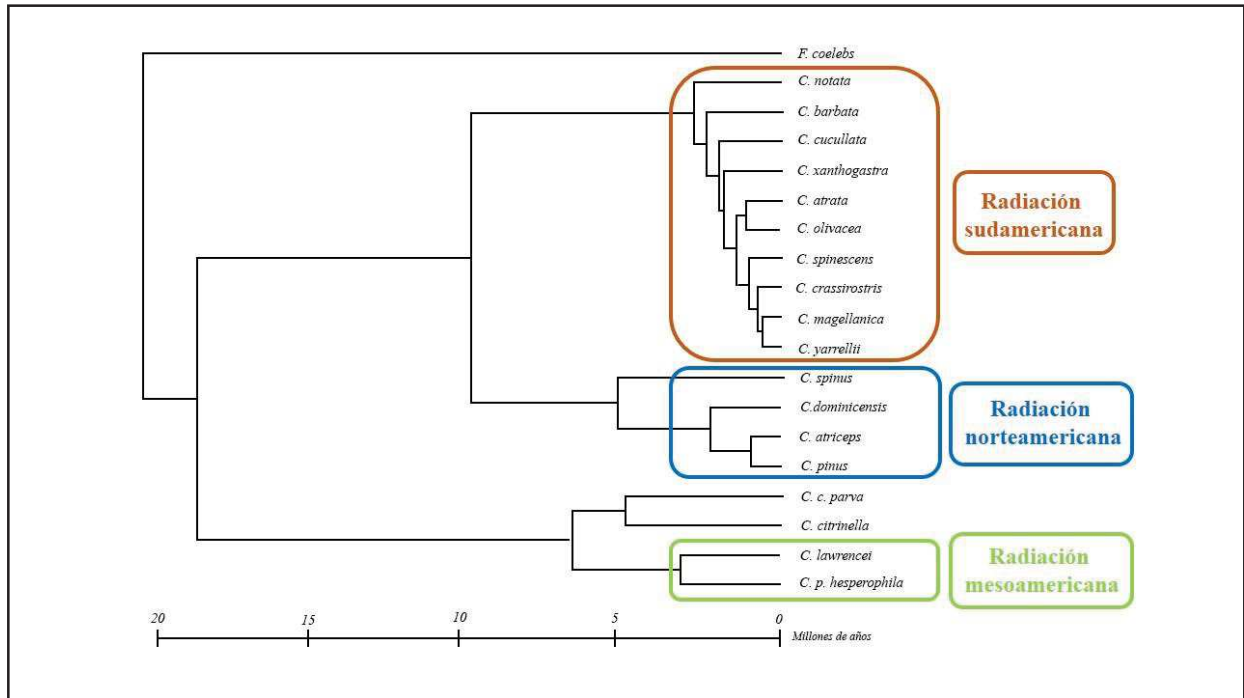


Figura D1. Dendrograma realizado mediante el método ML basado en secuencias de citocromo b mitocondrial con las tres radiaciones americanas de *Carduelis* señaladas.

El jilguero euroasiático (*C. spinus*) apareció en la tierra hace 5 MA y podría haber sido el antepasado de todos los lúganos americanos. Hoy en día, el jilguero euroasiático no se distribuye por América, sino que se encuentra en la Eurasia más oriental y más occidental. Una de las características más importantes de esta especie es que no siempre sigue el mismo patrón de migración, por ello, es posible que migrara a América, bien por el estrecho de Bering/Islands Aleutianas o bien por Europa (Arnaiz-Villena *et al.* 2012). Poco después de esta época, una glaciación cubrió América del Norte con un escudo de hielo de un kilómetro de espesor y *C. spinus* pudo haberse refugiado en zonas sin hielo como las islas del Caribe, las montañas mexicanas y el altiplano guatemalteco-mexicano (Arnaiz-Villena *et al.* 2008). Como las temperaturas en estas zonas eran cálidas, el jilguero euroasiático no tenía la necesidad de migrar a zonas más templadas y pudo haber dado origen al endémico lúgano de las Antillas (*C. dominicensis*) en la isla La Española, poco después de 2 MA. Posteriormente, hace unos 200.000 años, *C. spinus* podría haber dado lugar al jilguero de los pinos en las sierras mexicanas y, aproximadamente al mismo

tiempo, el jilguero capirotado (*C. atriceps*) podría haber aparecido en el altiplano guatemalteco-mexicano (Arnaiz-Villena *et al.* 2009). Estudios recientes han demostrado que *C. spinus* y *C. pinus* tienen una relación claramente genética ya que ambas especies comparten la misma secuencia de una proteína del MHC (Arnaiz-Villena *et al.* 2017). Los resultados de nuestros dendogramas coinciden con las relaciones filogenéticas de las especies del género *Carduelis* de la radiación norteamericana, con *C. spinus* como ancestro común a *C. dominicensis*, *C. pinus* y *C. atriceps*.

La radiación mesoamericana dio lugar al jilguero amarillo (*C. tristis*), al jilguero de Lawrence (*C. lawrencei*) y a dos subespecies del jilguero dominico (*C. psaltria hesperophila* y *C. psaltria colombiana*). *C. spinus* y *C. pinus* están emparentados con estos jilgueros, pero la especie que los relaciona debe de estar extinta. La distribución geográfica y el fenotipo de *C. psaltria* sugieren que podría haber compartido un ancestro común con los lúganos sudamericanos; sin embargo, Arnaiz-Villena y colaboradores (2012) demuestran que el jilguero dominico es un ave norteamericana que ha colonizado hábitats sudamericanos y que ha sido sometida a ciertos cambios fenotípicos, como presentar la cabeza y la espalda con un plumaje más oscuro (Arnaiz-Villena *et al.* 2012). Nuestro trabajo incluye, únicamente, a *C. lawrencei* y *C. p. hesperophila* como representantes de la radiación mesoamericana, y los dendogramas obtenidos coinciden con las relaciones filogenéticas propias de esta radiación, además de demostrar que ambas especies son un grupo monofilético que no está relacionado con los lúganos sudamericanos (Arnaiz-Villena *et al.* 2007).

Los jilgueros sudamericanos, también llamados lúganos sudamericanos, aparecieron en la Tierra hace 3 MA (Arnaiz-Villena *et al.* 2010a). El lúgano de cabeza negra (*C. notata*) dio lugar a la radiación de estas aves que coincide, más o menos, con la unión del Istmo de Panamá, es decir, con la unión de América del Norte con América del Sur. Posiblemente, al cerrarse este istmo, las plantas silvestres mesotérmicas pasaron a los Andes y sirvieron de alimento a los lúganos sudamericanos, dando lugar a esta dispersión aviaria. La especie que relaciona a *C. notata* y *C. spinus* se desconoce, pero recientemente se ha visto que los lúganos sudamericanos comparten proteínas del MHC con el jilguero euroasiático (Arnaiz-Villena *et al.* 2017). Los resultados obtenidos en nuestro trabajo coinciden con las relaciones filogenéticas de la radiación sudamericana, con *C. notata* como ancestro común a todos los lúganos sudamericanos. Todas estas especies, al estar genéticamente muy relacionadas entre sí, presentan fenotipos muy similares debido a la

rápida especiación durante la radiación sudamericana. Esto origina un gran parecido entre las especies y la posible confusión a la hora de identificar su pariente más próximo. No obstante, gracias a los estudios filogenéticos es posible conocer cómo evolucionaron estas especies y cuáles son sus ancestros más cercanos. Estudios genéticos recientes demuestran que *C. xanthogastra stegnejeri* es una subespecie o una nueva especie muy cercana a *C. atrata*, cuando antes se pensaba que era una subespecie de *C. xanthogastra* porque fenotípicamente los machos eran semejantes (Arnaiz-Villena *et al.* 2020).

En la Figura R5 se observa que las relaciones entre las distintas especies de *Carduelis* obtenidas mediante Inferencia bayesiana no se corresponden con las observadas en la Figura R4. Este hecho demuestra que las secuencias del MHC no son una buena opción a la hora de estudiar relaciones filogenéticas, y que es mejor emplear otros genes cuyas secuencias estén más conservados entre las especies, como es el caso del gen del citocromo b (Arnaiz-Villena *et al.* 2014). Sin embargo, el estudio de la evolución del MHC entre especies es muy interesante para conocer si estas secuencias se mantienen o no a lo largo del tiempo, y si existe algún motivo para su conservación. Estudios recientes demuestran que existe evolución transespecífica de los genes de histocompatibilidad de clase I entre distintas especies del género *Carduelis*, ya que *C. spinus* comparte una proteína con cuatro especies del mismo género, y 8 lúganos sudamericanos comparten la misma proteína del MHC (Arnaiz-Villena *et al.* 2017).

CONCLUSIONES

Tras analizar el polimorfismo y la evolución del complejo principal de histocompatibilidad en 18 especies del género *Carduelis* a través del estudio del intrón 2 de estos genes, se puede concluir que:

1. La divergencia entre las secuencias del intrón 2 del MHC de las aves estudiadas del género *Carduelis* es muy pequeña, habiendo especies que incluso comparten el 100% de similaridad.
2. El tamaño del intrón 2 del MHC es muy similar en todas las aves *Passeriformes* estudiadas con una longitud media de 311 pb en los jilgueros y 293 pb en el pinzón común.
3. El MHC “mínimo esencial” descrito en *Gallus gallus* no se ajusta al MHC de las aves *Passeriformes* porque sus intrones son significativamente más largos que el que presenta el pollo.
4. El alto grado de conservación que presenta el intrón 2 de las especies del género *Carduelis* puede deberse a que estas secuencias tengan alguna función en el MHC o bien, presenten elementos reguladores de la expresión génica.
5. El contenido medio en GC del intrón 2 del MHC de las especies del género *Carduelis* es del 55,2%, lo que lleva a pensar que este alto contenido en GC sea un indicio de que presente elementos reguladores en su secuencia.
6. Los resultados filogenéticos obtenidos mediante el análisis de las secuencias de citocromo b no coinciden con los resultados que se obtienen al estudiar la filogenia de las secuencias de MHC, ya que estos genes evolucionan de forma diferente dependiendo del ambiente en el que se desarrollen y de los patógenos a los que se enfrenten.

BIBLIOGRAFÍA

- Abbas A.K., Lichtman A.H. & Pillai S. (2017) *Cellular and molecular immunology*. Elsevier, Philadelphia, PA (USA).
- Abbas A.K., Lichtman A.H. & Pillai S. (2019) *Basic immunology: functions and disorders of the immune system*. Elsevier, St. Louis, MO (USA).
- Anderson S., Bankier A.T., Barrell B.G., de Bruijn M.H., Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A., Sanger F., Schreier P.H., Smith A.J., Staden R. & Young I.G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457-65.
- Antoniou M., Geraghty F., Hurst J. & Grosveld F.(1998). Efficient 3'-end formation of human beta-globin mRNA *in vivo* requires sequences within the last intron but occurs independently of the splicing reaction. *Nucleic Acids Res* **26**, 721–9.
- Arnaiz-Villena A., Alvarez-Tejado M., Ruiz-del-Valle V., Garcia-de-la-Torre C., Varela P., Recio M.J., Ferre S. & Martinez-Laso J. (1998) Phylogeny and rapid northern and southern hemisphere speciation of goldfinches during the Miocene and Pliocene epochs. *Cell Mol Life Sci* **54**, 1031-41.
- Arnaiz-Villena A., Areces C., Rey D., Enriquez-de-Salamanca M., Alonso-Rubio J. & Ruiz-del-Valle V. (2012) Three different North American siskin/goldfinch evolutionary radiations (genus *Carduelis*): pine siskin green morphs and European siskins in America. *Open Ornithol J* **5**, 73-81.
- Arnaiz-Villena A., Gomez-Prieto P. & Ruiz-del-Valle V. (2010a) El género *Carduelis*. *Ornitología Práctica* **42**, 1-9.
- Arnaiz-Villena A., Ruiz-del-Valle V. & Gomez-Prieto P. (2009) Phylogeography of finches and sparrows. In: (ed. by Rechi L), pp. 1-54. Nova Science, New York, NY (USA).
- Arnaiz-Villena A., Ruiz-del-Valle V., Gomez-Prieto P., Rey D., Enriquez-de-Salamanca M., Marco J., Muñoz E., Martín-Villa J.M. & Areces C. (2014) *Carduelini* new systematics: Crimson-winged finch (*Rhodopechys sanguineus*) is included in "arid-zone" *carduelini* finches by mitochondrial DNA phylogeny. *Open Ornithol J* **7**, 55-62.
- Arnaiz-Villena A., Ruiz-del-Valle V., Moscoso J., Serrano-Vela J.I. & Zamora J. (2007) mtDNA phylogeny of North American *Carduelis pinus* group. *Ardeola* **54**, 1-14.

- Arnaiz-Villena A., Ruiz-del-Valle V., Muñiz E., Palacio-Gruber J., Campos C., Gómez-Casado E., Martín-Villa J.M. & Serrano-Vela I. (2017) Major histocompatibility complex allele persistence in Eurasia and America in the genus *Carduelis* (*Spinus*) during million years. *Open Ornithol J* **10**, 92-104.
- Arnaiz-Villena A., Ruiz-del-Valle V., Reche P., Gomez-Prieto P., Lowy E., Zamora J., Areces C., Rey D., Parga C. & Serrano-Vela J. (2010b) Songbirds conserved sites and intron size of MHC Class I molecules reveal a unique evolution in vertebrates. *Open Ornithol J* **3**, 156-65.
- Arnaiz-Villena A., Ruiz-del-Valle V., Reguera R., Gomez-Prieto P. & Serrano-Vela J.I. (2008) What might have been the ancestor of New World siskins? *Open Ornithol J* **1**, 46-7.
- Arnaiz-Villena A., Ruiz-del-Valle V., Suarez-Trujillo F, Lopez-Nares A, Callado A, Gomez-Casado E, Crespo-Yuste E, Campos C. (2020) Description of new American *Carduelis/Spinus* bird species in La Paz (Bolivia): *C./S. lapazensis*. *Open Ornithol J* **13**: 24-33.
- Balakrishnam C.N., Ekblom R., Völker M., Westerdahl H., Godínez R., Kotkiewicz H., Burt D.W., Graves T., Griffin D.K., Warren W.C. & Edward S.V. (2010) Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biol* **8**, 29.
- Berget S., Moore C. & Sharp P. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA* **74**, 3171-5.
- Bjorkman P.J., Saper M.A., Samraoui B., Bennett W., Strominger J. & Wiley D.C. (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* **329**, 512-8.
- Bonneaud C., Pérez-Tris J., Federici P., Chastel O. & Sorci G. (2006) Major histocompatibility alleles associated with local resistance to malaria in a passerine. *Evolution* **60**, 383-9.
- Bonneaud C., Sorci G., Morin V., Westerdahl H., Zoorob R. & Wittzell H. (2004) Diversity of MHC class I and IIB genes in house sparrows (*Passer domesticus*). *Immunogenetics* **55**, 855-65.
- Bradnam K.R. & Korf I. (2008) Longer first introns are a general property of eukaryotic gene structure. *PloS ONE* **3**, e3093.

- Briles W., McGibbon W. & Irwin D.M. (1950) On multiple alleles affecting cellular antigens in the chicken. *Genetics* **35**, 633-40.
- Briles W.E., Goto R.M., Auffray C. & Miller M.M. (1993) A polymorphic system related to but genetically independent of the chicken major histocompatibility complex. *Immunogenetics* **37**, 408-14.
- Briles W.E., Stone H.A. & Cole R.K. (1977) Marek's disease: effects of B histocompatibility alloalleles in resistant and susceptible chicken lines. *Science* **195**, 193-5.
- Brown T. (2017) *Genomes 4*. Garland Science, New York (USA).
- Carmel L., Wolf Y.I., Rogozin I.B. & Koonin E.V. (2007) Three distinct modes of intron dynamics in the evolution of eukaryotes. *Genome Res* **17**, 1034-44.
- Castillo-Davis C.I., Mekhedov S.L., Hartl D.L., Koonin E.V. & Kondrashov F.A. (2002) Selection for short introns in highly expressed genes. *Nat Genet* **31**, 415-8.
- Cepellini R., Curtini E.S., Mattiuz P.L., Miggiano V., Scudeller G. & Serra A. (1967) Genetics of leukocyte antigens. A family study of segregation and linkage. In: *Histocompatibility Testing 1967* (eds. by Curtini ES, Mattiuz PL & Tosi RM), pp. 149-87. Munksgaard, Copenhagen (Denmark).
- Clement P., Harris P. & Davies J. (1993) *Finches and Sparrows*. Princeton University Press, London (UK).
- Chaves L.D., Krueth S.B. & Reed K.M. (2009) Defining the turkey MHC: sequence and genes of the B locus. *J Immunol* **183**, 6530-7.
- Chen L.C., Lan H., Sun L., Deng Y.L., Tang K.Y. & Wan Q.H. (2015) Genomic organization of the crested ibis MHC provides new insight into ancestral avian MHC structure. *Sci Rep* **5**, 7963.
- Chorev M., Bekker A.J., Goldberger J. & Carmel L. (2017) Identification of introns harboring functional sequence elements through positional conservation. *Sci Rep* **7**, 4201.
- Chorev M. & Carmel L. (2012) The function of introns. *Frontiers in Genetics* **3**, 1-15.
- Chow L., Gelinas R., Broker T. & Roberts R. (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1-8.
- Dausset J. (1958) Iso-leuko-antibodies. *Acta Haematol.* **20**, 156-66.
- Dausset J., Ivanyi P. & Ivanyi D. (1965) Tissue alloantigens in humans: identification of a complex system (Hu-1). In: *Histocompatibility Testing 1965* (eds. by Dausset J,

- Rappaport FJ, Ivanyi P & Colombani J), pp. 51-62. Munksgaard, Copenhagen (Denmark).
- Dvir H., Wang J., Ly N., Dascher C.C. & Zajonc D.M. (2010) Structural basis for lipid-antigen recognition in avian immunity. *J Immunol* **184**, 2504-11.
- Edwards S.V., Arctander P. & Wilson A.C. (1991) Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. *Proc Biol Sci* **243**, 99-107.
- Farlow A., Meduri E. & Schlötterer C. (2011) DNA double-strand break repair and the evolution of intron density. *Trends Genet* **27**, 1-6.
- Felsenstein J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**, 368-76.
- Frazer K.A., Elnitski L., Church D.M., Dubchak I. & Hardison R.C. (2003) Cross-species sequence comparisons: a review of methods and available resources. *Genome Res* **13**, 1-12.
- Gaunt A. & Wells M. (1973) Models of syringeal mechanisms. *Am Zool* **13**, 1227-47.
- Gazave E., Marqués-Bonet T., Fernando O., Charlesworth B. & Navarro A. (2007) Patterns and rates of intron divergence between humans and chimpanzees. *Genome Biol* **8**, R21.
- Gomez-Casado E., Vargas-Alarcon G., Martinez-Laso J., Granados J., Varela P., Alegre R., Longas J., Gonzalez-Hevilla M., Martin-Villa J.M. & Arnaiz-Villena A. (1999) Evolutionary relationships between HLA-B alleles as indicated by an analysis of intron sequences. *Tissue Antigens* **53**, 153-60.
- Gorer P.A. (1936) The detection of a hereditary antigenic difference in the blood of mice by means of human group A serum. *Journal of Genetics* **32**, 17-31.
- Graur D. & Li W. (2000) *Fundamentals of molecular evolution*. Sinauer Associates, Inc, Sunderland, MA (USA).
- Greenewalt C. (1969) How birds sing. *Sci Am* **221**, 126-39.
- Guillemot F., Billault A., Pourquoié O., Béhar G., Chaussé A.M., Zoorob R., Kreibich G. & Auffray C. (1988) A molecular map of the chicken major histocompatibility complex: the class II b genes are closely linked to the class I genes and the nucleolar organizer. *EMBO J* **7**, 2775-85.
- Haddrill P.R., Charlesworth B., Halligan D.L. & Andolfatto P. (2005) Patterns of intron sequence evolution in *Drosophila* are dependent upon length and GC content. *Genome Biol* **6**, R67.

- Helm-Bychowski K.M. & Wilson A.C. (1986) Rates of nuclear DNA evolution in pheasant-like birds: evidence from restriction maps. *PNAS* **83**, 688-92.
- Horton R., Wilming L., Rand V., Lovering R.C., Bruford E.A., Khodiyar V.K., Lush M.J., Povey S., Talbot C.C., Jr., Wright M.W., Wain H.M., Trowsdale J., Ziegler A. & Beck S. (2004) Gene map of the extended human MHC. *Nature Rev Genet* **5**, 889-99.
- Huang Y.F. & Niu D.K. (2008) Evidence against the energetic cost hypothesis for the short introns in highly expressed genes. *BMC Evol Biol* **8**, 154.
- Huelsenbeck J.P. & Ronquist F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-5.
- Hughes A.L. & Hughes M.K. (1995) Small genomes for better flyers. *Nature* **377**, 391.
- Irwin D.M., Kocher T.D. & Wilson A.C. (1991) Evolution of the cytochrome b gene of mammals. *J Mol Evol* **32**, 128-44.
- Jeffares D.C., Moirurier T. & Penny D. (2006) The biology of intron gain and loss. *Trends Genet* **22**, 16-22.
- Jeffares D.C., Penkett C.J. & Bähler J. (2008) Rapidly regulated genes are intron poor. *Trends Genet* **24**, 375-8.
- Jordan B.R., Caillol D., Damotte M., Delovitch T., Ferrier P., Kahn-Perles B., Kourilsky F., Layet C., Le Bouteiller P., Lemonnier F.A., Malissen M., N'Guyen C., Sire J., Sodoyer R., Strachan T. & Trucy J. (1985) HLA class I genes: from structure to expression, serology and function. *Immunol Rev* **84**, 73-92.
- Kappes D.J., Arnot D., Okada K. & Strominger J.L. (1984) Structure and polymorphism of the HLA Class II SB light chain genes. *EMBO J* **3**, 2985-93.
- Karlsson M., Schroeder J., Nakagawa S., Smith H.G., Burke T. & Westerdahl H. (2015) House sparrow *Passer domesticus* survival is not associated with MHC-I diversity, but possibly with specific MHC-I alleles. *J Avian Biol* **45**, 167-74.
- Karlsson M. & Westerdahl H. (2013) Characteristics of MHC class I genes in house sparrows *Passer domesticus* as revealed by long cDNA transcripts and amplicon sequencing. *J Mol Evol* **77**, 8-21.
- Kaufman J., Jacob J., Shaw I., Walker B., Milne S., Beck S. & Salomonsen J. (1999) Gene organisation determines evolution of function in the chicken MHC. *Immunol Rev* **167**, 101-17.

- Kaufman J., Volk H. & Wallny H.J. (1995) A "minimal essential MHC" and an "unrecognized MHC": two extremes in selection for polymorphism. *Immunol Rev* **143**, 63-88.
- Kent W.J. & Zahler A.M. (2000) Conservation, regulation, synteny, and introns in a large-scale *C. briggsae-C. elegans* genomic alignment. *Genome Res* **10**, 1115-25.
- Kocher T.D., Thomas W.K., Meyer A., Edwards S.V., Paabo S., Villablanca F.X. & Wilson A.C. (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *PNAS* **86**, 6196-200.
- Komiyama T., Ikeo K. & Gojobori T. (2003) Where is the origin of the Japanese gamecocks? *Gene* **317**, 195-202.
- Kumar S., Stecher G. & Tamura K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870-4.
- Lynch M. (2002) Intron evolution as a population-genetic process. *PNAS* **99**, 6118-23.
- Madden D.R. (1995) The three-dimensional structure of peptide-MHC complexes. *Annu Rev Immunol* **13**, 587-622.
- Malissen M., Malissen B. & Jordan B.R. (1982) Exon/intron organization and complete nucleotide sequence of an HLA gene. *PNAS* **79**, 893-7.
- Miggiano V., Birgen I. & Pink J.R.L. (1974) The mixed leukocyte reaction in chickens. Evidence for control by the major histocompatibility complex. *Eur J Immunol* **4**, 397-401.
- Miller M.M., Goto R., Bernot A., Zoorob R., Auffray C., Bumstead N. & Briles W.E. (1994) Two MHC class I and two MHC class II genes map to the chicken Rfp-Y system outside the B complex. *PNAS* **91**, 4397-401.
- Miller M.M., Goto R., Taylor Jr. R.L., Zoorob R., Auffray C., Briles R.W., Briles W.E. & Bloom S.E. (1996) Assignment of Rfp-Y to the chicken major histocompatibility complex/NOR microchromosome and evidence for high-frequency recombination associated with the nucleolar organizer region. *PNAS* **93**, 3958-62.
- Miller M.M., Goto R., Young S., Liu J. & Hardy J. (1990) Antigens similar to major histocompatibility complex B-G are expressed in the intestinal epithelium in the chicken. *Immunogenetics* **32**, 45-50.
- Miller M.M., Robinson C.M., Abernathy J., Goto R.M., Hamilton M.K., Zhou H. & Delany M.E. (2014) Mapping genes to chicken microchromosome 16 and

- discovery of olfactory and scavenger receptor genes near the major histocompatibility complex. *J Hered* **105**, 203-15.
- Miller M.M. & Taylor Jr. R.L. (2016) Brief review of the chicken Major Histocompatibility Complex: the genes, their distribution on chromosome 16, and their contributions to disease resistance. *Poult Sci* **95**, 375-92.
- Moon D., Veniamin S., Parks-Dely J. & Magor K. (2005) The MHC of the duck (*Anas platyrhynchos*) contains five differentially expressed class I genes. *J Immunol* **175**, 6702-12.
- Mungall A.J., Palmer S.A., Sims S.K., Edwards C.A., Ashurst J.L., Wilming L., Jones M.C., Horton R., Hunt S.E., Scott C.E., Gilbert J.G., Clamp M.E., Bethel G., Milne S., Ainscough R., Almeida J.P., Ambrose K.D., Andrews T.D., Ashwell R.I., Babbage A.K., Bagguley C.L., Bailey J., Banerjee R., Barker D.J., Barlow K.F., Bates K., Beare D.M., Beasley H., Beasley O., Bird C.P., Blakey S., Bray-Allen S., Brook J., Brown A.J., Brown J.Y., Burford D.C., Burrill W., Burton J., Carder C., Carter N.P., Chapman J.C., Clark S.Y., Clark G., Clee C.M., Clegg S., Cobley V., Collier R.E., Collins J.E., Colman L.K., Corby N.R., Coville G.J., Culley K.M., Dhimi P., Davies J., Dunn M., Earthrowl M.E., Ellington A.E., Evans K.A., Faulkner L., Francis M.D., Frankish A., Frankland J., French L., Garner P., Garnett J., Ghorri M.J., Gilby L.M., Gillson C.J., Glithero R.J., Grafham D.V., Grant M., Gribble S., Griffiths C., Griffiths M., Hall R., Halls K.S., Hammond S., Harley J.L., Hart E.A., Heath P.D., Heathcott R., Holmes S.J., Howden P.J., Howe K.L., Howell G.R., Huckle E., Humphray S.J., Humphries M.D., Hunt A.R., Johnson C.M., Joy A.A., Kay M., Keenan S.J., Kimberley A.M., King A., Laird G.K., Langford C., Lawlor S., Leongamornlert D.A., Leversha M., Lloyd C.R., Lloyd D.M., Loveland J.E., Lovell J., Martin S., Mashreghi-Mohammadi M., Maslen G.L., Matthews L., McCann O.T., McLaren S.J., McLay K., McMurray A., Moore M.J., Mullikin J.C., Niblett D., Nickerson T., Novik K.L., Oliver K., Overton-Larty E.K., Parker A., Patel R., Pearce A.V., Peck A.I., Phillimore B., Phillips S., Plumb R.W., Porter K.M., Ramsey Y., Ranby S.A., Rice C.M., Ross M.T., Searle S.M., Sehra H.K., Sheridan E., Skuce C.D., Smith S., Smith M., Spraggon L., Squares S.L., Steward C.A., Sycamore N., Tamlyn-Hall G., Tester J., Theaker A.J., Thomas D.W., Thorpe A., Tracey A., Tromans A., Tubby B., Wall M., Wallis J.M., West A.P., White S.S., Whitehead S.L.,

- Whittaker H., Wild A., Willey D.J., Wilmer T.E., Wood J.M., Wray P.W., Wyatt J.C., Young L., Younger R.M., Bentley D.R., Coulson A., Durbin R., Hubbard T., Sulston J.E., Dunham I., Rogers J. & Beck S. (2003) The DNA sequence and analysis of human chromosome 6. *Nature* **425**, 805-11.
- Nei M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY (USA).
- Newhouse D.J. & Balakrishnan C.N. (2015) High major histocompatibility complex class I polymorphism despite bottlenecks in wild and domesticated populations of the zebra finch (*Taeniopygia guttata*). *BMC Evol Biol* **15**, 265.
- Parham P., Lomen C.E., Lawlor D.A., Ways J.P., Holmes N., Coppin H., Salter R.D., Wan A.M. & Ennis P.D. (1988) Nature of polymorphism in HLA-A, -B, and -C molecules. *PNAS* **85**, 4005-9.
- Park S.G., Hannehalli S. & Choi S.S. (2014) Conservation in first introns is positively associated with the number of exons within genes and the presence of regulatory epigenetic signals. *BMC Genomics* **15**, 526.
- Promerová M., Albrecht T. & Bryja J. (2009) Extremely high MHC class I variation in a population of a long-distance migrant, the scarlet rosefinch (*Carpodacus erythrinus*). *Immunogenetics* **61**, 451-61.
- Regal P.J. (1977) Ecology and evolution of flowering plant dominance. *Science* **196**, 622-9.
- Rivero-de-Aguilar J., Westerdahl H., Martinez-de la Puente J., Tomas G., Martinez J. & Merino S. (2016) MHC-I provides both quantitative resistance and susceptibility to blood parasites in blue tits in the wild. *J Avian Biol* **47**, 669-77.
- Roger S.L., Göbel T.W., Viertlboeck B.C., Milne S., Beck S. & Kaufman J. (2005) Characterization of the chicken C-type lectin-like receptors B-NK and B-lec suggests that the NK complex and the MHC share a common ancestral region. *J Immunol* **174**, 3475-83.
- Saitou N. & Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-25.
- Sakharkar M.K., Chow V.T. & Kanguane P. (2004) Distributions of exons and introns in the human genome. *In Silico Biol* **4**, 387-93.

- Salomonsen J., Dunon D., Skjodt K., Thorpe D., Vainio O. & Kaufman J. (1991) Chicken major histocompatibility complex-encoded B-G antigens are found on many cell types that are important for the immune system. *PNAS* **88**, 1359-63.
- Sanger F., Nicklen S. & Coulson A.R. (1977) DNA sequencing with chain-terminating inhibitors. *PNAS* **74**, 5463-7.
- Sanz J. (1999) *Los dinosaurios voladores: historia evolutiva de las aves primitivas*. Libertarias Prodhufi, Madrid (España).
- Sarter K., Leimgruber E., Gobet F., Agrawal V., Dunand-Sauthier I., Barras E., Mastelic-Gavillet B., Kamath A., Fontannaz P., Guéry L., Do Valle Duraes F., Lippens C., Ravn U., Santiago-Raber M.L., Magistrelli G., Fischer N., Siegrist C.A., Stéphanie Hugues S. & Reith W. (2016) Btn2a2, a T cell immunomodulatory molecule coregulated with MHC class II genes. *J Exp Med* **213**, 177-87.
- Schierman L. & Nordskog A. (1961) Relationship of blood type to histocompatibility in chickens. *Science* **134**, 1008-9.
- Schierman L.W., Watanabe D.H. & McBride R.A. (1977) Increased growth of Rous sarcomas in chickens pretreated with formalinized syngeneic tumor cells. *Eur J Immunol* **7**, 710-3.
- Schut E., Aguilar J.R., Merino S., Magrath M.J., Komdeur J. & Westerdahl H. (2011) Characterization of MHC-I in the blue tit (*Cyanistes caeruleus*) reveals low levels of genetic diversity and trans-population evolution across European populations. *Immunogenetics* **63**, 531-42.
- Schwartz S., Meshorer E. & Ast G. (2009) Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* **16**, 990-6.
- Shiina T., Shimizu C., Oka A., Teraoka Y., Imanishi T., Gojobori T., Hanzawa K., Watanabe S. & Inoko H. (1999) Gene organization of the quail major histocompatibility complex (MhcCoja) class I gene region. *Immunogenetics* **49**, 384-94.
- Shiina T., Shimizu S., Hosomichi K., Kohara S., Watanabe S., Hanzawa K., Beck S., Kulski J.K. & Inoko H. (2004) Comparative genomic analysis of two avian (quail and chicken) MHC regions. *J Immunol* **172**, 6751-63.
- Sibley, G.C. (1996) *Birds of the world*. Thayer birding software, Naples, FL (USA).

- Snell G.D. (1948) Methods for the study of histocompatibility genes. *Journal of Genetics* **49**, 87-108.
- Solinhaç R., Leroux S., Galkina S., Chazara O., Fève K., Vignoles F., Morisson M., Derjusheva S., Bed'hom B., Vignal A., Fillon V. & Pitel F. (2010) Integrative mapping analysis of chicken microchromosome 16 organization. *BMC Genomics* **11**, 611.
- Stern L.J., Brown J.H., Jardetzky T.S., Gorga J.C., Urban R.G., Strominger J.L. & Wiley D.C. (1994) Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* **368**, 215-21.
- Stewart C.A., Horton R., Allcock R.J., Ashurst J.L., Atrazhev A.M., Coggill P., Dunham I., Forbes S., Halls K., Howson J.M., Humphray S.J., Hunt S., Mungall A.J., Osoegawa K., Palmer S., Roberts A.N., Rogers J., Sims S., Wang Y., Wilming L.G., Elliott J.F., de Jong P.J., Sawcer S., Todd J.A., Trowsdale J. & Beck S. (2004) Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res* **14**, 1176-87.
- Swofford D.L. (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods) version 4*. Sinauer Associates, Sunderland, MA (USA).
- Takashima Y., Ohtsuka T., Gonzalez A., Miyachi H. & Kageyama R. (2011) Intronic delay is essential for oscillatory expression in the segmentation clock. *PNAS* **108**, 3300-5.
- The MHC sequencing consortium. (1999) Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* **401**, 921-923.
- Thorne J.L., Kishino H. & Painter I.S. (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* **15**, 1647-57.
- Thummel C.S., Burtis K.C. & Hogness D.S. (1990) Spatial and temporal patterns of E74 transcription during *Drosophila* development. *Cell* **61**, 101-11.
- Trowsdale J., Ragoussis J. & Campbell R.D. (1991) Map of the human MHC. *Immunol Today* **12**, 443-6.
- Vainio O., Koch C. & Toivanen A. (1984) B-L antigens (class II) of the chicken major histocompatibility complex control T-B cell interaction. *Immunogenetics* **19**, 131-40.

- Vinogradov A. (2006) "Genome design" model: Evidence from conserved intronic sequence in human–mouse comparison. *Genome Res* **16**, 347-54.
- Vinogradov A.E. (1999) Intron-genome size relationship on a large evolutionary scale. *J Mol Evol* **49**, 376-84.
- Wang B., Ekblom R., Strand T., Portela-Bens S. & Höglund J. (2012) Sequencing of the core MHC region of black grouse (*Tetrao tetrix*) and comparative genomics of the galliform MHC. *BMC Genomics* **13**, 553.
- Westerdahl H., Waldenström J., Hansson B., Hasselquist D., Schantz T. & Bensch S. (2005) Associations between malaria and MHC genes in a migratory songbird. *Proc R Soc* **272**, 1511-8.
- Westerdahl H., Wittzell H. & von Schantz T. (1999) Polymorphism and transcription of MHC class I genes in a passerine bird, the great reed warbler. *Immunogenetics* **49**, 158-70.
- Wittzell H., Bernot A., Auffray C. & Zoorob R. (1999) Concerted evolution of two MHC class II B loci in pheasants and domestic chickens. *Mol Biol Evol* **16**, 479-90.
- Ye Q., He K., Wu S.Y. & Wan Q.H. (2012) Isolation of a 97-kb minimal essential MHC B locus from a new reverse-4D BAC library of the golden pheasant. *PLoS ONE* **7**, e32154.
- Zajonc D.M., Striegl H., Dascher C. & Wilson I.A. (2008) The crystal structure of avian CD1 reveals a smaller, more primordial antigen-binding pocket compared to mammalian CD1. *PNAS* **105**, 17925-30.
- Zhang Q. & Edwards S.V. (2012) The evolution of intron size in amniotes: a role for powered flight? *Genome Biol Evol* **4**, 1033-44.
- Zwirner N.W., Dole K. & Stastny P. (1999) Differential surface expression of MICA by endothelial cells, fibroblasts, keratinocytes, and monocytes. *Hum Immunol* **60**, 323-30.

ANEXOS

ANEXO 1: LISTA DE ABREVIATURAS

ADN	Ácido desoxirribonucleico
ARN	Ácido Ribonucleico
CPH	Complejo Principal de Histocompatibilidad
MHC	<i>Major Histocompatibility Complex</i>
ARNm	Ácido Ribonucleico mensajero
ML	<i>Maximum Likelihood</i>
IB	Inferencia Bayesiana
Pb	Pares de bases
G	Guanina
C	Citosina
A	Adenina
T	Timina
U	Uracilo
Mb	Megabase
xMHC	MHC extendido
HLA	<i>Human Leukocyte Antigen</i>
ARNt	Ácido Ribonucleico transferente
SLC	<i>Solute Carrier</i>
BTN	Genes de las butirofilinas
Asn	Asparagina
Cys	Cisteína
VNR	Receptores vomeronasales
HFE	Gen de la hemocromatosis
TNF	Factor de Necrosis Tumoral
HSP	<i>Heat Shock Proteins</i>
α	Alfa
β	Beta
β 2m	Beta-2-microglobulina
UTR	<i>UnTranslated Region</i>
RE	Retículo Endoplasmático
ARS	<i>Antigen Recognition Site</i>

TCR	Receptor de linfocitos T
NK	<i>Natural Killer</i>
NOR	Región Organizadora Nuclear
Rfp-Y	<i>Restriction fragment pattern-Y</i>
Py	Nucleótido pirimidínico
ARNr	Ácido Ribonucleico ribosómico
ARNnp	Ácido Ribonucleico nucleolar pequeño
ARNnc	Ácido Ribonucleico no codificante
ARNmi	microARN
ARNip	ARN interferente pequeño
ARNpi	ARN asociados a Piwi
ARNlnc	ARN largos no codificantes
EJC	Complejo de unión de exones
MA	Millones de años
PCR	<i>Polymerase Chain Reaction</i>
NJ	<i>Neighbor-Joining</i>
ADNmt	Ácido desoxirribonucleico mitocondrial
DMSO	dimetilsulfóxido
DO	Densidad Óptica
A	Absorbancia
dNTP	desoxinucleótidos trifosfato
ddNTP	didesoxinucleótidos trifosfato
Amp ^r	Resistencia a ampicilina
ADNc	ADN complementario
RH	Recombinación Homóloga
NHEJ	<i>NonHomologous End Joining</i>
CDSs	<i>CoDing Sequence</i>

ANEXO 2: CURRICULUM VITAETÍTULOS ACADÉMICOS

- Máster de Microbiología y Parasitología: I+D. Universidad Complutense de Madrid. 2012/2013.
- Licenciatura de Biología. Universidad de Sevilla. 2006/2012.

PUBLICACIONES CIENTÍFICAS

1. Arnaiz-Villena A, Muñiz E, **Campos C**, Gomez-Casado E, Tomasi S, Martinez-Quiles N, Martin-Villa JM, Palacio-Grüber J (2015) Origin of ancient Canary Islanders (Guanches): presence of atlantic/Iberian HLA and Y chromosome genes and Ancient Iberian language. *Int J Mod Anthropol* 1 (8): 67-93.
2. Recio MJ, Muñiz E, Palacio-Grüber J, Martinez-Laso J, Gomez-Casado E, Martin-Villa JM, Martinez-Quiles N, **Campos C**, Arnaiz-Villena A (2015) Major Histocompatibility complex-DMB allelic diversity in old and new world nonhuman primates: intraspecies pattern of evolution. *Int J Mod Anthropol* 1 (8): 25-47.
3. Arnaiz-Villena A, Enriquez-de-Salamanca M, Palacio-Grüber J, **Campos C**, Camacho A, Martin-Villa JM, Martinez-Quiles N, Gomez-Casado E, Muñiz E (2016) Characterisation and functional implications of the two new HLA-G alleles found in Amerindian and Caribbean populations. *Hum Immunol* 77 (9): 812-816.
4. Arnaiz-Villena A, Palacio-Grüber J, Muñiz E, Rey D, Recio MJ, **Campos C**, Martinez-Quiles N, Martin-Villa JM, Martinez-Laso J (2016) HLA-DMB in Amerindians: specific linkage of DMB*01:03:01/DRB1 alleles. *Hum Immunol* 77 (5): 389–394.
5. Arnaiz-Villena A, Palacio-Grüber J, Muñiz E, **Campos C**, Alonso-Rubio J, Gomez-Casado E, Cruz-Robles D, Martin-Villa JM, Silvera C (2016) HLA genes in Chimila Amerindians (Colombia), the peopling of America and medical implications. *Int J Mod Anthropol* 1 (9): 91-116.

6. Arnaiz-Villena A, Muñiz E, Palacio-Grüber J, **Campos C**, Alonso-Rubio J, Gomez-Casado E, Lopez-Pacheco F, Martin-Villa JM, Silvera C (2016) Ancestry of Amerindians and its impact in anthropology, transplantation, HLA pharmacogenomics and epidemiology by HLA study in Wiwa Colombian population. *Open Med J* 3: 269-285.
7. Arnaiz-Villena A, **Campos C**, Ruiz-del-Valle V (2016) Los gorriones del Viejo Mundo (Género *Passer*): genética, origen y evolución. *Ornitología Práctica* 80: 67-70.
8. Arnaiz-Villena A, Palacio-Grüber J, Muñiz E, **Campos C**, Alonso-Rubio J, Gomez-Casado E, Salih SF, Martin-Villa JM, Al-Qadi R (2017) Genetic HLA study of Kurds in Iraq, Iran and Tbilisi (Caucasus, Georgia): relatedness and medical implications. *PLoS ONE* 12 (1): e0169929.
9. Arnaiz-Villena A, Carballo A, Juarez I, Muñiz E, **Campos C**, Tejedor B, Martín-Villa JM, Palacio-Gruber J (2017) HLA Genes in Atlantic Celtic populations: Are Celts Iberians? *Int J Mod Anthropol* 1 (10): 50-72.
10. Arnaiz-Villena A, Ruiz-del-Valle V, Muñiz E, Palacio-Grüber J, **Campos C**, Gómez-Casado E, Martín-Villa JM, Serrano-Vela I (2017) Major histocompatibility complex allele persistence in Eurasia and America in the genus *Carduelis* (*Spinus*) during million years. *Open Ornithol J* 10: 92-104.
11. Arnaiz-Villena A, Palacio-Gruber J, Muñiz E, Rey D, Nikbin B, Nickman H, **Campos C**, Martín-Villa JM, Amirzargar A (2017) Origin of Azeris (Iran) according to HLA genes. *Int J Mod Anthropol* 1 (10): 115-13
12. Arnaiz-Villena A, Enriquez-de-Salamanca M, Palacio-Gruber J, Juarez I, Muñiz E, Nieto J, **Campos C**, Martin-Villa JM (2018) HLA-G in Amerindians: Epidemiology and Worldwide Population Comparison. *Open Med J* 5:1-12.
13. Arnaiz-Villena A, Palacio-Grüber J, Enriquez-de-Salamanca M, Juárez I, **Campos C**, Nieto J, Muñiz E, Martin-Villa JM (2018) HLA-G, -A haplotypes in Amerindians (Ecuador): HLA-G*01:05N World distribution. *Hum Immunol* 79 (2): 89-90.

14. Arnaiz-Villena A, Bayona B, Palacio-Grüber J, Hernández E, Muñiz E, **Campos C**, Juárez I, Gomez-Casado E, Martín-Villa JM, Silvera C. (2018) HLA genes in Barranquilla (North Colombia): searching for cryptic Amerindian genes. *Hum Immunol* 79: 3-4.
15. Arnaiz-Villena A, Palacio-Gruber J, Juárez I, Hernández E, Muñiz E, Bayona B, **Campos C**, Nieto J, Martín-Villa M, Silvera C (2018) HLA in North Colombia Chimila Amerindians. *Hum Immunol* 79 (4): 189-190.
16. Arnaiz-Villena A, Palacio-Gruber J, Juárez I, Muñiz E, Hernández E, Bayona B, **Campos C**, Nuñez J, Lopez-Nares A, Martín-Villa M, Silvera C (2018) Study of Colombia North Wiwa El Encanto Amerindians HLA- genes: Pacific Islanders relatedness. *Hum Immunol* 79 (7): 530-531.
17. Arnaiz-Villena A, Juárez I, Palacio-Grüber J, Muñiz E, **Campos C**, Martínez-Laso J, Nieto J, López-Nares A, Martín-Villa JM, Silvera C. (2018). The first free Africans in America: HLA study in San Basilio de Palenque (Colombia). *Hum Immunol* 79 (8): 585-586.
18. Arnaiz-Villena A, Palacio-Grüber J, Juárez I, López-Nares A, Nieto J, **Campos C**, Martín-Villa JM. (2019). HLA in Uros from Peru Titikaka Lake: Tiwanaku, Easter and Pacific Islanders. *Hum Immunol* 80 (2): 91-92.
19. Arnaiz-Villena A, Ruiz-del-Valle V, Suarez-Trujillo F, Lopez-Nares A, Callado A, Gomez-Casado E, Crespo-Yuste E, **Campos C**. (2020) Description of new American *Carduelis/Spinus* bird species in La Paz (Bolivia): *C./S. lapazensis*. *Open Ornithol J* 13: 24-33.

COMUNICACIONES A CONGRESOS

1. Negro MJ, **Campos C**, Ballesteros I, Sáez F, Ballesteros M, Manzanares P, Oliva JM. Producción de bioetanol a partir de la poda de olivo sometida a un pretratamiento hidrotérmico utilizando ácido fosfórico como catalizador. XVI Reunión de la Red Telemática Española Lignocel “Biotecnología de Materiales Lignocelulósicos: retos enzimáticos, químicos y moleculares para su aplicación industrial, energética y medioambiental”. Pontevedra, España. 2013.

2. Muñiz E, Palacio-Grüber J, **Campos C**, Rey D, Tomasi S, Martín Villa JM, Arnaiz-Villena A. Easter Island and Aleuts relatedness to Amerindians and Pacific Islanders according to HLA genes. 39 Congreso de la Sociedad Española de Inmunología. Alicante, España. 2016.
3. Palacio-Grüber J, Muñiz E, Amirzargar A, **Campos C**, Rey D, Martín Villa JM, Arnaiz-Villena A. Kurds HLA genes: its implications in transplantation and pharmacogenomics. 39 Congreso de la Sociedad Española de Inmunología. Alicante, España. 2016.
4. Arnaiz-Villena A, Muñiz E, Palacio-Grüber J, **Campos C**, Martín Villa JM, Gomez-Casado E. MHC variability in wild songbirds: normal intron size and molecule evolution suggest that a minimal essential MHC does not exist in birds. 39 Congreso de la Sociedad Española de Inmunología. Alicante, España. 2016.
5. Vargas-Alarcón G, Cruz-Robles D, López-Pacheco F, Muñiz E, Palacio-Grüber J, **Campos C**, Rey D, Martín-Villa JM, Arnaiz-Villena A. Mixtec Mexican Amerindians: an HLA alleles study for America peopling, pharmacogenomics and transplantation. 39 Congreso de la Sociedad Española de Inmunología. Alicante, España. 2016.
6. Arnaiz-Villena A, Palacio-Grüber J, Jose Recio M, Martinez-Laso J, Gomez-Casado E, Martín Villa JM, Martínez-Quiles N, **Campos C**, Muñiz E. Major Histocompatibility complex-DMB allelic diversity in old and new world non human primates: intraspecies pattern of evolution. 30th European Immunogenetics and Histocompatibility Conference. Kos, Grecia. 2016.
7. Arnaiz-Villena A, Muñiz E, **Campos C**, Gomez-Casado E, Tomasi S, Martínez-Quiles N, Martín Villa JM, Palacio-Grüber J. Origin of Ancient Canary Islanders Guanches: presence of Atlantic/Iberian Human Leukocyte Antigen and Y chromosome genes and Ancient Iberian language. 30th European Immunogenetics and Histocompatibility Conference. Kos, Grecia. 2016.
8. Arnaiz-Villena A, Amirzargar A, Muñiz E, **Campos C**, Enríquez-de-Salamanca M, Rey D, Martínez-Quiles N, Martín Villa JM, Palacio-Grüber J. HLA in Kurds (Iran): Its Implications in Transplantation and Pharmacogenomics. 30th European Immunogenetics and Histocompatibility Conference. Kos, Grecia. 2016.

9. Arnaiz-Villena A, **Campos C**, Muñiz E, Gomez-Casado E, Martín-Villa J.M, Palacio-Gruber J. First Population of Canary Island is not only North African: Presence of European Atlantic Genes and Ancient Iberian Language Scripts. 40 Congreso de la Sociedad Española de Inmunología. Zaragoza, España. 2017.
10. Arnaiz-Villena A, Muñiz E, Palacio-Gruber J, **Campos C**, Alonso-Rubio J, Gomez-Casado E, Tejedor B, Martín-Villa J.M, Silvera C. Pacific and European Ancestry of Amerindians: a HLA Relatedness Study in Wiwa (Arsario) Colombian Population. 40 Congreso de la Sociedad Española de Inmunología. Zaragoza, España. 2017.
11. Arnaiz-Villena A, Palacio-Gruber J, Muñiz E, **Campos C**, Alonso-Rubio J, Gomez-Casado E, Martín-Villa J.M, Silvera C. HLA Genes in Chimila Amerindians (Colombia), the Peopling of America and Relatedness to Europeans and Pacific Islanders. 40 Congreso de la Sociedad Española de Inmunología. Zaragoza, España. 2017.
12. Arnaiz-Villena A, Muñiz E, **Campos C**, Palacio-Gruber J, Martín-Villa J.M, Tejedor B, Gomez-Casado E, Enriquez de Salamanca M. Characterization and Functional Implications of the Two New HLA-G Alleles Found in Amerindians (HLA-G*01:08:02) and Caribbean (HLA-G*01:20) Populations. 40 Congreso de la Sociedad Española de Inmunología. Zaragoza, España. 2017.
13. Recio M.J, Palacio-Gruber J, Martinez-Laso J, Muñiz E, Gomez-Casado E, Martín-Villa J.M, **Campos C**, Arnaiz-Villena A. Major Histocompatibility Complex-DMB Allelic Diversity in Old and New World Primates: Intraspecies Pattern of Evolution. 40 Congreso de la Sociedad Española de Inmunología. Zaragoza, España. 2017.
14. Arnaiz-Villena A, Palacio-Gruber J, Muñiz E, Recio M.J, **Campos C**, Tejedor B, Martín-Villa J.M, Martinez-Laso J. HLA-DMB in Amerindians: Specific Linkage of DMB*01:03:01/DRB1 Alleles. 40 Congreso de la Sociedad Española de Inmunología. Zaragoza, España. 2017.
15. Arnaiz-Villena A, Muñiz E, **Campos C**, Palacio-Gruber J, Martín-Villa J.M, Tejedor B, Gomez-Casado E, Enriquez de Salamanca M. Characterization and Functional Implications of the Two New HLA-G Alleles Found in Amerindians (HLA-G*01:08:02) and Caribbean (HLA-G*01:20) Populations. 31st European Immunogenetics and Histocompatibility Conference. Mannheim, Germany. 2017.

16. Arnaiz-Villena A, Palacio-Gruber J, Muñiz E, Recio M.J, **Campos C**, Tejedor B, Martín-Villa J.M, Martinez-Laso J. HLA-DMB in Amerindians: Specific Linkage of DMB*01:03:01/DRB1 Alleles. 31st European Immunogenetics and Histocompatibility Conference. Mannheim, Germany. 2017.
17. Recio M.J, Palacio-Gruber J, Martinez-Laso J, Muñiz E, Gomez-Casado E, Martín-Villa J.M, **Campos C**, Arnaiz-Villena A. Major Histocompatibility Complex-DMB Allelic Diversity in Old and New World Nonhuman Primates: Intraspecies Pattern of Evolution. 31st European Immunogenetics and Histocompatibility Conference. Mannheim, Germany. 2017.
18. Arnaiz-Villena A, **Campos C**, Muñiz E, Gomez-Casado E, Martín-Villa J.M, Palacio-Gruber J. First Population of Canary Island is not only North African: Presence of European Atlantic Genes and Ancient Iberian Language Scripts. 31st European Immunogenetics and Histocompatibility Conference. Mannheim, Germany. 2017.
19. Arnaiz-Villena A, Muñiz E, Palacio-Gruber J, **Campos C**, Alonso-Rubio J, Gomez-Casado E, Tejedor B, Martín-Villa J.M, Silvera C. Pacific and European Ancestry of Amerindians: a HLA Relatedness Study in Wiwa (Arsario) Colombian Populations. 31st European Immunogenetics and Histocompatibility Conference. Mannheim, Germany. 2017.
20. Arnaiz-Villena A, Palacio-Gruber J, Muñiz E, **Campos C**, Alonso-Rubio J, Gomez-Casado E, Martín-Villa J.M, Silvera C. HLA Genes in Chimila Amerindians (Colombia), the Peopling of America and Relatedness to Europeans and Pacific Islanders. 31st European Immunogenetics and Histocompatibility Conference. Mannheim, Germany. 2017.
21. Arnaiz-Villena A, Muñiz E, **Campos C**, Martín-Villa JM, Palacio-Gruber J. Evidence of Prehistorical Atlantic and Pacific Transoceanic Genetic and Cultural Contacts with America. The 86th Annual Meeting of the American Association of Physical Anthropologists. New Orleans, Louisiana. 2017.
22. Arnaiz-Villena A, Muñiz E, Palacio-Gruber J, **Campos C**, Tejedor B, Martín-Villa JM, Ruiz-del-Valle V. Striking convergent phenotype evolution in an African Canary (Oriole Finch) and a Caribbean Siskin (Antillean Siskin). 135th Meeting of American Ornithology. East Lansing, Michigan. 2017.

23. Arnaiz-Villena A, Muñiz E, Palacio-Gruber J, **Campos C**, Tejedor B, Martín-Villa JM, Ruiz-del-Valle V. Intron study and evolution of wild songbird MHC genes: minimum essential MHC does not exist. 135th Meeting of American Ornithology. East Lansing, Michigan. 2017.
24. Arnaiz-Villena A, Palacio-Gruber J, Muñiz E, Rey D, Nikbin B, Nickman H, **Campos C**, Martín-Villa JM, Amirzargar A. Origin of Azeris (Iran) according to HLA genes. 32nd European Immunogenetics and Histocompatibility Conference. Venecia, Italy. 2018.
25. Arnaiz-Villena A, Ruiz-del-Valle V, Muñiz E, Palacio-Gruber J, **Campos C**, Gomez-Casado E, Martín-Villa JM, Serrano-Vela JI. Major histocompatibility complex allele persistence in Eurasia and America in the genus *Carduelis* (*Spinus*) during million years. 32nd European Immunogenetics and Histocompatibility Conference. Venecia, Italy. 2018.
26. Arnaiz-Villena A, Palacio-Gruber J, Enrique-de-Salamanca M, Juarez I, **Campos C**, Nieto J, Muñiz E, Martín-Villa JM. HLA-G, -A haplotypes in Amerindians (Ecuador): HLA-G*01:05N World distribution. 32nd European Immunogenetics and Histocompatibility Conference. Venecia, Italy. 2018.
27. Arnaiz-Villena A, Bayona B, Palacio-Grüber J, Hernández E, Muñiz E, **Campos C**, Juarez I, Gomez-Casado E, Martín-Villa JM, Silvera C. HLA genes in Barranquilla (North Colombia): Searching for cryptic Amerindian genes. 32nd European Immunogenetics and Histocompatibility Conference. Venecia, Italy. 2018.
28. Arnaiz-Villena A, Carballo A, Juarez I, Muñiz E, **Campos C**, Tejedor B, Martín-Villa JM, Palacio-Grüber J. HLA Genes in Atlantic Celtic Populations: Are Celts Iberians? 32nd European Immunogenetics and Histocompatibility Conference. Venecia, Italy. 2018.
29. Arnaiz-Villena A, Palacio-Grüber J, Muñiz E, **Campos C**, Rubio JA, Juarez I, Nieto J, Salih SF, Martín-Villa JM, Al-Qadi R. Genetic HLA study of Kurds in Iraq, Iran and Tbilisi (Caucasus, Georgia): relatedness and medical implications. 32nd European Immunogenetics and Histocompatibility Conference. Venecia, Italy. 2018.
30. Palacio-Grüber J, **Campos C**, López-Nares A, Juarez I, Nieto J, Martín-Villa JM, Arnaiz-Villena A. HLA-G-A haplotypes in Amerindians: HLA-G*01:05N world

distribution. 33rd European Immunogenetics and Histocompatibility Conference.
Lisboa, Portugal. 2019.