

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE MEDICINA
Departamento de Microbiología I



TESIS DOCTORAL

**Estudio en aves silvestres de canto ("Passeriformes") de las moléculas
MHC y reclasificación molecular de la familia "Carduelini"**

MEMORIA PARA OPTAR AL GRADO DE DOCTORA

PRESENTADA POR

Cristina Areces Viña

Director

Antonio Arnaiz Villena

Madrid, 2016

**UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE MEDICINA
DPTO. MICROBIOLOGÍA I**



**ESTUDIO EN AVES SILVESTRES DE CANTO
(*PASSERIFORMES*)
DE LAS MOLÉCULAS MHC Y RECLASIFICACIÓN
MOLECULAR DE LA FAMILIA *CARDUELINI***

TESIS DOCTORAL

CRISTINA ARECES VIÑA

**DIRECTOR: Prof. Dr. Antonio Arnaiz Villena
MADRID, 2015.**

TESIS DOCTORAL

ESTUDIO EN AVES SILVESTRES DE CANTO (*PASSERIFORMES*) DE LAS MOLÉCULAS MHC Y RECLASIFICACIÓN MOLECULAR DE LA FAMILIA *CARDUELINI*

AUTOR

Cristina Areces Viña

DIRECTOR

Prof. Dr. Antonio Arnaiz Villena

Coordinador I+D, Centro de Transfusión de la Comunidad de Madrid.

Catedrático de Inmunología, Universidad Complutense de Madrid.

LUGAR DE REALIZACIÓN

- Departamento de Inmunología, Facultad de Medicina. Universidad Complutense de Madrid.
- Departamento de Investigación y Desarrollo. Centro de Transfusión de la Comunidad de Madrid.

Don Antonio Arnaiz Villena, Coordinador I+D del Centro de Transfusión de la Comunidad de Madrid y Catedrático de Inmunología de la Universidad Complutense de Madrid

CERTIFICA

Que **Doña Cristina Areces Viña** ha realizado bajo su dirección el trabajo de Tesis Doctoral que lleva por título: **“Estudio en aves silvestres de canto (*Passeriformes*) de las moléculas MHC y reclasificación molecular de la familia *Carduelini*”**.

Revisado el presente trabajo, considero que tiene la debida calidad para su defensa y calificación.

Prof. Dr. Antonio Arnaiz Villena

AGRADECIMIENTOS

En primer lugar me gustaría agradecerle al Dr. Antonio Arnaiz Villena, quien ha sido mi tutor durante todos estos años, la tutela y dirección de este proyecto de Tesis, así como la confianza puesta en mí, en este y otros muchos trabajos llevados a cabo en mi periodo como doctorando y por supuesto, por la recolección de cada una de las muestras de aves que ha ido recopilando, con enorme esmero y paciencia durante muchos años. Gracias, por permitirme el lujo de trabajar con ellas y transmitirme tu pasión por estos pequeños “dinosaurios voladores”.

También me gustaría agradecer a mis compañeros, en especial a Mercedes y a Diego, todo su cariño y su apoyo. Hemos caminado pasito a pasito en este largo viaje juntos, me llevo muchas cosas aprendidas a nivel profesional y sobretodo personal, muchos buenos momentos, de esos que hacían que un día complicado fuese mucho mejor, gracias de corazón, por vuestra amistad.

A la Dra. Narcisa Quiles, por su ayuda en los últimos empujones para acabar este trabajo.

Y por último al personal del Centro de Transfusión de la Comunidad de Madrid, lugar en el que se ha desarrollado este proyecto de Tesis, por su amabilidad, gracias por estar siempre dispuestos a ayudar, por vuestros ánimos y por vuestras sonrisas.

A todos vosotros, GRACIAS.

Dedicatoria

Podría ir nombrando una a una y agradeciendo, a las personas que han formado parte de mi vida, no sólo desde que empecé este largo trayecto que acaba con este trabajo, si no a las personas que han formado y siguen formando parte de mi crecimiento personal, esas personas importantes y esenciales para mí, como son mis padres, mi hermana, mis abuelos, mi compañero, mis amigos y un largo etcétera.

En fin, podría escribir hojas y hojas agradeciéndoos vuestro cariño, vuestro apoyo, vuestras sonrisas, vuestros abrazos, dándoos las gracias por caminar junto a mí, siempre de la mano, por enseñarme y por hacerme crecer, pero creo que más importante que nombraros uno a uno en esta dedicatoria, y que sean simplemente bonitas palabras escritas de corazón, es haber podido demostraros a todos y cada uno de vosotros, lo afortunada que me siento de que forméis parte de mí, espero haber podido conseguir que sepáis lo importantes que sois y que simplemente leer esto sea un recordatorio de lo que ya sabéis.

Gracias a todos, porque con vosotros todo se hace más fácil, gracias porque este camino también lo habéis hecho vuestro. Gracias por acompañarme.

Para ti.

ÍNDICE

RESUMEN.....	1
INTRODUCCIÓN.....	7
I. EL SISTEMA PRINCIPAL DE HISTOCOMPATIBILIDAD	8
1. Función de los Genes de Histocompatibilidad	8
1.1 Historia y definición de los Genes de Histocompatibilidad	8
1.2 Genes de Histocompatibilidad: enfermedad, farmacogenómica y trasplante.....	10
1.2.1 Asociación entre HLA y enfermedad.....	10
HLA y autoinmunidad	10
HLA y patógenos.....	12
1.2.2 HLA y trasplante.....	12
1.2.3 HLA y farmacogenómica	13
1.2.4 Asociación del gen HFE a hemocromatosis.....	13
2. Estudio de los Genes de Histocompatibilidad	14
2.1 Modelos de laboratorio.....	14
2.1.1 Sistema Principal de Histocompatibilidad en el pollo	14
El complejo B (locus B)	15
Región B-F/B-L	16
Estructura proteica de los genes de clase I (B-F) y de clase II (B-L)	17
Función y estructura proteica de las moléculas B-G	18
Locus Rfp-Y	18
Susceptibilidad a enfermedades del pollo.....	19
Hipótesis del MHC "mínimo-esencial".....	20
2.1.2 Sistema Principal de Histocompatibilidad en humanos (HLA)	20
La región HLA.....	20
Las moléculas HLA de clase I	26
Las moléculas HLA de clase II	28
Región de clase III	30
2.2 Modelos silvestres (aves de canto)	30
3. Polimorfismo y evolución de los genes MHC	31
3.1 Mecanismos generadores de polimorfismo: la conversión génica	31
3.2 Mantenimiento del polimorfismo	32
A. Selección de balance.....	32
Selección dependiente de frecuencia	33
Ventaja de los heterocigotos	33
Selección variable en el tiempo y en el espacio.....	33
B. Emparejamientos no al azar	34
C. Interacción materno-fetal.....	34
3.4 Evolución transespecífica	35
3.5 Homogeneización de alelos y pérdida de ortología	35

II. LA CLASE AVES	37
4. Orden Passeriformes	38
4.1 Relaciones Intra-Orden	40
4.1.1 Oscines o Passeri	40
4.1.2 Suboscines o Tyranni	41
5. Características de las aves estudiadas	41
5.1 Carduelis atrata (Lúgano negro)	41
5.2 Carduelis carduelis (Jilguero común)	42
5.3 Carduelis citrinella (Verderón serrano)	42
5.4 Carduelis crassirostris (Lúgano de pico grueso)	42
5.5 Carduelis cucullata (Cardenalito de Venezuela)	42
5.6 Carduelis dominicensis (Pinero antillano)	43
5.7 Carduelis lawrencei (Dominguito gris)	43
5.8 Carduelis magellanicus (Lúgano encapuchado)	44
5.9 Carduelis notata (Lúgano de cabeza negra)	44
5.10 Carduelis olivacea (Lúgano oliváceo)	44
5.11 Carduelis pinus (Jilguero de los pinos)	45
5.12 Carduelis psaltria (Dominguito)	45
5.13 Carduelis spinescens (Lúgano de los Andes)	46
5.14 Carduelis spinus (Lúgano euroasiático)	46
5.15 Carduelis xanthogastra (Lúgano de vientre amarillo)	47
5.16 Carduelis yarrellii (Lúgano de cara amarilla)	47
5.17 Serinus sulphuratus (Canario sulfúreo)	47
5.18 Serinus atrogularis (Canario de obispillo amarillo)	48
5.19 Serinus canaria (Canario silvestre)	48
5.20 Serinus citrinelloides (Canario africano)	48
5.21 Serinus dorsostriatus (Canario de vientre blanco)	48
5.22 Serinus flaviventris (Canario amarillo)	49
5.23 Serinus gularis (Canario de cabeza estriada)	49
5.24 Serinus mozambicus (Canario de frente amarilla)	49
5.25 Serinus striolatus (Canario estriado)	50
5.26 Serinus thibetanus (Canario tibetano)	50
5.27 Rhodopechys sanguineus (Camachuelo Alirrojo)	51
III. FILOGENIA MOLECULAR	52
6. El ADN mitocondrial	52
6.1 Generalidades	52
6.2 Citocromo mitocondrial	52
6.2.1 Citocromo b y análisis filogenéticos	54
7. Filogenia molecular	55
7.1 Métodos cladísticos	56
7.2 Métodos basados en distancias genéticas	56
Modelos de sustitución nucleotídica	57
Métodos de reconstrucción de árboles filogenéticos	58
OBJETIVOS	60
MATERIAL Y MÉTODOS	63
1. Muestras empleadas	64
1.1 Especies estudiadas	64
Estudio de MHC clase I	64

Estudio fiogenético basado en el gen Citocromo b	65
1.2 Obtención de las muestras.....	67
2. Extracción de ADN	68
3. Obtención de secuencias de ADN de los genes de MHC de clase I	70
3.1 Amplificación de secuencias de MHC de clase I	70
3.2 Electroforesis en gel de agarosa de los fragmentos amplificados	73
3.3 Purificación de la banda de la secuencia amplificada	74
3.4 Secuenciación de los fragmentos amplificados.....	75
3.5 Clonación de los fragmentos amplificados.....	76
3.6 Obtención de las secuencias de ADN de los fragmentos clonados	79
4. Obtención de secuencias de ADN del gen de citocromo b	81
5. Análisis de secuencias obtenidas	82
5.1 Caracterización de las secuencias de ADN	82
5.2 Caracterización de las secuencias de aminoácidos	83
5.3 Análisis de la variabilidad	84
Diagramas de entropía.....	84
Cálculo de las tasas de sustitución sinónima y no sinónima	84
5.4 Análisis de conservación de todas las secuencias estudiadas	85
Cálculo de distancias genéticas.....	85
Rango de conservación	85
Porcentaje de similitud	86
6. Filogenia basada en el gen citocromo b	86
RESULTADOS.....	89
1. Secuencias MHC de clase I	90
1.1 Composición nucleotídica de las secuencias de ADN en aves <i>Passeriformes</i>	91
1.2 Distribución de la variabilidad: diagramas de entropía.....	92
1.3 Análisis de los tipos de sustituciones	93
1.4 Secuencias de aminoácidos en aves <i>Passeriformes</i>	95
1.5 Cálculo de distancias genéticas entre las aves silvestres de canto estudiadas ...	100
1.6 Análisis de las posiciones que determinan el sitio de reconocimiento antigénico (ARS).....	102
1.7 Secuencias de aminoácidos del resto de vertebrados a estudio	104
1.8 Cálculo de distancias genéticas entre las aves silvestres de canto y el resto de vertebrados	107
1.9 Análisis de las posiciones aminoacídicas conservadas en ambos grupos (passerines y resto de vertebrados).....	109
1.10 Cálculo del grado de conservación en cada una de las posiciones	111
1.11 Comparativa de medias de rango de conservación en los dos grupos	112
1.12 Porcentaje de similitud entre las secuencias estudiadas	113
1.13 Estudio de intrón 2 de MHC de clase I en aves <i>Passeriformes</i>	117
1.14 Porcentaje medio de similitud entre los intrones de las muestras estudiadas.	122
1.15 Comparación de la longitud del intrón 2 en las muestras estudiadas	123
2. Filogenia mitocondrial.....	125
DISCUSIÓN.....	129
1. Sistema Principal de Histocompatibilidad en aves silvestres de canto	130
2. Filogenia y evolución de aves <i>Passeriformes</i>	141

CONCLUSIONES	144
BIBLIOGRAFÍA	147
ANEXOS	163
1. Listado de las abreviaturas más empleadas en este trabajo.....	164
2. Clasificación de las aves (Sibley, 1996).....	165
3. Simbología y código genético.....	167
4. Información sobre aminoácidos.....	168
5. Publicaciones relacionadas.....	169
CURRICULUM VITAE	172

RESUMEN

STUDY OF MHC MOLECULES IN WILD SONGBIRDS (*PASSERIFORMES*) AND MOLECULAR RECLASSIFICATION OF *CARDUELINI* FAMILY

INTRODUCTION

The Major Histocompatibility Complex (MHC) comprises the most polymorphic loci in animals. MHC plays an important role during the first steps of the immune response in vertebrates.

In humans, MHC molecules (also named human leukocyte antigens, HLA) were initially regarded as class I or class II molecules. Each of them, presents to different T cells subsets. MHC class I molecules, are heterodimers in which the heavy chain (alpha) has three extracellular domains, two of which (alpha 1 and alpha 2) are polymorphic and conform the antigen recognition sites (ARS).

The ARS is thought to be subjected to balancing selection for variability, which is the cause of the very high polymorphism of the MHC molecules. Different pathogenic epitopes would be the evolutionary force causing balancing selection.

MHC class I genes have been completely sequenced ($\alpha 1$ and $\alpha 2$ protein domains) and thoroughly studied in *Gallus gallus* (chicken) as well as in mammals. In fact, the MHC locus was first defined in chicken, specifically in the highly consanguineous variety 'Leghorn'. It has been found that, in the case of chickens the MHC genetic region is considerably smaller than it is in mammals (remarkably shorter introns were found in chickens), and is organized quite differently.

The noteworthy presence of short introns in chickens; supported the hypothesis that chicken's MHC represented a 'minimal essential MHC'. Until now, it has been assumed that chicken (order *Galliformes*) MHC was similar to all species included in the whole class *Aves*.

Songbirds MHC is a relevant case of study given its characteristics. Songbirds's evolution (*Aves*, *Passeriformes*, *Passeri*) has been studied and it has been found that their phenotypic and molecular evolutions are not always correlated.

Similarities in morphology and behaviour may originate shared features among genetically unrelated species occurring in similar environments; conversely, different features may be present among genetic sister taxa thriving under different environments. This phenotypic plasticity has already been described for other bird species. Thus, some genetic and/or phenotypic traits may not correlate with the evolutionary histories of these birds.

Mitochondrial cytochrome b gene (cyt-b) sequencing has been widely used for molecular systematics studies. Cyt-b gen has been proved to be helpful for defining evolutionary relationships among distant or closely related taxa.

AIMS OF THE RESEARCH PROJECT

Class I MHC genes and their corresponding proteins in wild songbirds, have been studied and compared to those of other birds, and to other available vertebrate genes. We focused on those domains interacting with T-cell receptors and antigen peptides which was subjected to variability by balancing selection ($\alpha 1$ and $\alpha 2$ domains). Focus was made on highly conserved positions which are kept unchanged throughout vertebrate history; some of them are related to interaction with antigen or with the rest of heavy and $\beta 2$ -microglobulin chains. Study of histocompatibility genes in wild birds *Passeriformes*, is important because the evolution of MHC system compares these natural (wild birds) models with "artificial models" (humans, mice, chickens) and other vertebrates.

A phylogenetic study was carried out with the aim of analyzing genetic relationships among different *Carduelini* family tribes, using mitochondrial DNA coding for cytochrome b, previously obtained by us, in order to contrast the existing evolutionary relationship between them with the phylogenetic relationships observed in the sequences of MHC class I of the species we analyzed.

Main objectives of this study were:

- To make an evolutionary comparison of Major Histocompatibility Complex class I in wild *Passeriformes* birds mainly of the *Carduelis* and *Serinus* genera.
- To study the variability of these genes by comparing the sequences of DNA and aminoacids in order to determine location of variable positions and substitutions type.
- To determine the evolutionary pressures affecting histocompatibility molecules by analyzing rates of synonymous and nonsynonymous substitutions and number of transitions and transversions, in different regions of the DNA sequences.
- To study aminoacid residues conserved and non-conserved in vertebrate species, and compare the existence of these residues in selected *Passeriformes* birds.
- To compare the obtained intron sequences with sequences from other vertebrates, including birds of order *Galliformes*, in order to check whether songbirds MHC is similar to the "minimum essential MHC" found in chicken (*Gallus gallus*).
- To review how *Passeriformes* birds classical phenotypic grouping coincide with

molecular phylogenetic grouping, by performing an analysis of cytochrome b of the following *Passeriformes* genera: *Rhodopechys*, *Carpodacus*, *Bullfinch*, *Haematospiza*, *Pinicola*, *Uragus*.

RESEARCH RESULTS AND DISCUSSION

In this work, molecular data have been obtained from wild songbirds (*Passeriformes*). It have been sequenced 26 species, collected over world geography. Sequencing of exon 2, intron 2 and exon 3 of the gene coding for the heavy chain of MHC class I molecules were sequenced for every sample. Results, were compared with 20 different vertebrate sequences obtained from the GenBank database.

A study of evolutionary relationships among sampled species of wild songbirds was conducted. This was done in order to find out evolutionary process in this birds histocompatibility class I sequences. It was carried out by sequencing and performing a cytochrome b phylogenetic analysis.

Analysis of the substitution rate

Results of substitution rate sequences of exons 2 and 3 of MHC class I in songbirds, show how number of synonymous substitutions is greater than nonsynonymous in all regions: ARS (antigen recognition sites) and non-ARS.

There is no evidence that balancing selection (characterized by a high content of nonsynonymous substitutions in the sequences) is acting on these MHC class I genes. Therefore, observed selection type in these sequences is not the typical one that can be found in classical MHC class I molecules. However, this type of analysis requires a larger sample in order to avoid possible sampling errors.

Analysis of aminoacid variability

It was observed that none of sequences variable positions have more than two alternatives, so that no hypervariability in the MHC class I protein sequences of these wild birds was found, as was defined in human classical HLA class I molecules. Results show that most of aminoacid variations are located outside the antigen recognition site, which was well preserved. .

Analysis of conserved aminoacid positions in vertebrates

Our research verified 6 of the 7 positions identified as conserved positions for domains alpha 1 and alpha 2 of MHC class I sequences. Two of these 6 show different aminoacids in our

wild songbirds samples. Position 10 in other aminoacid vertebrates has a threonine which is substituted for a valine in *Passeriformes* birds except *S. canaria* y *C. crassirostris* species and position 96 in vertebrates is a glutamine and in *Passeriformes* birds changes to leucine.

The exception *Serinus canaria*

The results show that *Serinus canaria* MHC sequences have been one of the constant exceptions throughout all the work, which corroborates previous research. The variability found in this molecule when compared to that of other *Passerine* birds could be explained by isolation and different in existing pathogens in the Canary, Madeira and Azores Islands compared with continental microbes.

Analysis of intron sequences and “Minimal essential” theory

Intron 2 has been sequenced in different genera of *Passeriformes* birds (*Carduelis*, *Fringilla*, *Passer*, and *Serinus*). Results show that average length of this intron in our small wild songbirds (303pb) is very similar than other mammals (except *Homo sapiens* -241 bp) and birds of the same Order (*Coturnix japonica*) as well as the chicken. This suggests that, a "Minimal essential MHC" is not found in songbirds.

On the other hand, results show that intron 2 sequences are highly conserved in analyzed songbirds. Intron 2 conservation for long periods of time (million years) found by us, might suggest that introns are influenced by external unknown factors and could be due to evolutionary selection.

Mitochondrial phylogeny

Finally, we performed a phylogenetic study for analyzing genetic relationships between different *Carduelini* family tribes. This was done by using mitochondrial DNA sequences. It was possible to compare the existing evolutionary relationship among them with phylogenetic relationships observed in the MHC class I sequences. *Rhodopechys sanguineus* mitochondrial DNA has been sequenced for the first time for this work. In summary results show that *Rhodopechys obsoleta*, is a Greenfinches ancestor. On the other hand, “Arid-Zone” finches group includes (*Leucosticte arctoa tephrocotis*, *Leucosticte arctoa arctoa*, *Carpodacus nipalensis*, *Rhodopechys githaginea*, *Rhodopechys mongolica* and *Rhodopechys sanguineus*). Also *Carpodacus erythrinus* and *Haematospiza sipahi* form a single phylogenetic clade and Genus *Loxia* (Crossbills) are both placed as monophyletic and close to *Carduelis flammea* and *hornemanni*. Finally note that *Pinicola enucleator* should be classified together with Bullfinches (*Pyrrhula* species).

CONCLUSIONS

From the results obtained in this work, it can be concluded that:

1. Molecules of histocompatibility class I of *Passeriformes* birds studied bear specific aminoacid changes which do not exist in other birds (*Gallus gallus*, *Coturnix japonica*), or in other vertebrates studied.
2. Stability of these molecules and their ability to bind to β 2-microglobulin is very similar to the other vertebrates which were studied.
3. This suggests that small songbirds (*Passeriformes*) which mainly are air thriving, followed a different evolutionary pathway than other terrestrial birds (i.e: *Galliformes*).
4. Intron 2 size of MHC class I molecules, is similar in vertebrates and *Passeriformes* (average 304 bp). This intron is significantly longer in wild songbirds than in chicken (203 pb) and humans (241 pb) but not in quail (belonging to order *Galliformes* like chickens).
5. Intron 2 nucleotide sequence has been found to be kept in the *Carduelinae* family for at least 9 million years. This suggests the presence of a positive evolutionary selection acting on this intron 2 in *Passeriformes* birds.
6. In addition, this work presents a new systematic molecular phylogeny for many species of *Passeriformes* birds from 5 continents. Substantial variations are found with existing phenotypic phylogeny as it correlates with climatic variations which have been found on Earth since the Miocene (23 million years ago).

INTRODUCCIÓN

I. EL SISTEMA PRINCIPAL DE HISTOCOMPATIBILIDAD

1. Función de los Genes de Histocompatibilidad

1.1 Historia y definición de los Genes de Histocompatibilidad

El Sistema Principal de Histocompatibilidad (MHC, *Major Histocompatibility Complex*) está constituido por una serie de genes implicados en la activación de la respuesta inmunológica adaptativa. Se trata de un sistema altamente polimórfico que está presente en todos los vertebrados, salvo en los peces no mandibulados (Flajnik *et al.*, 1999) y codifica proteínas de superficie cuya función es presentar antígenos (péptidos derivados principalmente de agentes infecciosos) a los linfocitos T, células inmunitarias que una vez activadas son capaces de poner en marcha una respuesta inmunológica eficaz y específica contra el agente causante de su activación (Zinkernagel y Doherty, 1979).

El descubrimiento de los genes de histocompatibilidad requirió de la unión de dos campos de la investigación inicialmente separados: uno dedicado al estudio del cáncer mediante el trasplante de tumores en ratones y otro relacionado con la identificación de grupos sanguíneos. Ya en los primeros años del siglo XX se había comprobado que los tumores trasplantados entre ratones pertenecientes a cepas diferentes, como también tejido sano (Little y Johnson, 1922), eran destruidos por el receptor (Jensen, 1903;Loeb, 1908).

En la década de 1930, J. B. Haldane sugirió la idea de que la destrucción de tejidos trasplantados y la lisis o aglutinación eritrocitaria eran fenómenos similares, por lo que propuso la combinación de ambas disciplinas. Pocos años después, un discípulo suyo identificó varios antígenos eritrocitarios en el ratón y comprobó que uno de ellos (el antígeno II) estaba implicado en la supervivencia del tumor trasplantado, ya que en las cepas de ratón utilizadas la ausencia de dicho antígeno se correlacionaba con la destrucción del tumor, lo que le permitió formular una teoría inmunológica del trasplante (Gorer, 1936). Paralelamente, Robert Irwin comenzaba a estudiar los grupos sanguíneos en aves junto a otros investigadores, entre ellos Elwood Briles.

Años más tarde, los genes implicados en la supervivencia de tumores trasplantados en ratones, cuyos productos proteicos podían ser detectados mediante pruebas serológicas, fueron denominados genes de histocompatibilidad, y en el caso del ratón genes H-2, en referencia al primer antígeno descubierto (Snell, 1948). En ese mismo año fueron identificados los antígenos eritrocitarios en el pollo, comprobándose que estaban codificados en 2 loci independientes (A y B) (Briles *et al.*, 1950;Briles y McGibbon, 1948); más tarde se determinó la implicación del locus B en la viabilidad de tejidos trasplantados, por lo que se

concluyó que el locus B representaba el Sistema Principal de Histocompatibilidad del pollo (Schierman y Orskog, 1961).

El descubrimiento de los genes de histocompatibilidad en humanos fue precedido, como en los casos anteriores, de la identificación de antígenos sanguíneos, aunque en esta ocasión estaban ubicados en los leucocitos. Fueron descubiertos por Jean Dausset cuando, investigando el origen de la leucopenia en pacientes que habían recibido múltiples transfusiones sanguíneas, detectó la presencia de anticuerpos que provocaban la aglutinación de leucocitos del donante, pero no del receptor (Dausset, 1958; Dausset y Nenna, 1952). Después se verificó que dichos anticuerpos no eran responsables de la leucopenia y sí el resultado de la inmunización contra los antígenos leucocitarios de los donantes, anticuerpos que también fueron detectados en mujeres multíparas que habían desarrollado inmunidad contra los antígenos leucocitarios paternos presentes en el feto (Payne y Rolfs, 1958; van Rood *et al.*, 1958). En la década de 1960 se comprobó que los antígenos leucocitarios representaban el Sistema Principal de Histocompatibilidad, que fue denominado HLA (*human leukocyte antigen*) (Janeway y Travers, 1996; Klein, 1986).

Los primeros genes que se descubrieron en humanos, fueron HLA-A, HLA-B y HLA-C (Thorsby, 2009), el primer gen que se identificó con la “reacción leucocitaria mixta” fue el adyacente al locus del HLA, por lo que se denominó HLA-D, luego se identificó la molécula de este gen y se la bautizó HLA-DR (*HLA-D related*). Finalmente se descubrieron dos genes más, HLA-DQ y HLA-DP.

El Sistema Principal de Histocompatibilidad presenta una serie de características fundamentales que dificultan la evasión por parte de los patógenos de las respuestas inmunes mediadas por el MHC:

- Es **poligénico**, por lo que existen varios genes codificantes para moléculas MHC de clase I y de clase II.
- Es altamente **polimórfico**, lo que provoca que existan multitud de alelos para cada gen. De hecho, el MHC es el sistema más polimórfico conocido.
- Posee **desequilibrio de ligamiento**, propiedad de algunos genes de no segregarse de forma independiente por lo que poseen una frecuencia de recombinación menor del 50%. Esto se debe a que los loci implicados se encuentran en el mismo cromosoma, lo que imposibilita su transferencia a la progenie de manera aleatoria. De hecho, existen zonas en la región de clase II en las que no se ha observado nunca recombinación a pesar del elevado número de familias que se ha estudiado. Este es el caso del tramo presente entre DRA y DQA, fragmento génico que se

hereda como un bloque. No obstante, a pesar de que existen zonas del MHC en las que nunca se produce recombinación parece ser que hay otras en las que este fenómeno es bastante frecuente. Por ejemplo, se ha encontrado un "punto caliente" de recombinación entre DP y DQ. Se cree que podrían existir otras zonas de recombinación en la región de clase I y de clase III (Thomsen *et al.*, 1994).

- Su expresión es **codominante**, por lo que se expresan los alelos de procedencia paterna y materna.

Por estas características y por la experiencia acumulada, es indudable la implicación sanitaria de este sistema en cuanto a su relación con las enfermedades infecciosas y con los tumores, pero no es menor su relevancia, en el caso humano, en los trasplantes de órganos y en su estrecha vinculación con algunas enfermedades de tipo autoinmune.

1.2 Genes de Histocompatibilidad: enfermedad, farmacogenómica y trasplante

1.2.1 Asociación entre HLA y enfermedad

La primera vez que se estableció relación entre el sistema HLA y una enfermedad fue en pacientes con linfoma de Hodgkin (cáncer del sistema inmunitario que se caracteriza por la presencia de un tipo de célula maligna llamada célula de Reed Sternberg), que presentaban una frecuencia de antígenos HLA-B por encima de la media (Amiel, 1967). Desde entonces, numerosas variantes alélicas de genes HLA han sido asociadas con enfermedades autoinmunes, inflamatorias e infecciosas. Sin embargo, el carácter multifactorial de muchas de estas enfermedades sumado al elevado polimorfismo de algunos de estos genes y el fuerte desequilibrio de ligamiento existente entre ellos dificulta en la mayoría de los casos la determinación de la variante alélica implicada, así como su grado de participación en la enfermedad. De hecho, siempre existe la posibilidad de que los factores responsables de estas patologías pueden no residir en genes HLA, pero sí en otros estrechamente ligados a ellos.

HLA y autoinmunidad

La mayoría de enfermedades autoinmunes están asociadas a factores HLA, aunque los mecanismos responsables de esta asociación no están claros. La hipótesis más aceptada es que se rompe la tolerancia inmunológica hacia proteínas propias cuando moléculas HLA de clase II activan linfocitos T auto-reactivos al presentarles péptidos propios o extraños.

Una de las asociaciones más fuertes es la que relaciona la espondilitis anquilosante (inflamación crónica que afecta principalmente a la columna vertebral) con el grupo de alelos

HLA-B27, presentes en el 95% de los pacientes. Otras enfermedades que tienen el sistema HLA como principal factor asociado, son la esclerosis múltiple (inflamación crónica del sistema nervioso central asociada a variantes alélicas del antígeno HLA-DR2), la diabetes tipo 1 (destrucción de células beta productoras de insulina en los islotes de Langerhans del páncreas, mediada por linfocitos T, con un 50% de asociación con haplotipos HLA de clase II), el lupus eritematoso sistémico (proceso autoinmune generalizado asociado a variantes alélicas de los antígenos HLA-DR2 y HLA-DR3), la colitis ulcerosa junto con la enfermedad de Crohn (procesos inflamatorios intestinales asociados a diversos alelos y haplotipos HLA de clase II) y la artritis reumatoide (proceso inflamatorio que afecta a las articulaciones de forma generalizada asociado a las variantes alélicas del locus HLA-DRB1) (Tabla 1).

La mayoría de estas asociaciones se han establecido casi exclusivamente en poblaciones de origen europeo y el número de genes analizados dentro del sistema HLA no supera la veintena (Fernando *et al.*, 2008; Hill, 2001).

Enfermedad	Alelos HLA
Espondilitis anquilosante	B27
Síndrome de Reiter	B27
Uveítis aguda anterior	B27
Hemocromatosis hereditaria	A3, B14
Enfermedad de Behcet	B5
Narcolepsia esporádica	DR2, DQ6
Esclerosis múltiple	DR2, DQ6
Enfermedad celiaca	DR3, DQ2
Glomerulitis idiopática membranosa	DR3
Hepatitis crónica	DR3
Lupus eritematoso sistémico	DR3
Diabetes mellitus insulino dependiente	DR3, DR4,
Artritis reumatoide	DRB1
Dermatitis herpetiforme	DR3
Síndrome de Goodpasture	DR2
Enfermedad de Hodgkin	DP3
Enfermedad de Crohn	B27

Tabla 1. Ejemplo de algunas enfermedades inmunes asociadas a alelos HLA.

HLA y patógenos

Una hipótesis muy extendida es que la asociación entre el sistema HLA y enfermedades infecciosas se debe a la distinta capacidad que tienen los alelos de los genes implicados en la presentación antigénica, para activar respuestas eficaces contra los patógenos. De ahí se derivaría que unos alelos confieran resistencia y otros sean factores de susceptibilidad. Nuevamente, el elevado polimorfismo de estos genes y los desequilibrios de ligamiento entre alelos hacen muy difícil atribuir estas cualidades de resistencia o susceptibilidad a alelos concretos. De hecho son muy pocos los alelos considerados asociados a enfermedades infecciosas en comparación con el gran número de variantes alélicas existentes. Además, la eficacia de unas proteínas de histocompatibilidad puede estar condicionada por otras que se coexpresan con ellas.

Los casos más estudiados de asociación están relacionados con la malaria, la hepatitis B y el SIDA. Los individuos afectados de malaria (enfermedad producida por un protozoo parásito que invade los eritrocitos causando fiebre y anemia) son más resistentes a la enfermedad cuando poseen la variante alélica HLA-B*5301 (Hill *et al.*, 1991). La hepatitis B (causada por una infección viral que, una vez establecida de forma crónica, puede derivar en cirrosis y carcinoma hepatocelular) es prevenida por los antígenos HLA-DR13, mientras que los alelos HLA-DRB1*11/*12 y HLA-DQB1*0301 actúan como factores de susceptibilidad (Singh *et al.*, 2007). En el caso del SIDA (inmunodeficiencia provocada por la invasión viral de células T CD4+) está comprobado que los individuos con mayor probabilidad de supervivencia frente a la infección por el virus HIV-1 son, en general, los que presentan mayor grado de heterocigosidad en los genes HLA y ausencia de los antígenos HLA-B*35 y HLA-Cw*04 (Carrington *et al.*, 1999); en cualquier caso, se han identificado varios alelos y haplotipos de clase I y de clase II que confieren resistencia o susceptibilidad frente a la transmisión del virus y frente a la progresión de la enfermedad en individuos infectados tanto por HIV-1 (variante más virulenta) como por HIV-2 (variante menos virulenta) en distintos grupos humanos de estudio (Trachtenberg y Erlich, 2001).

1.2.2 HLA y trasplante.

Se ha comprobado con estudios realizados tras el seguimiento de pacientes con órganos trasplantados, que la compatibilidad entre donante y receptor en las moléculas HLA es necesaria para la mayor supervivencia del injerto en los casos de riñón, páncreas, pulmón, corazón, córnea, hígado e intestino. Esta compatibilidad, debe ser al menos con los antígenos pertenecientes a los loci HLA-B y –DR. La compatibilidad debe de ser máxima cuando se

habla de trasplante de médula ósea para evitar que se produzca la llamada enfermedad del “injerto contra huésped” de hecho un único cambio en un aminoácido es suficiente para que se produzca esta patología (Zachary *et al.*, 2010).

1.2.3 HLA y farmacogenómica

Entendemos por farmacogenómica, al estudio de la variabilidad de la expresión génica en respuesta a determinados fármacos. En los estudios farmacogenómicos, se llevan a cabo abordajes que tienen en cuenta las características de las secuencias genómicas, mediante una visión integradora, que incluye las interacciones entre genes.

Su objetivo principal es la creación de fármacos a medida para cada paciente y adaptados a sus condiciones genéticas.

Dentro de los genes estudiados unos de los que son objeto directo de estos estudios son los que componen el HLA, cuyas moléculas han sido ampliamente analizadas en trabajos de asociación a enfermedades (Becquemont, 2010;Charron, 2012). En este orden, debido a que la mayoría de medicamentos producen efectos adversos en determinados individuos, los genes HLA están empezando a jugar un papel central en la identificación precoz de estas manifestaciones.

De ahí que se hayan detectado, por ejemplo, algunos casos en que los medicamentos producen enfermedades como el síndrome de Stevens-Johnson producto del uso del anticonvulsivante Carbamazepina, asociado a la presencia de ciertos alelos HLA como el HLA-B*1502, o reacciones cutáneas debidas al uso de Alopurinol en presencia del alelo HLA-B*5801 (Tabla 2).

1.2.4 Asociación del gen HFE a hemocromatosis

La hemocromatosis, enfermedad que produce acumulación de hierro en los tejidos (con las consiguientes complicaciones cutáneas y hepáticas) y es causada por varias mutaciones en el gen HFE. Su constitución genética y proteica es similar a HLA de clase I y mapea telomérico a ellos. Su producto proteico interacciona con el receptor de transferrina reduciendo la capacidad de transporte de hierro desde la sangre hacia el interior de las células. Afecta a población de origen europeo y la mutación responsable, asociada al haplotipo HLA-A3-B7, es C282Y (presente en heterocigosis en 1/8 individuos y en homocigosis en 1/200 individuos) e impide que la proteína se pliegue correctamente y se ancle a la membrana, permitiendo así que el receptor de transferrina transporte más hierro. Otras dos mutaciones (H63D, 25% y S65C, 3%) tienen efectos similares en combinación con

la anterior (Distante *et al.*, 2004).

Medicamento	Uso	Efecto adverso	Alelo HLA asociado al efecto
Carbamazepina (Arnaiz-Villena <i>et al.</i> , 2007e)	Anti-convulsivante	Síndrome de Steven Johnson	B*1502
Fenitoina (Grahovac <i>et al.</i> , 1998)	Antiepiléptico	Síndrome de Steven Johnson	B*1502
Alopurinol (Munkhbat <i>et al.</i> , 1997)	Gota e Hiperuricemia	Síndrome de Steven Johnson y otras reacciones cutáneas	B*5801
Abacavir (Arnaiz-Villena <i>et al.</i> , 2003; Vargas-Alarcon <i>et al.</i> , 2007)	Inhibidor de la Transcriptasa (VIH)	Fiebre, rash cutáneo, náusea, vómito diarrea	B*5701
Flucloxacilina (Arnaiz-Villena <i>et al.</i> , 1999b; Martinez-Laso <i>et al.</i> , 2006)	Antibiótico betalactámico	Daño hepático	DRB1*0701 B*5701
Nevirapina (Kapustin <i>et al.</i> , 1999)	Inhibidor de la Transcriptasa	Hepatitis, fiebre, rash	DRB1*0101
Ximelagatran (Leffell <i>et al.</i> , 2004)	Anticoagulante inhibidor de la trombina	Eleva niveles séricos de alanina amino transferasa	DRB1*0701
Betalactámicos (García-Ortiz <i>et al.</i> , 2006)	Antibiótico	Alergias medicamentosa	A1, B8, DR3, DQ2
Aminopenicilina (Lazaro <i>et al.</i> , 1999)	Antibiótico	Alergia medicamentosa	A2, DRw52
Coamoxiclav (Martinez-Laso <i>et al.</i> , 2011)	Combinación de antibióticos	Ictericia	DRB1*1501
Clozapina (Martinez-Laso <i>et al.</i> , 1996)	Antipsicótico	Agranulocitosis	DRB5*0201
Sulfametoxazol (Roitberg-Tambur <i>et al.</i> , 1995)	Antibiótico	Steven-Johnson	B*3802
Lamotrigina (Roitberg-Tambur <i>et al.</i> , 1995)	Antiepiléptico	Steven-Johnson	B*3801
Oxicam (Roitberg-Tambur <i>et al.</i> , 1995)	Anti-Inflamatorio no esteroideo	Steven-Johnson	B*7301
Aspirina (Martinez-Laso <i>et al.</i> , 2001)	Anti-Inflamatorio, analgésico	Urticaria, angioedema, Alergia.	DRB1*1302, DQB1 *060, DPB1*0201

Tabla 2. Marcadores HLA asociados a efectos adversos de medicamentos

2. Estudio de los Genes de Histocompatibilidad

2.1 Modelos de laboratorio

2.1.1 Sistema Principal de Histocompatibilidad en el pollo

El Sistema Principal de Histocompatibilidad del pollo (*Gallus gallus*) es uno de los más estudiados y mejor conocidos fuera de los mamíferos. Fue inicialmente descrito como parte de un sistema de grupos sanguíneos codificados en dos loci independientes (Briles *et al.*, 1950; Briles y McGibbon, 1948). Se denominó **complejo B**, o locus B, y fue el primer MHC identificado fuera de los mamíferos.

Estudios moleculares posteriores detectaron además otra región de histocompatibilidad, independiente de la anterior, que fue denominada sistema **Rfp-Y** (*restriction fragment pattern Y*), o locus Y (Miller *et al.*, 1994; Miller *et al.*, 1996). Ambas regiones, fueron secuenciadas a partir de pollos pertenecientes a una estirpe denominada “Leghorn” (la más

empleada en investigación), obteniéndose así el primer mapa completo de la región MHC del pollo (Kaufman *et al.*, 1991); sin embargo se cuestiona que este mapa sea representativo de la diversidad real de la especie, ya que la estirpe Leghorn descende de un grupo reducido de individuos (Simonsen *et al.*, 1989).

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), de manera que la mitad de los genes están ubicados en los macro cromosomas y la otra mitad en los micro cromosomas (Guillemot *et al.*, 1988). Uno de éstos últimos, el micro cromosoma 16, con una longitud de 8.000 kb, aloja el complejo B y el sistema Rfp-Y, independientes entre sí pero estrechamente ligados a una región organizadora nucleolar (NOR, *nuclear organization region*), que abarca 6.000 kb y que contiene unas 150 unidades transcripcionales ribosomales (ADNr) (Bloom *et al.*, 1987; Guillemot *et al.*, 1988). Ambas regiones contienen genes de histocompatibilidad de clase I y de clase II, distribuidos en 5 "clusters" con múltiples genes codificantes para la cadena α de clase I y para la cadena β de clase II, entremezclados a su vez con otros genes que no tienen función inmunológica conocida (Guillemot *et al.*, 1988; Miller *et al.*, 1994; Miller *et al.*, 1996). Las principales causas de la denominación "MHC mínimo esencial" al Sistema Principal de Histocompatibilidad del pollo, fueron las observaciones comparativas con HLA y otros MHC de mamíferos, de las que podemos destacar, en general, las siguientes características (Kaufman *et al.*, 1995; Kaufman *et al.*, 1999a):

1. Proximidad en sus genes.
2. No se encuentran todos los genes contenidos en la región MHC.
3. Sólo se encuentra el gen C4 del complemento, representando la región de clase III.
4. Posee intrones más cortos (200 pb).

El complejo B (locus B)

El locus B está situado en el micro cromosoma 16 y codifica para los antígenos de clase I (B-F) y de clase II (B-L), éstos van a ser muy polimórficos y su patrón de expresión varía según el tipo celular (Figura 1). Las moléculas de clase I, al igual que en mamíferos, tienen una distribución muy amplia y las moléculas de clase II tienen una distribución mucho más restringida (Crone y Simonsen, 1987; Guillemot *et al.*, 1988).

Por otro lado, los genes de clase I y clase II son virtualmente inseparables por recombinación (Simonsen *et al.*, 1982). Esta situación puede estar provocada por el mínimo espacio existente entre los genes de estas 2 clases. Parece ser que el bajo nivel de recombinación en la región B-F/B-L provoca que evolucionen como un grupo alélico, dando lugar a diferentes

haplotipos que son relativamente estables en la evolución.

Por último, decir que además de B-F y B-L, también encontramos otra familia multigénica altamente polimórfica denominada B-G que va a presentar un ligamiento muy fuerte con el MHC (Figura 1).

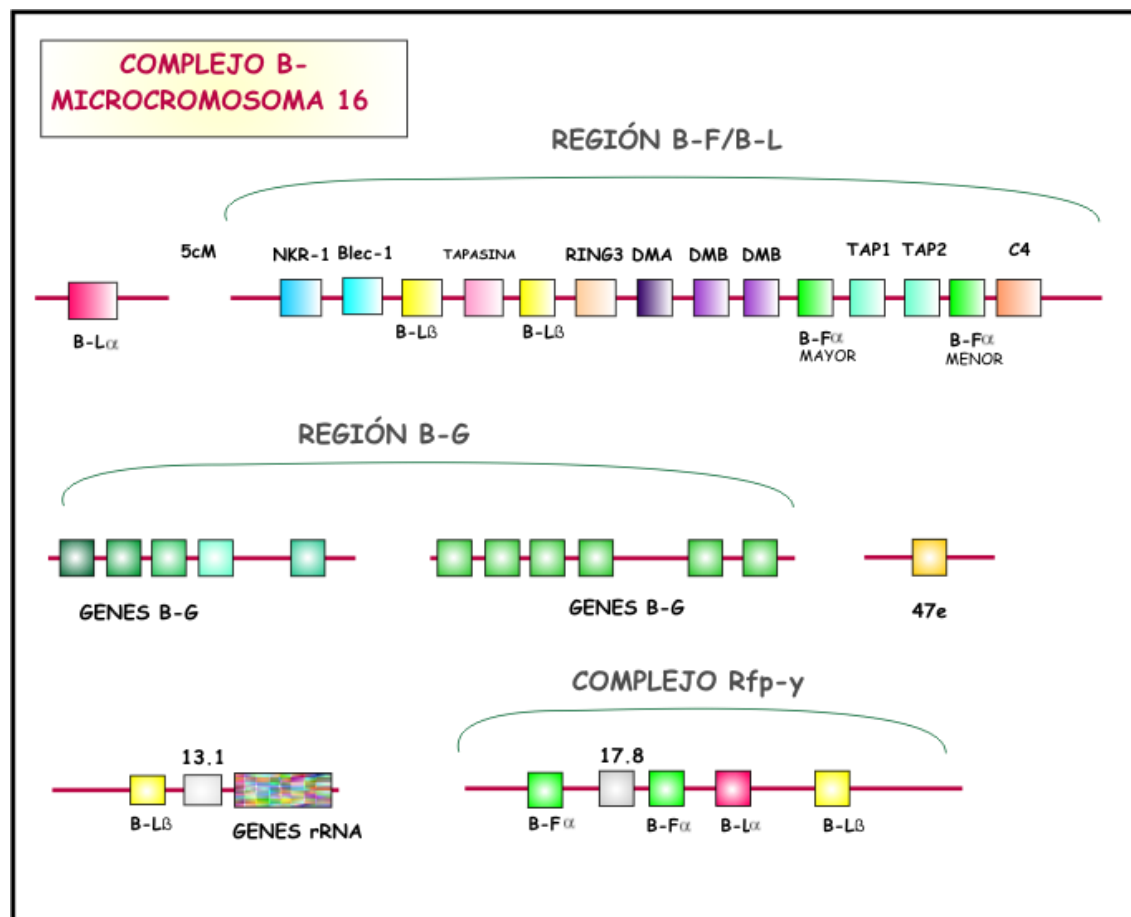


Figura 1. Mapa genético del Sistema Principal de Histocompatibilidad del pollo.

Región B-F/B-L

Como características principales de esta región podemos destacar:

- Es simple y compacta, con 19 genes presentes en 92 Kb.
- En la mayoría de haplotipos sólo se expresa a gran nivel una proteína de clase I presente en células sanguíneas y bazo denominada **B-F principal**. Por otra parte también se ha definido otra molécula de clase I cuya expresión es más débil y que se denomina **B-F menor** (Figura 1).
- Los 2 genes codificantes para el transportador asociado con la presentación antigénica (TAP) están localizados entre 2 genes de clase I.

Esta característica es diferente en pollos que en mamíferos ya que en éstos últimos

aparecen en la región de clase II. Por otra parte, el gen para tapasina aparece entre 2 genes de clase II (B-L) mientras que en humanos su localización es diferente ya que aparece a 100 Kb desde el centrómero respecto al MHC.

- En el cluster también aparece un gen para la proteína G (en mamíferos no aparece ligada al MHC) y un gen para B-G, proteína cuya expresión se limita a las células B (Guillemot *et al.*, 1989a; Kaufman *et al.*, 1991) El único gen para la cadena α de clase II se localiza a 5 cM del cluster I. Esta situación es muy diferente a la que se da en mamíferos donde los genes codificantes para las 2 cadenas de la molécula MHC de clase II aparecen estrechamente ligados.
- Secuenciando el cluster I, se ha visto que la porción central de la secuencia (desde los genes β para clase II (B-L) hasta los genes α de clase I (B-F) es muy compacta, con 11 genes en 39 Kb separados por unas decenas de nucleótidos solamente (Kaufman *et al.*, 1999a) y con intrones de 100 nucleótidos, dando lugar a genes con un 1/3 de tamaño respecto a los de mamíferos. Este fenómeno se puede comprobar al secuenciar el gen codificante para la cadena β de las moléculas B-L en el que se observa una estructura exón-intrón idéntica a la de las cadenas β de clase II de mamíferos pero con unos intrones que tienen un tamaño medio de sólo 100 pb y que dan lugar a genes unas 10 veces más pequeño. Solamente se ha encontrado un gen perteneciente a la región de clase III, que sería el componente del complemento C4 aunque es factible que aparezcan en un futuro más genes de esta clase cerca de la región secuenciada.

Estos loci B-F y B-L se parecen en parte a los genes no clásicos por las siguientes características (Pharr *et al.*, 1996):

- Polimorfismo muy bajo.
- Divergencia de secuencia relativamente pequeña.
- Nivel de expresión bastante reducido.

Estructura proteica de los genes de clase I (B-F) y de clase II (B-L)

Los dominios extracelulares proximales (respecto a la membrana) de B-L ($\alpha 2$ y $\beta 2$) y de B-F ($\alpha 3$ y $\beta 2m$) (Figura 2) son homólogos a los dominios constantes de las inmunoglobulinas. Esta homología no se presenta en los dominios más distales ($\alpha 1$ y $\alpha 2$ en B-F y $\alpha 1$ y $\beta 1$ en B-L), caracterizados por ser polimórficos y ser capaces de unir el péptido (Guillemot *et al.*, 1989b).

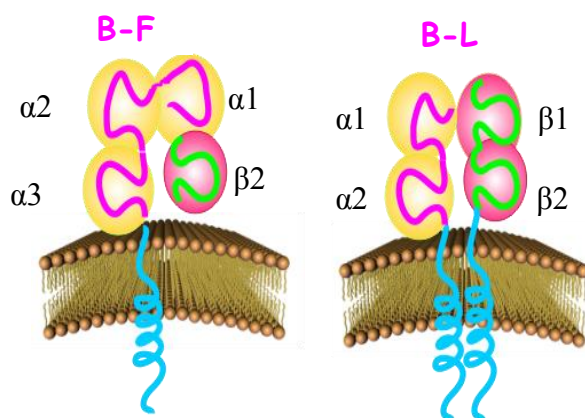


Figura 2. Representación esquemática de B-F (izquierda) y B-L (derecha). En azul los motivos carbohidrato. Nótese la homología entre la molécula B-F y la molécula HLA de clase I (figura 5) y B-L y la molécula HLA de clase II (figura 7).

Función y estructura proteica de las moléculas B-G

Las moléculas B-G son sintetizadas por trombocitos, eritrocitos y también pueden ser encontradas en linfocitos y en ciertas células tímicas. Estas moléculas van a presentar un gran polimorfismo.

En la actualidad, todavía no se han encontrado en mamíferos moléculas equivalentes a las codificadas por la región B-G. No obstante, estas proteínas presentan cierta semejanza con una proteína denominada glucoproteína oligodendrocitaria de la mielina (MOG), cuyo gen se localiza en la región de clase I del sistema HLA.

La función de estas moléculas no está muy clara y todos los estudios dedicados a asignar un papel para estas moléculas se han limitado a experimentos enfocados a su expresión en eritrocitos. En este contexto se ha visto que los antígenos B-G tienen cierta actividad adyuvante en la respuesta humoral contra las moléculas de clase I del locus B (Hala *et al.*, 1981).

Las moléculas B-G tienen un peso molecular que varía de 30 a 70 Kd, esta gran variación en el tamaño de sus cadenas se debe a variaciones en la longitud de su tallo citoplasmático.

Se han llevado a cabo estudios con técnicas electroforéticas que han demostrado que las moléculas B-G en su estado nativo se disponen formando dímeros unidos por puentes disulfuro y que los 2 polipéptidos se enrollan en su porción citoplasmática para dar lugar a un tallo en forma de α hélice fibroso y estable.

Locus Rfp-Y

También denominado locus Y, va a corresponder a los clusters II/IV y III (Guillemot *et al.*, 1988). Al contrario que el locus B no es capaz de producir respuestas potentes de rechazo de alotrasplantes, de reacción mixta de linfocitos y de resistencia a enfermedades (Juul-Madsen *et al.*, 1997).

Esta región contiene al menos 2 loci para la cadena α de clase I y 2 loci para la cadena β de clase II junto a un gen codificante para una lectina tipo C (17.5) y un gen (17.8) cuya función es desconocida (Miller *et al.*, 1994).

Susceptibilidad a enfermedades del pollo

En los apartados anteriores se han esbozado las características del MHC en una especie tan importante para el hombre como es el pollo. Parece que presenta algunas similitudes con el MHC de mamíferos aunque otros rasgos de este sistema solamente aparecen en el pollo y lo diferencian en gran medida de todos los MHC estudiados hasta la actualidad. Estas características se resumen en:

- Es más pequeño y simple que el MHC de mamíferos. La mayor parte de los haplotipos MHC sólo expresan predominantemente una molécula de clase I y una de clase II.
- La consecuencia de poseer un MHC más simple va a ser que la resistencia y la susceptibilidad a un agente infeccioso en particular está determinada por la especificidad de unión al péptido de los alelos MHC presentes en haplotipos concretos (hipótesis discutible, pero mantenida por el principal experto (Kaufman *et al.*, 1999a)).

La situación en el segundo punto anteriormente citado, contrasta claramente con lo que ocurre en mamíferos (fundamentalmente en humanos).

En *G. gallus* (pollo) existen haplotipos MHC que van a determinar claramente resistencia o susceptibilidad a enfermedades provocadas por patógenos virales bastante comunes y de gran importancia económica (Plachy *et al.*, 1992a; Plachy *et al.*, 1992b; Schat, 1987), mientras que en mamíferos, las asociaciones importantes MHC/enfermedad suelen ocurrir con enfermedades autoinmunes o con defectos bioquímicos, de manera que es muy difícil encontrar asociaciones claras entre el MHC y enfermedades infecciosas.

Kaufman y colaboradores (Kaufman *et al.*, 1995) intentaron explicar este fenómeno aduciendo que la gran complejidad del MHC de mamíferos, con familias multigénicas codificantes para moléculas de clase I y II y cuyos componentes presentan un elevado nivel de polimorfismo, provoca que los haplotipos MHC confieran resistencia a mayor o menor nivel para la mayoría de los patógenos. Esto significa, que los patógenos ejercen una selección relativamente débil para el mantenimiento del polimorfismo y que, según estos autores, las teorías anteriormente mencionadas para explicar la presencia y el mantenimiento del elevado número de alelos presentes en el MHC de mamíferos que reivindicaban el papel

fundamental de los patógenos en el mantenimiento del polimorfismo (selección sobredominante y selección dependiente de frecuencia) no serían válidas, ya que estos autores defienden que la selección ejercida por los virus y bacterias es muy débil y que el elevado polimorfismo presente en el MHC de humanos provoca que siempre exista una molécula MHC capaz de presentar los antígenos procedentes de los patógenos a los que se encuentra sometida una población particular. Por último, decir que Kaufman y colaboradores (Kaufman *et al.*, 1995) postularon la existencia de otros mecanismos generadores de variabilidad en la naturaleza diferentes a los implicados en la selección de balance (selección sobredominante y selección dependiente de frecuencia) que explicarían el mantenimiento del polimorfismo en humanos. En resumen, en el pollo, la selección provocada por el patógeno depende de la especificidad para el péptido de una sola molécula, al contrario que en mamíferos en los que dependería de muchas moléculas MHC.

Hipótesis del MHC "mínimo-esencial"

Como ya se ha comentado, las características únicas detectadas en el MHC de pollo (pocos genes, distancias intergénicas muy cortas, intrones pequeños) llevaron a Kaufman y col. a proponer la provocativa hipótesis del MHC mínimo-esencial (Kaufman *et al.*, 1995) por la cual se cree que este sistema contiene sólo el mínimo número de genes absolutamente esenciales para la función del MHC en un sentido evolutivo; como consecuencia de este hecho se cree que cada gen MHC presente en el pollo está expuesto a una intensidad selectiva mucho más potente que en mamíferos. Además, este MHC tan atípico se propuso como general en toda la clase Aves, generalización que más tarde, con la llegada de datos nuevos sobre el sistema MHC en otros órdenes diferentes, se comprobó que no era correcta (Arnaiz-Villena *et al.*, 2010b; Kaufman *et al.*, 1995; Kaufman *et al.*, 1999a)

2.1.2 Sistema Principal de Histocompatibilidad en humanos (HLA)

La región HLA

Los genes de histocompatibilidad humanos se encuentran ubicados en el brazo corto del cromosoma 6, en una región que se denomina HLA y contiene genes que no están directamente relacionados con la presentación antigénica ni, en ocasiones, con la función inmunológica.

El primer mapa genético de esta región se completó en 1993 (Campbell y Trowsdale, 1993) Unos años más tarde se elaboró otro mapa a partir de secuencias parciales obtenidas de individuos diferentes, quedando así representado un haplotipo virtual según el cual la región

HLA ocupa 3.6 Mb y contiene 224 genes de los que se expresan algo más de la mitad, unos 128 genes (The MHC sequencing consortium, 1999).

Este mapa fue posteriormente actualizado con una secuencia completa obtenida a partir de una línea celular denominada PGF, que es consanguínea y homocigótica para todos los genes del sistema (Stewart *et al.*, 2004).

Por otro lado, la secuenciación completa del cromosoma 6 reveló la existencia de nuevos genes relacionados con el MHC más allá de las 3.6 Mb descritas inicialmente, por lo que se elaboró un nuevo mapa que recibió el nombre de MHC extendido (xMHC, del inglés *extended MHC*) (Mungall *et al.*, 2003) para diferenciarlo del mapa de 1999/2004, que ahora es aludido como mapa clásico (Horton *et al.*, 2004).

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 (Figura 3).

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 4). 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) (Figura 4).

Los genes de las histonas codifican proteínas que participan en el plegamiento y estabilización de la molécula de ADN formando parte del nucleosoma. La mayor concentración de estos genes 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.

Los genes de ARN transferente (ARNt) son muy cortos (75-90 nucleótidos) y codifican moléculas de ARN implicadas en la síntesis proteica, ya que reconocen específicamente los tripletes del ARN mensajero (ARNm) y aportan el aminoácido correspondiente. Los genes transportadores de solutos (SLC, *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 cuyo papel fisiológico se desconoce.

Los genes de las butirofilinas (BTN), presentes mayoritariamente en la región de clase I extendida, 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 B-G del Sistema Principal de Histocompatibilidad del pollo (Olivo *et al.*, 1996). Codifican proteínas de la superfamilia de las inmunoglobulinas y tienen una función desconocida

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.

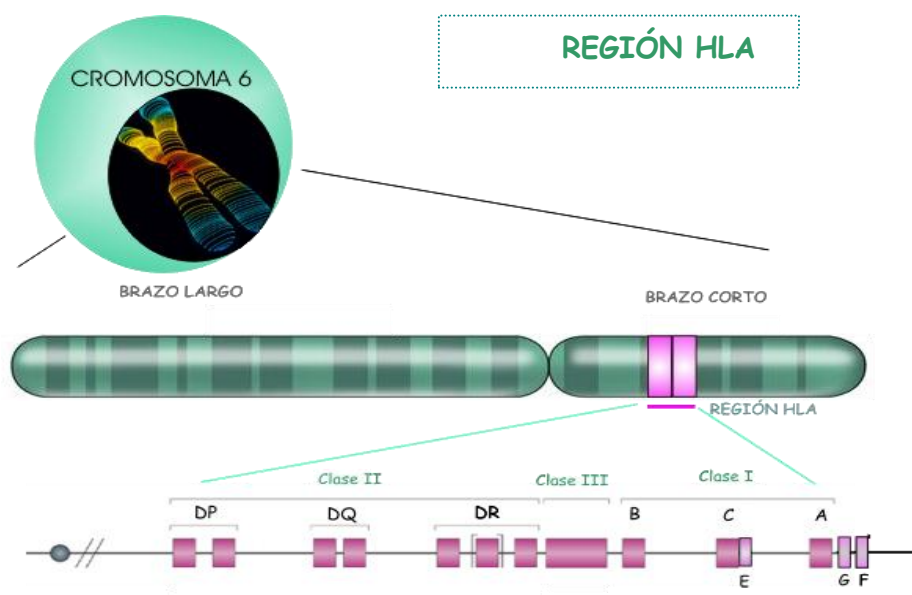


Figura 3. Mapa clásico HLA

Los genes *zinc-finger* codifican proteínas muy diversas que tienen en común su capacidad para unir iones de zinc. 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).

Los genes HLA-I se caracterizan por expresar proteínas estructuralmente muy similares, aunque diferentes 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 y HLA-G, también conocidos como HLA-Ib), genes tipo HLA-I (MIC-A y MIC-B) y gen de la hemocromatosis (HFE). Excepto éste último, que se encuentra

en el extremo telomérico de la región de clase I extendida, muy alejado del resto, todos están localizados en la región de clase I clásica, en la que también existen 17 pseudogenes.

Los genes HLA-Ia son muy polimórficos y expresan proteínas implicadas en el reconocimiento de péptidos antigénicos de origen intracelular (derivados 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 (*natural killer*). Se expresan en todas las células nucleadas del organismo. El patrón de expresión de los genes HLA-Ib es mucho más reducido y la función exacta de sus productos proteicos no está muy clara. Su polimorfismo es mucho menor y podrían estar relacionados con fenómenos de tolerancia inmunológica. Los genes MIC expresan proteínas que son reconocidas por los receptores activadores NKG2D de las células NK y que pueden tener una función inmunológica asociada a la mucosa intestinal. Por su parte, el gen HFE está implicado en el transporte de hierro. Los genes de los factores de necrosis tumoral (TNF) codifican 3 tipos de citoquinas (TNF, LTA y LTB) que participan en procesos inflamatorios y se encuentran situados en la región de clase III.

Los genes del antígeno 6 linfocitario (LY6), que codifican proteínas de membrana asociadas a glicosil-fosfatidil-inositol que pueden tener una función inmunológica, son unos genes muy conservados presentes también en el ratón y están localizados en la región de clase III.

Los genes de las proteínas de choque térmico (HSP, *heat shock proteins*) aumentan su expresión en situaciones de estrés celular colaborando en los procesos de síntesis, plegamiento, ensamblaje, transporte y degradación de proteínas. Concretamente este grupo que se encuentra en la región de clase III se expresa activamente cuando el sistema inmune es activado para eliminar células dañadas, infectadas o tumorales. Algunos genes de complemento de la vía clásica (C2, C4A, C4B) y de la vía alternativa (Bf) también están presentes en la región de clase III.

Los genes HLA-II están ubicados en la región de clase II y dan lugar proteínas polimórficas implicadas en la presentación antigénica que se denominan HLA-DM, HLA-DO, HLA-DP, HLA-DQ y HLA-DR. Se han descrito numerosos genes funcionales y también pseudogenes. Se expresan en células especializadas en la presentación antigénica (denominadas APC, *antigen presenting cells*), fundamentalmente células dendríticas (DC, *dendritic cells*), macrófagos y linfocitos B. Su función es unir péptidos de origen extracelular (generalmente procedentes de bacterias) y presentarlos a linfocitos T cooperadores (Bembridge *et al.*, 1995) para poner en marcha una respuesta inmunológica de tipo humoral, mediada por anticuerpos.

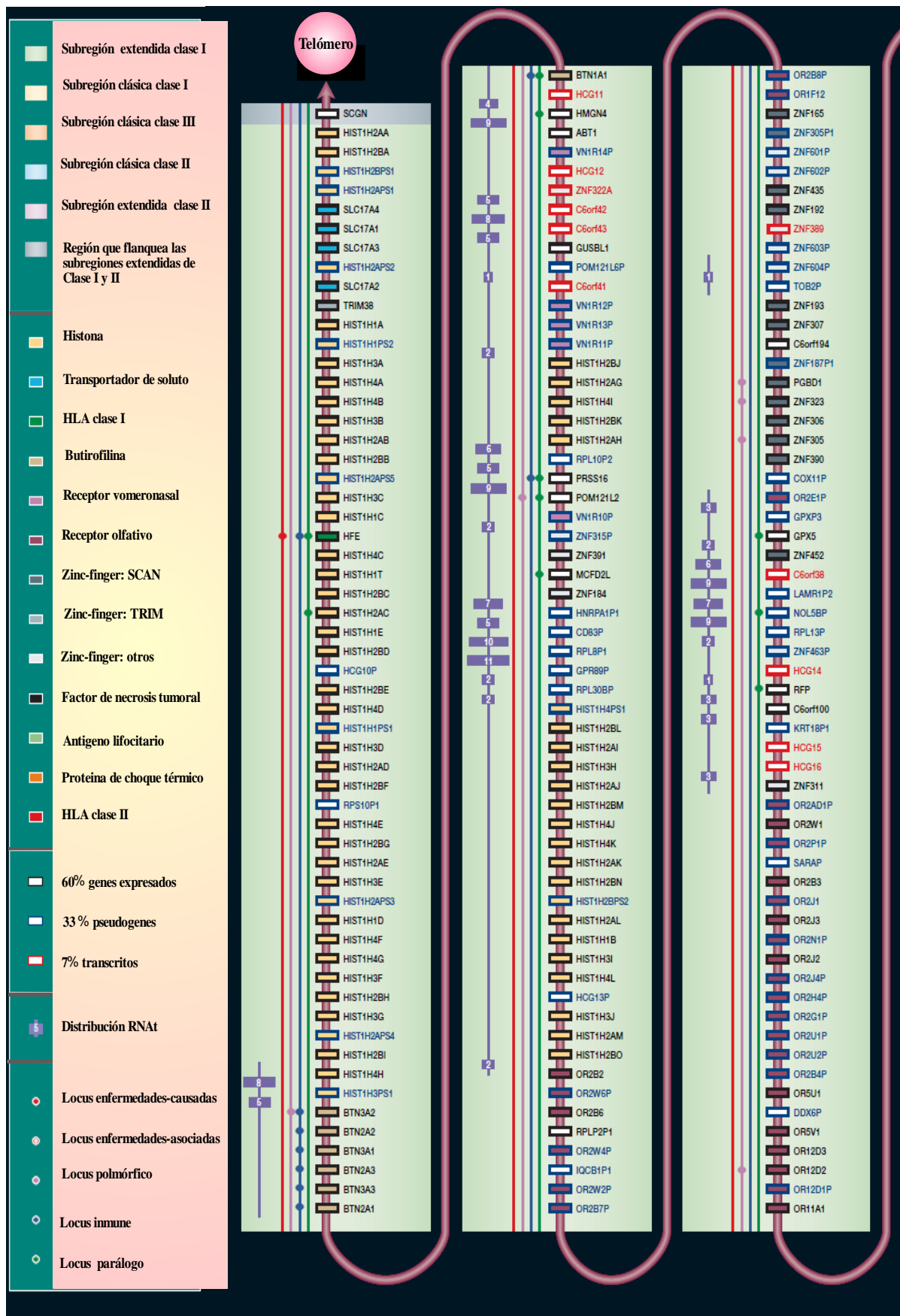
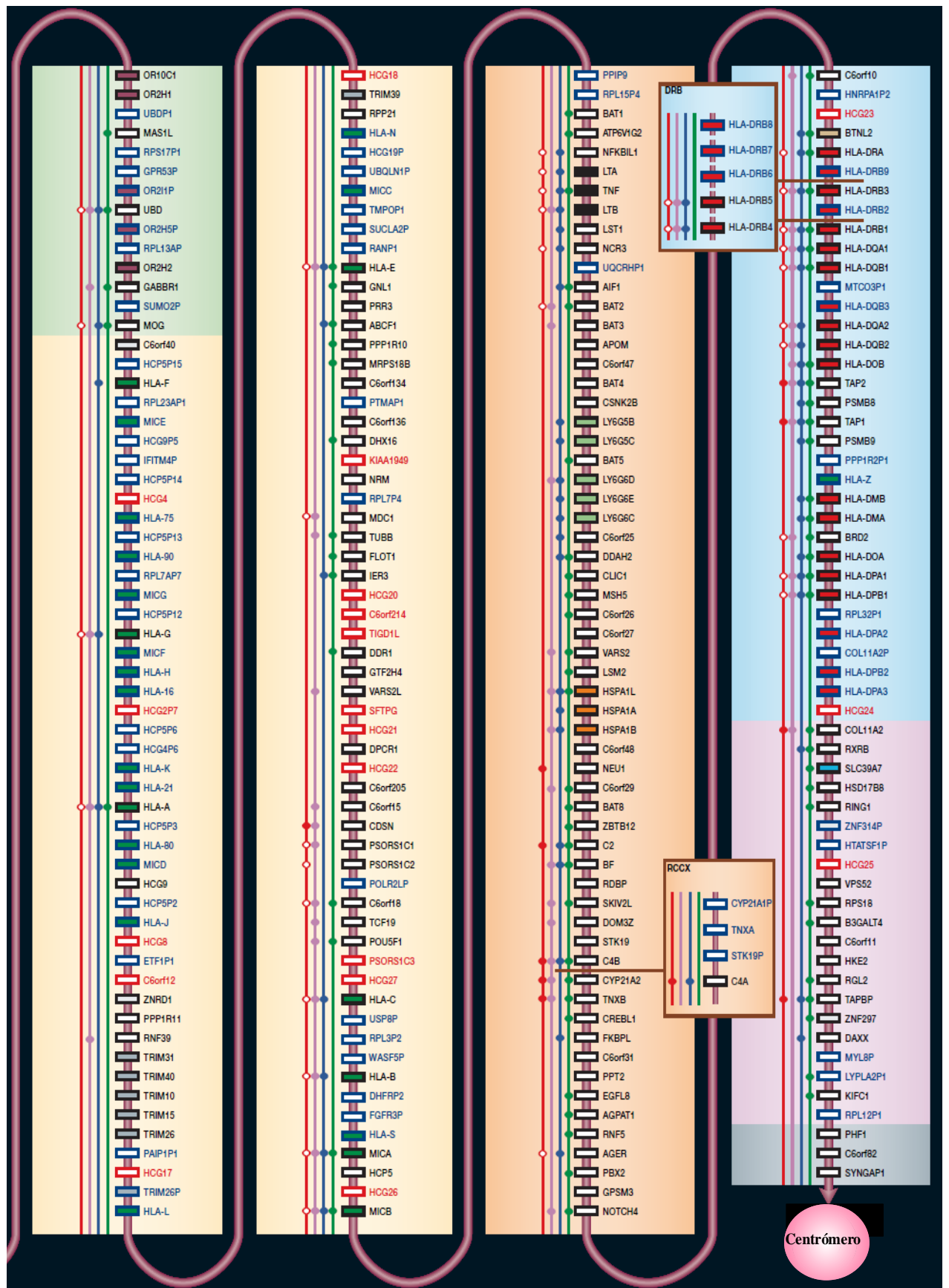


Figura 4 . Mapa de la región HLA extendida (Horton *et al.*, 2004).



Cont. Figura 4 . Mapa de la región HLA extendida (Horton *et al.*, 2004).

Una característica fundamental de la región HLA es que los genes que contiene, están muy próximos entre sí, generando desequilibrios de ligamiento muy acentuados (la recombinación entre ellos es altamente improbable) que provocan la existencia de haplotipos (combinaciones de alelos concretos situados en loci contiguos) que se transmiten como unidades genéticas indestructibles a lo largo de muchas generaciones.

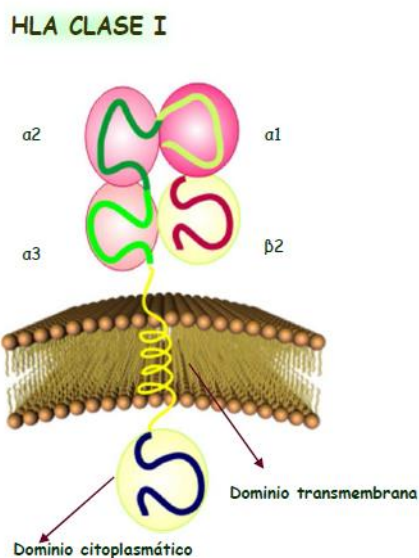
Las moléculas HLA de clase I

Normalmente, al hablar de moléculas de clase I se está haciendo alusión a las 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.

Centrando el tema en las 6 primeras, se trata de glicoproteínas de membrana resultantes de la asociación entre un polipéptido denominado cadena alfa y la proteína beta-2 microglobulina (Tysoe-Calnon *et al.*, 1991). Las cadenas alfa están codificadas en la región de clase I del sistema HLA (brazo corto del cromosoma 6), mientras que la $\beta 2m$ está codificada en el cromosoma 15. En la cadena alfa se distinguen 3 regiones: una extracelular formada por 3 dominios (alfa 1, alfa 2, alfa 3), otra transmembrana y una tercera citoplásmica. La $\beta 2m$ está acoplada al dominio alfa 3 mediante una unión no covalente (formando una subestructura que recuerda mucho a la de las inmunoglobulinas) y todo el complejo proteico está anclado a la membrana plasmática, de manera que el extremo amino terminal (NH_3^+) de la cadena alfa queda en el exterior celular y el extremo carboxilo terminal (Hughes, 2003) en el citoplasma celular (Figura 5).

La gran similitud estructural entre las distintas proteínas HLA de clase I se debe a que la estructura de los genes en los que están codificadas sus respectivas cadenas alfa (formados por 8 exones) es también muy parecida. Las únicas diferencias remarcables afectan a la longitud de la cola citoplásmica de las moléculas, lo cual puede tener implicaciones funcionales. La molécula HLA-G prácticamente no tiene cola citoplásmica a causa de un codón de terminación al comienzo del exón 6, mientras que HLA-B y HLA-E tienen una región citoplásmica más corta debido a la presencia de un codón de terminación al final del exón 7. El gen HLA-F, por su parte, transcribe pero no traduce el exón 7 al tener alterado el sitio de *splicing* en el extremo 3' del intrón 6, por lo que la proteína resultante también presenta una cola citoplásmica más corta. La estructura de las moléculas HLA de clase I fue descrita por primera vez a partir del antígeno HLA-A2, cristalizado y analizado por difracción de rayos X (Bjorkman *et al.*, 1987a). El resultado de este análisis reveló la existencia de dos dominios tipo inmunoglobulina próximos a la membrana (alfa 3 y $\beta 2m$) y

otros dos dominios más alejados (alfa 1 y alfa 2) asociados a un pequeño péptido no identificado ajeno a la molécula de histocompatibilidad



Esos dos dominios distales forman una especie de concha (comúnmente denominada ‘valva’), cuyo fondo está constituido por 8 láminas beta (4 de cada dominio) y limitado por 2 hélices alfa (1 de cada dominio), y son capaces de alojar un pequeño péptido (8-10 aminoácidos). La función de estas moléculas no depende sólo de su estructura, sino también de su polimorfismo.

Figura 5. Representación esquemática de la molécula de MHC de clase I

Las moléculas no clásicas tienen poca variabilidad, no presentan antígenos a los linfocitos T, sino que se unen a receptores inhibidores de las células NK, por este motivo, por ejemplo, las proteínas HLA-G se denominan inmunosupresoras y se expresan en el citotrofoblasto del feto. Se piensa que esta expresión evita que el feto sea rechazado como un trasplante.

Las moléculas clásicas, son muy polimórficas y su capacidad para unir péptidos está relacionada con la presentación antigénica, que es consecuencia de la interacción específica entre las moléculas HLA asociadas a un péptido y el receptor de los linfocitos T citotóxicos (TCR-CD8). La capacidad que tienen estas moléculas para presentar cualquier péptido radica en su elevado polimorfismo. De la comparación de numerosas secuencias de histocompatibilidad de clase I clásicas se determinó que la mayor variabilidad se concentra precisamente 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) constituido por aminoácidos cuyas cadenas laterales son accesibles y están implicadas en la unión del péptido y en la interacción con el receptor de la célula T (TCR)(Bjorkman *et al.*, 1987b). En esta comparación se observó también que la mayor variabilidad se concentra en la hélice alfa del dominio alfa 1 y en las láminas beta del dominio alfa 2 (Parham *et al.*, 1988) (Figura 6).

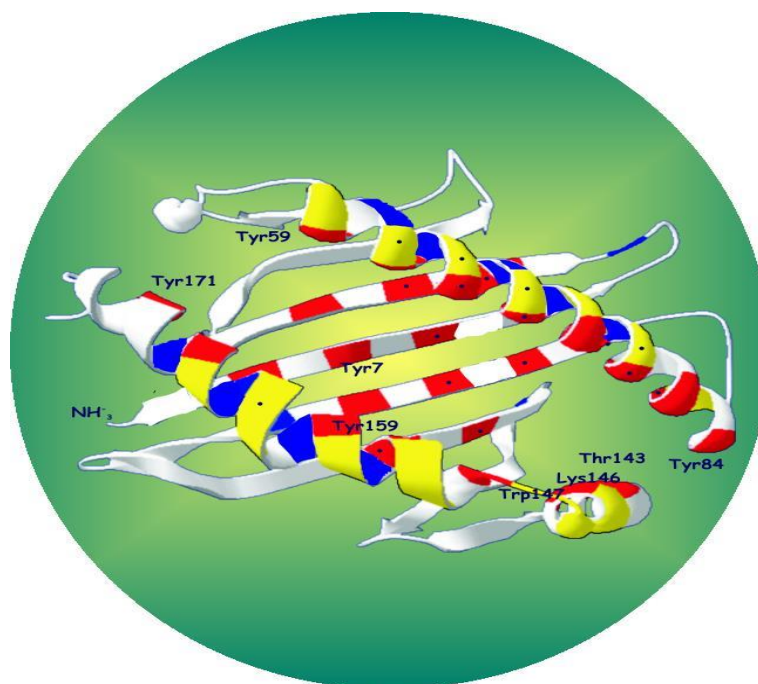


Figura 6. Modelo tridimensional de los dominios alfa 1 y alfa 2 de una molécula HLA de clase I. Los aminoácidos que constituyen el sitio de reconocimiento antigénico (ARS) aparecen coloreados, y sus cadenas laterales están dirigidas hacia el interior de la valva (en rojo, participan en la unión del péptido), 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.*, 1987b). Los puntos señalan las posiciones hipervariables descritas en las moléculas HLA de clase I clásicas (Parham *et al.*, 1988). También se indican los aminoácidos que dentro del ARS se mantienen invariantes en las secuencias humanas y que son responsables del anclaje del péptido a través de sus extremos (Bjorkman *et al.*, 1987b).

Las moléculas HLA de clase II

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. Cada cadena tiene dos dominios extracelulares (alfa 1 y alfa 2 la cadena alfa; beta 1 y beta 2 la cadena beta), un dominio transmembrana y una región citoplásmica. Los dominios extracelulares alfa 1 y beta 1 contienen los extremos amino terminales (NH_3^+), mientras que los extremos carboxilo terminales (Hughes, 2003) se encuentran en la porción citoplásmica de cada cadena (Figura 7). Nuevamente, los dominios extracelulares más cercanos a la membrana (alfa 2 y beta 2) forman una subestructura similar a la de las inmunoglobulinas.

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 (más de 13 aminoácidos).

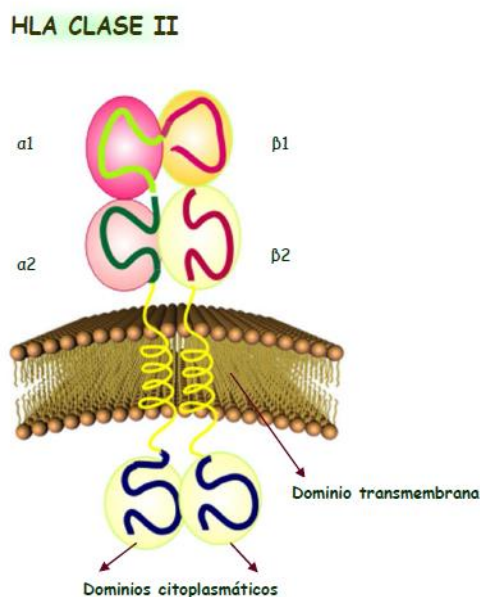


Figura 7. Representación esquemática de la molécula de MHC de clase II.

En concreto hay identificados: 3 genes DPA (denominados DPA1, DPA2, DPA3) y 2 genes DPB (DPB1, DPB2) que codifican cadenas alfa y beta de las proteínas HLA-DP; 1 gen DRA y 9 genes DRB (DRB1, DRB2, ..., DRB9) que codifican cadenas alfa y beta de las proteínas HLA-DR; 2 genes DQA (DQA1, DQA2) y 3 genes DQB (DQB1, DQB2, DQB3) que codifican cadenas alfa y beta de las proteínas HLA-DQ. Es destacable que muchas de esas copias extra no son funcionales. De hecho solo se expresan DPA1, DPB1, DRB1, DRB3, DRB4, DRB5, DQA1, DQA2, DQB1, DQB2 y por supuesto los genes que aparecen en copia única (DMA, DMB, DOA, DOB, DRA) (Horton *et al*, 2004). El fenómeno más notable en esta región afecta a los genes HLA-DRB y se denomina polimorfismo de dosis génica. Significa que no todos los individuos poseen todas las copias descritas de genes codificantes de cadenas beta de las moléculas HLA-DR. Si bien todos los individuos poseen el gen HLA-DRB1 (y por supuesto el gen HLA-DRA, que es de copia única), la presencia o no de los otros depende del alelo expresado por el gen HLA-DRB1.

En definitiva, la presencia de las copias extra del gen HLA-DRB depende del haplotipo del individuo.

El polimorfismo también se concentra en el sitio de reconocimiento antigénico (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 (TCRCD4) para activar la respuesta inmune.

Región de clase III

Situada entre la región de clase II y la región de clase I, contiene un gen cada 15 kb lo que provoca que en esta región los genes estén más próximos que en otras zonas del genoma (Kendall *et col.* 1990). Contiene los genes de algunos de los componentes del complemento como C4A y C4B, el factor B (Bf) y C2. Muy cerca de los genes C4, encontramos el gen codificante para la 21-hidroxilasa (enzima involucrada en la síntesis de esteroides).

Se ha descubierto en la región de clase III, un grupo de genes (TNF- α , TNF- β , familia MIC, HSP70 etc...) supuestamente involucrados en la inflamación, estrés o infección. Algunos autores postulan que este grupo de genes es lo suficientemente distinto como para considerarlos una región independiente que ha de ser denominada región de clase IV.

La región de clase III tiene varias características que la diferencian del resto de las regiones localizadas en el MHC. De hecho, al contrario que en la región de clase I y de clase II, existen en ella muchos genes con funciones que no tienen nada que ver con la presentación antigénica o el procesamiento antigénico; además, se ha podido comprobar que esta región tiene un contenido G+C muy diferente al de la región de clase II. Por todo esto se cree que la región de clase III podría haberse originado por una inserción en bloque de material genético entre la región de clase II y la región de clase I. Esta hipótesis se ve reforzada por la existencia de largos bloques de ADN repetitivo (*Alu* etc) entre la región de clase II y la de clase I.

2.2 Modelos silvestres (aves de canto)

El estudio de los genes de histocompatibilidad en las aves de canto es interesante a la hora de comparar el polimorfismo MHC de los modelos naturales, adquirido tras millones de años de evolución, con el polimorfismo descrito en modelos considerados “artificiales” (humanos, ratones, pollos), así como de otras especies de vertebrados.

La facilidad en la identificación de especies y la comodidad del manejo de muestras hace de las aves de canto un buen modelo natural de estudio.

Además, la amplia distribución biogeográfica a nivel mundial de estas aves y el gran número de especies que las representan, permiten realizar un estudio evolutivo de los genes MHC y un análisis de los diferentes procesos selectivos-ambientales (diversos patógenos) que inducen su variación.

3. Polimorfismo y evolución de los genes MHC

Como ya se ha comentado, el MHC es el sistema genético más polimórfico descrito hasta ahora en vertebrados, y en humanos es el que registra mayor variabilidad de todo el genoma (Tabla 3). Parece lógico pensar que esa variabilidad deriva de la función presentadora de antígenos de las moléculas expresadas por estos genes, que deben tener la capacidad de reconocer un repertorio de antígenos lo más amplio posible (Zinkernagel y Doherty, 1979). La necesidad de tanta variabilidad empezó a ser cuestionada cuando empezaron a conocerse sistemas menos polimórficos en diversos mamíferos (Klein, 1986). El elevado polimorfismo sería por tanto un lastre evolutivo resultante de fenómenos excepcionales ocurridos en el pasado como cuellos de botella que habrían dado lugar a un grado de consanguinidad elevado.

HLA Clase I		HLA Clase II	
GEN	ALELOS	GEN	ALELOS
A	2735	DRA	7
B	3459	DRB	1569
C	2259	DQA1	52
E	15	DQB1	594
F	22	DPA1	38
G	50	DPB1	344
		DMA	7
		DMB	13
		DOA	12
		DOB	13
TOTAL	8576	TOTAL	2649

Tabla 3. Resumen de número de alelos encontrados de moléculas de MHC de clase I y II.

<http://www.ebi.ac.uk/ipd/>

3.1 Mecanismos generadores de polimorfismo: la conversión génica

Como norma general se asume que los genes evolucionan al acumular mutaciones en posiciones concretas del ADN. Sin embargo, en el caso del MHC este fenómeno no es capaz de explicar por sí solo la enorme diversidad alélica encontrada, ya que además existen evidencias de que se produce recombinación entre alelos de un mismo locus e incluso entre alelos situados en diferentes loci. Este fenómeno, denominado conversión génica, provoca el intercambio de pequeños fragmentos de ADN entre unos alelos y otros, y afecta sobre todo a los exones implicados en el reconocimiento del antígeno.

Las primeras evidencias se obtuvieron al comparar entre sí secuencias de distintos alelos. En

un estudio en el que se comparaban alelos HLA de clase I se comprobó que en algunos casos dos alelos determinados eran idénticos entre sí salvo por un pequeño segmento (de no más de 35 nucleótidos) que parecía pertenecer a un tercer alelo (Parham *et al.*, 1995).

Otros estudios han revelado que numerosas poblaciones nativas de Sudamérica presentan un elevado porcentaje de alelos de nueva creación que no se encuentran en ninguna otra parte del mundo (Arnaiz-Villena *et al.*, 2000; Arnaiz-Villena *et al.*, 2005; Arnaiz-Villena *et al.*, 2007d; Arnaiz-Villena *et al.*, 2007e; Arnaiz-Villena *et al.*, 2009c; Arnaiz-Villena *et al.*, 2009d; Arnaiz-Villena *et al.*, 2010a; Arnaiz-Villena *et al.*, 2011; Arnaiz-Villena *et al.*, 2013; Arnaiz-Villena *et al.*, 2014b; Martínez-Laso *et al.*, 2011; Parga-Lozano *et al.*, 2011; Rey *et al.*, 2013; Silvera *et al.*, 2011; Vargas-Alarcon *et al.*, 2010; Vargas-Alarcon *et al.*, 2011) y se descarta que se hayan originado sólo por mutaciones puntuales ya que no han tenido tiempo, al estimarse que dichas poblaciones no tienen una antigüedad superior a 20.000 años (unas 1.000 generaciones), momento en el que se supone se separaron de su grupo original (tal vez asiático) para introducirse en el continente americano.

Como estos hay numerosos estudios que corroboran que el principal mecanismo generador de variabilidad es la conversión génica.

3.2 Mantenimiento del polimorfismo

Una vez conocidos los mecanismos que generan la variabilidad en los genes MHC es importante analizar qué otros mecanismos (en este caso de tipo selectivo) están actuando sobre ellos para mantener las variantes alélicas a lo largo del tiempo. Los principales agentes selectivos propuestos como responsables del mantenimiento de la variabilidad son los patógenos, que favorecen la permanencia de alelos en la población ejerciendo un tipo de selección denominada selección de balance (balancing selection).

Sin embargo se han sugerido otros mecanismos alternativos que podrían contribuir al mantenimiento del polimorfismo, como son las interacciones materno-fetales (al menos en mamíferos) y los emparejamientos no al azar (*negative-assortative mating*).

A. Selección de balance

La selección de balance fue propuesta inicialmente como una estrategia evolutiva encaminada a mantener las distintas variantes alélicas de genes polimórficos en una población con el fin de incrementar sus opciones de supervivencia al favorecer el predominio de individuos heterocigotos, supuestamente más aptos que los homocigotos (Hedrick, 2007; Lewontin y Hubby, 1966). En el caso de los genes MHC, la abundancia de alelos

distintos en una población garantizaría la abundancia de individuos heterocigotos que expresarían más proteínas de histocompatibilidad diferentes, que serán capaces de presentar una mayor variedad de antígenos.

Se han propuesto 3 modelos de selección de balance:

Selección dependiente de frecuencia

(Bodmer, 1972;Kojima, 1971): esta teoría (denominada originalmente *frequency-dependent selection*) defiende la importancia evolutiva de los llamados alelos raros. En una población en la que existe una variedad de alelos, cada alelo está presente con una frecuencia determinada. Teóricamente, los alelos más frecuentes son los que más posibilidades de supervivencia aportan a los individuos que los poseen y por ello han ido incrementando (o manteniendo) su frecuencia generación tras generación. Por el contrario, los alelos menos frecuentes lo son bien porque resultan deletéreos o bien porque no aportan una ventaja adicional. Estos alelos raros no llegan a desaparecer porque eventualmente pueden resultar ventajosos ante la presencia de un nuevo agente infeccioso que haya desarrollado resistencia frente a los alelos más frecuentes por un fenómeno de coevolución, de manera que empezarán a aumentar su frecuencia a la par que los otros serán cada vez menos frecuentes. Este proceso, repetido indefinidamente, favorecería la permanencia de alelos en la población, y el número de alelos iría aumentando a medida que se generaran nuevas variantes.

Ventaja de los heterocigotos

(Black y Salzano, 1981;Doherty y Zinkernagel, 1975): este modelo también denominado selección sobredominante (*overdominant selection*), sostiene que un heterocigoto es capaz de reconocer más antígenos que un homocigoto, por lo que los heterocigotos son seleccionados favorablemente. Sin embargo, la eficacia biológica de cada genotipo es difícil de establecer ya que no todos los posibles heterocigotos son igualmente ventajosos ni tienen por qué ser más aptos que cualquiera de los posibles homocigotos.

Selección variable en el tiempo y en el espacio

(Gillespie, 1978;Hedrick *et al.*, 1976): Según este modelo (*variable selection in time and space*), la presión ejercida por los patógenos no es constante a lo largo del tiempo ni está presente en todos los ambientes, de manera que los alelos que aporten resistencia o susceptibilidad a tales patógenos aumentarán o reducirán su frecuencia en función de si el patógeno está o no presente en cada momento y en cada lugar.

Es difícil determinar cuál de los tres modelos de selección de balance es más responsable del mantenimiento del polimorfismo y de la existencia de una proporción de individuos heterocigotos superior a la esperada en condiciones de neutralidad. La exposición variable de los individuos frente a diversos patógenos provoca que los alelos que garantizan mejor la supervivencia también varíen, de manera que a la larga los individuos más aptos suelen ser heterocigotos, a pesar de que ocasionalmente el grado de heterocigosidad pueda ser inferior a lo esperado. Por otra parte, la eventual ventaja que supone la presencia de alelos raros en una población garantiza la permanencia de dichos alelos durante mucho tiempo.

B. Emparejamientos no al azar

Como resultado de algunos estudios experimentales, se ha propuesto que antes de la selección de balance ya ha existido otra etapa selectiva que favorece el exceso de individuos heterocigotos en una población, y que se debe a emparejamientos no al azar (*negative assortative mating*). Se llegó a esta conclusión al observarse que en algunas especies los individuos que se emparejan presentan menos coincidencias de alelos MHC de las esperables por azar, por lo que se supuso que existía algún mecanismo que permitía seleccionar a las parejas que poseían alelos diferentes a los propios.

Este fenómeno se ha observado en humanos (Ober *et al.*, 1992), en ratones (Egid y Brown, 1989; Potts *et al.*, 1991; Yamazaki *et al.*, 1976) y en peces (Landry *et al.*, 2001; Reusch *et al.*, 2001) al contrastar el nivel de coincidencia de alelos MHC en cada pareja con el esperado por azar. Sin embargo, otros estudios realizados en humanos (Hedrick y Black, 1997) y en ovejas (Paterson y Pemberton, 1997) no han detectado este sesgo. En aves se han detectado emparejamientos no al azar en gorriones, observándose que los machos con pocos alelos tenían más dificultad para encontrar pareja, aunque por otro lado las parejas que no tenían ningún alelo en común eran menos frecuentes de lo esperado por azar. Una explicación podría ser que las parejas son seleccionadas para aumentar la diversidad alélica y al mismo tiempo para evitar la ruptura de combinaciones de alelos que se han coadaptado (Bonneaud *et al.*, 2006).

C. Interacción materno-fetal

Otro fenómeno selectivo que se ha propuesto resulta de la observación de que la incidencia de abortos espontáneos parece afectar más a parejas que comparten varias proteínas de histocompatibilidad (Alberts y Ober, 1993; Thomas *et al.*, 1985) y se concluye que es necesaria una respuesta inmunológica en la placenta para la correcta implantación del feto.

Estas interacciones materno-fetales, los emparejamientos no al azar y la selección ejercida por patógenos son los tres mecanismos selectivos a los que se atribuye la gran variabilidad observada en el sistema MHC y el predominio de individuos heterocigotos (Garrigan y Hedrick, 2003; Hedrick, 1998).

3.4 Evolución transespecífica

La evolución transespecífica se produce cuando especies evolutivamente relacionadas poseen variantes proteicas idénticas. Esto ocurre cuando una especie ancestral transmite sus alelos a las especies descendientes. Aunque no es muy habitual, pueden encontrarse alelos idénticos en especies que han divergido hace relativamente poco tiempo. Sin embargo, este fenómeno se observa en especies más distantes sólo cuando los genes implicados están sometidos a un proceso de selección de balance. Uno de los casos más representativos es el Sistema Principal de Histocompatibilidad de los vertebrados (MHC) donde se han identificado linajes de alelos que se separaron hace muchos millones de años y han sobrevivido a varios procesos de especiación.

3.5 Homogeneización de alelos y pérdida de ortología

En mamíferos es frecuente encontrar relaciones de secuencias ortólogas (secuencias homólogas que se han separado por un evento de especiación) entre genes de histocompatibilidad de clase I de especies pertenecientes a un mismo orden, aunque nunca se ha encontrado tal relación entre órdenes diferentes. Esto significa que los distintos genes de clase I se han ido originando por duplicación genética dentro de cada orden. Así, al comparar filogenéticamente secuencias MHC-I de distintos loci (MHC-A, -B, -C, -E, -F, -G) de algunos primates (humanos, chimpancés, gorilas y orangutanes) se observa que las secuencias pertenecientes a loci ortólogos agrupan juntas sin diferenciar entre especies. En cambio, cuando en la comparación se incluyen secuencias de otros órdenes (por ejemplo de roedores) los diferentes loci (ya sean clásicos o no clásicos) de un mismo orden aparecen agrupado juntos (Hughes y Nei, 1989a). Las relaciones de ortología en aves son más difíciles de encontrar, debido a la alta tasa de duplicación genética y a un fenómeno evolutivo denominado *concerted evolution*, que es consecuencia de la conversión génica entre alelos de distintos loci y afecta a los genes MHC de las aves, especialmente *Passeriformes*. El resultado es que alelos situados en genes parálogos en una especie se parecen más entre sí que alelos pertenecientes a genes ortólogos de distintas especies, incluso cuando éstas

pertenecen al mismo orden. Esta situación puede responder a 2 fenómenos: o bien los genes parálogos se han ido originado por duplicación dentro de cada especie o bien surgieron en un antepasado común a todas ellas y tras el proceso de especiación sufrieron conversión génica entre distintos loci provocando la denominada homogeneización de alelos y la pérdida de la ortología que inicialmente debía existir. Varios estudios ejemplifican estos fenómenos. En uno de ellos se describieron dos genes de clase II en el faisán (DABI y DAB2) que eran ortólogos, respectivamente, a los genes de clase II del pollo B-LBI y B-LBII, lo que significa que ambos genes se originaron por duplicación génica antes de la separación evolutiva del pollo y el faisán, hace 20 millones de años (Wittzell *et al.*, 1999). En el otro estudio, realizado con aves de canto, se comprueba la pérdida de ortología al comparar entre sí secuencias de clase II de la urraca (*Aphelocoma coerulescens*), el carpodaco común (*Carpodacus mexicanus*) y el turpial (Gasper *et al.*, 2001) y se sugiere que o bien existen antepasados comunes recientes dentro de cada grupo o bien el ritmo de homogeneización de alelos (Wittzell *et al.*, 1999) es muy elevado. En un estudio posterior se estimó que los genes de clase II DAB1 y DAB2 del turpial son el resultado de un proceso de duplicación ocurrido hace 40 millones de años; hoy, el gen DAB1 es polimórfico y el gen DAB2 no es funcional (Edwards *et al.*, 2000). En otro trabajo más reciente realizado con las especies de tordos de Nueva Zelanda (*Petroica australis*) y de la Isla Chatham (*Petroica traversi*) se han encontrado 4 secuencias de clase II en el primero y 8 en el segundo que parecen estar ligadas y muestran una gran diversidad en el exón 2, que parece estar sujeto a selección de balance; en el estudio encuentran evidencias de fenómenos de duplicación, conversión génica y pérdida de ortología (Miller y Lambert, 2004).

II. LA CLASE AVES

Las aves representan una de las cinco clases de vertebrados dentro del filo Cordados del Reino Animal. Según la clasificación de Sibley (Sibley, 1996; Sibley y Monroe, 1990) consta de 9.946 especies distribuidas en 23 órdenes y 2.068 géneros. Son animales homeotermos cuya característica principal y que les distingue del resto de los animales es que sus cuerpos están cubiertos de plumas (Encyclopaedia Britannica, 1999). Su esqueleto es resistente y liviano gracias a la fusión de sus elementos y a la presencia de sacos aéreos, poseen extremidades delanteras modificadas que dan lugar a las alas y su capacidad para volar les ha permitido extenderse por todo el planeta. Además, producen huevos que tienen una cáscara calcárea, tienen una visión muy aguda y la mayoría se caracteriza por sus hábitos diurnos. El registro fósil de las aves es reducido comparado con el de peces y reptiles. Sin embargo los antecesores de las aves se remontan al Mesozoico, hace más de 150 millones de años. La historia evolutiva de las aves comienza con la transformación de sus ancestros reptiles en un ave voladora con plumas. Este origen de las aves en los reptiles fue propuesto por Thomas H. Huxley (Huxley T, 1867), quien clasificó a ambos grupos dentro de la categoría taxonómica Saurópodos (*Sauropsida*) en base a características craneales. Además, tanto aves como reptiles ponen huevos en los que se desarrolla el embrión. Más aún, las hembras de aves y algunas hembras de reptiles son el sexo heterogamético; así, estas hembras tienen inactivo el cromosoma sexual W y solamente un cromosoma Z mientras que los machos tienen dos cromosomas Z (ZZ) (Buckley, 1987). El hecho de que, al contrario que los mamíferos, aves y reptiles tengan los glóbulos rojos nucleados hace pensar aún más en un origen común para ambos. Algunos autores sostienen que la llamada extinción de los dinosaurios, hace 65 millones de años, no afectó a la totalidad de ellos. Al igual que sobrevivieron los mamíferos y algunas formas de reptiles (como lagartos, serpientes, cocodrilos y tortugas), un grupo de dinosaurios voladores denominados neornites, que ya se habían diversificado 5 millones de años antes, también sobrevivieron; no así otros grupos voladores como los hesperornitiformes y los ictiornitiformes, que se extinguieron al igual que los dinosaurios terrestres y acuáticos. Por tanto, desde su aparición en el Triásico hace 200 millones de años y pese a la gran extinción, los dinosaurios han sido un grupo dominante que ha llegado hasta nuestros días (aves) como únicos colonizadores del medio aéreo con una diversidad (9.600 especies) que duplica la diversidad alcanzada por los mamíferos (4.200 especies), que también aparecieron en el Triásico y empezaron a dominar el medio terrestre una vez desaparecidos los dinosaurios terrestres hace 65 millones de años (Sanz, 1999).

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 metros. Los órdenes de las aves modernas divergieron hace aproximadamente 60 millones de años o incluso antes en el periodo terciario. Las aves acuáticas como los colimbos, gaviotas, patos, grullas, petreles y el extinto *Pinguinus impennis* aparecieron durante el Eoceno, hace entre 37 y 54 millones de años. Las formas primitivas de pájaros carpinteros así como los grupos afines a ellos también aparecieron en el comienzo del Eoceno y llegaron a ser las aves de percha más abundantes durante el Mioceno. Las carracas (familia *Coraciidae*), martines pescadores (familia *Alcedinidae*) y cálaos (familia *Bucerotidae*) se diversificaron durante el Oligoceno. En el Mioceno, la rápida evolución de plantas con flores e insectos generó nuevos nichos para pájaros que se alimentan de insectos, fruta o néctar; esta diversidad ecológica dio origen a la amplia radiación de las aves de canto (Allende *et al.*, 2001;Arnaiz-Villena A *et al.*, 2012;Arnaiz-Villena *et al.*, 1998;Arnaiz-Villena *et al.*, 2001;Arnaiz-Villena *et al.*, 2007b;Arnaiz-Villena *et al.*, 2007c;Regal, 1977;Zamora *et al.*, 2006a) . A finales del Terciario, hace entre 5 y 10 millones de años, las aves se diversificaron dando lugar a una gran variedad de formas que incluían ya muchos de los géneros actuales.

4. Orden Passeriformes

El orden *Passeriformes* comprende alrededor de 83 familias, 1170 géneros y unas 5875 especies (Gill, 1999), lo que supone aproximadamente el 60% de las especies de la clase aves reconocidas por las clasificaciones de (Sibley y Monroe, 1990;Sibley y Monroe, 1993). La mayoría de los passeriformes o passerines son moradores del suelo que se alimentan de insectos, semillas, fruta y néctar. Los de mayor tamaño son las aves-lira australianas (género *Menura*) y el cuervo (*Corvus corax*) mientras que los más pequeños son los sastrecillos (género *Psaltriparus*), reyezuelos (género *Regulus*), ratonas australianas (género *Malurus*), y el sastrecito pigmeo (*Psaltria exilis*).

Los passerines son morfológicamente similares, difiriendo fundamentalmente en la estructura de la siringe, pico y patas. Presentan una glándula uropigial con una única estructura en forma de tetilla. Esta glándula produce un aceite que las aves esparcen en su plumaje al acicalarse. Además, las patas, que están especializadas con un gran hálux (dedo dirigido hacia atrás), la disposición de los tendones profundos, y los músculos simplificados les facilitan posarse en los árboles (Raikow, 1982). Estas características indican que los miembros del orden *Passeriformes* evolucionaron desde un ancestro común, o sea, son un

grupo monofilético. Esta relativa uniformidad, comparada además con la de los no *Passeriformes*, llevó a algunos sistemáticos a clasificarlos dentro de una única o unas pocas familias. Otros los han contemplado dentro de 50 -104 familias. Según los datos de hibridación de ADN el orden *Passeriformes* se divide en 45 familias. Se da la circunstancia de que ciertas familias de otras clasificaciones han pasado a tener estatus de subfamilia o tribu atendiendo a estos datos de hibridación de ADN. Esta circunstancia se puede producir debido a la evolución convergente que lleva a que especies genéticamente próximas que se distribuyan en hábitat diferentes, desarrollen características fenotípicas dispares, y al revés, especies no estrechamente emparentadas pueden llegar a presentar gran similitud morfológica al compartir el mismo hábitat. Los datos resultantes del análisis del ADN de estas especies han revelado algunas de estas situaciones y han hecho posible clasificar juntas muchas especies genéticamente hermanas (Sibley, 1995).

Este grupo monofilético de los *Passeriformes* ha sido constatado, dentro de las Neoaves, como un grupo monofilético tanto en base a caracteres morfológicos (Prum, 1993;Raikow, 1982) como de sus secuencias de ADN (Irestedt *et al.*, 2001). No obstante, no se ha logrado conocer el linaje que engloba a los *Passeriformes* con su “grupo hermano” o evolutivamente más cercano dentro de las Neoaves. Así, los *Passeriformes* parecen más bien ser un grupo basal respecto a los demás grupos de Neoaves. Quizá por ello, se les ha relacionado con grupos como los pájaros carpinteros (Olson, 1983), aunque según los resultados obtenidos de la hibridación ADN-ADN, los *Passeriformes* formarían parte de un grupo basal que incluye a su vez, otros grupos con adaptaciones tan diversificadas como grullas, gaviotas, raptores, piqueros, alcatraces y pingüinos (orden Ciconiformes) (Sibley y Ahlquist, 1990). Por otra parte, los resultados del análisis del ADN mitocondrial señalan que el orden *Passeriformes* podría ser parafilético y basal dentro de la clase aves (Härlid *et al.*, 1998;Härlid y Arnason, 1999;Mindell *et al.*, 1997). Es posible que esta conclusión esté sesgada por un muestreo insuficiente de especies, una alta tasa de sustitución nucleotídica así como la utilización de controles externos (*outgroups*) demasiado distantes como pudieren ser los cocodrilos (Garcia-Moreno y Mindell, 2000). Contrariamente, de los marcadores de ADN nuclear se extraen conclusiones más concordantes con los datos morfológicos aunque tampoco establecen el grupo “hermano” de los *Passeriformes* (Garcia-Moreno y Mindell, 2000;Groth y Barrowclough, 1999;Johnson, 2001;van Tuinen *et al.*, 2000). La datación del origen de su radiación puede inferirse del registro fósil. El registro fósil más antiguo procede del Eoceno temprano y fue encontrado en Queensland. Si estos fósiles son de oscines o suboscines no se ha podido determinar. Los fósiles más antiguos de oscines están datados en el Oligoceno

superior (25- 30MA), (Boles, 1995) y fueron encontrados en Francia (Caceres,2015;Mourer-Chauvire *et al.*, 1989 Jarvis *et al.*, 2014).

4.1 Relaciones Intra-Orden

El orden *Passeriformes* o Passerines se divide a su vez en dos subórdenes, a saber, Suboscines y Oscines. Esta subdivisión de los *Passeriformes*, en estos dos subórdenes, atiende a diferencias en la posición y estructura de la siringe, o aparato vocal responsable de las vocalizaciones. La siringe es un órgano característico de las aves, localizado en una cavidad en la unión de la tráquea con los dos bronquios primarios. La siringe no contiene cuerdas vocales si no que sirve para abrir y cerrar la glotis manteniendo así el agua y la comida fuera del tracto respiratorio. El aire es impulsado desde el saco de aire interclavicular por la contracción de los músculos torácicos y abdominales y así llega hasta la siringe a través de los bronquios. El sonido se produce cuando el aire que recorre los pasillos de la siringe vibra a su paso por esta. La vibración de la membrana timpaniforme interna determina las características del sonido (Gaunt y Wells, 1973;Greenewalt, 1968;Greenewalt, 1969). Los músculos de la siringe controlan los detalles durante la producción del sonido alterando la tensión de la membrana timpaniforme. Por ello, las especies con una musculatura simple de la siringe únicamente emiten gruñidos, silbidos, o sonidos similares. A parte de las diferencias a nivel de la siringe, la divergencia entre oscines y suboscines ha sido constatada con análisis del mtDNA, y la divergencia entre ambos parece ser anterior a la mayoría del resto de radiaciones dentro de las neoaves (Johnson, 2001;Mindell *et al.*, 1999).

4.1.1 Oscines o Passeri

Los oscines, o verdaderas aves de canto, poseen una musculatura de la siringe más compleja (hasta seis pares de músculos intrínsecos), y aprenden a emitir sus vocalizaciones a partir del aprendizaje de sus padres, en lugar de tratarse de vocalizaciones heredadas de forma innata. Es posible que ésta elaborada musculatura de la siringe aumente el abanico de frecuencias empleado en sus cantos (Brackenbury, 1982).

Dentro del suborden *Passeri* (Oscines) se pueden distinguir a su vez otros dos grupos, así aquellas aves emparentadas con los cuervos constituyen el parvorden *Corvida* y el resto de especies se engloban en el parvorden *Passerida*. Este último grupo comprende varias superfamilias, una de ellas, la superfamilia *Passeroidea*, contiene a los representantes del género *Carduelis* (comúnmente llamados jilgueros) en los que se centra este estudio, y también a otros muy relacionados como los canarios (Lowy, 2003), y demás géneros de la

tribu *Carduelini*.

Los pinzones (tribu *Fringillini*), también están dentro de la superfamilia *Passeroidea* y una de las tres especies de este grupo, el pinzón vulgar (*Fringilla coelebs*) ha sido incluida en este trabajo como se explica más adelante.

4.1.2 Suboscines o Tyranni

Los suboscines muestran pocos caracteres que los unifiquen aparte de los músculos de la siringe. Estos músculos son más simples que los que poseen los *Passeriformes* del suborden Passeri (oscines), siendo generalmente menos de cuatro pares de músculos intrínsecos. Se pueden considerar dos grupos según si se distribuyen en el nuevo o en el viejo mundo (Sibley 1996). Los suboscines del viejo mundo comprenden a los eurilaimos (familia *Eurylaimidae*), las pitas de África y Asia (familia *Pittidae*), las filepitas de Madagascar (familia *Philepittidae*) y los reyezuelos de Nueva Zelanda (familia *Acanthisittidae*). En base a las características morfológicas de estos últimos y a los resultados de los análisis de hibridación ADN-ADN, se cree que podrían ser el linaje más antiguo de los *Passeriformes* (Raikow, 1984; Sibley y Ahlquist, 1982), o bien ser un grupo relicto no estrechamente relacionado con el resto de *Passeriformes*, mereciendo por tanto ser considerado en un tercer suborden. Los suboscines del nuevo mundo se dividen en dos grandes grupos o radiaciones, el primero comprende a los mosqueros, papamoscas atrapamoscas, cotingas y manequines entre otros (familia *Tyrannidae*) y el segundo engloba a los horneros y pájaros carpinteros (familia *Furnariidae*), hormigueros (familias *Thamnophilidae* y *Formicariidae*), y tapaculos (familia *Rhinocryptidae*) entre otros.

5. Características de las aves estudiadas

5.1 *Carduelis atrata* (Lúgano negro)



Mide unos 12-13 cm de longitud. El macho adulto se distingue por el color negro muy oscuro y el amarillo muy intenso, mientras que la hembra es más clara, de un tono entre negro y marrón oscuro y el color amarillo se encuentra más extendido por el vientre. Su hábitat natural en Perú y Bolivia, en altas mesetas y páramos andinos ("la Puna"), caracterizados por pendientes, con rocas y hierba salpicadas de arbustos (Clement *et al.*, 1993).

5.2 Carduelis carduelis (Jilguero común)

Su tamaño está entre los 13 y 15 cm de longitud. El macho de esta especie posee una máscara de color rojo anaranjado, con las plumas de alrededor del pico, la nuca y los lados del cuello negras, su vientre y la parte central de su pecho son de color blanco.



Las alas, cola y obispillo son de color negro con las típicas “perlas” blancas y el espejuelo alar amarillo. La hembra en cambio, posee una máscara también roja pero más redondeada sin superar el ojo, y sus coberteras alares pequeñas son marrones oscuras y no negras. El jilguero vive en Europa y norte de África. Frecuenta zonas cultivadas, parques, jardines e incluso bosques (Clement *et al.*, 1993).

5.3 Carduelis citrinella (Verderón serrano)

De longitud de unos 11,5-12 cm. Es un pájaro verde-grisáceo que vive en la montaña alpina. Su obispillo es de color intermedio entre el amarillo y el verde. Las plumas que utilizan para el vuelo y la cola son negras. Los ejemplares adultos se diferencian fácilmente porque la parte trasera de su cabeza, la nuca y los laterales del cuello son de color gris. Las hembras de esta especie, poseen unos tonos menos brillantes. Aparece en bosques subalpinos de piceas, alerces y pino montano aunque también lo podemos encontrar en praderas. Habita en España, Francia, Italia, Austria y Alemania y la subespecie *C. citrinella consicanus* en Córcega y Cerdeña (Clement *et al.*, 1993).



5.4 Carduelis crassirostris (Lúgano de pico grueso)

Mide en torno a 14 cm de longitud. Posee una cabeza y cuello negros su dorso es de tono verde-amarillento. El pecho, abdomen son de color amarillo. Las alas de tonos negro-parduscos poseen una banda transversal amarilla. De esta especie hay que destacar el pico notoriamente más grande y ancho en comparación a las otras especies de *Carduelis*. La hembra es de colores más apagados y no tiene la cabeza negra. Habita en Bolivia, centro de Chile y en el norte de Argentina (Clement *et al.*, 1993; Renzo Esuperanzi, 2008).



5.5 Carduelis cucullata (Cardenalito de Venezuela)

El cardenalito mide de unos 10 cm de largo. El macho es principalmente de color rojo profundo, con negro en la cabeza, la garganta, las plumas de vuelo y la punta de la cola. La

parte baja del vientre y cloaca son blanquecinas. La hembra posee una base de color gris en la cabeza, espalda y pecho. El pecho además tiene los flancos rojizos, y el resto de la parte



ventral, las alas y la cola se asemejan a las áreas correspondientes de la masculina. Las hembras inmaduras son más pálidas que los adultos, y los machos inmaduros son de color marrón en lugar de rojo. Este *Carduelis* es un residente de la cría de aves en las zonas tropicales de América del Sur en el norte de Colombia y el norte de Venezuela (Clement *et al.*, 1993;Renzo Esuperanzi, 2008).

5.6 Carduelis dominicensis (Pinero antillano)

Presenta una longitud de entre 11 y 12cm. Es una especie endémica de La Española (República Dominicana y Haití) y la única especie de *Carduelis* del Caribe. El macho es amarillo con la cabeza negra, mientras que la hembra está tintada de verde oliva con blanco y marrón por encima, y un rayado oscuro por debajo. En ambos sexos el pico es marrón pálido.



Suele habitar en bosques de pinos de tierra altas (por encima de los 1500m), donde se encuentra en bandadas dispersas.

Se posa en los árboles y encuentra el alimento en arbustos, en plantas con una cantidad importante de semillas, aunque fundamentalmente prefiere semillas de acedera (Clement *et al.*, 1993;Renzo Esuperanzi, 2008).

5.7 Carduelis lawrencei (Dominiquito gris)

Presenta una longitud de entre 10 y 11cm. Es de color marrón grisáceo con alas amarillo brillantes, grupa amarillenta y cola negra. El macho tiene la cara negra y, en invierno, presenta una franja negra en el ala, además el pecho y la barriga son de color amarillo brillante. En otros periodos del año el amarillo es menos extenso. La parte ventral es como la de la hembra, blanca o blanquecina-ante con los costados marrones. Es una especie monotípica.



Se distribuye por el suroeste de Estados Unidos, sur de California (condado de Los Ángeles), este de Sierra Nevada, este de Arizona y noroeste de México (norte de Baja California y Sierra de San Pedro Mártir)(Clement *et al.*, 1993;Renzo

Esuperanzi, 2008).

5.8 Carduelis magellanicus (Lúgano encapuchado)

Posee una longitud de entre 11-12 cm. El macho tiene la cabeza, alas y cola de color negro brillante el dorso verde brillante con tonos negros. El pecho, cuello, vientre, región subcaudal y obispillo de color amarillo, el pico y las patas son negras. En las hembras el color de la cabeza, el cuello, el dorso y los flancos es verde oliva y la banda alar y la cola son de tonos amarillentos. Su vientre es blanquecino. El Lúgano de cabeza negra habita en gran parte de América del Sur (Ecuador, Perú, Argentina, Bolivia, Colombia, Venezuela, Brasil Y Chile (Clement *et al.*, 1993;Renzo Esuperanzi, 2008).



5.9 Carduelis notata (Lúgano de cabeza negra)



De longitud de unos 11 cm, El macho posee una cabeza y la parte superior del pecho negras en forma de “v”. De este color son también las alas y la cola, aunque salpicadas con tonos dorados. El resto del pecho, el vientre, la región subcaudal y el obispillo son de color amarillo dorado intenso. El pico y las patas son de color gris plomo. La hembra es similar al macho pero la “v” del pecho es mucho más redondeada, sus tonos son algo menos brillantes. El Lúgano de pecho negro está extendido por toda América Central (Honduras, Costa Rica, Guatemala, El Salvador y Nicaragua) y en Mexico (Clement *et al.*, 1993;Renzo Esuperanzi, 2008).

5.10 Carduelis olivacea (Lúgano olivaceo)

Especie muy parecida y a veces confundida con el *Carduelis magellanicus*, presenta una longitud de entre 10 -11 cm. Posee las mismas diferencias entre sexos que el Lúgano de cabeza negra y un tono amarillo mucho más pálido, presentando el pájaro un color general verde aceituna. Habita en Ecuador (Santiago, Zamora y Loja), Perú (San Martín, Cuzco, Puno), Bolivia (La Paz, Cochabamba y Santa Cruz) (Clement *et al.*, 1993;Renzo Esuperanzi, 2008).



5.11 Carduelis pinus (Jilguero de los pinos)

Presenta una longitud de 11-13 cm. Se trata de un pequeño pinzón norteamericano de rayado marrón. Tiene un pico fino y apuntado, alas y cola marrones, con una extensión variable de color amarillo en las alas y en la base de la cola. Es característica su llamada (simple o doble). Su abundancia es irregular (o localmente regular), de forma que puede anidar abundantemente un año y no hacerlo el siguiente.



Habita en zonas de coníferas (fundamentalmente píceas), bosques mixtos (caducifolios y perennes), matorrales de aliso, árboles caducifolios y ornamentales, y se alimenta en arbustos en áreas suburbanas. Es muy social y colonial, aparece en bandadas de diferente tamaño durante todo el año, hasta llegar al millar en invierno. Se alimenta de diferentes semillas o brotes, sobre todo de píceas, cicuta, cedro blanco, abedul y aliso, que toman de los propios árboles. También se alimenta de semillas de plantas en el suelo y ocasionalmente de pequeños insectos. Se distribuye desde el sur de Alaska a través de Canadá, hacia el centro de Ontario, centro de Quebec y hacia el este hasta Terranova y Nueva Escocia, hacia el sur de Estados Unidos hasta el norte de Nueva Inglaterra, llegando al sur de California, sur de Arizona, sur de Nuevo Mexico y suroeste de Texas. Es parcialmente migratorio. (Clement *et al.*, 1993;Renzo Esuperanzi, 2008).

5.12 Carduelis psaltria (Dominiquito)



De tamaño de 10 centímetros el macho presenta un negro brillante desde la frente hasta la cola; garganta, pecho y bajo vientre amarillo intenso, alas y cola negra, manchas blancas en las rémiges terciarias. Apenas presenta banda alar y ésta es de color blanco y con habas del mismo color en su cola. Pico color hueso claro y en época de celo tono amarillento, este más bien grueso, cónico, puntiagudo. En la hembra el color negro es reemplazado por el verde oliva claro en y los tonos amarillos del pecho son mucho más suaves incrustados de verde en zonas de la garganta, pecho y flancos, contiene las áreas blancas más pequeñas que las del macho. Habita regiones templadas y tropicales, en zonas de matorrales, arbustos, jardines y campos abiertos. Habita desde la costas del estado de Washington y el suroccidente de Estados Unidos, así como en México y Centroamérica hasta Venezuela, Colombia, Ecuador y Perú. En invierno migra de los lugares fríos de Estados Unidos (Clement *et al.*, 1993;Renzo

Esuperanzi, 2008).

5.13 Carduelis spinescens (Lúgano de los Andes)

El Lúgano de los Andes mide entre 11-12 cm de longitud. El macho posee una frente de color negro, un dorso de tono verde oliva. Su pecho y vientre son amarillos con tonalidades marrones verdosas y las alas y la cola negras. La hembra de esta especie tienen una cabeza, cuello y dorso y flancos de color verde botella, el pecho es verde oliváceo con la región central amarillenta y el vientre blanco. El obispillo y los laterales de la cola son amarillos. Habita en las regiones andinas de Colombia y Venezuela. Y se nutre de semillas verdes y de pequeñas bayas (Clement *et al.*, 1993; Renzo Esuperanzi, 2008).



5.14 Carduelis spinus (Lúgano euroasiático)



Presenta una longitud de 12 cm y una envergadura de 20-23 cm, es un poco más pequeño que el jilguero (*Carduelis carduelis*), con la cola notablemente más pequeña y de forma más compacta. Es bastante rechoncho aunque elegante con la cola bifurcada. Su plumaje es amarillo-verdoso y rayado. En todas las edades muestra un pico fino así como una acusada combinación de banda amarilla

en el ala y parches amarillos en la base de la cola.

El macho es de color verde y amarillo, con corona negra, buena parte de las alas y de la cola es negra, presenta marcas amarillas, y la rabadilla es amarillo-verdosa.

La hembra es más verdosa, con un rayado distintivo por encima y por debajo, especialmente desde los lados del pecho hasta los costados traseros. Común o localmente común, se suele alimentar en coníferas, principalmente piceas, pero también en alisos (especialmente en invierno), alerce y abedul. Se observa en cualquier hábitat que tenga los árboles adecuados, especialmente brezales, bosques, matorrales, huertos y jardines. Es una especie monotípica de distribución discontinua que habita en el norte de Europa y Rusia, existiendo una población separada en Asia oriental. En el sur de Europa hay poblaciones aisladas en los Pirineos, Cerdeña, Suiza, centro y sur de Italia, Yugoslavia, Hungría, Bulgaria y Rumania, Crimea, Chipre, norte de Turquía, norte del Cáucaso y norte de Irán hacia la costa sureste del mar Caspio. Las poblaciones situadas al norte de su área de distribución se desplazan hacia el sur para pasar el invierno. En la parte occidental del área de distribución llegan hasta España,

Italia, los Balcanes, Grecia, Chipre, Turquía, centro de Irán sur de Asia central, y en el rango más oriental llegan hasta Mongolia, Corea, sur de China, Taiwán y Japón (Clement *et al.*, 1993;Renzo Esuperanzi, 2008).

5.15 Carduelis xanthogastra (Lúgano de vientre amarillo)

De longitud de unos 11 cm, posee la cabeza, cuello, parte superior del pecho, obispillo, alas y cola de color negro brillante. La parte inferior del pecho, vientre y región subcaudal de color amarillo intenso. El pico es gris oscuro con la punta negra y las patas negras. La hembra en cambio posee una parte superior de color verde botella. Las alas y la cola son negras. La parte inferior del pecho y el vientre también de color amarillo intenso. Habita en las zonas de clima templado de Venezuela, Bolivia, Panamá, Ecuador y Colombia. Su dieta se compone de semillas verdes de plantas silvestres y cultivadas, así como de algunos insectos (Clement *et al.*, 1993;Renzo Esuperanzi, 2008).



5.16 Carduelis yarrellii (Lúgano de cara amarilla)



De longitud de unos 11 cm, el macho de esta especie posee un característico casquete de color negro que llega hasta la nuca pasando por los ojos. El dorso es verde claro brillante con tonalidades amarillas. La región subcaudal y el obispillo poseen un tono amarillo intenso. El pico es de color carne con la punta negra. La hembra posee la cabeza, dorso y cuello de color verde oliva brillante, el pecho

y el obispillo de color amarillo con tonalidades en verde y el vientre blanquecino.

El Lúgano de Yarrel, vive en las regiones de clima templado de Brasil en zonas no muy elevadas. Habita en sabanas de hierbas altas y arbustos, así como en las plantaciones que están cerca de los bosques (Clement *et al.*, 1993;Renzo Esuperanzi, 2008).

5.17 Serinus sulphuratus (Canario sulfúreo)

De medidas que rondan los 16 cm, este canario es uno de los de mayor tamaño de su género. El macho posee el vientre, garganta y región subcaudal de color amarillo muy intenso, el pecho y el obispillo es de color verde oliva, las alas y la cola son de tonos marrones oscuro. La hembra posee colores más claros y mates con tendencia al gris. Este canario suele localizarse en regiones esteparias de África meridional



(Clement *et al.*, 1993).

5.18 Serinus atrogularis (Canario de obispillo amarillo)

Posee una longitud de entre 11-12 cm. Se trata de un pájaro pequeño color marrón grisáceo. Posee un pico corto de forma cónica, una mancha negra en la mejilla y en la garganta, rayas en los laterales del pecho y en los flancos y un obispillo color amarillo brillante.



Los 2 sexos son más o menos similares. Su área de distribución va desde el este a sur de África. Es una especie que vive en una amplia gama de hábitats, desde sabana con maleza hasta bosques con árboles de hoja ancha, cultivos etc. (Clement *et al.*, 1993).

5.19 Serinus canaria (Canario silvestre)



Posee un tamaño de 13 cm. El canario silvestre posee la garganta, los hombros, el pecho y el obispillo de color amarillo intenso, la nuca y el dorso son de tonos gris verdoso con dibujos negros. La cola y las plumas son de color amarillo ocre. El pico y las patas grises carnosas. La hembra de esta especie presenta un color amarillo exclusivamente en la zona facial y en el obispillo. El resto es de color gris-marrón con algunas pinceladas de verde. El canario silvestre es originario de las Islas canarias. Vive tanto a nivel del mar como en zonas montañosas (Clement *et al.*, 1993).

5.20 Serinus citrinelloides (Canario africano)



Mide entre 11,5-12 cm. Se trata de un pájaro pequeño con un plumaje brillante pero variable. Existe una variación considerable entre las diferentes razas de esta especie. Puede presentar una cara negra con un supercilio amarillo o no presentar estas características; la parte superior es verde y suele presentar un rayado negro. La parte inferior de los machos es color amarillo brillante o amarillo-verdoso apagado con rayas en el pecho y en los flancos.

Se puede encontrar en el este de África, Etiopía hasta Zimbabue (Clement *et al.*, 1993).

5.21 Serinus dorsostriatus (Canario de vientre blanco)

De medidas que van entre 11,5–13 cm, este canario posee un patrón facial parecido al de *Serinus mozambicus* aunque no tan marcado. La parte inferior del pecho y la parte superior del vientre son amarillas La parte inferior del vientre y los flancos son blancos.

Las hembras son similares a los machos aunque cabe destacar que sus colores son un poco más apagados y que tienen bandas en el pecho y en los flancos. Habita en Etiopía, Somalia, Uganda, Kenia y Tanzania, en matorrales áridos y semi-áridos, en bosques, en áreas de sabana herbácea y también lo hallamos en zonas de cultivos. Puede aparecer a alturas superiores a los 2100 m aunque normalmente se encuentra por debajo de los 1400 m (Clement *et al.*, 1993).



5.22 Serinus flaviventris (Canario amarillo)

Mide de 13-14 cm. Sus partes superiores (incluyendo el obispillo) son verdes o amarillas. Los machos que viven en la zona noroeste del área de distribución tienen las partes superiores de color verde pálido y el obispillo de color amarillo brillante; los que viven en el sudeste son más verdosos. Las hembras son más pálidas y carecen de partes color amarillo brillante, suelen tener bandas.



Se distribuye en el sur de África. Se encuentra en matorrales y matorrales, en zonas secas y semiáridas y a los pies de las montañas. Se alimenta de varios tipos de semillas y ocasionalmente come insectos (termitas) (Clement *et al.*, 1993).

5.23 Serinus gularis (Canario de cabeza estriada)

Tiene un supercilio blanco y un píleo rayado; posee mejillas y auriculares color marrón oscuro, el mentón es blanco y el obispillo es de color marrón. Los 2 sexos son bastante similares. Se distribuye en el este y sur de África. Común, localmente común o poco común. Se encuentra en una gran variedad de hábitats, desde bosques de *Brachystegia* a maleza, sabana, jardines y huertas.

Normalmente se suele encontrar en parejas o en pequeñas bandadas, pero ocasionalmente se puede encontrar en grandes acumulaciones.

Es un pájaro bastante tranquilo que se alimenta de semillas, fruta, brotes de flores y árboles (Clement *et al.*, 1993).



5.24 Serinus mozambicus (Canario de frente amarilla)

Mide unos 11-13 cm. Es un canario pequeño y verde ampliamente distribuido por toda África. Tiene un patrón facial muy marcado, la frente y el píleo son verdes o verde-

amarillentos. El obispillo y las partes bajas son de color amarillo. Los dos sexos son muy similares aunque las hembras y los juveniles tienen un color ligeramente más apagado, los juveniles tienen rayas en las partes bajas. Se puede encontrar en Kenia, Zambia, Malawi, Tanzania, Mozambique, Zimbabue, Botsuana, Namibia y Sudáfrica. Ampliamente distribuido, se trata de una especie bastante castigada por el comercio con pájaros de jaula. Vive en las mesetas de la sabana o en claros que presenten árboles ocasionales, también se encuentra en bosques de sabana o en los límites de áreas cultivadas, de jardines y alrededor de las casas (Clement *et al.*, 1993).



5.25 Serinus striolatus (Canario estriado)

De longitud de unos 15 cm aproximadamente, es una especie de tamaño pequeño con un pico bastante grande. Es fácil de identificar por su supercilio blanco, su parte inferior es rayada, las alas y cola son completamente marrones. Los 2 sexos son muy similares. La podemos encontrar en el sureste de África, en Sudán, Etiopía, Kenia y Tanzania, Uganda y Zaire. Vive a campo abierto, en el límite de bosques con alturas que oscilan entre los 1300 m y los 4300 m, en bosques húmedos secundarios, matorrales, maleza, jardines y en los límites de áreas cultivadas. Normalmente aparece solo, en parejas o en pequeños grupos. Se alimenta en vegetación de baja altura o en el suelo (Clement *et al.*, 1993; Renzo Esuperanzi, 2008).



5.26 Serinus thibetanus (Canario tibetano)



Posee una longitud de 12 cm. Los 2 sexos son diferentes. Los machos tienen el píleo y las partes superiores color verde oliva, el supercilio aparece amarillo, las partes inferiores no aparecen rayadas y son de un color verde amarillento. Las hembras son más oscuras y tienen rayas oscuras en las partes superiores y en los flancos. El juvenil es parecido a la hembra adulta. Su distribución es monotípica. Desde el Himalaya oriental hasta el sudeste del Tíbet. También aparece en China occidental. Habita en laderas abiertas y en bosques de coníferas, abedules, alisos, cicutas y abetos a alturas que oscilan entre 2800 m y 4000 m. Ocasionalmente se alimenta en el suelo pero normalmente lo suele hacer en arbustos. Su alimentación consiste en diferentes semillas, principalmente de aliso y abedul (Clement *et al.*, 1993).

5.27 Rhodopechys sanguineus (Camachuelo Alirrojo)

Mide entre 15 y 18 cm de longitud. El macho posee una frente y casquete de color marrón muy oscuro, así como su dorso y flancos. La nuca, los lados de la cabeza y del cuello son de color marrón grisáceo. La garganta y el vientre son de color marrón con dibujo en negro. El vientre es blanquecino y el obispillo rosado. Las alas y la cola rosas. Su pico es amarillo anaranjado y suele tener una pequeña máscara rosa. La hembra es muy parecida pero de colores más mates, menos intensos, carece de máscara y de dibujo negro en el pecho. El camachuelo trompetero de alas rosas vive en Turquía, Marruecos y China. Vive en zonas semidesérticas donde los árboles y matorrales son bastante escasos (Clement *et al.*, 1993; Renzo Esuperanzi, 2008)



III. FILOGENIA MOLECULAR

6. El ADN mitocondrial

6.1 Generalidades

Las mitocondrias son orgánulos presentes en el citoplasma de casi la totalidad de las células eucariotas. La mitocondria está limitada por dos membranas, externa e interna, que tienen un papel esencial en la función de este orgánulo y determinan los dos compartimentos mitocondriales, el espacio intermembranoso y el espacio de la matriz.

La membrana externa es permeable a todas las moléculas de menos de 10.000 daltons, incluidas pequeñas proteínas. Por el contrario, una vez llegado al espacio intermembranoso, la mayoría de estas moléculas no pueden atravesar la membrana interna siendo esta únicamente permeable a pequeñas moléculas necesarias para las enzimas mitocondriales que se encuentran en la matriz. Estas enzimas realizan procesos oxidativos. El NADH constituye la fuente principal de electrones que circulan a través de la cadena de transporte (cadena respiratoria). Mediante este transporte electrónico se bombean protones al espacio intermembranoso generándose un gradiente electroquímico. Este gradiente impulsa los protones de nuevo a la matriz a través del complejo proteico transmembrana ATP sintetasa que utiliza el flujo de protones para sintetizar ATP a partir de ADP y fosfato inorgánico (Pi) en la matriz. De ahí que éstos orgánulos sean “la fábrica de energía” de toda célula.

En vertebrados, el ADN mitocondrial (mtDNA) está presente en múltiples copias (1000-10.000 copias/célula) y codifica para trece proteínas de la membrana interna mitocondrial, participantes de la cadena respiratoria. El resto de los componentes mitocondriales son codificados por genes situados en el núcleo y dirigidos a la mitocondria (Glick y Schatz, 1991; Pfanner y Neupert, 1990; Schatz, 1996). La molécula de mtDNA también codifica para dos moléculas de RNA ribosomal (rRNA) y los 22 RNA transferentes (tRNA) necesarios para la traducción de los mRNAs codificados por el mtDNA en la matriz mitocondrial. El genoma mitocondrial de vertebrados es extremadamente compacto, de forma que cada gen que codifica para proteína o rRNA). El tamaño del mtDNA en vertebrados oscila entre 16 y 19kb y posee forma circular. Las dos hebras, pesada (H o *Heavy*) y ligera (L o *Light*) se diferencian en su mayor y menor contenido respectivo de G+C (Guaninas y Timinas).

6.2 Citocromo mitocondrial

Los citocromos son hemoproteínas de color pardo o rojo cuyos grupos hemo, se diferencian de los que están presentes en la hemoglobina y la mioglobina, en que los átomos de Fe

sufren reacciones de oxidación/reducción entre Fe^{++} y Fe^{+++} . Existen 3 tipos de citocromos (a, b y c) que se clasifican según sus diferencias en los espectros de absorción de la luz.

Los citocromos transportan electrones dentro de la cadena respiratoria.

El complejo b- c_1 , acepta electrones de la ubiquinona y los transfiere al citocromo c. Cada citocromo, cuando se halla en estado férrico (Fe_{3+}), acepta un electrón para alcanzar el estado ferroso (Fe_{2+}). El citocromo b forma, junto con el citocromo c_1 y un centro ferro-sulfato específico, el complejo III de la cadena respiratoria (Figura 8).

El gen del citocromo b mitocondrial está codificado por la cadena pesada H (“Heavy”, rica en guaninas) y se localiza entre las bases 14747 y 15887 (Anderson *et al.*, 1981;Wallace *et al.*, 1994). Este gen de herencia materna (Case y Wallace, 1981;Giles *et al.*, 1980), comprende 924pb y codifica para un único polipéptido sin intrones. El mRNA tiene una región 5’ no codificante seguida del codón de inicio AUG, y finaliza con el U del codón de terminación UAA (Anderson *et al.*, 1981;Montoya *et al.*, 1981;Ojala *et al.*, 1981;Wallace *et al.*, 1994). Se transcribe como parte del mRNA policistrónico de la cadena pesada que está flanqueado por transcritos de tRNA para ácido glutámico y treonina.

La comparación de secuencias del gen del citocromo b, muestra que las transversiones (sustituciones de una purina por una pirimidina o viceversa) se acumulan linealmente con el tiempo en las terceras posiciones de los codones, de forma similar a lo que ocurre en los genes de rRNA. La tasa de divergencia para las transversiones es menor que la correspondiente para las transiciones (sustitución de una base púrica por otra base púrica, o una base pirimidínica por otra base pirimidínica). Por otra parte, la tasa de divergencia de las terceras posiciones es mayor que en la primera y segunda posiciones; consecuencia de que las transversiones silenciosas (que no suponen cambio de aminoácido) pueden darse lugar en muchas de las terceras posiciones de los codones (Irwin *et al.*, 1991).

El citocromo b es una proteína altamente conservada que, dependiendo de los autores, presenta nueve (Saraste, 1984;Widger *et al.*, 1984) u ocho (Crofts *et al.*, 1987;Howell y Gilbert, 1988) dominios transmembrana apolares en forma de hélice (Figura 8). En este estudio se considera el modelo de ocho dominios. La proteína contiene dos grupos hemo (b_K y b_T). Cada grupo hemo se halla ligado a dos dominios transmembrana 2 y 4 mediante dos histidinas invariantes en posiciones 96 y 182 (hemo b_K) y 82 y 198 (hemo b_T) (Barber, 1984). A nivel de aminoácidos, y dentro de mamíferos, se ha comprobado que el 59.1% de los residuos son invariantes. La mayoría de las posiciones variables están localizadas dentro de los segmentos transmembrana, o en los extremos carboxilo y amino de la proteína. La parte exterior de la proteína parece evolucionar más lentamente que las regiones transmembrana e interna que parecen llevar una tasa de cambio similar.

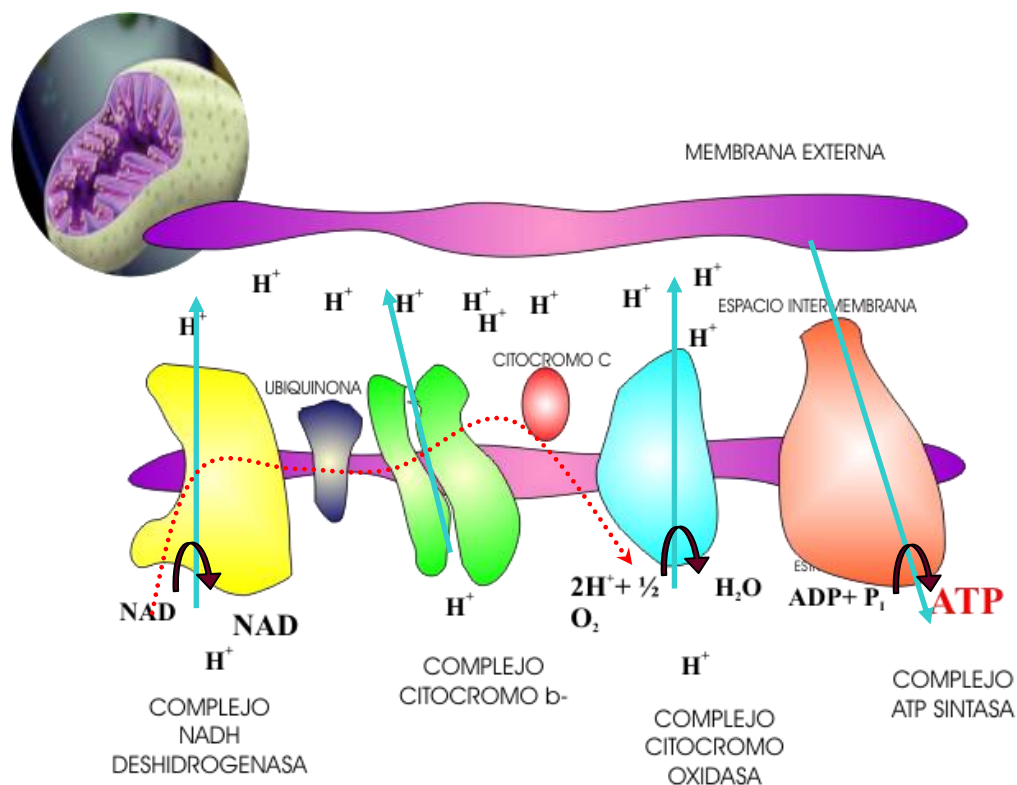


Figura 8. Esquema del flujo de electrones a través de los tres complejos enzimáticos respiratorios principales, durante la transferencia de dos electrones desde el NADH hasta el oxígeno. Cuando un electrón es transferido a lo largo de la cadena de transporte electrónico, tres complejos enzimáticos respiratorios bombean protones hacia el exterior del espacio de la matriz. Este bombeo genera un gradiente electroquímico de protones a través de la membrana mitocondrial interna que impulsa a los protones a volver hacia la matriz a través de la ATP sintasa, un complejo proteico transmembrana que utiliza la energía del flujo de protones para sintetizar ATP a partir de ADP y Pi en la matriz. La ubiquinona y el citocromo c actúan como transportadores entre los complejos.

La mayoría de cambios variables en la región transmembrana son cambios entre residuos hidrofóbicos (leucina, isoleucina, y valina). La mayoría de la parte de la superficie interna está compuesta por los extremos amino y carboxilo terminales, y no parece tener un papel funcional muy importante. Las regiones definidas como Q_o y Q_i mediante estudios mutacionales (Howell y Gilbert, 1988) o mediante comparaciones entre organismos evolutivamente distantes (Howell, 1989), están muy conservadas en mamíferos. El centro redox Q_i supone una pequeña porción del primer segmento transmembrana. Sin embargo, la mayoría de la superficie exterior está implicada en el centro redox Q_o (Howell, 1989) y quizás de ahí la reducida tasa evolutiva de la superficie exterior (Irwin *et al.*, 1991).

6.2.1 Citocromo b y análisis filogenéticos

A la hora de llevar a cabo un estudio filogenético basado en el ADN hay que escoger muy bien la secuencia que será utilizada como representante de cada especie, y para ello es necesario conocer el ritmo y el modo de evolución de la secuencia elegida. Las secuencias no codificantes no están sujetas a selección, por lo que se supone que las mutaciones ocurren al

azar en cualquier posición y provocando cualquier tipo de cambio con igual probabilidad. Sin embargo, el ritmo de acumulación de mutaciones es rápido y puede llevar a subestimar la distancia genética entre especies, sobre todo si están alejadas evolutivamente.

Las secuencias codificantes, por el contrario, al estar sujetas a selección, evolucionan más lentamente ya que sólo acumulan las mutaciones que no han afectado negativamente a la función proteica. También hay que tener en cuenta si la secuencia de estudio se encuentra en el núcleo o en la mitocondria, ya que el ADN mitocondrial tiene una tasa de mutación 10 veces superior a la del ADN nuclear. Por otro lado hay que tener en cuenta los mecanismos de variación que afectan a cada gen además de la mutación. Los genes que presentan alelos y sufren recombinación (genes nucleares) o bien que experimentan fenómenos de conversión génica o están sujetos a selección de balance (como ocurre con los genes MHC) pueden complicar los análisis filogenéticos y ofrecer resultados poco fiables. Por todo ello, los genes más utilizados en los estudios filogenéticos se encuentran en la mitocondria y de ellos el que se usa con mayor frecuencia es el de citocromo b (Questiau *et al.*, 1998; Wink, 1995).

La composición nucleotídica del gen, varía notablemente entre las tres posiciones de los codones. Así, las terceras posiciones contienen pocas guaninas, las segundas posiciones son ricas en timinas, y las primeras posiciones parecen no tener ningún sesgo.

Las terceras y segundas posiciones muestran respectivamente la mayor y menor variabilidad interespecífica. Las sustituciones en las primeras posiciones tienden a conservar el aminoácido en mayor medida que las sustituciones en las segundas posiciones; por tanto, se esperan más cambios en las primeras posiciones que en las segundas. Como los cambios en las terceras posiciones tienden a ser sinónimos, estas posiciones han de ser las más variables (Irwin *et al.*, 1991).

La utilidad de este gen se debe por tanto a su modo particular de evolución (bien conocido) y por otro a las características propias del ADN mitocondrial, como son ausencia de alelos y ausencia de recombinación.

Además, el hecho de aparecer en múltiples copias (tantas como mitocondrias haya en cada célula) y de carecer de intrones facilita mucho su manipulación y estudio.

7. Filogenia molecular

La filogenia molecular basada en el ADN tiene como objetivo determinar las relaciones evolutivas existentes entre diversos taxones mediante la comparación de secuencias de ADN ortólogas de cada uno de ellos.

Existen dos tipos de metodologías que comparan secuencias y representan las relaciones

entre taxones mediante árboles filogenéticos (dendrogramas): los métodos cladísticos y los métodos basados en distancias genéticas.

7.1 Métodos cladísticos

La cladística es una rama de la biología que define las relaciones evolutivas entre los organismos basándose en caracteres derivados o apomorfos (originados a partir de un carácter primitivo) y en sinapomorfías (conjunto de caracteres apomorfos compartidos por un grupo de taxones).

El análisis cladístico es la base de la mayoría de los sistemas modernos de clasificación biológica que pretenden agrupar los organismos en función de sus relaciones evolutivas.

Los primeros análisis filogenéticos que se empezaron a realizar eran de tipo cladístico y estaban basados en los métodos de **máxima parsimonia** (MP, *maximum parsimony*). Este sistema establece una relación entre los grupos, considerando el mínimo número de transformaciones necesarias para pasar de los caracteres primitivos (plesiomórficos) a los rasgos derivados de un taxón ancestral filogenéticamente próximo (apomórficos).

El método de parsimonia más empleado es el que considera que cualquier cambio de nucleótido es igualmente probable, y esa probabilidad es la misma en cualquier región de la secuencia (Fitch, 1971). Por tanto, los métodos cladísticos asumen que la evolución sigue el camino más sencillo y directo, lo cual no tiene por qué ocurrir de esa manera.

Este tipo de metodologías analiza todos los posibles árboles que se pueden construir a partir de una matriz de secuencias de ADN y ofrece como resultado final el árbol (o árboles) más parsimonioso, esto es, el árbol que puede explicarse con un menor número de cambios de nucleótido (Fitch, 1971).

7.2 Métodos basados en distancias genéticas

Cuando hablamos de distancia genética nos referimos a la divergencia genética entre especies o entre poblaciones de una misma especie. Las distancias genéticas menores, indican una estrecha relación genética mientras que las grandes distancias genéticas indican una relación genética más distante.

Los métodos de distancias alcanzan el resultado en dos etapas: primero calculan una matriz de distancias genéticas resultante de la comparación de todas las secuencias por parejas y después construyen un árbol filogenético basado en las distancias genéticas anteriores.

Existen varios modelos que permiten calcular distancias genéticas entre secuencias y varios métodos que permiten construir árboles filogenéticos a partir de las distancias.

Modelos de sustitución nucleotídica

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 calcular una distancia 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, 1972; Nei, 1987). La distancia p asume que todos los cambios de nucleótido ocurren con igual frecuencia y en cualquier parte de la molécula. Sin embargo, el proceso evolutivo no es tan sencillo como eso. La experiencia nos dice que determinados tipos de mutaciones ocurren con mayor frecuencia que otros, y las mutaciones no siempre se distribuyen homogéneamente.

Por ello se han ido desarrollando los denominados modelos de sustitución nucleotídica (o modelos evolutivos), que tienen en cuenta una serie de variables a la hora de realizar el cálculo de la distancia genética entre dos secuencias.

Entre estas variables podemos destacar:

- **Frecuencias nucleotídicas:** se puede asumir que la frecuencia de cada nucleótido es la misma (0,25) o bien que es diferente, lo que se aproxima más a la realidad.
- **Número de tipos de sustituciones:** también se puede asumir que cualquier cambio nucleotídico es igualmente probable o bien que cada tipo de cambio nucleotídico tiene su propia probabilidad. En el primer caso, el número de tipos de sustituciones (denominado nst, *number of substitution types*) sería 1 y en el segundo caso sería 6. De todas las posibilidades intermedias, la más frecuentemente utilizada es la que considera dos tipos de cambios (nst=2), denominados transiciones (ts) y transversiones (tv). Las transiciones son cambios entre nucleótidos cuyas bases nitrogenadas son del mismo tipo (púrica o pirimidínica) (A/G = C/T) y las transversiones se producen entre nucleótidos cuyas bases son de distinto tipo (A/C = A/T = G/C = G/T).
- **Tasas de los tipos de sustitución:** Los tipos de sustituciones definidos en el apartado anterior ocurren con una probabilidad (tasa) que puede ser definida por el usuario, calculada empíricamente o bien estimada como un valor de probabilidad. Los programas filogenéticos denominan esta variable como *substitution rate matrix* y en ella se incluyen tantas tasas (*rates*) como tipos de sustituciones definidas por la variable anterior (nst).

- **Proporción de sitios invariantes:** Un número elevado de posiciones invariantes reduce las posibilidades de obtener un resultado filogenético fiable (ya que son pocos los elementos variables de 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, fenómeno denominado saturación.

Métodos de reconstrucción de árboles filogenéticos

Una vez establecidas las distancias genéticas que separan las secuencias de estudio de acuerdo con un modelo de sustitución nucleotídica, el resultado se muestra gráficamente en forma de dendrograma o árbol filogenético, reflejando las relaciones evolutivas que vinculan a las secuencias consideradas.

Entre los métodos más utilizados destacamos:

- **Método UPGMA (*Unweighted pair-group method using arithmetic mean*) (Sneath y Sokal, 1973):** este método comienza agrupando las dos secuencias más cercanas de acuerdo con la matriz de distancias genéticas. A partir de ahí, cada nueva incorporación se efectuará añadiendo la secuencia que muestre menor distancia con el grupo ya incorporado, considerándose que la distancia genética entre éste y las demás secuencias es la media de las distancias que existían entre las secuencias que lo integran.
- **Método NJ (*Neighbor-joining*) (Saitou y Nei, 1987):** este método comienza analizando 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.
- **Método ML (*Maximum likelihood*) (Felsenstein, 1981):** este método es similar al anterior, aunque en este caso no se evalúa la longitud del árbol final, sino la **probabilidad** de obtenerlo a partir de la 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*), que se expresa en forma de logaritmo neperiano negativo ($-\ln L$), de manera que este valor disminuye al aumentar la probabilidad del árbol. Los valores que toman las variables que definen el modelo evolutivo pueden ser estimadas durante el proceso o bien pueden ser tomadas de un árbol NJ previamente construido.

- **Método Bayesiano (BI, *bayesian inference*) (Huelsenbeck y Ronquist, 2001):** Este método expresa lo bueno que es el resultado con un valor de probabilidad y supone un avance frente a otros métodos similares porque es capaz de modificar los parámetros de partida (los valores de las variables que definen el modelo de sustitución nucleotídica) *a posteriori* (una vez alcanzado el resultado) con el fin de mejorar ese valor de probabilidad final. Para ello comienza construyendo un árbol a partir de la matriz de secuencias de ADN y de acuerdo con un modelo de sustitución nucleotídica programado por el usuario cuyas variables toman un valor cualquiera. A continuación calcula la probabilidad de ese árbol y a partir de ahí tiene lugar una secuencia de análisis denominada *Markov Chain Monte Carlo*, que consiste en repetir el proceso inicial un número indeterminado de veces (generaciones), en cada una de las cuales se modifican ‘intencionadamente’ los parámetros de partida con el fin de mejorar el resultado final (el árbol), o más bien su valor de probabilidad.

OBJETIVOS

El estudio de los genes de histocompatibilidad en aves silvestres *Passeriformes*, es importante ya que se compara la evolución del sistema en éstos modelos naturales con los modelos “artificiales” (humanos, ratones, pollos) y otros vertebrados. Entendemos como “artificiales” a poblaciones pequeñas que por consanguinidad han crecido rápidamente, haciéndose poblaciones relativamente grandes.

En este trabajo nos hemos centrado en las aves silvestres del orden *Passeriformes*, principalmente de los géneros *Carduelis* y *Serinus*.

Estos géneros, han sido elegidos como modelos animales naturales de estudio ya que:

- Presentan un número elevado de especies (35 y 32 especies respectivamente) (Sibley y Ahlquist, 1990; Sibley y Monroe, 1990; Sibley y Monroe, 1993).
- Su distribución geográfica es muy amplia.
- Presenta unos rasgos fenotípicos que, en general, permiten una fácil diferenciación e identificación de las especies.

Los objetivos básicos de este trabajo son:

❖ **A.-**

1. Comparación evolutiva del Sistema Principal de Histocompatibilidad de clase I en aves silvestres *Passeriformes* de los géneros *Carduelis* y *Serinus*.
2. Estudiar la variabilidad de estos genes comparando las secuencias de ADN y de aminoácidos entre sí, para determinar la localización de las posiciones variables y el tipo de sustituciones que experimentan.
3. Determinar el tipo de presión evolutiva al que están sometidas las moléculas de histocompatibilidad estudiadas, analizando las tasas de sustitución sinónima y no sinónima y número de transiciones y transversiones, en las distintas regiones de las secuencias de ADN.
4. Estudiar los residuos aminoacídicos exónicos conservados y no conservados en las especies de vertebrados escogidas, y comparar la existencia de estos residuos en las aves *Passeriformes* seleccionadas.

5. Comparar los intrones de las secuencias obtenidas con las secuencias de otros vertebrados, entre ellas aves del orden *Galliformes*, con el fin de comprobar si el MHC en estas aves de canto es similar al “MHC mínimo esencial” encontrado en el pollo (*Gallus gallus*).

❖ B.-

6. Comprobar mediante análisis filogenético del citocromo b mitocondrial, la concordancia de la agrupación filogenética molecular con la clásica agrupación fenotípica en los siguientes géneros de aves *Passeriformes*:
 - *Rhodopechys*
 - *Carpodacus*
 - *Pyrrhula*
 - *Haematospiza*
 - *Pinicola*
 - *Uragus*
7. Obtener mediante los análisis de ADN mitocondrial unos datos más precisos que los de la filogenia fenotípica, para estudiar la evolución molecular del Sistema Principal de Histocompatibilidad."

MATERIAL Y MÉTODOS

1. Muestras empleadas

1.1 Especies estudiadas

Estudio de MHC clase I

En este trabajo se han utilizado muestras de 28 especies de aves *Passeriformes* diferentes (26 muestras propias). Los individuos utilizados fueron muestreados a lo largo de la geografía mundial. En la Tabla M1 se observa un resumen de los individuos considerados y de la procedencia de los mismos. Para la comparación del MHC de clase I con otras especies de vertebrados se utilizaron secuencias obtenidas de la base de datos GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>), en la tabla M1 están indicados los códigos de acceso a cada una de las secuencias.

Nombre	Clase/Orden	GenBank	Procedencia
<i>Danio rerio</i>	Actinopterygii	AAF20179	*GenBank
<i>Ameiva ameiva</i>	Sauropsida	AAA48519	*GenBank
<i>Xenopus laevis</i>	Amphibia	AF185582	*GenBank
<i>Bos taurus x Bos indicus</i>	Mammalia	ABW70136	*GenBank
<i>Canis familiaris</i>	Mammalia	NP_001014767	*GenBank
<i>Equus caballus</i>	Mammalia	NP_001075976	*GenBank
<i>Mus musculus</i>	Mammalia	AAY85367	*GenBank
<i>Ovis aries</i>	Mammalia	CAJ57269	*GenBank
<i>Rattus norvegicus</i>	Mammalia	CAA74333	*GenBank
<i>Sus scrofa</i>	Mammalia	ACA33862	*GenBank
<i>Hylobates lar</i>	Mammalia	AAB08074	*GenBank
<i>Pongo pygmaeus</i>	Mammalia	AAK67485	*GenBank
<i>Gorilla gorilla</i>	Mammalia	CAA43100	*GenBank
<i>Pan troglodytes</i>	Mammalia	BAC78189	*GenBank
<i>Pan paniscus</i>	Mammalia	AAY59433	*GenBank
<i>Homo sapiens (HLA-A)</i>	Mammalia	BAA07530	*GenBank
<i>Homo sapiens (HLA-B)</i>	Mammalia	CAA06616	*GenBank
<i>Homo sapiens (HLA-C)</i>	Mammalia	CAB02408	*GenBank
<i>Homo sapiens (HLA-E)</i>	Mammalia	CAA40172	*GenBank
<i>Homo sapiens (HLA-F)</i>	Mammalia	BAA32747	*GenBank
<i>Homo sapiens (HLA-G)</i>	Mammalia	CAA43298	*GenBank
<i>Gallus gallus</i>	Aves-Galliformes	NP_001026509	*GenBank
<i>Coturnix japonica</i>	Aves-Galliformes	BAA83672	*GenBank
<i>Acrocephalus arundinaceus</i>	Aves -Passeriformes	CAA06566	*GenBank
<i>Carduelis atrata</i>	Aves -Passeriformes	DQ257462	Sucre, Bolivia
<i>Carduelis carduelis parva</i>	Aves -Passeriformes	FJ266447	Alcalá de Henares, España
<i>Carduelis citrinella</i>	Aves -Passeriformes	DQ257482	Alcalá de Henares, España
<i>Carduelis crassirostris</i>	Aves -Passeriformes	DQ257464	Pisac, Perú
<i>Carduelis cucullata</i>	Aves -Passeriformes	DQ257465	Mérida, Venezuela
<i>Carduelis dominicensis</i>	Aves -Passeriformes	HM041048	Constanza, Rep. Dominicana

<i>Carduelis lawrencei</i>	Aves -Passeriformes	HM041049	California, USA
<i>Carduelis magellanica</i>	Aves -Passeriformes	DQ257467	Lima, Perú
<i>Carduelis notata</i>	Aves -Passeriformes	DQ257468	Chiapas, México
<i>Carduelis olivacea</i>	Aves -Passeriformes	DQ257470	Lima, Perú
<i>Carduelis pinus</i>	Aves -Passeriformes	FJ266376	Jackson, USA
<i>Carduelis psaltria</i>	Aves -Passeriformes	HM041050	Sacramento, USA
<i>Carduelis spinescens</i>	Aves -Passeriformes	DQ257472	Mérida, Venezuela
<i>Carduelis spinus</i>	Aves -Passeriformes	FJ266399	Madrid, España
<i>Carduelis xanthogastra</i>	Aves -Passeriformes	DQ257473	San José, Costa Rica
<i>Carduelis yarrellii</i>	Aves -Passeriformes	DQ257475	Recife, Brasil
<i>Serinus atrogularis</i>	Aves -Passeriformes	DQ257479	Cape Town, Sudáfrica
<i>Serinus canaria</i>	Aves -Passeriformes	DQ257480	Isla de la Palma, España
<i>Serinus citrinelloides</i>	Aves -Passeriformes	DQ257484	Nairobi, Kenia
<i>Serinus dorsostriatus</i>	Aves -Passeriformes	DQ257486	Dar es Salam, Tanzania
<i>Serinus flaviventris</i>	Aves -Passeriformes	DQ257487	Cape Town, Sudáfrica
<i>Serinus gularis</i>	Aves -Passeriformes	DQ257489	Cape Town, Sudáfrica
<i>Serinus mozambicus</i>	Aves -Passeriformes	DQ257491	Dar es Salam, Tanzania
<i>Serinus striolatus</i>	Aves -Passeriformes	DQ257493	Nairobi, Kenia
<i>Serinus sulphuratus</i>	Aves -Passeriformes	DQ257494	Cape Town, Sudáfrica
<i>Serinus thibetanus</i>	Aves -Passeriformes	DQ257496	Hong Kong, China
<i>Taeniopygia guttata</i>	Aves -Passeriformes	XM_002186531	*GenBank

Tabla M1. Muestras utilizadas en este estudio. *Para hacer una comparativa del Sistema Principal de Histocompatibilidad de aves *Passeriformes* con el resto de vertebrados se utilizaron algunas muestras obtenidas en la base de datos Genbank.

Estudio filogenético basado en el gen Citocromo b

El estudio de las relaciones evolutivas existentes entre las especies estudiadas es necesario para interpretar el proceso evolutivo al que han estado sometidas las secuencias de histocompatibilidad obtenidas en este trabajo. Ya que las secuencias MHC no están recomendadas para analizar relaciones filogenéticas, debido a su peculiar modo de evolución, se recurrió al gen mitocondrial, citocromo b, para analizar la filogenia. En la tabla M2 están reflejadas todas las especies de la colección del Dr. Arnaiz, utilizadas.

Se puede destacar que en este trabajo se ha secuenciado por primera vez el citocromo b mitocondrial de la especie *Rhodopechys sanguineus* (Arnaiz-Villena *et al.*, 2014a).

Tabla M2. Especies utilizadas en el estudio filogenético. A esta tabla se le han añadido muestras de otras especies, con el fin de completar el estudio. La especie sombreada en verde (*Rhodopechys sanguineus*, ha sido por primera vez secuenciada en este trabajo (Arnaiz-Villena *et al.*, 2014a).

Nombre	GenBank	Procedencia
<i>Carduelis atrata</i>	L76385	Sucre, Bolivia
<i>Carduelis ambigua</i>	U78322	Szechwan, China
<i>Carduelis atriceps</i>	AF342863	Quetzaltenango, Guatemala
<i>Carduelis barbata</i>	L77868	Magallanes, Chile
<i>Carduelis carduelis caniceps</i>	L76388	Katmandu, Nepal

<i>Carduelis carduelis parva</i>	L76387	Madrid, España
<i>Carduelis chloris</i>	L76297	Madrid, España
<i>Carduelis citrinella citrinella</i>	L77872	Madrid, España
<i>Carduelis citrinella corsicanus</i>	AY583725	Sardinia, Italia
<i>Carduelis crassirostris</i>	L77869	Pisac, Perú
<i>Carduelis cucullata</i>	L76299	Mérida, Venezuela
<i>Carduelis dominicensis</i>	AF342864	Constanza, República Dominicana
<i>Carduelis flammea</i>	L76386	Bruselas, Bélgica
<i>Carduelis hornemanni</i>	U83201	Antwerp, Bélgica
<i>Carduelis lawrencei</i>	L76392	San Diego, USA
<i>Carduelis magellanica</i>	U79016	Lima, Perú
<i>Carduelis notata</i>	U79019	Xalapa, México
<i>Carduelis olivacea</i>	L77871	Lima, Perú
<i>Carduelis pinus pinus</i>	U79020	Jackson, USA
<i>Carduelis pinus perplexus</i>	DQ246804	Quetzaltenango, Guatemala
<i>Carduelis psaltria colombiana</i>	U78324	Maracay, Venezuela
<i>Carduelis psaltria hesperophilus</i>	L76390	Sacramento, USA
<i>Carduelis sinica</i>	L76592	Szechwan, China
<i>Carduelis spinescens</i>	U79017	Mérida, Venezuela
<i>Carduelis spinoides</i>	U79018	Katmandú, Nepal
<i>Carduelis spinus</i>	L76391	Madrid, España
<i>Carduelis tristis</i>	U79022	San Francisco, USA
<i>Carduelis xanthogastra</i>	L76389	San José, Costa Rica
<i>Carduelis yarrellii</i>	U83200	Recife, Brasil
<i>Carpodacus erythrinus</i>	AF342883	Islamabad, Pakistan
<i>Carpodacus nipalensis</i>	AF342866	Katmandú, Nepal
<i>Carpodacus roseus</i>	AF342867	Beijing, China
<i>Carpodacus rubicilloides</i>	AF342868	Katmandu, Nepal
<i>Carpodacus thura</i>	AF342869	Katmandu, Nepal
<i>Carpodacus trifasciatus</i>	AF342870	Szechwan, China
<i>Fringilla coelebs</i>	L76609	Madrid, España
<i>Haematospiza sipahi</i>	AF342875	Katmandú, Nepal
<i>Leucosticte arctoa arctoa</i>	DQ257460	Ulaanbaatar, Mongolia
<i>Leucosticte arctoa tephrocotis</i>	AY156385	Islas Pribilof , USA
<i>Loxia curvirostra curvirostra</i>	AF342876	Alcalá de Henares, España
<i>Loxia curvirostra japonica</i>	AF342877	Beijing, China
<i>Loxia leucoptera</i>	AF342878	Siberia, Rusia
<i>Pinicola enucleator</i>	AF342882	Novosibirsk, Rusia
<i>Pyrrhula erythaca</i>	AF342862	Beijing, China
<i>Pyrrhula nipalensis</i>	AF342884	Katmandu, Nepal
<i>Pyrrhula pyrrhula cineracea</i>	AF342886	Novosibirsk, Rusia
<i>Pyrrhula pyrrhula griseiventris</i>	AF342881	Beijing, China
<i>Pyrrhula pyrrhula iberiae</i>	AF342885	Santander, España
<i>Rhodopechys githaginea</i>	AF342887	Isla de Gran Canaria, España
<i>Rhodopechys mongolica</i>	AF342888	Gilgit, Pakistan
<i>Rhodopechys obsoleta</i>	AF342889	Kabul, Afghanistan
<i>Rhodopechys sanguineus</i>	KJ608059	Monte Hermon, Israel
<i>Serinus alario</i>	L76276	Capetown, Sudáfrica

<i>Serinus albogularis</i>	L78705	Capetown, Sudáfrica
<i>Serinus atrogularis</i>	L76267	Islas Canarias, España
<i>Serinus canaria</i>	L76266	Islas Canarias, España
<i>Serinus canicollis</i>	L78706	Capetown, Sudáfrica
<i>Serinus citrinelloides</i>	L77555	Nairobi, Kenya
<i>Serinus citrinipectus</i>	L78707	Maputo, Mozambique
<i>Serinus dorsostrigatus</i>	L76278	Dar es Salam, Tanzania
<i>Serinus flaviventris</i>	L76280	Capetown, Sudáfrica
<i>Serinus gularis</i>	L77556	Capetown, Sudáfrica
<i>Serinus leucopygius</i>	L76264	Dakar, Senegal
<i>Serinus mozambicus</i>	L76265	Dar es Salam, Tanzania
<i>Serinus pusillus</i>	L77873	Sin Wiang, China
<i>Serinus serinus</i>	L76263	Madrid, España
<i>Serinus striolatus</i>	L77557	Nairobi, Kenya
<i>Serinus sulphuratus</i>	L76294	Capetown, Sudáfrica
<i>Serinus syriacus</i>	AY570547	Mount Hermon, Israel
<i>Serinus totta</i>	AY570548	Cape Town, Sudáfrica
<i>Uragus sibiricus</i>	AF365877	Beijing, China

1.2 Obtención de las muestras

Todas las muestras empleadas para el desarrollo del presente estudio, pertenecen a la colección que el Dr. Arnaiz Villena, que ha ido reuniendo a lo largo de los últimos 20 años y corresponden a individuos salvajes capturados en sus entornos naturales.

Se trata de muestras de sangre (unas tres gotas) que fueron obtenidas mediante el corte de una uña tras anestesiarse la zona con una crema de lidocaína (EMLA, Laboratorios Astra, Suecia).

La sangre fue depositada en tubos estériles de 3 ml que contenían una solución salina formada por cloruro sódico 5M y dimetilsulfóxido (DMSO) en una proporción de 3:1. Las muestras se conservaron a 4°C hasta ser almacenadas a -20°C.

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, lo que la mantiene erguida y tensa. Las aves vuelan hacia la red y generalmente caen en las bolsas y se enredan. El tamaño de malla utilizado fue de 35 mm, que es apropiado para la captura de aves de tamaño pequeño a moderado. 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, la proximidad al sitio de procesado, la pendiente y el tipo de suelo. La colocación de las redes se hizo entre dos personas. Se utilizaron postes de

metal para sostener las redes. El procedimiento general para levantar los postes para la red se pueden consultar en la Guía de Estudio del Anillador de Norteamérica (NABC, 2003). Una vez instaladas, las redes eran revisadas cada 20 minutos. La extracción de las aves de las redes fue realizada por miembros de los grupos locales de anillamiento y de acuerdo con las leyes vigentes en cada país.

2. Extracción de ADN

Como se sabe, los glóbulos rojos son las células sanguíneas mayoritarias. En el caso de las aves son células nucleadas, de manera que la mayor parte del ADN presente en la sangre se obtiene de ellos. Gracias a ésta característica se puede obtener una buena cantidad de ADN a partir de volúmenes muy pequeños de sangre. Las muestras que se han procesado para este trabajo contaban con un volumen máximo de unos 2 ml de (en ocasiones bastante coagulada), de forma que se optó por emplear un método de extracción de ADN automático para garantizar el mayor rendimiento posible con el mínimo gasto de muestra.

El proceso se desarrolla en un extractor automático (*Nucleic Acid Extraction System, QuickGene-810, FUJIFILM, Japón*) después de someter las muestras a un tratamiento previo con un kit comercial (*QuickGene Whole Blood Extraction Kit S, FUJIFILM, Japón*) indicado para la extracción de ADN a partir de sangre total. El proceso de extracción consiste en:

- Diluir 100 µl de sangre en 100 µl de solución salina (*PBS pH 7.4, GIBCO, INVITROGEN, Auckland, USA*) en un tubo de 1.5 ml.
- Añadir 30 µl de proteasa (*DNA whole blood Protease EDB-01, FUJIFILM, Japón*).
- Añadir 250 µl de solución de lisis (*DNA whole blood Lysis Buffer LDB-04, FUJIFILM, Japón*). Mezclar bien.
- Incubar en un baño a 56°C hasta que desaparezcan los posibles coágulos (2-10 min).
- Añadir 250 µl de etanol absoluto (*Ethanol Absolute PRS - CH₃CH₂OH, PANREAC, España*). Mezclar bien.

En la mayoría de los casos sobre todo en las muestras de aves donde existe una gran cantidad de ADN, la adición de etanol absoluto provoca 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, ya que

quedaría retenida en la membrana y se perdería.

La medusa debe, por tanto, ser procesada manualmente sometiéndola a un lavado con etanol absoluto (*Ethanol Absolute PRS - CH₃CH₂OH, PANREAC, España*) y otro con etanol 70% (*Ethanol 70% v/v BP CODEX - CH₃CH₂OH, PANREAC, España*). 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, Japón*) suministrada con el kit de extracción.

- 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 sus características hidrófilas descarta proteínas y lípidos.
- Introducir la columna en el extractor automático, donde será sometida a 3 ciclos de lavado (*DNA whole blood Wash Buffer WDB-03, FUJIFILM, Japón*) y posteriormente será diluida en 200 µl de una solución de elución (*DNA whole blood Elution Buffer CDB-02, FUJIFILM, Japón*). Una vez completado el proceso, se obtiene en el tubo colector suministrado por la casa, entre unos 150 a 200 µl de ADN.

La concentración y la pureza del ADN resultante fue medida en espectrofotómetro (*ND-1000 spectrophotometer, NanoDrop, DE, USA*), utilizando para ello tan solo 2 µl del material genético obtenido.

Este espectrofotómetro mide la absorbancia de la muestra a 260, 280 y 230 nm. Como sabemos, 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:

$$1 A_{260} \text{ nm} = 1 \text{ DO ADN} = 50 \text{ ng}/\mu\text{l}$$

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.

3. Obtención de secuencias de ADN de los genes de MHC de clase I

Como ya se ha descrito anteriormente, el Sistema Principal de Histocompatibilidad se caracteriza por estar constituido por varios genes que pueden presentar un polimorfismo elevado. El estudio de la variabilidad implica identificar las diferentes secuencias de ADN (Olivo *et al.*, 1996) presentes en cada individuo (y en cada población), para lo cual se recurrió a técnicas de amplificación, secuenciación y clonación.

3.1 Amplificación de secuencias de MHC de clase I

El estudio de secuencias de ADN requiere disponer de un número de copias muy superior al obtenido tras el proceso de extracción, para lo cual se hace necesario obtener numerosas réplicas de la secuencia deseada. Este proceso es conocido como amplificación.

La reacción en cadena de la polimerasa, conocida como PCR (*polymerase chain reaction*), es 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. La PCR está diseñada según el principio natural de replicación del ADN. Es un proceso varios pasos, designado como un ciclo, que se repite un número específico de veces (Figura M3).

- ***Inicio***

Este paso consiste en llevar la reacción hasta una temperatura de 94-96 °C (ó 98 °C si se está usando una polimerasa termoestable extrema), que se mantiene durante 1-9 minutos. Esto sólo es necesario para ADN polimerasas que requieran activación por calor.

- ***Desnaturalización***

Es el proceso mediante el cual se separa la doble hebra de ADN en dos cadenas. Este paso puede realizarse de diferentes modos, siendo el calentamiento (94-95 °C) de la muestra la forma más habitual. La temperatura a la cual se decide realizar la desnaturalización depende, por ejemplo, de la proporción de G+C que tenga la cadena, como también del largo de la misma.

- ***Anillamiento o hibridación***

Es la fase de unión de los cebadores con sus secuencias complementarias. Los cebadores o primers son pequeñas cadenas de nucleótidos a partir de las cuales la ADN polimerasa inicia la síntesis de una molécula nueva de ADN.

Se trata de secuencias sintéticas de oligonucleótidos que son utilizadas para reconocer por apareamiento complementario secuencias blanco en ADN. Suelen ser utilizados para definir los extremos del producto que se desea amplificar, y a partir de ellos la ADN polimerasa inicia la polimerización en dirección 5' - 3'.

Esta fase ocurre a una temperatura que oscila entre 50°C y 65°C en función de la composición nucleotídica y especificidad de los cebadores empleados. El tiempo requerido para ello suele ser superior a 30 segundos.

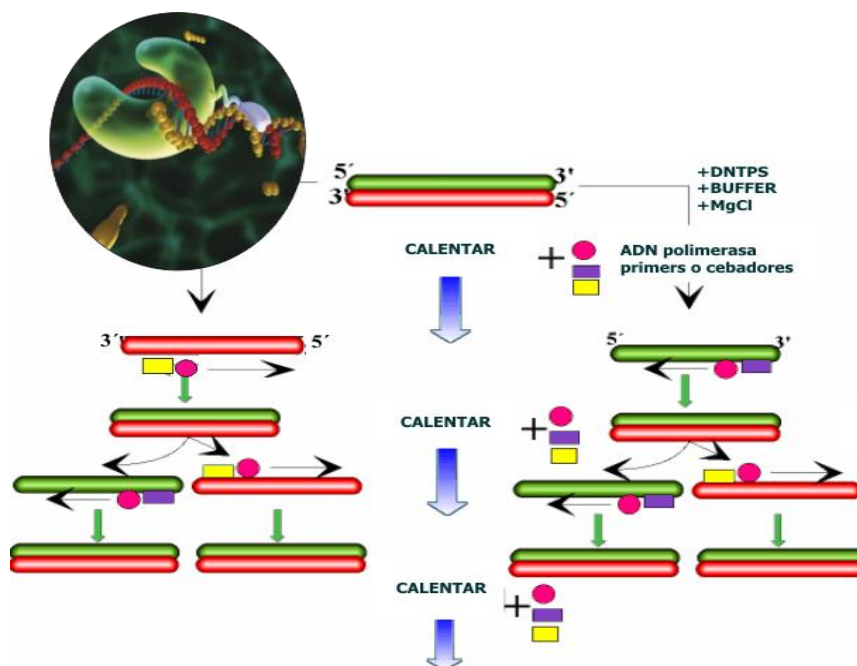


Figura M3. Esquema de Reacción en Cadena de la Polimerasa.

- **Elongación**

Es la fase donde actúa la ADN polimerasa, tomando el ADN molde para sintetizar la cadena complementaria y partiendo del cebador como soporte inicial necesario para la síntesis de nuevo ADN.

La polimerasa sintetiza una nueva hebra de ADN complementaria a la hebra molde añadiendo los dNTPs complementarios en dirección 5'→ 3', uniendo el grupo 5'-fosfato de los dNTPs con el grupo 3'-hidroxilo del final de la hebra de ADN creciente.

La temperatura para este paso depende del ADN polimerasa que usemos. Para la polimerasa Taq, la temperatura de máxima actividad está en 75-80 °C (comúnmente 72 °C). El tiempo de extensión depende tanto del ADN polimerasa usada como de la longitud del fragmento de ADN que se va a amplificar.

Como sabemos las proteínas de histocompatibilidad de clase I son polimórficas y registran la mayor variabilidad en los dominios extracelulares alfa 1 y alfa 2, encargados de la unión del péptido que será presentado a los linfocitos T.

Dichos dominios están codificados, respectivamente, en los exones 2 y 3 de los genes de clase I. Por ello, el estudio de la variabilidad de estos genes se centra en el análisis de las secuencias de ADN que incluyan dichos exones. En este trabajo se han amplificado secuencias que contienen gran parte del exón 2, el intrón 2 completo y casi todo el exón 3 (Figura M4).



FiguraM4. Esquema del fragmento del gen MHC clase I amplificado.

Las reacciones de amplificación se llevaron a cabo en un termociclador (*Mastercycler epgradient S, Eppendorf, Alemania*). La ADN polimerasa utilizada fue *Taq DNA Polymerase (Invitrogen, USA)*, suministrada junto con el tampón (*10x PCR Buffer, Minus Mg⁺⁺, Invitrogen, USA*), el cloruro de magnesio (*50mM Magnesium Chloride, Invitrogen, USA*) y los dNTPs (*100 mM dNTP Set, Invitrogen, USA*). El cebador directo (DIR) utilizado fue: DIR-5'-GTTCTCCACTCCCTGGATTACC-3' y el cebador inverso (INV) fue 5'- GCGCTCCAGCTCCTTCTGCCC_AGTA-3' (Westerdahl *et al.*, 1999). Estos cebadores fueron suministrados por la casa Integrated DNA Technologies (<http://eu.idtdna.com>).

Preparado mezcla para la reacción PCR

<i>Reactivos</i>	<i>Concentración</i>	<i>Volumen</i>
dNTPs	25 mM	1.6 µl
Buffer	10x	2 µl
MgCl₂	2.5 mM	2 µl
Cebador DIR	10 µM	0.8 µl
Cebador INV	10 µM	0.8 µl
Taq polimerasa	5U/ µl	0.3 µl
ADN	100-200 ng/ µl	1 µl
Agua		11.5 µl
	Total	20 µl

Tabla M5. Condiciones de amplificación de las secuencias de histocompatibilidad de clase I estudiadas. En esta tabla se indica la cantidad de cada reactivo necesaria (según la concentración dada) para un volumen de reacción de 20 µl por muestra.

<i>Nº ciclos</i>	<i>Tª</i>	<i>Tiempo</i>	<i>Proceso</i>
1	95°C	5 min	<i>Desnaturalización inicial</i>
	95 °C	10 seg	<i>Desnaturalización</i>
42	65°C	30 seg	<i>Hibridación</i>
	72 °C	60seg	<i>Elongación</i>
1	72 °C	10min	<i>Elongación final</i>

Tabla M6. Características del programa de amplificación empleado en el proceso.

3.2 Electroforesis en gel de agarosa de los fragmentos amplificados

Los procesos de amplificación no siempre son específicos, de forma que secuencias de ADN no deseadas pueden ser también amplificadas.

Por ello se hace necesario utilizar la técnica de 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.

En nuestro caso, los productos de la amplificación fueron sometidos al proceso de electroforesis en gel de agarosa (*Biotoools HR Agarose, High resolution, BIOTOOLS, España*) al 2% en tampón 1X TBE (*10X TBE Buffer, Invitrogen, UK*). En el proceso se incluyó un marcador de peso molecular (*DNA Molecular Weight Marker XIV - 100 bp ladder, ROCHE, Alemania*) 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 (*Olerup SSP GelRed Dropper Bottle, Suecia*) 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.

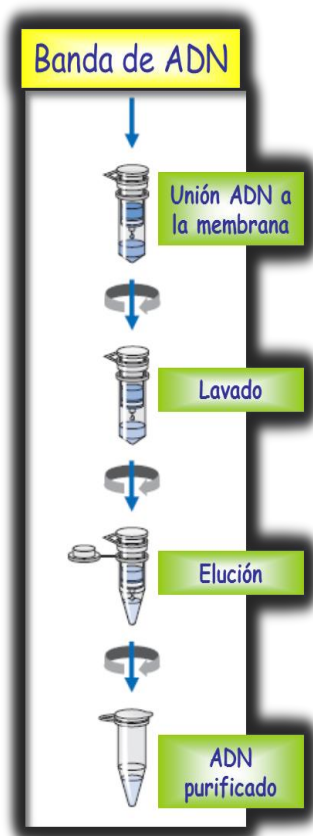
3.3 Purificación de la banda de la secuencia amplificada

El fragmento de ADN fue extraído del gel con un kit comercial de purificación (*MinElute Gel Extraction Kit (250), QIAGEN, Alemania*) (Figura M7).

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 tubo colector tras varios lavados.

El protocolo de purificación es el siguiente:

- Añadir 300-400 ml de Buffer QG - Solubilization Buffer (*QIAGEN, Alemania*) 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 minutos).
- Pasar el contenido del tubo a una columna del kit. Centrifugar 1 minuto a 10000 G y desechar el filtrado.
- Añadir 500 ml de Buffer QG - Solubilization Buffer (*QIAGEN, Alemania*). Centrifugar 1 minuto a 10000 G y desechar el filtrado.
- Añadir 750 ml de Buffer PE - Wash Buffer (*QIAGEN, Alemania*). Esperar 5 minutos.
- Centrifugar 1 minuto a 10000 G y desechar el filtrado.



- Centrifugar 1 minuto a 10000 G, desechar el filtrado y tirar el colector.
- Poner la columna en un tubo de 1.5 ml.
- Añadir 50 μ l de agua destilada y centrifugar 1 minuto a 10000 G.
- Tirar la columna y guardar el ADN purificado a 4°C o -20°C hasta su uso en secuenciación directa o en clonación.

Figura M7. Esquema del proceso de purificación de una banda de electroforesis mediante el kit comercial *MinElute Gel Extraction Kit* (QIAGEN, Alemania) basado en una columna con una membrana de sílice que “atrapa” el ADN, y limpia la muestra de todos los productos de la electroforesis.

3.4 Secuenciación de los fragmentos amplificados

La mayoría de los métodos de secuenciación automática utilizados actualmente son modificaciones del método desarrollado por Sanger y colaboradores a finales de los años 70. En este método, el ADN que va ser secuenciado funciona como un molde para la síntesis enzimática de un nuevo ADN que comienza en un sitio definido por la unión de un primer. En la reacción se utiliza una mezcla de desoxinucleótidos y didesoxinucleótidos a unas concentraciones que crean una probabilidad finita de que un didesoxinucleótido se incorpore en lugar del correspondiente desoxinucleótido en cada posición de la cadena que está siendo sintetizada.

La incorporación de un didesoxinucleótido bloquea la elongación de la cadena y esto da como resultado una población de fragmentos de ADN truncados de diferente longitud. La identidad del nucleótido que termina la cadena en cada posición se puede determinar bien realizando 4 reacciones de elongación separadas cada una de las cuales con un didesoxinucleótido (ddATP, ddCTP, ddTTP ddGTP) o con una única reacción de elongación combinando los 4 didesoxinucleótidos pero marcando cada uno de estos específicamente.

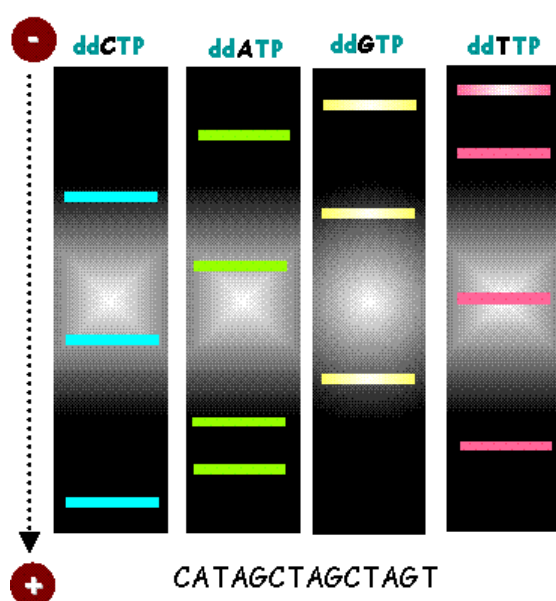


Figura M8. Ejemplo del resultado de las reacciones de secuenciación tradicional por el método de Sanger. Las reacciones se desarrollan en 4 tubos y en cada uno se añade, además de todos los reactivos necesarios, uno de los 4 nucleótidos dideoxi marcados (con radioactividad, por ejemplo). Los productos de secuenciación de cada tubo se cargan en una calle diferente en un gel de electroforesis. Una vez revelado el gel, se puede leer la secuencia 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 población de moléculas resultantes se separan por tamaños mediante electroforesis en acrilamida y la secuencia se obtiene correlacionando el orden de los fragmentos en la electroforesis con el didesoxinucleótido que termina cada uno de ellos (Figura M8).

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 4 tipos de nucleótidos dideoxi generalmente por fluorescencia; dicho marcaje es detectado directamente por el secuenciador.

En nuestro caso, se recurrió al servicio de secuenciación de la empresa SECUGEN S.L. (www.secugen.es), situada en el Centro de Investigaciones Biológicas (CIB-CSIC) de Madrid. Para desarrollar el proceso requieren el envío de 15 μ l del fragmento de ADN amplificado y purificado a una concentración de 10-20 ng/ μ l (para fragmentos de más de 500 pb), y 1.5 μ l de cebador a una concentración de 5 μ M. En nuestro caso se solicitó la secuenciación utilizando los mismos cebadores empleados en la amplificación (DIR e INV) ajustando la concentración a la requerida.

3.5 Clonación de los fragmentos amplificados

La calidad de los resultados obtenidos tras la secuenciación, depende en gran medida de la pureza y correcta cuantificación de las muestras de ADN analizadas. En muchas ocasiones los resultados nos dan ambigüedades, lo que significa que en determinadas posiciones ofrecen dos (o más) posibles alternativas.

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.

Estas ambigüedades se representan en el cromatograma como un doble pico (Figura M9).

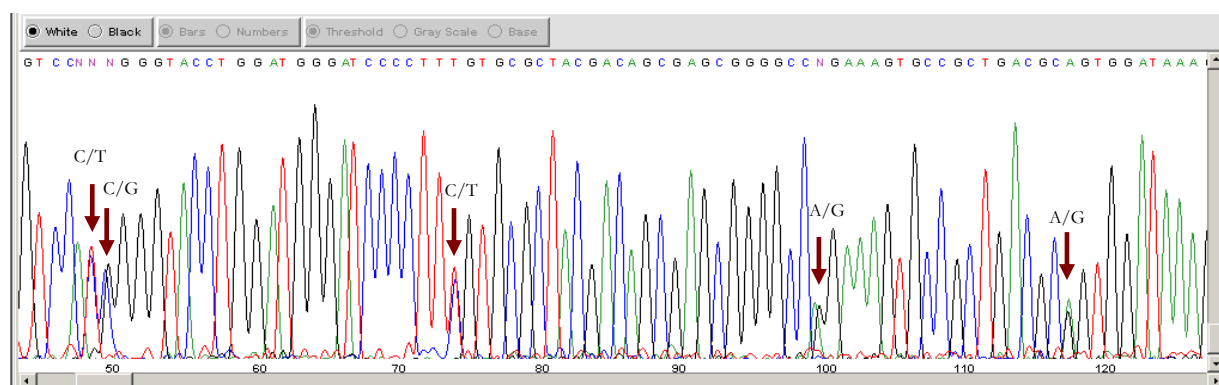


Figura M9. Ejemplo de ambigüedades representadas por dobles picos en un cromatograma.

Para resolver dichas ambigüedades, los productos de amplificación fueron sometidos a un proceso de clonación antes de ser definitivamente secuenciados. La clonación se basa en la capacidad natural que tienen algunas bacterias para incorporar pequeñas moléculas circulares de ADN (plásmidos) presentes en su entorno en un proceso denominado 'transformación bacteriana'. El ADN exógeno se encuentra en el ambiente y se introduce a través de la membrana de la célula bacteriana. Para que ocurra la transformación, la bacteria debe estar en un estado de competencia. La transformación es uno de los tres procesos por los que el material genético exógeno se puede introducir en una célula bacteriana. Los otros dos son la conjugación y la transducción.

Los plásmidos a menudo contienen genes o paquetes de genes que les confieren una ventaja selectiva, lo que les da la habilidad entre otras de hacer a la bacteria resistente a los antibióticos. Cada plásmido contiene al menos una secuencia de ADN que sirve como un origen de replicación u ORI (un punto inicial para la replicación del ADN), lo cual habilita al ADN para ser duplicado independientemente del ADN cromosómico. Estas propiedades bacterianas, sumadas a la posibilidad de diseñar plásmidos artificiales (denominados vectores de forma genérica) e insertar en ellos fragmentos de ADN mediante técnicas de ingeniería genética hacen de la clonación una herramienta muy útil para aislar y caracterizar secuencias de ADN de forma individual.

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 M10. La utilidad de este

vector se debe a la presencia de dos marcadores; el primero es un gen de resistencia a ampicilina (que permite seleccionar las bacterias que han incorporado el plásmido, con o sin fragmento); el segundo es el lugar de inserción localizado en la mitad del gen *lacZ*.

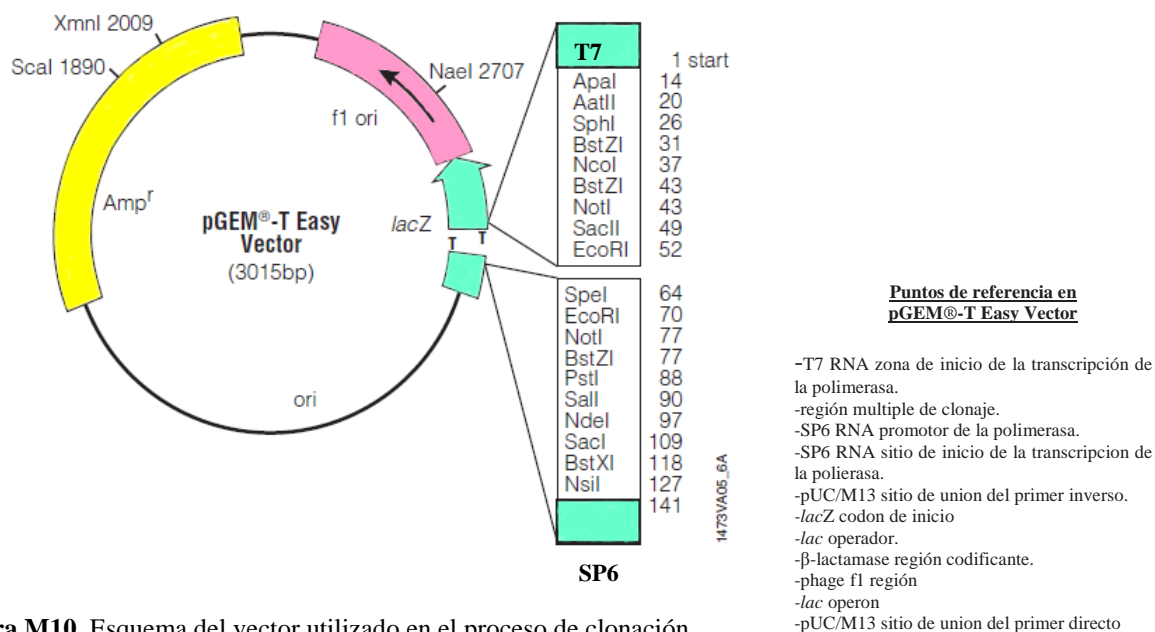


Figura M10. Esquema del vector utilizado en el proceso de clonación.

Este gen codifica la enzima β-galactosidasa, que cataliza la reacción de hidrólisis a galactosa y 5-bromo-4-cloro-3-hidroxiindol. Este último es oxidado a 5,5'-dibromo-4,4'-dicloro-índigo, un compuesto azul insoluble. De este modo, si X-gal y un inductor de la β-galactosidasa (normalmente IPTG) son disueltos en el medio de agar de una placa de cultivo, las colonias crecidas en la placa que posean un gen *lacZ* funcional, podrán ser claramente distinguidas por su coloración azul.

El vector plasmídico que porta el operón *lac*, se puede diseñar de modo que la secuencia del gen de interés a clonar se coloque específicamente en el marco de lectura del gen *lacZ*. Al insertar la secuencia de esta manera, se interrumpe el gen *lacZ* y el operón se vuelve incapaz de sintetizar la enzima β-galactosidasa funcional. Habitualmente se utilizan para este fin, la bacteria *E. coli*, que no pueden producir la enzima β-galactosidasa.

Una vez transformadas, las bacterias que portan el plásmido con el inserto no son capaces de producir la enzima β-galactosidasa debido a que la región que sintetiza la enzima se parte para dar paso a la secuencia recombinante.

Las bacterias que no tengan la secuencia recombinante pueden romper el X-gal presente en el agar del medio, tornándose sus colonias de color azul.

El protocolo de clonación empleado en este estudio fue el siguiente (Figura M11):

- 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 90 minutos a temperatura ambiente o bien 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 gr/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). 1 litro de medio contiene 35 gr de LB Agar (*LB AGAR LENNOX*, PRONADISA), 10 ml de ampicilina 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”

3.6 Obtención de las secuencias de ADN de los fragmentos clonados

Una vez incubadas las placas se seleccionaron 20 colonias blancas por muestra para extraer el ADN bacteriano siguiendo este protocolo:

- 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 (*Mastercycler epgradient S*, Eppendorf, Alemania).
- Incubar durante 5 minutos en hielo.
- Centrifugar 3 minutos a 12000 G.
- 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 3.1). Los productos de amplificación fueron sometidos a electroforesis para separar el fragmento correspondiente (ver apartado 3.2), se realizó una nueva purificación (ver apartado 3.3) y finalmente fueron secuenciados (ver apartado 3.4).

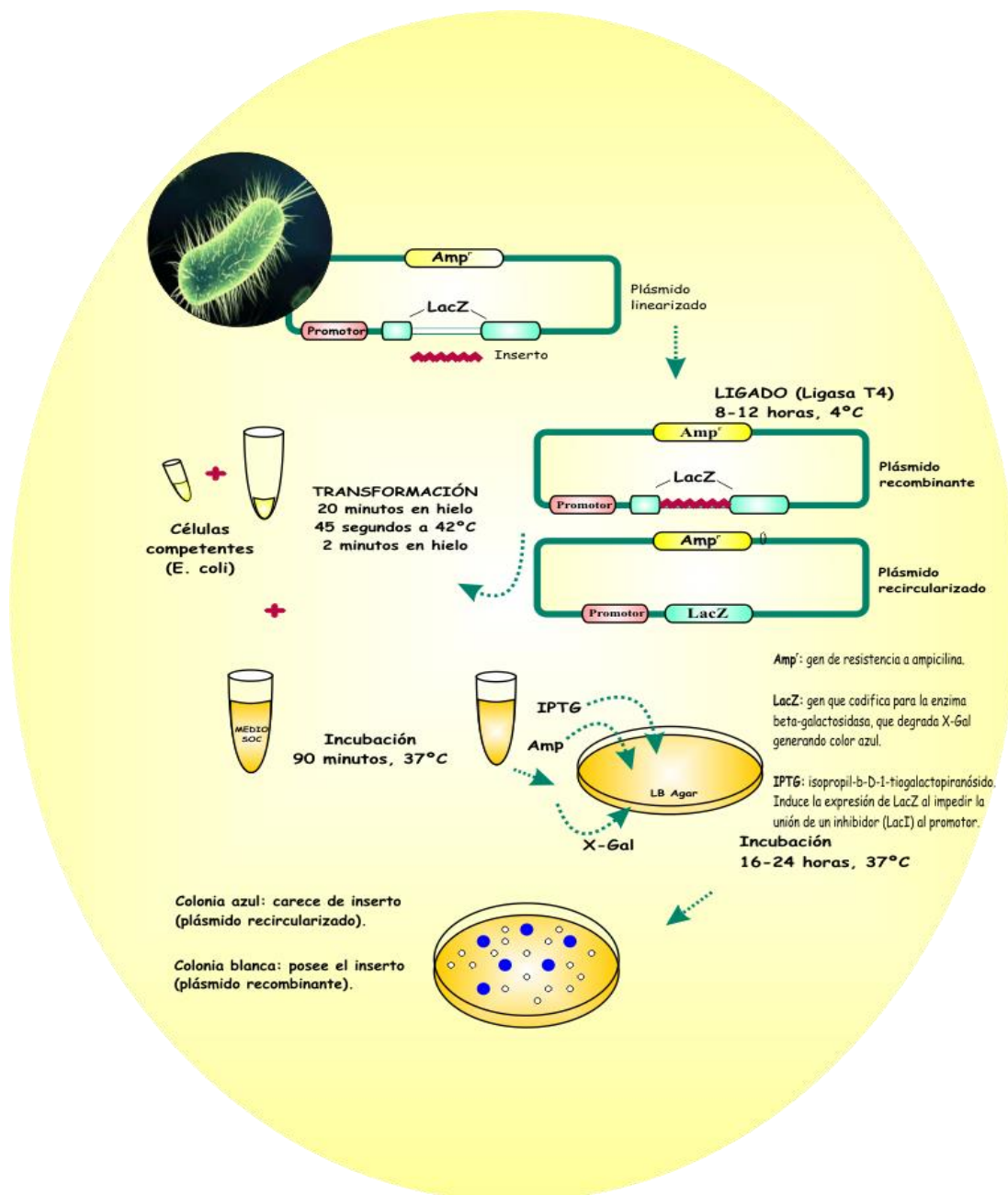


Figura M11. Esquema del proceso de clonación.

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

El estudio de las relaciones evolutivas existentes entre las especies estudiadas es necesario para interpretar el proceso evolutivo al que han estado sometidas las secuencias de histocompatibilidad obtenidas en este trabajo. Ya que las secuencias MHC no están recomendadas para analizar relaciones filogenéticas debido a su peculiar modo de evolución comentado anteriormente, se recurrió al gen mitocondrial citocromo b para analizar la filogenia de algunas de las muestras empleadas en este trabajo.

La obtención de las secuencias de citocromo b se llevaron a cabo en un termociclador (*Mastercycler egradient S Eppendorf, Alemania*).

La ADN polimerasa utilizada fue *Taq DNA Polymerase (Invitrogen, USA)*, suministrada junto con el tampón (*10x PCR Buffer, Minus Mg⁺⁺, Invitrogen, USA*), el cloruro de magnesio (*50mM Magnesium Chloride, Invitrogen, USA*) y los dNTPs (*100 mM dNTP Set, Invitrogen, USA*).

El cebador directo (DIR) utilizado fue: en el extremo 5' (Kocher *et al.*, 1989) 5'-AAAAAGCTTCCATCCAACATCTCAGCATGATGAAA-3' y el inverso (INV) (Edwards *et al.*, 1991): 5'ATGAAGGGATGTTCTACTGGTTG-3' Estos cebadores fueron suministrados por la casa Integrated DNA Technologies (<http://eu.idtdna.com>).

Preparado mezcla para la reacción PCR

<i>Reactivos</i>	<i>Concentración</i>	<i>Volumen</i>
dNTPs	2,5 mM	1.6 µl
Buffer	10x	2 µl
MgCl₂	25 mM	2 µl
Cebador DIR	10 µM	1.2 µl
Cebador INV	10 µM	1.2 µl
Taq polimerasa	5U/ µl	0.3 µl
ADN	100-200 ng/ µl	1 µl
Agua		10.7 µl
	Total	20 µl

Tabla M12. Condiciones de amplificación de secuencias de citocromo b mitocondrial. En esta tabla se indica la cantidad de cada reactivo necesaria (según la concentración dada) para un volumen de reacción de 20 µl por muestra.

Nº ciclos	Tª	Tiempo	Proceso
1	95°C	5 min	Desnaturalización inicial
	96 °C	30 seg	Desnaturalización
45	50°C	35 seg	Hibridación
	72 °C	1 min 10seg	Elongación
1	72 °C	20min	Elongación final

Tabla M13. Características del programa de amplificación empleado en el proceso.

Tras el proceso de amplificación bajo las condiciones descritas en las tablas M12 y M13, los productos de amplificación fueron sometidos a electroforesis para separar el fragmento deseado siguiendo el mismo protocolo descrito en el apartado 3.2.

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 3.3 y finalmente fueron secuenciadas (ver apartado 3.4).

5. Análisis de secuencias obtenidas

5.1 Caracterización de las secuencias de ADN

Cualquier estudio de filogenético o de evolución molecular basado en secuencias necesita de un alineamiento múltiple para determinar las correspondencias de homología a nivel de los residuos individuales o caracteres.

La verificación de que las secuencias de ADN obtenidas corresponden a genes MHC de clase I, se obtuvo mediante el alineamiento con otras secuencias previamente descritas con las que comparten una gran identidad. Para ello se utilizó el programa MEGA 6 (Tamura *et al.*, 2013) bajo la metodología de alineamiento MUSCLE.

Además, se hizo un cálculo de las frecuencias nucleotídicas ya que está comprobado que los animales vertebrados de sangre caliente poseen regiones de hasta 300 kb ricas en nucleótidos de guanina y citosina (G+C), y dichas regiones son más abundantes en los genomas de las aves que en los de mamíferos (Bernardi *et al.*, 1997).

Se ha comprobado que las regiones ricas en G+C son más estables frente a la desnaturalización térmica, lo que representa una ventaja para las aves, que muestran temperaturas y tasas metabólicas más elevadas (Bernardi y Bernardi, 1986). En grupos de aves de género *Serinus* el contenido G+C en el exón 2 supera el 60% (Lowy, 2003), y en el complejo B del pollo el contenido en G+C alcanza el 98% (Kaufman *et al.*, 1992). Sin

embargo, en la región HLA de clase I de humanos el porcentaje de G+C no llega al 50%.

Para analizar la distribución de frecuencias nucleotídicas en las secuencias estudiadas se ha utilizado el programa MEGA 6 (Tamura *et al.*, 2013), que permite calcular las frecuencias de cada nucleótido en cada región de la secuencia (exón 2 / intrón 2 / exón 3) y además posibilita el cálculo de las frecuencias nucleotídicas independizando las posiciones 1^a, 2^a y 3^a de los tripletes.

5.2 Caracterización de las secuencias de aminoácidos

Las proteínas de histocompatibilidad de clase I obtenidas en este trabajo fueron analizadas desde el punto de vista estructural y funcional. Las posiciones aminoacídicas importantes para la estructura y la función de la proteína fueron identificadas a partir de secuencias humanas en las que están bien definidos el sitio de reconocimiento antigénico (ARS) (Bjorkman *et al.*, 1987b), los aminoácidos conservados en vertebrados (Grossberger y Parham, 1992) y los que interaccionan con CD8 (Gao *et al.*, 1997; Sun *et al.*, 1995).

Para analizar la variabilidad en el sitio de reconocimiento antigénico se alinearon las secuencias de aminoácidos de nuestras muestras con las secuencias consenso de las proteínas clásicas (HLA-Ia) y no clásicas (HLA-Ib) humanas. Sobre ellas se marcaron, tal como fue descrito para el antígeno HLA-A2 (Bjorkman *et al.*, 1987b), los aminoácidos cuyas cadenas laterales son accesibles y conforman el denominado sitio de reconocimiento antigénico (ARS), en el que unos aminoácidos participan en la unión del péptido (los que tienen las cadenas laterales dirigidas hacia el interior de la valva) y otros interaccionan con el receptor de la célula T (TCR) (los que tienen las cadenas laterales dirigidas hacia arriba y hacia el exterior de la valva). Además se señalaron las posiciones que fueron definidas como hipervariables en las secuencias HLA de clase I clásicas humanas (HLA-Ia) (Parham *et al.*, 1988) y aquellas que dentro del ARS no varían y son responsables del anclaje del péptido a través de sus extremos (Bjorkman *et al.*, 1987b) (tablas R6 y R7).

La estructura tridimensional (Figura D1) se obtuvo con el programa SwissModel3.7 (<http://swissmodel.expasy.org/SWISS-MODEL.html>), que ofrece el modelo que mejor se ajusta a una secuencia de aminoácidos dada de entre todos los modelos que tiene registrados y que se han obtenido por difracción de rayos X a partir de moléculas cristalizadas de una secuencia conocida. Para comprobar que dicha estructura puede ser estable en las proteínas obtenidas se verificaron las posiciones que en los dominios alfa 1 y alfa 2 contribuyen a mantener la estabilidad conformacional de la molécula y que se encuentran conservados en

vertebrados (Grossberger y Parham, 1992). Dichas posiciones son en humanos: Thr10, Asp29, Gln96, Cys101, Gly120, Cys164; no se ha podido analizar una séptima posición conservada (His3) situada en el tramo inicial del exón 2 ya que aquí no ha sido secuenciada en todas las muestras.

5.3 Análisis de la variabilidad

Para estudiar el tipo de variabilidad que soportan las secuencias analizadas se construyeron diagramas de entropía y se analizaron las tasas de cambios sinónimos y no sinónimos en distintas regiones de la secuencia de ADN.

Diagramas de entropía

En un alineamiento de secuencias de ADN o de aminoácidos se puede definir la entropía en cada posición como un indicador de variabilidad. La entropía es una medida de la falta de contenido informativo de una posición dada, lo que implica la incapacidad de predecir qué nucleótido (o aminoácido) ocuparía esa posición si se añadiera una nueva secuencia. Por tanto, en las posiciones conservadas la entropía (y la variabilidad) es nula, y en las posiciones variables es tanto mayor cuanto mayor es el número de alternativas diferentes que presenta esa posición y más parecida la frecuencia con la que aparece cada alternativa. Los diagramas de entropía se construyeron a partir de las secuencias de ADN y de aminoácidos con el programa BioEdit (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>), en ellos aparecen representados los puntos de la secuencia que registran mayor variabilidad.

Cálculo de las tasas de sustitución sinónima y no sinónima

En teoría, cualquier mutación es igualmente probable en cualquier región de una secuencia de ADN. Sin embargo, cuando se trabaja con secuencias codificantes se observa un sesgo en la distribución de mutaciones y en el tipo.

Lo normal es que al comparar secuencias ortólogas (pertenecen al mismo gen en especies distintas) se observe un predominio de diferencias de tipo sinónimo (no provocan cambio de aminoácido). Esto es debido a que la mayoría de los genes expresan proteínas cuya función frecuentemente se ve alterada negativamente cuando sufre cambios de secuencia, por lo que desde el punto de vista evolutivo las mutaciones responsables de esos cambios no prosperan, mientras que las otras se mantienen y acumulan. En el caso de los sistemas genéticos polimórficos, como el MHC, ocurre lo contrario, dando la impresión de que la evolución

tiende a incrementar la variabilidad, al menos en las regiones de la molécula en las que la variabilidad es vital para su función.

Por este motivo es interesante el estudio de las tasas de sustitución de cada tipo en secuencias de histocompatibilidad, en las que está comprobado que en las posiciones que componen el sitio de reconocimiento antigénico (ARS) la tasa de cambios no sinónimos supera a la de cambios sinónimos (Hughes y Nei, 1988; Hughes y Nei, 1989b). Las tasas se calculan como el cociente entre el número de sustituciones sinónimas o no sinónimas y el número de sitios potencialmente sinónimos o no sinónimos.

El cálculo de los sitios sinónimos y no sinónimos se realiza con la fórmula de Nei-Gojobori (Nei y Gojobori, 1986), incluida en el programa MEGA 6 (Tamura *et al.*, 2013), que además de calcular el número de sitios de cada tipo calcula las tasas de sustituciones no sinónimas (dn) y sinónimas (ds).

A su vez fueron analizadas con este mismo programa las tasas de sustitución nucleotídica de transición y transversión en estas moléculas, para ello se empleó el modelo de sustitución nucleotídica Tamura-Nei (TN93), (Tamura y Nei, 1993).

5.4 Análisis de conservación de todas las secuencias estudiadas

Cálculo de distancias genéticas

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 calcular una distancia 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 6 (Tamura *et al.*, 2013).

Rango de conservación

Con el fin de obtener unos resultados más precisos sobre el grado de conservación de cada uno de los residuos de las secuencias de MHC de clase I estudiadas se calcularon en cada una de las posiciones este parámetro. Entendemos por rango de conservación (RC) al número total de residuos aminoacídicos observados por posición dividido entre el número de los diferentes residuos encontrados en cada posición. Los resultados se pueden observar en la tabla R10.

Porcentaje de similitud

Por último para poder completar el análisis de la conservación y evolución de las secuencias de MHC de clase I estudiadas, calculamos el porcentaje de similitud entre ellas con el programa BioEdit (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>).

6. Filogenia basada en el gen citocromo b

El estudio de las relaciones evolutivas existentes entre las especies estudiadas es necesario para interpretar el proceso evolutivo al que han estado sometidas, las secuencias de histocompatibilidad obtenidas en este trabajo.

Ya que las secuencias MHC no están recomendadas para analizar relaciones filogenéticas debido a su peculiar modo de evolución, se recurrió al gen mitocondrial citocromo b para analizar la filogenia.

Las secuencias utilizadas forman parte del archivo del grupo (ver Tabla M2) exceptuando una nueva, secuenciada por primera vez en este trabajo, de la especie *Rhodopechys sanguineus* (Arnaiz-Villena *et al.*, 2014a).

Para ello se han recurrido a dos métodos, uno es el método ML (Felsenstein, 1981) cuyo árbol ML fue construido con el programa PAUP4.0b10 (Swofford, 2002).

Como ya hemos comentado en la introducción, este método es similar al NJ (*Neighbor-joining*) (Saitou y Nei, 1987), aunque en este caso no se evalúa la longitud del árbol final, sino la probabilidad de obtenerlo a partir de la 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 de máxima verosimilitud 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

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.*, 2014a; Arnaiz-Villena *et al.*, 2008).

Posteriormente, para corroborar estos resultados, se hizo además un análisis bayesiano (Huelsenbeck y Ronquist, 2001).

Como sabemos, la inferencia bayesiana 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 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 cuales 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.

RESULTADOS

1. Secuencias MHC de clase I

En este trabajo se ha conseguido secuenciar un fragmento génico que equivale a casi todo el exón 2 (dominio $\alpha 1$ de la cadena pesada), al intrón 2 completo y a la mayor parte del exón 3 (dominio $\alpha 2$ de la cadena pesada) de moléculas de clase I del Sistema Principal de Histocompatibilidad.

Un total de 28 individuos de los géneros *Carduelis*, *Serinus*, *Acrocephalus* y *Taeniopygia* fueron escogidos para el estudio genético y evolutivo de esta molécula, aunque nos hemos centrado en los dos primeros sobretodo, ya que pertenecen a muestras propias de nuestro grupo.

A partir de muestras de sangre y otros tejidos, se obtuvo ADN, que fue amplificado y secuenciado como mínimo dos veces, obteniéndose idéntico resultado de cada réplica para cada individuo. Las secuencias de ADN obtenidas, una vez eliminados los cebadores, mostraban una longitud aproximada de 800 nucleótidos.

Para verificar que estas secuencias de ADN obtenidas correspondían a genes MHC de clase I fueron alineadas con otras secuencias previamente descritas en estos géneros, con las que comparten una gran identidad.

El alineamiento de estas secuencias se realizó con el programa MEGA 6 (Tamura *et al.*, 2013), con el método de alineamiento múltiple de secuencias MUSCLE.

Además, se hizo un cálculo de las frecuencias nucleotídicas ya que está comprobado que los animales vertebrados de sangre caliente poseen regiones de hasta 300 kb ricas en nucleótidos de guanina y citosina (G+C), y dichas regiones son más abundantes en los genomas de las aves que en los de mamíferos (Bernardi *et al.*, 1997).

En grupos de aves el contenido G+C en el exón 2 supera el 60% (Lowy, 2003), y en el complejo B del pollo el contenido en G+C alcanza el 98% (Kaufman *et al.*, 1992). Sin embargo, en la región HLA de clase I de humanos el porcentaje de G+C no llega al 50%.

Para analizar la distribución de frecuencias nucleotídicas en las secuencias estudiadas se utilizó el programa MEGA 6 que permite calcular las frecuencias de cada nucleótido en cada región de la secuencia (exón 2 / intrón 2 / exón 3) y además posibilita el cálculo de las frecuencias nucleotídicas independizando las posiciones 1ª, 2ª y 3ª de los tripletes. Estos resultados se muestran en los siguientes apartados.

1.1 Composición nucleotídica de las secuencias de ADN en aves *Passeriformes*

Se analizó la frecuencia de los distintos nucleótidos para todas las posiciones de los codones en conjunto y también se calculó la frecuencia desglosada para cada posición de los codones (1ª, 2ª y 3ª). El resultado obtenido se muestra en las tablas R1 y R2.

En la tabla R1 podemos observar la frecuencia de las cuatro bases en los géneros *Carduelis* y *Serinus*, considerando el exón 2 y el exón 3 en conjunto y comprobar que el porcentaje medio de G (33,3 %) es superior al de C la diferencia respecto a una distribución teórica al azar y que sería equiprobable (25 % cada nucleótido) es significativa al 0.1 %. Las A y las T alcanzan porcentajes que rondan el 22,5 y 18.5 respectivamente.

Tabla R1. Frecuencias nucleotídicas de las secuencias estudiadas

	T	C	A	G
<i>Carduelis lawrencei</i>	14,0	26,7	22,2	37,0
<i>Carduelis dominicensis</i>	19,3	25,4	22,5	32,8
<i>Carduelis psaltria</i>	19,0	25,6	22,7	32,7
<i>Serinus canaria</i>	17,8	25,0	22,3	34,9
<i>Carduelis atrata</i>	19,0	25,6	22,1	33,3
<i>Carduelis carduelis parva</i>	19,1	24,8	23,2	33,0
<i>Carduelis citrinella</i>	19,8	24,5	22,9	32,7
<i>Carduelis crassirostris</i>	18,9	25,2	22,7	33,2
<i>Carduelis cuculata</i>	19,0	25,4	22,7	32,9
<i>Carduelis magellanicus</i>	19,0	25,4	22,7	32,9
<i>Carduelis notata</i>	19,1	25,4	22,5	33,0
<i>Carduelis olivacea</i>	19,0	25,4	22,7	32,9
<i>Carduelis pinus</i>	18,9	25,2	22,5	33,4
<i>Carduelis spinescens</i>	18,6	25,5	22,5	33,4
<i>Carduelis spinus</i>	19,1	25,2	22,3	33,4
<i>Carduelis xanthogastra</i>	19,0	25,4	22,7	32,9
<i>Carduelis yarrelli</i>	18,8	25,6	22,3	33,3
<i>Serinus atrogularis</i>	19,8	25,2	21,7	33,3
<i>Serinus citrinelloides</i>	18,6	25,6	22,7	33,1
<i>Serinus dorsostriatus</i>	19,2	25,4	22,3	33,1
<i>Serinus flaviventris</i>	18,8	25,8	22,3	33,1
<i>Serinus gularis</i>	19,2	25,2	22,3	33,3
<i>Serinus mozambicus</i>	19,2	24,9	22,3	33,5
<i>Serinus striolatus</i>	18,6	26,0	22,7	32,7
<i>Serinus sulphuratus</i>	18,4	26,0	22,7	32,9
<i>Serinus thibetanus</i>	18,8	26,2	21,5	33,5

La composición nucleotídica de estas secuencias se ajusta a lo esperado para este tipo de genes en aves de canto. El porcentaje medio de G+C es de 57,8 % cuando se considera la secuencia completa.

Si se analiza por separado cada región se mantiene la misma tendencia, el porcentaje medio de G+C en el exón 2 (63,6%) es prácticamente igual al del intrón 2 (63,39%) aunque se observa un predominio en comparación con el exón 3 (53,5%) (no se han encontrado diferencias significativas entre las cantidades de G+C entre los dos exones, $p \geq 0.05$).

Por último, en el análisis por posiciones en los tripletes, se observa que el porcentaje de

G+C en las terceras posiciones es mayor en el exón 2 (61,9%) en comparación con el resto de regiones y de posiciones (Tabla R2).

Tabla R2. Frecuencias nucleotídicas medias de las secuencias de MHC de clase I en las aves de canto estudiadas.

	EXON 2			INTRON 2			EXON 3			Ex2 + In2 + Ex3					
	1 ^a	2 ^a	3 ^a	A	C	G	T	1 ^a	2 ^a	3 ^a	A	C	G	T	TOTAL
A	21,6	33,5	32,1	A	19,54	A	33,1	14,1	21	A	21,6	27,2	14,9	21,2	
C	26,7	19,5	10,2	C	17,07	C	17	30,8	25,3	C	21,9	18,7	26,3	22,3	
G	35,4	25,6	51,7	G	38,30	G	27,9	27,9	31,7	G	35,5	31,1	39,3	35,5	
T	16,3	21,4	6	T	25,08	T	22	27,2	22	T	21	23	19,5	21	
G+C		63,6		G+C	63,39	G+C		53,5		G+C		57,8			

1.2 Distribución de la variabilidad: diagramas de entropía

Para estudiar el tipo de variabilidad que soportan las secuencias analizadas se construyeron diagramas de entropía.

Estos diagramas, muestran que la variabilidad se acumula en los exones 2 y 3, mientras que el intrón 2 (no mostrado en la Figura R1) se mantiene prácticamente constante en todas las secuencias.

El exón 2 muestra siempre una mayor variabilidad que el exón 3 tanto en la comparación de todas las especies como en la comparación por separado de los distintos géneros (R.1a, b, c)

La mayoría de los sitios de unión de péptidos y de interacción con TCR (#) que conforman el ARS no coinciden con los picos de variabilidad (Bjorkman *et al.*, 1987b; Parham *et al.*, 1988).

Figura R1a. Diagrama de entropía correspondiente a las secuencias de ADN la molécula de MHC clase I, de las 26 muestras de aves silvestres de canto secuenciadas de los géneros *Carduelis* y *Serinus*. Se indica la ubicación aproximada de los sitios implicados en la unión de péptidos (Bjorkman *et al.*, 1987b; Parham *et al.*, 1988) y de la interacción con el receptor de la célula T (#).

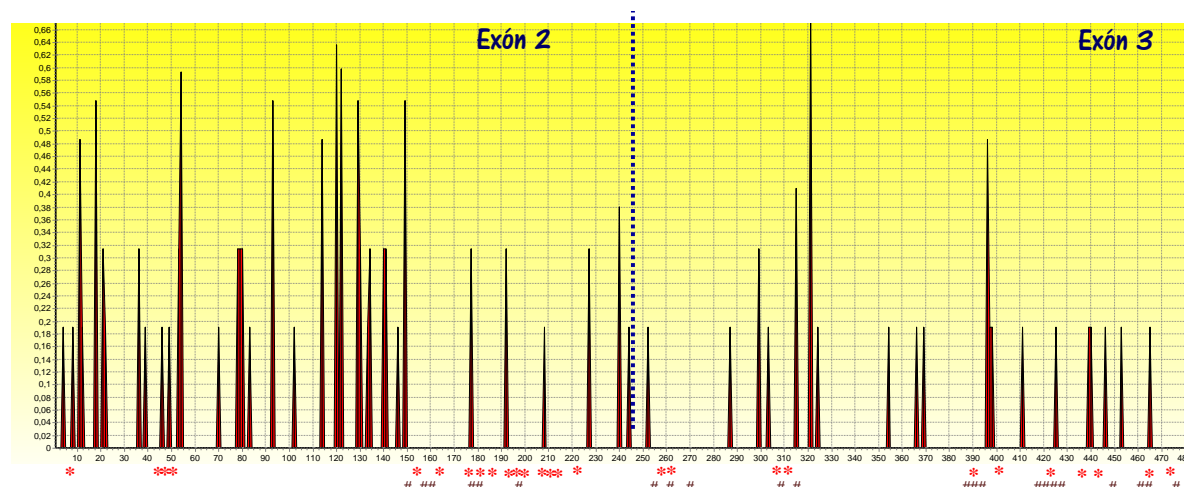


Figura R1b. Diagrama de entropía correspondiente a las secuencias de ADN la molécula de MHC clase I, de las muestras del género *Serinus*. Se indica la ubicación aproximada de los sitios implicados en la unión de péptidos (Bjorkman *et al.*, 1987b; Parham *et al.*, 1988) y de la interacción con el receptor de la célula T (#).

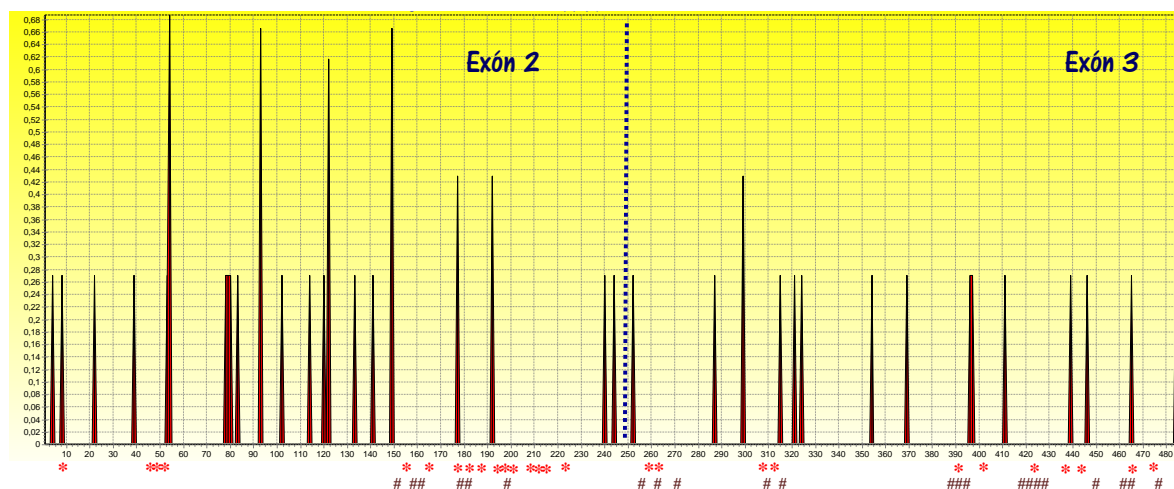
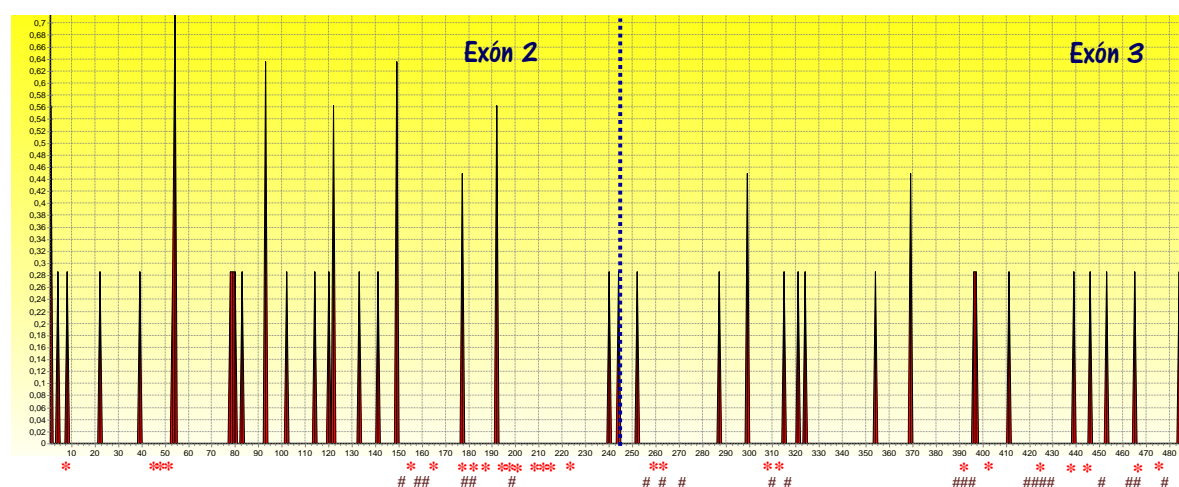


Figura R1c. Diagrama de entropía correspondiente a las secuencias de ADN la molécula de MHC clase I, de las muestras del género *Carduelis*. Se indica la ubicación aproximada de los sitios implicados en la unión de péptidos (Bjorkman *et al.*, 1987b; Parham *et al.*, 1988) y de la interacción con el receptor de la célula T (#).



1.3 Análisis de los tipos de sustituciones

Se han analizado los tipos de sustituciones (sinónimas y no sinónimas) comparando las secuencias por parejas y diferenciando entre regiones (exón 2 / exón 3 / exones 2 y 3 ó total). Los valores medios de dichas comparaciones se muestran en la tabla R3. Todas las regiones analizadas registran algo de variación. Esta variación es mayor en el exón 2 que en el exón 3. Si se diferencia entre variaciones sinónimas (d_s) y no sinónimas (d_n) se observa que la tasa de sustituciones no sinónimas es siempre inferior a la tasa de sustituciones sinónimas en el exón 2 y ligeramente superior en el exón 3. Se han encontrado diferencias significativas entre d_s de los dos exones ($p \leq 0.05$), pero no entre d_n .

Tabla R3. Tasas de sustituciones sinónimas y no sinónimas en distintas regiones de la secuencia MHC de clase I de las aves silvestres de canto estudiadas.

	Tasa de sustitución		
	EXÓN 2	EXÓN 3	TOTAL
Sinónimas	0,061 ± 0,014	0,024 ± 0,005	0,061 ± 0,008
No sinónimas	0,032 ± 0,005	0,030 ± 0,005	0,026 ± 0,002

Teniendo en cuenta la naturaleza del Sistema MHC como sistema genético polimórfico, donde la evolución tiende a incrementar la variabilidad, nos ha parecido interesante estudiar la tasa de sustitución sinónima y no sinónima en las secuencias nucleotídicas implicadas en el sitio de reconocimiento antigénico, de nuestras muestras de aves *Passeriformes*.

En la tabla R4 se pueden ver los resultados. En general se puede observar cómo d_s es significativamente superior en la región ARS (0,064) que en el resto de la secuencia (0,045) ($p \leq 0.05$).

Los dos lugares poseen sustituciones no sinónimas, pero hay que destacar que las posiciones que componen el sitio de reconocimiento antigénico tienen una tasa de sustitución significativamente mayor ($p \leq 0.05$) de (0,036) en comparación con resto de posiciones (NO ARS) (0,027).

Tabla R4. Comparativa de la tasa de sustitución sinónima y no sinónima de las posiciones que componen el sitio de reconocimiento antigénico (ARS), frente al resto de posiciones (NO ARS).

	Tasa de sustitución	
	ARS	NO ARS
Sinónimas	0,064 ± 0,027	0,045 ± 0,009
No sinónimas	0,036 ± 0,006	0,027 ± 0,004

También se han analizado las transiciones y transversiones en estas moléculas. Para ello se empleó el modelo de sustitución nucleotídica Tamura-Nei (TN93), (Tamura *et al.*, 2004; Tamura y Nei, 1993). En la tabla R5 se muestra la probabilidad de sustitución (r) de una base a otra base (columna). La probabilidad de las transiciones está sombreada en verde respecto a la probabilidad de las transversiones sombreada en amarillo. Las frecuencias de nucleótidos estimadas bajo este método son 22,44% (A), 18,55% (T / U), 25,48% (C), y 33,54% (G).

Tabla R5. Análisis de sustitución nucleotídica transición/transversión.

	A	T	C	G
A	-	2.17	2.98	20.41
T	2.62	-	24.62	3.92
C	2.62	17.92	-	3.92
G	13.66	2.17	2.98	-

En la tabla R5 podemos observar que el número de transiciones (en verde) es mucho más alto, con una media de 19,15 frente al número de transversiones (en amarillo) de media 2,92. La transición más frecuente es el cambio de timina por citosina seguida del cambio de adenina por guanina. La transición que menos aparece en nuestras secuencias, sería el cambio de guanina por adenina.

En las transversiones, se encuentra en mayor frecuencia la que cambia timina o citosina por guanina.

Por otro lado destacar que, los resultados de tasas de transición / transversión son $k1 = 5.205$ (purinas) y $k2 = 8,263$ (pirimidinas). El sesgo general transición / transversión es $R = 3,174$, donde $R = [A \times G \times k1 + T \times C \times k2] / [(A + T) \times (T + C)]$.

1.4 Secuencias de aminoácidos en aves *Passeriformes*

Se ha realizado la traducción de las secuencias nucleotídicas obtenidas del gen codificante para la proteína de MHC de clase I en las especies estudiadas, con el fin de comparar la variabilidad de los residuos.

Se pueden ver alineadas las secuencias que conforman el dominio alfa 1 y alfa 2 en las figuras R2 y R3 respectivamente.

Los números situados sobre las secuencias indican la posición aminoacídica. Los residuos conservados están sombreados en granate. Los (-) indican regiones que no se han conseguido secuenciar. Los lugares sombreados en rosa claro, hacen referencia a los residuos que difieren del resto. Las posiciones recuadradas en negro destacan seis de los siete lugares conservados en vertebrados (Grossberger y Parham, 1992) de los que posteriormente hablaremos en los apartados 1.9, 1.10, 1.11.

Cabe destacar que en la posición 10 se han encontrado dos excepciones en las aves estudiadas, una de la especie *S. canaria* y otra de la especie *C. crassirostris*.

Se pueden localizar un total de 37 posiciones conservadas en dominio alfa 1 y 42 posiciones conservadas en el dominio alfa 2.

A simple vista, se puede observar que la especie *S. canaria* es la que posee un mayor número de residuos variables, un total de 28 en el primer dominio y 21 en el segundo, en comparación con especies de su mismo género y de géneros distintos como son *Acrocephalus* y *Taeniopygia*.

Figura R2. Alineamiento de las secuencias de aminoácidos obtenidas por traducción de las secuencias nucleotídicas pertenecientes al dominio alfa 1 del gen de MHC de clase I (ver anexo 4, información sobre aminoácidos).

Posición	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
<i>Carduelis lawrencei</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Carduelis dominicensis</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Carduelis psaltria</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Serinus canaria</i>	L	K	L	V	V	S	E	P	S	P	G	V	P	Q	F	L	A	T	G	F	M	D	G	I	P	F	V	R
<i>Carduelis atrata</i>	L	S	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Carduelis carduelis parva</i>	-	T	V	A	V	S	E	A	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	T	H
<i>Carduelis citrinella</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Carduelis crassirostris</i>	L	T	E	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Carduelis cuculata</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Carduelis magellanicus</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Carduelis notata</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Carduelis olivacea</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Carduelis pinus</i>	-	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	M	G	Y	L	D	G	I	P	F	V	R
<i>Carduelis spinescens</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Carduelis spinus</i>	-	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Carduelis xanthogastra</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Carduelis yarrelli</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Serinus atrogularis</i>	L	T	V	V	V	S	E	P	S	P	G	V	P	Q	F	M	S	I	G	Y	L	D	G	I	F	F	V	R
<i>Serinus citrinelloides</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>serinus dorsostratus</i>	L	T	V	V	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Serinus flaviventris</i>	L	T	V	V	V	S	D	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Serinus gularis</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Serinus mozambicus</i>	L	T	V	V	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Serinus striolatus</i>	L	T	V	A	V	S	D	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Serinus sulphuratus</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	L	A	T	G	Y	L	D	G	I	P	F	T	R
<i>Serinus thibetanus</i>	L	T	V	A	V	S	E	P	S	P	G	V	P	Q	F	M	S	T	G	Y	L	D	G	I	P	F	V	R
<i>Acrocephalus arundina</i>	-	S	V	G	V	S	E	P	S	P	G	I	P	Q	F	M	E	M	G	F	V	D	G	I	P	F	V	R
<i>Taeniopygia guttata</i>	L	H	V	A	V	S	E	P	S	P	G	V	P	Q	F	T	S	I	G	F	V	D	G	I	P	F	V	R

36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	G	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	G	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	G	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	W	A	E	P	L	T	Q	W	M	K	D	R	A	E	P	G	Y	W	E	E	E	T	Q	K	F	V	E	H
Y	D	S	E	R	G	-	Q	K	V	P	L	T	R	W	I	K	D	G	A	E	P	G	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	G	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	N	G	A	E	P	G	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	R	W	I	K	D	G	A	E	P	G	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	E	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	E	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	E	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	E	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	E	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	E	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	E	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	E	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	E	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	E	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	E	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	E	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	E	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	Q	K	V	P	L	T	Q	W	I	K	D	G	A	E	P	E	Y	W	E	R	Q	T	Q	I	C	E	G	W
Y	D	S	E	R	G	-	R	M	E	P	L	T	E	W	I	K	D	S	A	D	P	G	Y	W	D	R	N	T	Q	N	A	V	G	S
Y	D	S	E	R	G	-	R	A	E	P	L	T	Q	W	M	K	D	G	A	E	P	G	Y	W	D	E	E	T	Q	I	S	V	R	N

119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	R	D	F	I	S	F	D	L	E	S	G	R	F	V	P	A	D	S	A	A	E	I	T	R	R	R	W	E	E	G	-	-	-	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	W	D	H	L	S	F	D	P	K	S	G	K	F	V	P	A	N	S	S	A	E	T	N	G	K	H	W	E	E	G	I	E	V	E	
D	G	R	D	F	I	S	F	D	L	E	S	G	R	F	V	A	A	D	S	A	A	E	I	T	R	R	H	W	E	E	G	I	V	A	E	
D	G	R	D	F	I	S	F	D	L	G	S	G	K	F	L	A	A	D	S	A	A	E	I	T	R	R	R	W	E	E	D	M	-	A	E	

155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R	W	T	N	-	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	R	T	N	Y	L	K	H	V	C	P	E	W	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	A	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	M	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
R	W	T	N	Y	L	K	H	L	C	P	E	S	L	R	K	Y	V	G
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R	K	T	N	Y	L	K	H	E	C	P	E	W	L	Q	R	H	V	R
R	L	K	N	Y	L	K	H	K	C	P	E	W	L	R	K	Y	V	G

1.5 Cálculo de distancias genéticas entre las aves silvestres de canto estudiadas

Entendemos el término distancia como una medida del grado de divergencia entre dos secuencias. En este caso lo primero que se hace es generar una matriz de distancias por parejas de secuencias a partir del alineamiento múltiple. En estos métodos toda la información de similitudes y diferencias entre dos secuencias queda resumida por un simple número.

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 calcularla 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, 1972). La distancia p asume que todos los cambios de nucleótido ocurren con igual frecuencia y en cualquier parte de la molécula.

En la figura R4 se pueden observar las distancias genéticas obtenidas con este método (Nei, 1987).

En ella se puede ver subrayados en verde los lugares donde la “distancia p” es igual a cero, localizándose dos grupos donde esta distancia es nula:

Grupo 1. *C. lawrencei*/ *C. dominicensis*/ *C. psaltria*/ *S. thibetanus*

Grupo 2. *C. magellanicus*/ *C. notata*/ *C. olivácea*/ *C. spinescens*/ *C. xanthogastra*

Sombreado en granate se puede ver la distancia genética de la especie *S. canaria* respecto al resto de aves.

Distancia media de todas las especies de $d=0.345$, una media respecto a las especies de su mismo género de $d=0.355$, una media respecto al género *Carduelis* de $d=0.353$ y una media de los géneros *Acrocephalus* y *Taeniopygia* de $d=0.264$. Lo que nos da una idea de que su molécula de MHC de clase I se asemeja más a otras especies de géneros distintos que a las del suyo propio.

Por otro lado, sombreado en rosa, se pueden ver las distancias de los géneros *Acrocephalus* y *Taeniopygia*. La media de distancias entre estos géneros y el resto de aves *Passeriformes* son de 0,335 y 0,287 respectivamente.

El resto de distancias que se pueden ver en la figura R4 oscilan entre los valores de 0,006 de mínima y 0,086 de máxima.

Figura R4. Matriz de distancias genéticas entre las muestras de aves *Passeriformes* estudiadas

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1 <i>Carduelis lawrencei</i>																											
2 <i>Carduelis dominicensis</i>	0,000																										
3 <i>Carduelis psaltria</i>	0,000	0,000																									
4 <i>Serinus canaria</i>	0,346	0,316	0,314																								
5 <i>Carduelis atrata</i>	0,025	0,012	0,012	0,351																							
6 <i>Carduelis carduelis parva</i>	0,038	0,031	0,031	0,377	0,050																						
7 <i>Carduelis citrinella</i>	0,012	0,031	0,031	0,365	0,046	0,066																					
8 <i>Carduelis crassirostris</i>	0,025	0,012	0,012	0,352	0,008	0,050	0,047																				
9 <i>Carduelis cuculata</i>	0,012	0,006	0,006	0,351	0,015	0,042	0,039	0,015																			
10 <i>Carduelis magellanicus</i>	0,012	0,006	0,006	0,351	0,019	0,046	0,042	0,019	0,004																		
11 <i>Carduelis notata</i>	0,012	0,006	0,006	0,352	0,019	0,046	0,043	0,019	0,004	0,000																	
12 <i>Carduelis olivacea</i>	0,012	0,006	0,006	0,351	0,019	0,046	0,042	0,019	0,004	0,000	0,000																
13 <i>Carduelis pinus</i>	0,013	0,012	0,012	0,361	0,027	0,046	0,043	0,027	0,019	0,023	0,023	0,023															
14 <i>Carduelis spinescens</i>	0,012	0,006	0,006	0,359	0,020	0,048	0,044	0,020	0,004	0,000	0,000	0,000	0,020														
15 <i>Carduelis spinus</i>	0,013	0,006	0,006	0,357	0,012	0,046	0,043	0,012	0,012	0,015	0,015	0,015	0,023	0,016													
16 <i>Carduelis xanthogastra</i>	0,012	0,006	0,006	0,351	0,019	0,046	0,042	0,019	0,004	0,000	0,000	0,000	0,023	0,000	0,015												
17 <i>Carduelis yarrelli</i>	0,012	0,006	0,006	0,351	0,008	0,050	0,046	0,008	0,015	0,011	0,012	0,011	0,027	0,012	0,012	0,011											
18 <i>Serinus atrogularis</i>	0,086	0,043	0,043	0,363	0,046	0,073	0,066	0,046	0,038	0,042	0,042	0,042	0,046	0,043	0,042	0,042	0,046										
19 <i>Serinus citrinelloides</i>	0,012	0,006	0,006	0,351	0,015	0,042	0,046	0,015	0,015	0,019	0,019	0,019	0,027	0,020	0,012	0,019	0,015	0,046									
20 <i>Serinus dorsostriatus</i>	0,037	0,019	0,018	0,347	0,034	0,062	0,054	0,035	0,027	0,031	0,031	0,031	0,038	0,032	0,031	0,031	0,034	0,034	0,034								
21 <i>Serinus flaviventris</i>	0,062	0,037	0,037	0,359	0,046	0,073	0,066	0,046	0,038	0,042	0,042	0,042	0,050	0,043	0,042	0,042	0,046	0,046	0,046	0,042							
22 <i>Serinus gularis</i>	0,012	0,012	0,012	0,351	0,011	0,046	0,039	0,012	0,011	0,015	0,015	0,015	0,023	0,016	0,008	0,015	0,011	0,042	0,011	0,031	0,042						
23 <i>Serinus mozambicus</i>	0,037	0,025	0,025	0,351	0,038	0,054	0,062	0,038	0,031	0,034	0,035	0,034	0,035	0,036	0,035	0,034	0,038	0,038	0,038	0,034	0,046	0,034					
24 <i>Serinus striolatus</i>	0,012	0,006	0,006	0,351	0,023	0,042	0,046	0,023	0,015	0,019	0,019	0,019	0,027	0,020	0,019	0,019	0,023	0,046	0,015	0,034	0,038	0,019	0,038				
25 <i>Serinus sulphuratus</i>	0,074	0,037	0,037	0,347	0,034	0,054	0,058	0,035	0,027	0,031	0,031	0,031	0,038	0,032	0,031	0,031	0,034	0,050	0,034	0,038	0,042	0,031	0,042	0,034			
26 <i>Serinus thibetanus</i>	0,000	0,000	0,000	0,371	0,021	0,047	0,043	0,021	0,013	0,017	0,017	0,017	0,021	0,017	0,017	0,017	0,021	0,047	0,021	0,034	0,043	0,017	0,034	0,021	0,034		
27 <i>Acrocephalus arundina</i>	0,300	0,335	0,333	0,264	0,325	0,362	0,350	0,340	0,337	0,337	0,340	0,337	0,331	0,323	0,331	0,337	0,331	0,362	0,331	0,344	0,356	0,331	0,350	0,337	0,362	0,324	
28 <i>Taeniopygia guttata</i>	0,238	0,281	0,280	0,264	0,288	0,309	0,301	0,296	0,288	0,288	0,290	0,288	0,290	0,297	0,290	0,288	0,288	0,301	0,288	0,288	0,301	0,288	0,288	0,288	0,301	0,285	0,256

1.6 Análisis de las posiciones que determinan el sitio de reconocimiento antigénico (ARS)

Las secuencias de aminoácidos de las proteínas MHC de clase I obtenidas en este trabajo han sido alineadas con las secuencias consenso de las proteínas HLA de clase I clásicas (HLA-Ia) y no clásicas (HLA-Ib) humanas (Bjorkman et al, 1987b) (tablas R6, R7).

Hemos destacado, dentro del ARS, 20 de las 21 posiciones que en humanos fueron definidas como hipervariables (!) (Parham *et al.*, 1988) y 7 aminoácidos que se mantienen invariables (=) en las secuencias humanas y que participan en el anclaje del péptido a través de sus extremos.

La mayoría de los cambios que se observan al comparar las distintas secuencias de aminoácidos obtenidos en este trabajo están fuera del sitio de reconocimiento antigénico (ARS).

Dentro de las 20 posiciones señaladas como hipervariables, podemos observar destacadas en rojo, azul y verde, las excepciones encontradas en aves *Passeriformes*. Cabe destacar que en casi todas estas excepciones, siempre aparece la especie *S. canaria* (Tablas R6, R7).

Tabla R6. Esquema de las posiciones que conforman el sitio de reconocimiento antigénico (ARS) y que presumiblemente están implicadas en la unión de péptidos (de los géneros *Carduelis* y *Serinus*). En verde, rosa y azul, están subrayadas las excepciones dentro del grupo de aves *Passeriformes* estudiadas. Las posiciones hipervariables se representan con (!) (Parham *et al.*, 1988) y las invariables (=) (Bjorkman et al, 1987b).

SITIO DE UNIÓN AL PÉPTIDO				SITIO DE UNIÓN AL PÉPTIDO			
DOMINIO ALFA 1				DOMINIO ALFA 2			
Posición	Residuo <i>Passeriformes</i>	Excepciones	Residuo HLA-A	Posición	Residuo	Excepciones	Residuo
9	T	K/S	Y !	95	L	R	L !
22	F		F !	97	W	R	R !
24	S	A	A !	99	Y	S	Y !
26	G		G !	113	S		H !
59	Y		Y =	115	R	W	Y !
63	Q	E	E !	143	N	T	T =
66	I	K	I !	147	W		W =
67	C	F	V !	152	E	G	V !
70	W	H	Q !	156	W	R	L !
73	V	T	T !	159	Y		Y =
74	E	F	D !	167	S	W	W =
77	N	D	S !	171	Y		Y =
80	T		T !				
81	L		L !				
85	Y		Y =				

 Serinus canaria y *Serinus sulphuratus*

 Serinus canaria

 Serinus canaria, *Carduelis atrata*

! Residuos que en humanos fueron definidos como hipervariables. (Parham et al, 1988).

= Aminoácidos que en las secuencias humanas se mantienen invariables dentro del ARS. (Bjorkman et al, 1987b).

De las 7 posiciones definidas como invariables en esta región, 5 también los son en las moléculas de estas aves dentro de los géneros *Carduelis* y *Serinus* (posiciones 59,85, 147, 159, 171). De las posiciones restantes una está dentro sitio de unión al péptido (143) donde se cambia treonina (T) por una asparagina (N) los dos de carácter polar neutro, a excepción de la

especie *S. canaria* que continúa teniendo en este lugar una N. La otra posición se encuentra dentro de la región que conforma el receptor de células T del dominio alfa 2 (tabla R7). Este lugar presenta también un aminoácido distinto suponiendo un cambio de lisina (K) por histidina (H), los dos de carácter polar básico. En esta posición se ha encontrado una excepción en la especie *S. canaria* que posee en este lugar una arginina (R), también de tipo polar básico (anexo 4). Por otro lado, teniendo en cuenta los otros dos géneros estudiados (*Taeniopygia* y *Acrocephalus*), se observa que coinciden en dos de las posiciones definidas como invariantes (143, 146) con las excepciones obtenidas en la especie *S. canaria* (Figura R3). En la posición 171 podemos considerar a la especie *Acrocephalus arundinaceus* como la excepción cambiando un aminoácido polar neutro (tirosina) por un aminoácido polar básico (histidina) (Figura R3). De las posiciones que conforman el receptor de células T (tabla R7), hemos encontrado cuatro posiciones iguales respecto a los residuos de HLA-A: 62, 65, 154, 166 de las 18 estudiadas.

Tabla R7. Esquema de las posiciones que conforman el receptor de células T. En verde, rosa y azul, están subrayadas las excepciones dentro del grupo de aves *Passeriformes* estudiadas. Las posiciones hipervariables se representan con (!) (Parham *et al.*, 1988) y las invariantes (=) (Bjorkman *et al.*, 1987b) (ver anexo 4, información sobre aminoácidos).

RECEPTOR CELULAS T				RECEPTOR CELULAS T			
DOMINIO ALFA 1				DOMINIO ALFA 2			
Posición	Residuo <i>Passeriformes</i>	Excepciones	Residuo HLA-A	Posición	Residuo	Excepciones	Residuo HLA-A
58	G	E	E	144	G	R/E	R
62	R	E	R	146	H	R	K
65	Q		Q	148	E	D	A
69	G	E	A	149	E		A
72	H		Q	150	G		R
76	R	K	E	154	E		E
82	Q	R	R	155	R		Q
				158	N		A
				162	H		G
				163	L	V	T
				166	E		E

	<i>C. xanthogastra, C. notata, C. spinescens, C. olivacea, C. cucullata, C. magellanicus</i>
	<i>Serinus canaria</i>
	<i>Carduelis citrinella</i>
	<i>Serinus canaria, Carduelis citrinella</i> y <i>Serinus gularis</i>
!	Residuos que en humanos fueron definidos como hipervariables. (Parham <i>et al.</i> , 1988)
=	Aminoácidos que en las secuencias humanas se mantienen invariantes dentro del ARS. (Bjorkman <i>et al.</i> , 1987b).

POSIBLE INTERACCION CON TCR				POSIBLE INTERACCION CON TCR			
DOMINIO ALFA 1				DOMINIO ALFA 2			
Posición	Residuo <i>Passeriformes</i>	Excepciones	Residuo HLA-A	Posición	Residuo <i>Passeriformes</i>	Excepciones	Residuo HLA-A
57	P	L	P	157	T	A/M	R
61	E		D	161	K		E
64	T		T	165	P		V
68	E	V	K	169	R		R
71	R	Q	S				
75	A	V	R				

	<i>Serinus sulphuratus</i>
	<i>Serinus canaria</i>

Carduelis carduelis parva, Serinus mozambicus

De las posiciones que posiblemente formen parte de la interacción con el TCR, se conservan en humanos y aves *Passeriformes* 3 de ellas: la 57 con la excepción (sombreada en azul en la tabla R7) de *S. sulphuratus* que cambia una prolina por una leucina también de origen apolar, la 64 y la 169, de los 10 lugares estudiados.

1.7 Secuencias de aminoácidos del resto de vertebrados a estudio

Con el fin de comparar los residuos aminoacídicos de las aves silvestres de canto secuenciadas en este trabajo, se han seleccionado y alineado una serie de muestras, obtenidas la gran mayoría de la base de datos GenBank (Figuras R5 y R6; material y métodos tabla M1). Para que el estudio fuese más completo se han elegido muestras de distintas Clases, como Actinopterygii (pez), Sauropsida (reptil), Amphibia (anfibio), Aves (*Galliformes*), Mammalia (mamíferos). Para una mejor comparativa, hemos dividido la clase Mammalia en tres grupos. Por un lado, tendríamos el grupo denominado primates (*Hylobates lar*, *Pongo pygmaeus*, *Gorilla gorilla*, *Pan troglodytes*, *Pan paniscus*), del que hemos sacado a *Homo sapiens*, para analizarlo por separado, y por último tendríamos al grupo llamado, en este trabajo, mamíferos que englobarían 7 especies de géneros distintos (*Bos Taurus*, *Canis familiaris*, *Mus musculus*, *Sus scrofa*, *Rattus norvegicus*, *Equus caballus*, *Ovi aries*). En las figuras R5 y R6 se muestran las secuencias alineadas, y divididas en los distintos grupos. Los números situados sobre las secuencias indican la posición aminoacídica. Los residuos conservados están sombreados en granate. Como se observa, existen un total de 16 en el dominio alfa 1 y un total de 14 en el dominio alfa 2. Los (-) indican regiones que no se han conseguido secuenciar. Los lugares sombreados en rosa claro, subrayan los residuos que difieren del resto, así podemos ver lugares muy variables como 17, 66, 67, 69, 74, 79, 90, 95, 97, 114, 116, 131, 138, 144, 174, 182 entre otros. Las posiciones recuadradas en negro destacan seis de los siete lugares conservados en vertebrados (Grossberger y Parham, 1992) de los que hablaremos posteriormente.

Para ser comparadas estas secuencias con todas las secuencias de aves *Passeriformes* del estudio se realizó una matriz con la mezcla de ambas, y posteriormente fueron alineadas para poder ser manejadas en el programa MEGA 6 (Tamura *et al.*, 2013) y BioEdit (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>) y así localizar todas las similitudes y diferencias.

Figura R5. Alineamiento de las secuencias de aminoácidos pertenecientes al dominio alfa 1 del gen de MHC de clase I del resto de muestras de vertebrados analizados procedentes de la base de datos GenBank (Tabla M1, ver anexo 4, información sobre aminoácidos).

Posición		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
<i>Dario rerio</i>	pez	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	D	G	
<i>Ameiva ameiva</i>	reptil	-	-	-	-	L	Q	Y	F	Y	T	G	V	S	E	P	G	Q	G	L	P	Q	F	I	V	V	G	Y	V	D	G	
<i>Xenopus laevis</i>	anfibio	G	S	H	S	L	R	N	Y	Y	T	A	V	S	D	R	A	F	G	L	P	E	F	Y	A	A	G	Y	V	D	D	
<i>Hylobates lar</i>	primates	G	S	H	S	M	R	Y	F	Y	T	S	V	S	R	P	G	R	G	E	P	R	F	I	T	V	G	Y	V	D	D	
<i>Pongo pygmaeus</i>	primates	G	S	H	S	M	R	Y	F	Y	T	S	V	S	R	P	G	R	G	E	P	R	F	I	A	V	G	Y	V	D	D	
<i>Gorilla gorilla</i>	primates	G	S	H	S	M	R	Y	F	Y	T	T	M	S	R	P	G	R	G	E	P	R	F	I	S	V	G	Y	V	D	D	
<i>Pan troglodytes</i>	primates	G	S	H	S	M	R	Y	F	D	T	A	V	S	R	P	G	R	G	E	P	R	F	I	A	V	G	Y	V	D	D	
<i>Bos taurus</i>	mamíferos	G	S	H	S	L	R	Y	F	L	T	A	V	S	R	P	G	F	G	E	P	R	F	I	A	V	G	Y	V	D	D	
<i>Canis familiaris</i>	mamíferos	G	S	H	S	L	R	Y	F	Y	T	S	V	S	R	P	G	R	G	D	P	R	F	I	A	V	G	Y	V	D	D	
<i>Mus musculus</i>	mamíferos	G	S	H	S	L	R	Y	F	T	T	A	V	S	R	P	G	L	G	E	P	R	F	I	I	V	G	Y	V	D	D	
<i>Sus scrofa</i>	mamíferos	G	P	H	S	L	S	Y	F	Y	T	A	V	S	R	P	D	R	G	D	S	R	F	I	A	V	G	Y	V	D	D	
<i>Rattus norvegicus</i>	mamíferos	G	S	H	S	L	R	Y	F	Y	T	A	V	S	R	P	G	L	G	E	P	R	F	I	A	V	G	Y	V	D	D	
<i>Homo sapiens (HLA-A)</i>	primates	G	S	H	S	M	R	Y	F	F	T	S	V	S	R	P	G	R	G	E	P	R	F	I	A	V	G	Y	V	D	D	
<i>Pan paniscus</i>	primates	G	S	H	S	M	R	Y	F	D	T	A	V	S	R	P	G	R	G	E	P	R	F	I	T	V	G	Y	V	D	D	
<i>Equus caballus</i>	mamíferos	G	S	H	S	M	R	Y	F	S	T	A	V	S	R	P	G	R	G	E	S	W	Y	L	E	V	G	Y	V	D	D	
<i>Ovis aries</i>	mamíferos	-	-	-	-	-	-	-	-	F	S	T	A	V	S	R	A	G	A	G	E	P	R	Y	L	E	V	G	Y	V	D	D
<i>Gallus gallus</i>	galliformes	-	-	-	-	-	-	-	-	F	-	T	A	M	T	D	P	G	P	G	L	P	W	F	V	D	V	G	Y	V	D	G
<i>Coturnix japonica</i>	galliformes	-	-	-	-	-	-	-	-	F	-	T	A	M	T	D	P	G	P	G	L	P	W	F	V	S	V	G	Y	V	D	G

31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64			
E	Q	F	V	Y	F	D	S	N	T	M	-	-	K	T	V	P	K	T	E	W	M	R	Q	N	-	-	V	G	E	D	Y	W	E	R	E	T
Q	L	F	V	Q	Y	D	S	N	T	R	-	-	E	M	L	P	R	V	S	W	I	K	D	N	-	-	E	D	S	K	Y	W	E	G	Q	T
T	L	-	V	R	Y	S	S	D	K	D	R	V	E	A	A	T	Q	-	-	W	M	K	D	K	-	-	A	G	P	E	Y	W	E	Q	Q	K
T	Q	F	V	R	F	D	S	D	A	A	S	P	R	M	E	P	R	A	P	W	I	E	Q	E	-	-	-	G	P	E	Y	W	D	Q	Q	T
T	Q	F	V	R	F	D	S	D	A	A	S	P	R	E	E	P	R	A	P	W	M	E	R	E	-	-	-	G	P	E	Y	W	D	R	N	T
T	Q	F	V	R	F	D	S	D	A	A	S	Q	R	M	E	P	R	A	P	W	I	E	Q	E	-	-	-	G	P	E	Y	W	D	E	E	T
T	Q	F	V	R	F	D	S	D	A	P	N	P	R	M	E	P	R	T	R	W	V	K	Q	E	-	-	-	G	P	E	Y	W	D	R	N	T
T	Q	F	V	R	F	D	S	D	A	A	T	G	R	M	E	P	R	A	P	W	M	E	Q	E	-	-	-	G	P	E	Y	W	D	R	E	T
T	Q	F	V	R	F	D	S	D	A	E	N	P	R	M	E	P	R	A	P	W	M	E	Q	E	-	-	-	G	P	E	Y	W	E	R	N	T
T	Q	F	V	R	F	D	S	D	A	E	N	P	R	E	E	P	R	A	P	W	I	Q	Q	E	-	-	-	G	Q	E	Y	W	D	R	N	T
T	E	F	V	R	F	D	S	D	A	E	N	P	R	M	E	P	R	A	R	W	M	E	R	E	-	-	-	G	P	E	Y	W	E	E	E	T
T	Q	F	V	R	F	D	S	D	A	A	S	Q	R	M	E	P	R	A	P	W	I	E	Q	E	-	-	-	G	P	E	Y	W	D	G	E	T
T	Q	F	V	R	F	D	S	D	A	A	S	P	R	E	E	P	R	A	P	W	M	E	Q	E	-	-	-	G	P	E	Y	W	D	R	N	T
T	Q	F	V	R	F	D	S	D	A	A	S	P	R	M	E	P	R	A	P	W	M	E	Q	E	-	-	-	G	P	E	Y	W	E	E	E	T
T	Q	F	V	R	F	D	S	D	A	P	D	P	K	M	E	Q	R	E	P	W	M	K	Q	V	-	-	-	G	P	E	Y	W	D	R	N	T
E	L	F	M	H	Y	N	S	T	A	V	R	R	R	A	V	P	R	T	E	W	I	A	A	N	-	-	T	D	Q	Q	Y	W	D	R	E	T
E	I	F	A	H	Y	D	S	T	T	R	-	-	R	V	V	P	R	T	E	W	I	K	A	P	G	A	V	D	P	E	Y	W	E	R	N	T

65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
Q	I	F	T	G	A	H	P	V	F	K	N	N	I	Q	V	I	K	E	R	F	N	Q	T	-	Q
Q	N	L	Q	G	A	E	P	V	F	R	G	N	I	N	T	A	M	N	R	Y	N	Q	T	-	G
R	E	M	K	G	T	E	P	V	F	K	H	N	V	K	T	A	M	E	R	F	N	Q	S	T	-
Q	I	S	K	T	N	A	Q	T	D	R	E	N	L	R	I	A	L	R	Y	Y	N	Q	S	E	D
R	N	V	K	A	H	A	Q	T	D	R	V	D	L	G	T	L	R	G	Y	Y	N	Q	S	E	D
Q	I	Y	K	A	Q	A	Q	T	D	R	V	D	L	E	T	L	R	G	Y	Y	N	Q	S	E	G
R	S	V	K	A	S	A	Q	T	D	R	V	D	L	G	T	L	R	G	Y	Y	N	Q	S	E	D
R	N	A	K	N	T	A	Q	T	F	R	V	N	L	N	N	L	R	G	Y	Y	N	Q	S	E	A
R	T	V	K	E	T	A	Q	R	Y	R	V	D	L	D	T	L	R	G	Y	Y	N	Q	S	E	A
Q	V	S	K	E	N	E	Q	S	F	R	V	S	L	G	T	A	L	S	Y	Y	N	Q	S	K	G
Q	I	Y	K	E	T	A	Q	T	Y	R	V	N	L	N	T	L	R	G	Y	Y	N	Q	S	E	A
Q	I	A	K	G	H	E	Q	I	Y	R	V	D	L	R	T	L	R	G	Y	Y	N	Q	S	E	G
R	K	V	K	A	H	S	Q	T	H	R	V	D	L	G	T	L	R	G	Y	Y	N	Q	S	E	A
Q	I	C	K	A	Q	A	Q	T	D	R	E	N	L	R	I	A	L	R	Y	Y	N	Q	S	E	A
Q	R	A	A	G	L	A	H	S	F	R	G	N	L	N	N	L	R	G	Y	Y	N	Q	S	E	A
R	N	P	K	G	N	A	Q	T	F	R	V	G	L	T	I	L	R	G	Y	Y	N	Q	S	E	T
Q	I	V	Q	G	S	E	Q	I	N	R	E	N	L	D	I	L	R	R	Y	Y	N	Q	T	-	G
Q	I	V	Q	R	N	E	Q	I	D	R	E	N	L	V	T	A	A	R	R	Y	N	Q	S	-	G

Figura R6. Alineamiento de las secuencias de aminoácidos pertenecientes al dominio alfa 2 del gen de MHC de clase I del resto de muestras de vertebrados analizados procedentes de la base de datos GenBank (Tabla M1, ver anexo 4, información sobre aminoácidos).

Posición		91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115		
<i>Danio rerio</i>	pez	G	V	H	T	F	Q	Q	M	Y	G	C	E	W	D	D	Q	T	E	A	K	N	G	F	D	Q		
<i>Ameiva ameiva</i>	reptil	G	L	H	T	W	Q	W	M	Y	G	C	E	L	R	G	D	G	-	S	K	G	G	Y	S	Q		
<i>Xenopus laevis</i>	anfibio	G	T	H	T	V	Q	W	M	Y	G	C	E	L	G	D	D	G	S	-	I	R	G	Y	D	Q		
<i>Hylobates lar</i>	primates	G	S	H	T	L	Q	T	M	Y	G	C	D	L	G	P	D	G	R	L	L	R	G	Y	R	Q		
<i>Pongo pygmaeus</i>	primates	G	S	H	T	I	Q	I	M	Y	G	C	D	V	G	P	D	G	R	F	L	L	R	G	Y	R	Q	
<i>Gorilla gorilla</i>	primates	G	S	H	T	I	Q	R	M	Y	G	C	E	V	G	P	D	G	R	F	L	L	R	G	Y	L	Q	
<i>Pan troglodytes</i>	primates	G	S	H	T	I	Q	L	M	F	G	C	D	V	G	S	D	G	R	F	L	L	R	G	Y	R	Q	
<i>Bos taurus</i>	mamíferos	G	S	H	T	V	Q	E	M	H	G	C	D	V	G	P	D	G	R	L	R	R	G	F	M	Q	Q	
<i>Canis familiaris</i>	mamíferos	G	S	H	T	R	Q	T	M	Y	G	C	D	L	G	P	G	G	R	L	L	R	G	Y	S	Q	Q	
<i>Mus musculus</i>	mamíferos	G	S	H	T	L	Q	W	L	V	G	C	D	L	G	P	D	G	S	L	L	L	R	G	Y	E	Q	Q
<i>Sus scrofa</i>	mamíferos	G	S	H	T	I	Q	I	M	Y	G	C	D	V	G	P	D	G	L	L	L	L	R	G	Y	R	Q	Q
<i>Rattus norvegicus</i>	mamíferos	G	S	H	T	I	Q	E	M	Y	G	C	D	V	G	S	D	G	S	L	L	L	R	G	Y	R	Q	Q
<i>Homo sapiens (HLA-A)</i>	primates	G	S	H	T	V	Q	R	M	C	G	C	D	V	G	S	D	W	R	F	L	L	R	G	Y	H	Q	Q
<i>Pan paniscus</i>	primates	G	S	H	T	L	Q	R	M	Y	G	C	D	V	G	P	D	G	R	L	L	L	R	G	Y	S	Q	Q
<i>Equus caballus</i>	mamíferos	G	S	H	T	L	Q	L	M	Y	G	C	D	V	G	P	H	G	R	L	L	S	A	S	F	Q	Q	Q
<i>Ovis aries</i>	mamíferos	G	S	H	T	W	Q	C	M	Y	G	C	D	V	G	P	D	G	R	L	L	R	G	F	M	Q	Q	Q
<i>Gallus gallus</i>	galliformes	G	S	H	T	V	Q	W	M	S	G	C	D	I	L	E	D	G	-	T	I	R	G	Y	H	Q	Q	Q
<i>Coturnix japonica</i>	galliformes	G	S	L	T	R	Q	Q	M	Y	G	C	D	I	L	E	D	G	-	T	T	R	G	Y	D	Q	Q	Q

116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150			
I	G	Y	D	G	E	D	F	L	S	L	D	L	K	E	I	R	W	I	S	P	V	Q	Q	G	L	I	T	T	Q	N	W	N	N	N	R		
F	G	Y	D	G	R	D	F	V	A	L	D	K	E	T	L	T	W	T	A	A	D	S	E	A	Q	V	T	K	S	K	W	D	A	L	G		
H	V	Y	D	G	R	D	F	F	A	L	D	T	E	E	W	V	V	P	S	V	F	Y	A	Q	L	T	T	Q	K	W	N	S	P	V			
A	Y	D	G	K	D	Y	I	A	L	N	E	D	L	S	S	W	T	A	A	D	T	A	A	Q	I	T	Q	R	K	W	E	A	-	A	A		
D	A	Y	D	G	K	D	Y	I	A	L	N	E	D	L	R	S	W	T	A	A	D	M	A	A	Q	I	T	Q	R	K	W	E	A	-	A	A	
D	A	Y	D	G	K	D	Y	I	A	L	N	E	D	L	R	S	W	T	A	A	D	M	A	A	Q	I	T	Q	R	K	W	E	A	-	A	A	
Y	G	Y	D	G	R	D	Y	L	A	L	N	E	D	L	R	S	W	T	A	V	E	T	V	A	Q	I	S	K	R	K	M	E	A	-	A	A	
D	A	Y	D	G	A	D	Y	I	A	L	N	E	D	L	R	S	W	T	A	A	D	T	A	A	Q	I	T	R	R	K	W	E	A	-	A	A	
S	A	Y	D	G	R	D	Y	L	A	L	N	E	D	L	I	T	W	T	A	A	D	L	A	A	L	K	T	R	S	K	L	E	Q	-	A	A	
D	A	Y	D	G	A	D	Y	I	A	L	N	E	D	L	R	S	W	T	A	A	D	T	A	A	Q	I	T	K	R	K	W	E	A	-	A	A	
D	A	Y	D	G	R	D	Y	I	A	L	N	E	D	L	R	K	T	W	T	A	A	D	F	A	A	Q	I	T	R	N	K	W	E	A	-	A	A
Y	A	Y	D	G	K	D	Y	I	A	L	N	E	D	L	R	S	W	T	A	A	D	M	A	A	Q	T	T	K	H	K	W	E	A	-	A	A	
S	A	Y	D	G	K	D	Y	I	A	L	N	E	D	L	S	S	W	T	A	A	D	T	A	A	Q	I	T	Q	R	K	W	E	A	-	A	A	
Y	A	Y	D	G	A	D	Y	L	A	L	N	E	D	L	R	S	W	T	A	A	D	T	A	A	Q	I	S	K	R	K	K	E	V	-	A	A	
F	G	Y	D	G	R	D	Y	I	A	L	N	E	D	L	R	S	W	T	A	A	D	T	A	A	Q	V	T	Q	R	K	W	E	K	-	E	A	
A	A	Y	D	G	R	D	F	V	A	F	D	K	G	T	M	T	L	T	A	A	V	P	E	A	V	P	T	K	R	K	W	E	E	-	G	A	
V	A	Y	N	G	R	D	F	I	A	F	D	K	D	T	M	T	F	T	A	A	V	P	E	A	V	P	T	K	R	A	W	E	E	-	G	A	

151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182			
G	F	L	Q	S	Y	G	G	Y	Y	S	T	V	C	I	E	W	L	Q	K	Y	I	Q	Y	G	K	S	S	L	K	K	T	-	-	-
A	M	N	Q	G	R	K	F	Y	L	E	K	I	C	I	E	W	L	Q	K	Y	L	R	Y	G	N	K	T	L	L	R	K	-	-	-
N	A	P	E	R	N	K	N	Y	L	E	N	I	C	I	E	D	L	K	K	Y	L	S	Y	G	Q	-	-	-	-	-	-	-	-	-
H	E	A	E	Q	M	R	A	Y	L	E	G	R	C	L	E	W	L	R	R	Y	L	E	N	G	K	E	T	L	Q	R	A	-	-	-
G	A	A	E	Q	R	R	A	Y	L	E	G	L	C	V	E	S	L	R	R	Y	L	E	N	G	K	E	T	L	Q	R	T	-	-	-
R	E	A	E	R	L	R	A	Y	M	E	G	T	C	V	E	W	L	R	R	H	L	E	N	G	K	E	T	L	Q	R	T	-	-	-
H	A	A	E	Q	L	R	A	Y	L	E	G	T	C	V	E	W	L	R	R	Y	L	E	N	G	K	E	T	L	Q	R	T	-	-	-
G	E	A	E	V	Q	R	N	Y	L	E	G	T	C	V	E	W	L	R	R	Y	L	E	N	G	K	E	T	L	L	R	A	-	-	-
G	T	A	E	H	D	R	N	Y	L	E	T	T	C	V	E	W	L	R	R	Y	L	E	M	G	K	D	T	L	L	R	A	-	-	-
G	L	A	E	K	R	R	A	Y	L	E	V	D	C	L	T	W	L	R	R	Y	L	E	L	G	K	E	T	L	L	H	T	-	-	-
D	V	A	E	Q	L	R	S	Y	L	Q	G	R	C	V	E	G	L	R	R	Y	L	E	M	G	K	D	T	L	Q	R	A	-	-	-
R	Y	A	E	R	L	R	A	Y	L	E	G	T	C	V	E	S	L	R	R	Y	L	E	L	G	K	E	T	L	L	R	S	-	-	-
H	V	A	E	Q	L	R	A	Y	L	E	G	T	C	V	E	W	L	R	R	Y	L	E	N	G	K	E	T	L	Q	R	T	-	-	-
R	V	A	E	Q	L	R	A	Y	L	E	G	L	C	V	E	W	L	R	R	Y	L	E	N	G	K	E	T	L	Q	R	A	-	-	-
G	N	A	K	Y	W	R	N	Y	L	E	G	T	C	V	K	W	L	S	R	Y	L	E	N	G	K	E	T	L	Q	R	A	-	-	-
G	A	A	D	H	Y	R	N	Y	V	E	G	T	C	V	E	C	-	R	R	Y	L	E	-	-	-	-	-	-	-	-	-	-	-	-
G	Y	A	E	G	L	K	Q	Y	L	E	E	T	C	V	E	W	L	R	R	Y	V	E	Y	G	K	A	E	L	G	R	R	-	-	-
G	V	P	E	R	R	K	H	Y	L	E	E	T	C	V	Q	W	L	R	R	Y	V	E	H	G	K	E	E	-	G	R	T	-	-	-

1.8 Cálculo de distancias genéticas entre las aves silvestres de canto y el resto de vertebrados

Como ya hemos dicho anteriormente, entendemos por distancia a una medida del grado de divergencia entre dos secuencias. Por ello, fueron calculadas las “distancias p” entre los distintos grupos analizados (Figura R7).

Figura R7. Matriz de distancias de las secuencias de MHC de clase I, entre los distintos grupos analizados.

	1	2	3	4	5	6	7
1 Passerines							
2 Pez	0,732						
3 Reptil	0,653	0,563					
4 Primates	0,705	0,686	0,557				
5 Mamíferos	0,702	0,674	0,554	0,260			
6 <i>Homo sapiens</i>	0,726	0,710	0,563	0,174	0,279		
7 Galliformes	0,684	0,639	0,466	0,509	0,521	0,500	
8 Anfibio	0,713	0,593	0,494	0,584	0,576	0,594	0,595

En la figura R7, podemos ver como el grupo de aves *Passeriformes* dista prácticamente lo mismo de peces, anfibios, mamíferos, primates (incluido *Homo sapiens*), que de reptiles y *Galliformes*, (0,653 y 0,684) siendo estos dos últimos grupos de los que menos distancia se encuentra, aunque no se han encontrado diferencias significativas con el resto (p-valor >0.001).

Por otro lado se observa que las aves *Galliformes* empleadas en el estudio, muestran una distancia más pequeña si se comparan con *Homo sapiens* y con el grupo de los reptiles (0,466) resultando llamativo que este grupo diste menos con especies de clases distintas que con la suya propia (Aves).

La secuencia de MHC clase I de la especie *Homo sapiens* dista 0,174 del resto de primates analizados (siendo esta la menor distancia encontrada en la tabla) y 0,279 del grupo de mamíferos.

En la tabla R8 podemos observar de forma más específica las distancias genéticas del grupo de passerines y de algunas de las especies utilizadas frente a cada uno de los grupos.

En la tabla R8 se puede ver de forma más clara como el grupo de “passerines” dista prácticamente lo mismo de todos los vertebrados a los que se ha comparado. La menor distancia encontrada en este grupo es la de *Coturnix japónica* y las máximas serían a *Homo sapiens* y a *Danio rerio* (0,73).

Tabla. R8 Matriz de distancias de las secuencias de MHC de clase I, entre las especies y los distintos grupos analizados. Las distancias del grupo “passerines” son el resultado de la media de todas las especies de aves silvestres de canto empleadas en este trabajo.

	Passerines	Pez	Reptil	<i>Homo sapiens</i>	Galliformes	Primates
Passerines	0,00	0,73	0,65	0,73	0,67	0,71
<i>Danio rerio</i>	0,73	0,00	0,64	0,71	0,65	0,69
<i>Ameiva ameiva</i>	0,65	0,56	0,00	0,56	0,48	0,56
<i>Hylobates lar</i>	0,70	0,68	0,53	0,22	0,51	0,00
<i>Pongo pygmaeus</i>	0,70	0,69	0,55	0,14	0,51	0,00
<i>Gorilla gorilla</i>	0,71	0,68	0,57	0,21	0,52	0,00
<i>Bos taurus</i>	0,72	0,67	0,56	0,27	0,52	0,27
<i>Canis familiaris</i>	0,72	0,67	0,55	0,21	0,51	0,26
<i>Mus musculus</i>	0,68	0,68	0,53	0,33	0,52	0,31
<i>Sus scrofa</i>	0,71	0,68	0,56	0,26	0,51	0,23
<i>Rattus norvegicus</i>	0,68	0,68	0,53	0,25	0,48	0,24
<i>Pan troglodytes</i>	0,72	0,70	0,57	0,10	0,51	0,00
<i>Homo sapiens</i>	0,73	0,71	0,56	0,00	0,52	0,00
<i>Pan paniscus</i>	0,70	0,67	0,56	0,21	0,49	0,00
<i>Equus caballus</i>	0,71	0,67	0,59	0,30	0,57	0,29
<i>Ovis aries</i>	0,70	0,67	0,57	0,33	0,56	0,30
<i>Gallus gallus</i>	0,68	0,64	0,47	0,50	0,00	0,51
<i>Coturnix japonica</i>	0,66	0,65	0,49	0,53	0,00	0,51
<i>Xenopus laevis</i>	0,71	0,59	0,49	0,59	0,59	0,58

El grupo “pez” que consta de una única especie, dista menos de *Ameiva ameiva* (reptil) y de *Xenopus laevis* (anfibio) (0,56; 0,59) que del resto. Podemos destacar, que la distancia que muestra respecto a *Homo sapiens* y al grupo de “passerines” es prácticamente similar (0,71; 0,73) y mayor respecto a las dos aves del grupo “Galliformes” (0,64;0,65).

Del grupo “reptiles” se puede subrayar, que la distancia genética de las secuencias de MHC de clase I, es mayor respecto al grupo de “passerines” y a *Danio rerio* (0,65; 0,64) que al resto de especies utilizadas, destacando las distancias de las aves *Galliformes* (0,47; 0,49), *Homo sapiens* (0,56) y a algunos mamíferos como *Hylobates lar*, *Mus musculus* y *Rattus norvegicus* (0,53).

Respecto a las distancias entre *Homo sapiens* y el resto de especies, se observa que dista muy poco su secuencia del resto del grupo de “mamíferos” y del grupo de “primates”, siendo la menor distancia con *Pan troglodytes* (0,10), como era de esperar. Cabe destacar que dista menos de las dos aves pertenecientes al grupo de “Galliformes” (0,50; 0,53) y del reptil *Ameiva ameiva* (0,56), que de nuestro grupo de aves *Passeriformes* (0,73).

La comparativa de distancias del grupo “Galliformes” nos vuelve a mostrar lo mismo, y es que la secuencia de MHC de clase I, dista menos con reptiles (0,58), con humanos (0,52) e incluso con anfibios (0,59) que con aves *Passeriformes* (0,67).

Por último destacar que, el grupo denominado “primates” sigue la misma tendencia que hemos visto hasta ahora, acercándose mucho más su distancia genética a las aves *Galliformes* (0,51), incluso a peces anfibios (0,56) y reptiles (0,58) que al grupo de las aves silvestres de canto.

1.9 Análisis de las posiciones aminoacídicas conservadas en ambos grupos (passerines y resto de vertebrados)

En el alineamiento de secuencias proteicas, el grado de similitud entre los aminoácidos que ocupan una posición concreta en la secuencia puede interpretarse como una medida aproximada de conservación en una región particular.

Para determinar las regiones aminoacídicas conservadas de los dos grandes grupos objeto de estudio, (aves silvestres de canto y resto de vertebrados elegidos), en las secuencias de los exones 2 y 3 de MHC de clase I, se han ido examinado minuciosamente cada una de las posiciones de las 46 muestras, previamente alineadas (Figuras R2, R3, R5 y R6). La tabla R9 nos muestra un resumen de todas las posiciones analizadas. En ella se observan por separado las regiones conservadas en aves *Passeriformes* (izquierda) y las conservadas en el resto de vertebrados (derecha).

Se pueden localizar un total de 79 posiciones conservadas en el grupo de aves *Passeriformes*, frente a un total de 30 en el grupo de vertebrados elegidos. Los números reflejados en la tabla coinciden con la posición en la secuencia, seguido de la letra que corresponde al residuo aminoacídico en ese lugar. En amarillo se encuentran destacadas las posiciones conservadas en cada uno de los grupos por separado. En verde y rojo, están subrayados los 21 lugares conservados en los dos grupos (10*, 18, 26, 29, 33, 38, 59, 60, 86, 87, 91, 94, 96, 100,101, 118,120,122, 159, 164, 168).

De estas 19 regiones, tan solo 2 en el grupo de aves *Passeriformes* difieren en los residuos respecto al resto de vertebrados (sombreado en rojo tabla R9).

Se puede observar, que las siete posiciones conservadas en vertebrados descritas por (Grossberger y Parham, 1992) (3,10, 29, 96, 101, 120, 164) hemos podido comprobar 6 de ellas.

Todas, están también conservadas en las aves silvestres de canto, pero existen dos que poseen residuos distintos a los esperados. Se localizan en la posición 10 donde hay una valina (V) y en la posición 96 donde hay leucina (L) en vez de treonina (T) y glutamina (Q)

respectivamente. En ambos casos se ha sustituido un aminoácido de tipo polar neutro por otro de naturaleza apolar (ver tabla R9 y anexo 4).

Tabla R9. Análisis de posiciones conservadas de los dos grandes grupos objeto de estudio. A la izquierda observamos las posiciones conservadas encontradas en las secuencias aminoacídicas, de nuestras muestras de pequeñas aves de canto. A la derecha se muestran las posiciones conservadas del resto de vertebrados escogidos para este trabajo. Los números reflejados en la tabla coinciden con la posición en la secuencia, seguido de la letra que corresponde al residuo aminoacídico en ese lugar. En verde, están destacadas las posiciones conservadas en ambos grupos. En rojo, las posiciones conservadas en ambos grupos pero con residuos distintos. En amarillo las posiciones conservadas sólo en cada grupo.

Posiciones conservadas en los exones 2 y 3 de MHC de las “Aves silvestres de canto” analizadas					Posiciones conservadas en los exones 2 y 3 de MHC de los Vertebrados analizados	
8. L	37. D	81. L	108. S	140. A	1. G	87. Q
10. V*	38. S	85. Y	112. G	141. E	3. H	91. G
12. V	39. E	86. N	113. S	147. W	4. S	94. T
13. S	40. R	87. Q	117. G	149. E	10. T	96. Q
16. S	41. G	88. S	118. Y	154. E	18. G	100. G
17. P	46. P	91. G	119. D	155. R	26. G	101. C
18. G	47. L	93. H	120. G	158. N	27. Y	115. Q
20. P	48. T	94. T	122. D	159. Y	28. V	118. Y
21. Q	50. W	96. L	125. S	150. L	29. D	120. G
22. F	59. Y	100. G	126. F	161. K	33. F	122. D
26. G	60. W	101. C	127. D	162. H	38. S	159. Y
29. D	64. T	103. L	130. S	164. C	51. W	164. C
30. G	65. Q	104. L	131. G	165. P	59. Y	168. L
32. P	72. H	105. S	133. F	166. E	60. W	175. Q
33. F	78. L	106. D	136. A	168. L	86. N	179. L
36. Y	79. E	107. G	138. S			

Posiciones conservadas en nuestras muestras de aves que no se encuentran en el resto de vertebrados

Posiciones conservadas en ambos grupos pero con residuos distintos.

*Existen dos excepciones en esta posición

Posiciones conservadas en ambos grupos.

Posiciones conservadas en el resto de vertebrados que no lo están en aves passeriformes.

Posiciones conservadas en ambos grupos pero con residuos distintos.

Posiciones conservadas en ambos grupos.

Cabe destacar que en este trabajo, tras varias repeticiones de secuenciación, nos han dado dos excepciones en la posición 10 (Figura R2) para las muestras *S. canaria* que en vez de una valina (V) presenta una leucina (L) y en la especie *C. crassirostris* donde se localiza ácido glutámico (E) (anexo 4).

1.10 Cálculo del grado de conservación en cada una de las posiciones

Con el fin de obtener unos resultados más precisos sobre el grado de conservación de cada uno de los residuos de las secuencias de MHC de clase I, se calcularon en cada una de las posiciones el rango de conservación tanto en las muestras de aves *Passeriformes*, como en las muestras de los vertebrados escogidos para este trabajo.

Entendemos por rango de conservación (RC) a:

$$\text{RANGO DE CONSERVACIÓN (RC)} = n/d$$

n = Nº de residuos aminoacídicos observados en cada posición.

d = Nº de residuos diferentes encontrados en cada posición.

En la tabla R10, se detallan, en porcentaje, el rango de conservación en cada una de las posiciones para el exón 2 y 3 de las secuencias de MHC de clase I estudiadas.

Gracias a la tabla R10 podemos ver de nuevo, que los lugares conservados al 100% coinciden con las posiciones reflejadas en la tabla R9 anterior (10*, 18, 26, 29, 33, 38, 59, 60, 86, 87, 91, 94, 96, 100,101, 118,120,122, 159, 164, 168).

Como se ha explicado en el apartado anterior, en nuestro estudio la posición 10 nos da un par de excepciones para las especies *S. canaria* y *C. crassirostris*. Si no tenemos en cuenta estas excepciones el RC pasaría de 33% a 100%.

En aves *Passeriformes* la posición con un menor grado de conservación con un 25% sería la 9.

En el resto de vertebrados los valores más bajos de RC, están entre 11-16%. En el exón 2 con un 11% estarían las posiciones 67 y 79; con un 12% tendríamos la posición 66; con un 14% estarían la 24, 41 y 70; por último con un 16% tendríamos las posiciones 9,17 y 55.

En el exón 3 tendríamos con un porcentaje de 11% la 153; con un 12% de RC las posiciones 97, 114, 116, 149, 156, 157; con 14 % estarían los lugares 131, 138 y 159 y por último con un rango de conservación de 16% tendríamos 95, 99, 142, 151, 152, 163.

Por lo que existen un total de 9 lugares de muy baja conservación en el exón 2 frente a 16 del exón 3.

Tabla R10. Tabla de porcentaje de rango de conservación (RA) en cada una de las posiciones de las secuencias aminoacídicas de MHC de clase I, para los exones 2 y 3, en los dos grandes grupos objeto de estudio (aves *Passeriformes* y resto de vertebrados). La columna “aa” muestra la posición de cada uno de los aminoácidos, en blanco se observan las regiones donde existe un RC del 100%.

RANGO DE CONSERVACIÓN MHC-I (%)						RANGO DE CONSERVACIÓN MHC-I (%)					
EXON 2						EXON 3					
aa	Passeriformes	Vertebrados	aa	Passeriformes	Vertebrados	aa	Passeriformes	Vertebrados	aa	Passeriformes	Vertebrados
1	-	100	46	100	25	91	100	100	136	100	25
2	-	50	47	100	33	92	50	25	137	50	33
3	-	100	48	100	33	93	100	50	138	100	14
4	-	100	49	33	25	94	100	100	139	50	20
5	-	50	50	100	25	95	33	16	140	100	50
6	-	33	51	50	100	96	100	100	141	100	33
7	-	50	52	50	33	97	50	12	142	50	16
8	100	50	53	33	25	98	50	50	143	50	50
9	25	16	54	33	20	99	50	16	144	33	25
10	100/33*	100	55	50	16	100	100	100	145	50	20
11	33	25	56	50	50	101	100	100	146	50	33
12	100	50	57	50	25	102	50	50	147	100	25
13	100	50	58	50	25	103	100	25	148	50	33
14	50	33	59	100	100	104	100	25	149	100	12
15	50	33	60	100	100	105	100	20	150	50	50
16	100	33	61	50	50	106	100	25	151	50	16
17	100	16	62	50	25	107	100	33	152	33	16
18	100	100	63	33	33	108	100	25	153	50	11
19	50	33	64	100	50	109	50	20	154	100	25
20	100	50	65	100	50	110	50	20	155	100	25
21	100	25	66	33	12	111	50	25	156	50	12
22	100	50	67	33	11	112	100	50	157	50	12
23	33	25	68	50	25	113	100	33	158	100	33
24	33	14	69	33	16	114	33	12	159	100	100
25	33	50	70	33	14	115	50	100	160	100	25
26	100	100	71	50	25	116	33	12	161	100	25
27	50	100	72	100	33	117	100	33	162	100	33
28	33	100	73	50	20	118	100	100	163	25	16
29	100	100	74	33	20	119	100	50	164	100	100
30	100	50	75	50	50	120	100	100	165	100	33
31	50	33	76	50	20	121	50	25	166	100	33
32	100	25	77	50	25	122	100	100	167	50	25
33	100	100	78	100	33	123	50	50	168	100	100
34	50	33	79	100	11	124	50	25	169	50	25
35	50	25	80	50	25	125	100	33	170	50	25
36	100	50	81	100	33	126	100	50	171	50	50
37	100	33	82	50	20	127	100	33	172	50	50
38	100	100	83	50	20	128	50	25	173	50	33
39	100	33	84	50	50	129	33	25	174		25
40	100	25	85	100	50	130	100	33	175		100
41	100	14	86	100	100	131	100	14	176		22
42		20	87	100	100	132	50	25	177		33
43	33	25	88	100	50	133	100	25	178		20
44	33	33	89	50	33	134	50	33	179		100
45	50	20	90	-	20	135	50	33	180		33
									181		25
									182		33
									183		20

1.11 Comparativa de medias de rango de conservación en los dos grupos

Por otro lado se han realizado una comparativa de las medias del RC entre aves *Passeriformes* y el resto de vertebrados analizados (tabla R11).

Cuando se comparan las medias de RC en el total de la secuencia observamos que existe un porcentaje mayor de conservación en aves *Passeriformes* que en el resto de vertebrados (70,78 frente a 40,97).

Si se analizan el exón 2 y 3 por separado, no se han encontrado diferencias significativas

entre el rango de conservación en ambos ($p \geq 0,05$).

Se observa que en aves *Passeriformes* existe un mayor porcentaje de conservación en el exón 3, corroborando los resultados obtenidos en los diagramas de entropía del apartado 1.2; mientras que en vertebrados la diferencia es al revés siendo el exón 2 ligeramente el más conservado de los dos. Considerando las dos comparativas siempre es mayor el porcentaje de rango de conservación (RC) en aves *Passeriformes* (habiéndose encontrado diferencias significativas, $p \geq 0,05$) pero debemos destacar que se emplean 4 géneros distintos con muchas especies frente a 17 géneros distintos con 1 especie cada uno. Para que esta comparación fuese más equilibrada se necesitarían un mayor número de secuencias de géneros distintos de aves silvestres de canto.

Tabla R11. Comparativa de las medias de porcentaje de rango de conservación (RC) en los dos grupos principales a estudio, Entendemos por SEM a:

$$\text{SEM} = \text{ERROR ESTANDAR DE LA MEDIA} = \frac{s}{\sqrt{N}}$$

s = es la desviación estándar (es decir, la estimación basada en la muestra de la desviación estándar de la población).

N = es el tamaño (en este caso, el número de posiciones de la muestra).

<u>TOTALES</u>			
<u>RANGO DE CONSERVACIÓN MHC-I (%)</u>			
	<u>PASSERIFORMES</u>	<u>VERTEBRADOS</u>	
MEDIA	70,78658537	40,97802198	
SEM	2,174075634	2,108389729	
N	173	183	

<u>EXON 2</u>			<u>EXON 3</u>		
<u>RANGO DE CONSERVACIÓN MHC-I (%)</u>			<u>RANGO DE CONSERVACIÓN MHC-I (%)</u>		
	<u>PASSERIFORMES</u>	<u>VERTEBRADOS</u>		<u>PASSERIFORMES</u>	<u>VERTEBRADOS</u>
MEDIA	68,34567901	43,48888889	MEDIA	73,1686747	38,52173913
SEM	3,079419159	3,034024469	SEM	3,067840031	2,899543754
N	90	90	N	83	93

1.12 Porcentaje de similitud entre las secuencias estudiadas

Como ya se ha comentado anteriormente, en el alineamiento de secuencias proteicas, el grado de similitud entre los aminoácidos que ocupan una posición concreta en la secuencia puede

interpretarse como una medida aproximada de conservación en una región particular.

En la tabla R12 hemos calculado la media porcentajes de similitud entre las secuencias aminoacídicas de las especies de vertebrados estudiadas y el conjunto de aves *Passeriformes* frente a distintos grupos (“passerines, pez, reptil, *Homo sapiens* y *Galliformes*”)

Tabla R12. Porcentaje medio de similitud entre las secuencias estudiadas. Sombreados en amarillo están los porcentajes más destacados.

	Passerines	Pez	Reptil	<i>Homo sapiens</i>	<i>Galliformes</i>
Passerines	100,0%	9,8%	17,5%	23,0%	23,1%
<i>Danio rerio</i>	9,8%	100,0%	46,5%	38,0%	40,5%
<i>Ameiva ameiva</i>	17,5%	46,5%	100,0%	54,0%	64,4%
<i>Hylobates lar</i>	20,1%	39,7%	53,8%	82,0%	57,5%
<i>Pongo</i>	22,5%	39,1%	52,1%	88,8%	56,5%
<i>Gorilla gorilla</i>	22,5%	40,3%	50,3%	85,0%	58,8%
<i>Bos taurus</i>	22,7%	41,9%	50,3%	79,4%	55,3%
<i>Canis familiaris</i>	22,7%	41,9%	51,5%	83,5%	57,8%
<i>Mus musculus</i>	22,5%	39,7%	52,1%	78,8%	57,1%
<i>Sus scrofa</i>	22,4%	42,3%	54,4%	80,4%	50,8%
<i>Rattus</i>	22,5%	42,1%	51,5%	82,1%	55,2%
<i>Pan troglodytes</i>	21,6%	42,7%	54,4%	90,6%	58,2%
<i>Homo sapiens</i>	23,0%	38,0%	54,0%	100,0%	55,6%
<i>Pan paniscus</i>	22,1%	46,9%	51,5%	82,4%	56,4%
<i>Equus caballus</i>	19,9%	46,9%	58,5%	75,1%	50,0%
<i>Ovis aries</i>	21,6%	41,4%	50,1%	71,1%	52,3%
<i>Gallus gallus</i>	22,2%	40,8%	65,0%	57,1%	100,0%
<i>Coturnix</i>	23,9%	40,1%	63,8%	54,1%	100,0%

En la tabla R12, se puede observar que el grupo denominado “passerines” (que representa la media del porcentaje de similitud de todas las especies de aves *Passeriformes* utilizadas en este estudio) tiene un porcentaje de similitud respecto al resto de especies muy parecido (una media de (21,8%) exceptuando a *Danio rerio* (9,8%). Cabe destacar que el porcentaje de similitud entre aves *Passeriformes* y aves *Galliformes* es prácticamente el mismo que entre el primer grupo y humanos.

Respecto al grupo “pez”, se puede ver como el porcentaje de similitud es mucho menor frente a *Passeriformes* que frente al resto de vertebrados cuya media oscila (41,92%).

Si hablamos del grupo “reptiles”, volvemos a observar lo mismo y es que su secuencia se asemeja más al resto de especies de vertebrados (*Homo sapiens* un 54% y con *Galliformes* de media un 64,4%) que con aves *Passeriformes*.

En el grupo denominado *Homo sapiens* se puede destacar de nuevo que su secuencia es bastante más similar a cualquier ave galliforme que a cualquiera de las aves *Passeriformes* estudiadas, incluso se asemeja más a las secuencias de *Danio rerio* (pez) y *Ameiva ameiva* (reptil). Por otro lado, dentro de lo esperado, vemos que la similitud con *Pan troglodytes* es de (90,6%).

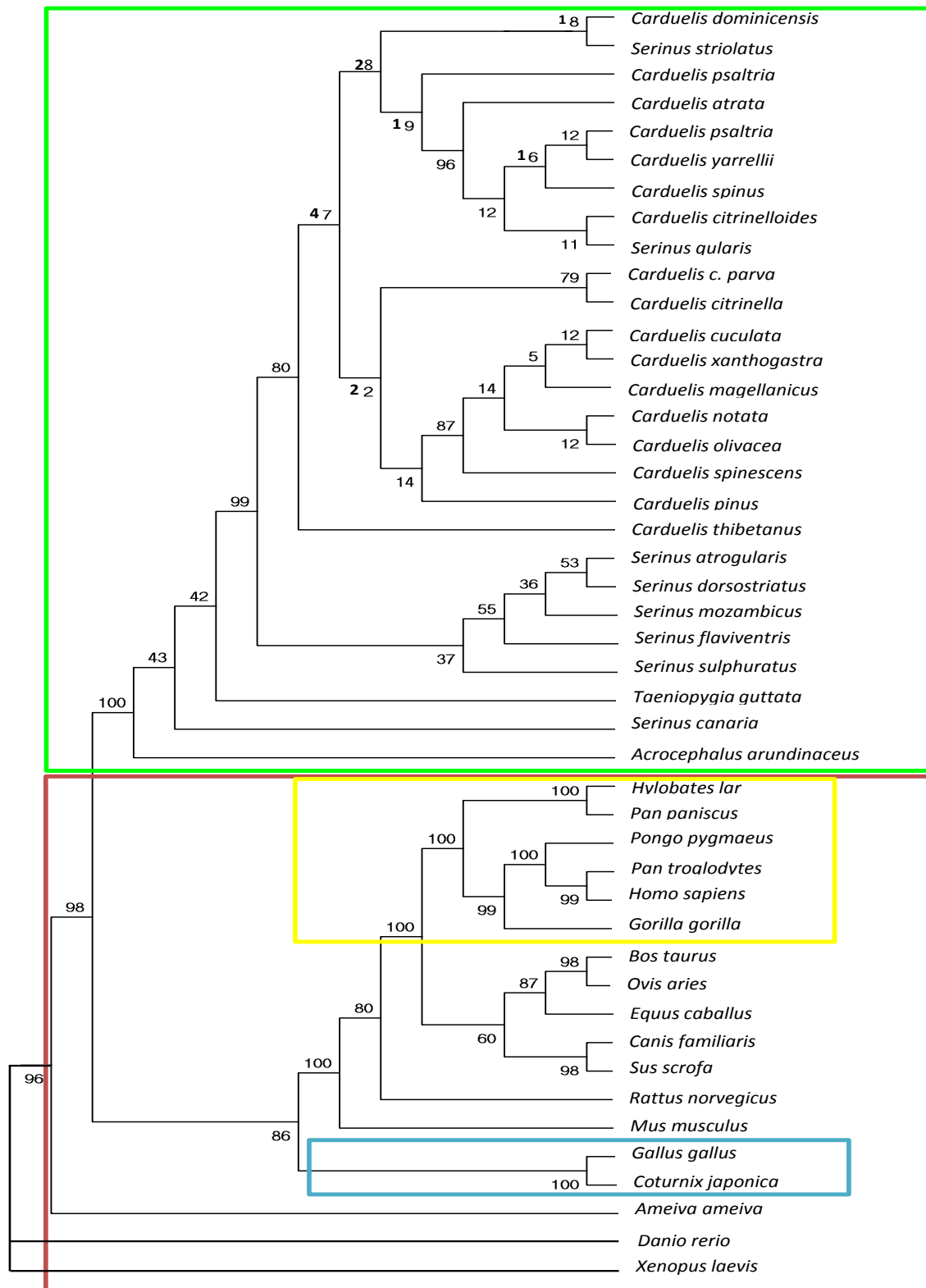
Por último en el grupo de las aves *Galliformes*, volvemos a encontrar lo mismo, una proporción media de similitud con *Homo sapiens* de 55,6 % frente a 23,1 % en las aves *Passeriformes*, y una proporción media de similitud de 55,09% respecto al resto de especies.

Para comprobar la tendencia de la secuencia aminoacídica de esta molécula, que tiende a separar el grupo de aves *Galliformes* del resto de aves del estudio, se recurrió a un análisis bayesiano, entendiendo que, como ya se ha comentado anteriormente, los genes que están sujetos a selección de balance (como ocurre con los genes MHC) pueden complicar los análisis filogenéticos.

En la figura R8, podemos observar los resultados de este análisis. En ella se ven tres grandes grupos, uno recuadrado en verde, que corresponde a algunas de las aves silvestres de canto estudiadas, otro recuadrado en granate, que corresponde a los vertebrados escogidos para este trabajo, y dentro de este y alejado del grupo de aves *Passeriformes*, estarían las aves *Galliformes* (*Coturnix japonica* y *Gallus gallus*) recuadradas en azul.

Una vez más, se muestra la distinta evolución con respecto a MHC de estos dos tipos de órdenes dentro de la clase Aves.

Figura R8: Árbol de inferencia bayesiana construido a partir de las secuencias de aminoácidos para los dominios alfa-1 y alfa-2. En él se pueden observar dos diferentes grupos, uno que engloba las aves silvestres de canto (verde) y otro que son el resto de vertebrados (granate). Dentro de este último grupo se ven, el grupo de primates (amarillo) separado del resto de mamíferos y el grupo de aves *Galliformes* (azul) alejado del resto de géneros dentro de esta clase.



1.13 Estudio de intrón 2 de MHC de clase I en aves *Passeriformes*

Como ya se ha comentado anteriormente, las características únicas detectadas en el MHC de pollo (pocos genes, distancias intergénicas muy cortas e intrones pequeños) llevaron a Kaufman y col. a proponer la hipótesis del MHC mínimo-esencial (Kaufman *et al.*, 1995). Nos hemos propuesto como objetivo analizar si el diseño tan compacto del Sistema Principal de Histocompatibilidad del pollo, se mantiene en otras aves y compararlo a su vez con el resto de animales vertebrados escogidos para este trabajo.

Para completar esta comparativa, hemos añadido a nuestras muestras, secuencias recogidas de otros trabajos (Shiina *et al.*, 1995; Arnaiz-Villén *et al.*, 2007a; Arnaiz-Villén *et al.*, 2010b; Karlsson *et al.*, 2013) así como del archivo del grupo.

Las secuencias del intrón 2 en aves silvestres de canto, que poseen aproximadamente 300 pb, fueron comparadas entre 4 géneros distintos (*Carduelis*, *Serinus*, *Fringilla*, *Passer*) con la región alineable del resto de vertebrados objeto de estudio.

La figura R9, muestra todas las secuencias alineadas. En ella se observan, sombreados en granate, los lugares conservados. Se han encontrados un total de 78 lugares variables si se comparan los intrones de aves silvestres de canto. Los porcentajes de nucleótidos para estas muestras son: 20,4 % A; 25,7% T; 19,1% C; 34,8% G. Los resultados de tasas de transición / transversión son $k1 = 4.42$ (purinas) y $k2 = 4.53$ (pirimidinas). El sesgo general transición / transversión es $R = 2.537$, donde $R = [A \times G \times k1 + T \times C \times k2] / [(A + T) \times (T + C)]$. (Tamura *et al.*, 2013)

En estas muestras se han encontrado cuatro posiciones que presentan ambigüedades (más de una clase de nucleótidos en esa posición), sombreadas en amarillo en la figura R9. En la tabla R13 se muestra un resumen de estas posiciones y las especies implicadas.

Tabla R13. Listado de ambigüedades encontradas en las secuencias del intrón 2.

AMBIGÜEDADES INTRÓN 2			
AVES PASSERIFORMES			
Posición	Ambigüedad	Nucleótidos	Muestras
68	R	G-A	<i>C. cucullata</i>
70	R	G-A	<i>C. cucullata</i>
73	R	G-A	<i>C. notata</i> ; <i>C. olivacea</i>
162	M	A-C	<i>C. olivacea</i> ; <i>C. yarrellii</i>

Figura R9. Secuencias alineadas del intrón 2 de aves *Passeriformes* y parte de las secuencias del resto de vertebrados. Sombreado en granate están los sitios conservados, en amarillo las ambigüedades observadas.

Posición	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
<i>Carduelis atrata</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	T	T	G	G	A	A	T
<i>Carduelis atriceps</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Carduelis carduelis parva</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Carduelis citrinella</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Carduelis crassirostris</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	T	T	G	G	A	A	T
<i>Carduelis cucullata</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	T	T	G	G	A	A	T
<i>Carduelis magellanica</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	T	T	G	G	A	A	T
<i>Carduelis notata</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	T	T	G	G	A	A	T
<i>Carduelis olivacea</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	T	T	G	G	A	A	T
<i>Carduelis psaltria</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Carduelis spinescens</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	T	T	G	G	A	A	T
<i>Carduelis spinus</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Carduelis xanthogastra</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Carduelis yarrelli</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	T	T	G	G	A	A	T
<i>Fringilla coelebs</i>	G	T	G	A	G	T	G	G	G	G	G	G	G	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T	
<i>Passer domesticus</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Serinus atrogularis</i>	G	T	G	A	G	T	G	C	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Serinus citrinelloides</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Serinus dorsostriatus</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Serinus flaviventris</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Serinus gularis</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Serinus mozambicus</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Serinus striolatus</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Serinus sulphuratus</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Serinus thibetanus</i>	G	T	G	A	G	T	G	G	G	G	G	A	-	A	T	C	T	G	T	G	C	A	G	T	T	G	G	A	A	T
<i>Ovis aries</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	T	T	G	G	A	A	T
<i>Rattus norvegicus</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	T	T	G	G	A	A	T
<i>HLA-A</i>	G	T	G	A	G	T	G	A	C	C	C	C	G	C	C	C	C	G	G	G	G	C	G	C	A	G	G	T	C	A
<i>Hylobates agilis</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Pan paniscus</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	C	T	G	G	A	A	T
<i>Pan troglodytes</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	T	T	G	G	A	A	T
<i>Pongo pygmaeus</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	T	T	G	G	A	A	T
<i>Coturnix coturnix</i>	G	T	G	A	G	T	G	G	G	G	G	G	-	A	T	C	T	G	T	G	C	A	G	T	T	G	G	A	A	T
<i>Gallus gallus</i>	G	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	T

G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	C	T	G	G	
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	A	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C	A	G	T	G	G	G	G	C	T	G	G
G	G	G	A	A	C	T	G	T	G	G	G	G	C	T	T	G	G	T	G	G	A	G	A	T	C</											

Respecto a la comparación con el resto de muestras de vertebrados, destacar que se han encontrado un total de 63 posiciones conservadas, subrayando que las secuencias de *Homo sapiens* y *Gallus gallus*, son las que más difieren del resto.

1.14 Porcentaje medio de similitud entre los intrones de las muestras estudiadas

Como ya se ha comentado en apartados anteriores (1.12) el grado de similitud entre secuencias puede interpretarse como una medida aproximada de conservación en una región particular, en este caso, del intrón 2 de la molécula MHC de clase I.

La tabla R14 muestra, en porcentaje, una comparación media de los valores de similitud entre especies y grupos. El grupo Passerines englobaría las especies del género *Carduelis* y *Serinus*, representando al grupo de mamíferos estarían las especies de *Rattus norvegicus* y *Ovis aries* y por último el grupo denominado primates estaría formado por *Hylobates lar*, *Pongo pygmaeus*, *Gorilla gorilla*, *Pan troglodytes*, *Pan paniscus*.

En ella, se puede observar cómo el porcentaje medio de similitud de la secuencia del intrón 2 del MHC de clase I entre aves silvestres de canto de los géneros *Carduelis* y *Serinus* es de 98,15%. A su vez, tendrían un alto porcentaje medio de similitud (91,8%) con las especies de mamíferos utilizadas para esta comparativa (*Rattus norvegicus* y *Ovis aries*), así como también con el grupo de primates (95,07%) y con el ave terrestre *Coturnix japonica* (93,57%). Sin embargo, como era de esperar, tan sólo poseen una media de un 20,79 % de similitud con *Gallus gallus*.

Respecto al grupo de mamíferos, podemos ver que existe una mayor semejanza de estas secuencias con el grupo de passerines (incluyendo la especie *Passer domesticus*) con el grupo de primates, y con la especie *Coturnix japonica*, que con el resto, destacando la baja similitud con *Homo sapiens* (18,85%) muy parecida a la que presentan respecto a la especie *Gallus gallus* (20,07%) (destacada en naranja).

Dentro de la comparación de la especie *Homo sapiens*, se puede observar que el grado medio de similitud con los otros grupos ronda el 21,37 %, destacando que el mayor porcentaje no es con el grupo de los primates, como era de esperar, si no con *Gallus gallus* (32,6%), ocurriendo lo mismo si comparamos este último con el resto de especies y de grupos. Podemos destacar también que esta especie posee mayor porcentaje de similitud en su intrón, con el ser humano que con la otra especie de su misma Clase aquí estudiada (*Coturnix japonica*) (13%).

Tabla R14. Porcentaje medio de similitud entre algunas especies y grupos estudiados de la secuencia de intrón 2 de MHC de clase I. El grupo Passerines engloba las especies del género *Carduelis* y *Serinus* de la figura 9, en el grupo mamíferos estarían las especies de *Rattus norvegicus* y *Ovis aries* y por último el grupo denominado primates estaría formado por *Hylobates lar*, *Pongo pygmaeus*, *Gorilla gorilla*, *Pan troglodytes*, *Pan paniscus*.

	Passerines	Mamíferos	Homo sapiens	Gallus gallus
Carduelis y Serinus	98,15	91,8	18,7	20,79
Passer domesticus	94,21	89,35	18,9	18,12
Mamíferos	91,8	98,6	18,85	20,07
Homo sapiens	18,7	18,85	-	32,6
Gallus gallus	20,79	20,07	32,6	-
Coturnix japonica	93,57	95,8	20,57	13
Primates	95,07	97,2	20,9	21,03

1.15 Comparación de la longitud del intrón 2 en las muestras estudiadas

Con el fin de comprobar, si el diseño compacto del intrón 2 del Sistema Principal de Histocompatibilidad del pollo se mantiene en nuestras muestras de aves *Passeriformes* y en el resto de vertebrados utilizados para este estudio, se ha hecho una comparación de la longitud de los mismos, reflejada en la tabla R15.

En esta tabla se puede ver, como la longitud media del intrón 2 en aves silvestres de canto que incluyen los géneros (*Carduelis*, *Serinus*, *Fringilla*, *Passer*) ronda las 303 pb.

Muy parecida, es la longitud media del grupo “mamíferos” (*Rattus norvegicus* y *Ovis aries*) de 304 pb, como también la del grupo “primates” (*Hylobates lar*, *Pongo pygmaeus*, *Gorilla gorilla*, *Pan troglodytes*, *Pan paniscus*) de media 305 pb.

Dentro del grupo de aves terrestres (*Galliformes*), sombreadas en naranja, se puede observar la diferencia entre el intrón compacto del pollo, de 203 pb de longitud, frente al intrón de la codorniz de 305 pares de bases, mucho más parecido al de *Passeriformes* que al de la otra especie del grupo “*Galliformes*”.

Por último, destacar la longitud del intrón en *Homo sapiens*, de 241 pb, mucho más próxima a la de *Gallus gallus*, que a la del resto de organismos pertenecientes a su misma Clase.

Para comprobar si existen diferencias significativas entre las longitudes tan próximas, del intrón 2 del ser humano y del pollo respecto al resto secuencias analizadas, se realizó un test de Mann-Whitney con el programa estadístico SPSS 22.0, dándonos como resultado que sí existen diferencias significativas entre éstas secuencias y el resto ($p\text{-valor} \leq 0,05$).

Tabla R15. Diferencias entre las longitudes del intrón 2 en los diferentes grupos estudiados

	<u>CLASE</u>	<u>ORDEN</u>	<u>FAMILIA</u>	<u>pb</u>
<i>Carduelis atrata</i>	Aves	Passeriformes	Fringillidae	308
<i>Carduelis atriceps</i>	Aves	Passeriformes	Fringillidae	290
<i>Carduelis carduelis parva</i>	Aves	Passeriformes	Fringillidae	290
<i>Carduelis citrinella</i>	Aves	Passeriformes	Fringillidae	303
<i>Carduelis crassirostris</i>	Aves	Passeriformes	Fringillidae	308
<i>Carduelis cucullata</i>	Aves	Passeriformes	Fringillidae	309
<i>Carduelis magellanica</i>	Aves	Passeriformes	Fringillidae	309
<i>Carduelis notata</i>	Aves	Passeriformes	Fringillidae	309
<i>Carduelis olivacea</i>	Aves	Passeriformes	Fringillidae	309
<i>Carduelis psaltria</i>	Aves	Passeriformes	Fringillidae	304
<i>Carduelis spinescens</i>	Aves	Passeriformes	Fringillidae	304
<i>Carduelis spinus</i>	Aves	Passeriformes	Fringillidae	304
<i>Carduelis xanthogastra</i>	Aves	Passeriformes	Fringillidae	309
<i>Carduelis yarrelli</i>	Aves	Passeriformes	Fringillidae	309
<i>Fringilla coelebs</i>	Aves	Passeriformes	Fringillidae	286
<i>Passer domesticus</i>	Aves	Passeriformes	Passeridae	304
<i>Serinus atrogularis</i>	Aves	Passeriformes	Fringillidae	303
<i>Serinus citrinelloides</i>	Aves	Passeriformes	Fringillidae	302
<i>Serinus dorsostratus</i>	Aves	Passeriformes	Fringillidae	304
<i>Serinus flaviventris</i>	Aves	Passeriformes	Fringillidae	304
<i>Serinus gularis</i>	Aves	Passeriformes	Fringillidae	304
<i>Serinus mozambicus</i>	Aves	Passeriformes	Fringillidae	304
<i>Serinus striolatus</i>	Aves	Passeriformes	Fringillidae	304
<i>Serinus sulphuratus</i>	Aves	Passeriformes	Fringillidae	304
<i>Serinus thibetanus</i>	Aves	Passeriformes	Fringillidae	304
<i>Ovis_aries</i>	Mammalia	Artiodactyla	Bovidae	305
<i>Rattus_norvegicus</i>	Mammalia	Rodentia	Muridae	303
HLA-A	Mammalia	Primates	Hominidae	241
<i>Hylobates_agilis</i>	Mammalia	Primates	Hylobatidae	305
<i>Pan_paniscus</i>	Mammalia	Primates	Hominidae	304
<i>Pan_troglodytes</i>	Mammalia	Primates	Hominidae	305
<i>Pongo_pygmaeus</i>	Mammalia	Primates	Hominidae	305
<i>Coturnix_coturnix</i>	Aves	Galliformes	Phasianidae	305
<i>Gallus_gallus</i>	Aves	Galliformes	Phasianidae	203

2. Filogenia mitocondrial

Para finalizar, se realizó un estudio filogenético con el objetivo, de analizar la relación genética entre distintas tribus de la familia *Carduelini*, utilizando ADN mitocondrial, codificante para citocromo b, obtenido previamente por nosotros (Arnaiz-Villena *et al.*, 2012; Arnaiz-Villena *et al.*, 2009a; Arnaiz-Villena *et al.*, 2009b ; Arnaiz-Villena *et al.*, 2008; Arnaiz-Villena *et al.*, 2007b; Arnaiz-Villena *et al.*, 2007c; Arnaiz-Villena *et al.*, 2001; Arnaiz-Villena *et al.*, 1999a; Arnaiz-Villena *et al.*, 1998; Zamora *et al.*, 2006a; Zamora *et al.*, 2006b) para contrastar la relación evolutiva existente entre ellas con las relaciones filogenéticas observadas en las secuencias de histocompatibilidad de clase I de las especies analizadas.

Por otro lado, también hemos querido comprobar la concordancia de la agrupación filogenética molecular con la clásica agrupación fenotípica de los géneros de aves *Passeriformes*: *Rhodopechys*, *Carpodacus*, *Pyrrhula*, *Haematospiza*, *Pinicola* y *Uragus*.

A su vez destacar que se ha secuenciado por primera vez en este trabajo, el ADN mitocondrial de la especie *Rhodopechys sanguineus* (Arnaiz-Villena *et al.*, 2014a)

En primer lugar, se aplicó el método ML, que fue linearizado para determinar la antigüedad relativa de cada grupo (Figura R10).

En general, encontramos varios clados monofiléticos, por un lado los piquituertos o género *Loxia*; (bootstrap ML= 100) y los pinzones (género *Pyrrhula*; bootstrap ML = 83).

Las especies del género *Carpodacus* forman un clado polifilético, como ya fue demostrado por (Arnaiz-Villena *et al.*, 2001).

Las especies de los géneros *Carduelis* y *Serinus*, sin embargo, son agrupadas en varios grupos parafiléticos (Zamora *et al.*, 2006b).

Por otro lado, algunos grupos monotípicos (grupo taxonómico con un solo tipo nomenclatura) aparecen con géneros diferentes: *Pinicola enucleator* (pájaro de pino) forma un clado con género *Pyrrhula* (bootstrap ML= 81) , *Uragus sibiricus* aparece junto a *con Carpodacus rubicilloides* (bootstrap ML= 81) y *Haematospiza sipahi* (Pinzón escarlata) está agrupado junto con *Carpodacus erythrinus* (ML=49); concordando con los anteriores resultados (Arnaiz-Villena *et al.*, 2008; Arnaiz-Villena *et al.*, 2001).

Podemos observar también, dentro de lo esperado, cómo la especie *Rhodopechys obsoleta*, se agrupa con los verderones (*C. ambigua*, *C. spinoides*, *C. sínica* y *C. chloris*, sombreados en el árbol en verde) (ML= 87) separando el género *Rhodopechys* (Zamora *et al.*, 2006a).

Finalmente, destacar un grupo filogenético que contiene dos subespecies de *Leucosticte arctoa* (*L. a. arcota* y *L. a. tephrocotis*), tres especies del género *Rhodopechys*: *R. mongólica*, *R. githaginea* y *R. sanguineus* y la especie *Carpodacus nipalensis* (ML =100); a los que hemos denominado grupo de “pinzones de zonas áridas” (destacados en rojo en el árbol).

Teniendo en cuenta las divergencias genéticas (según el modelo de máxima verosimilitud) entre los taxones analizados (Figura R10), parece que los géneros *Carduelis* y *Serinus* divergieron hace poco y de forma simultánea, mientras que el género *Carpodacus* y los pinzones (*Pyrrhula*), así como los “pinzones de la zonas áridas” divergieron incluso antes que lo indicado en la base del árbol.

Por último, se realizó un análisis bayesiano para contrastar los resultados obtenidos anteriormente (Figura R11).

En este árbol de inferencia bayesiana, se pueden observar cómo los resultados son iguales a los obtenidos con el método de ML, englobando a *Rhodopechys sanguineus* dentro de las especies de su mismo género (sombreado en rojo) exceptuando al *Rhodopechys obsoleta* que vuelve a estar dentro del grupo de los verderones (sombreado en verde).

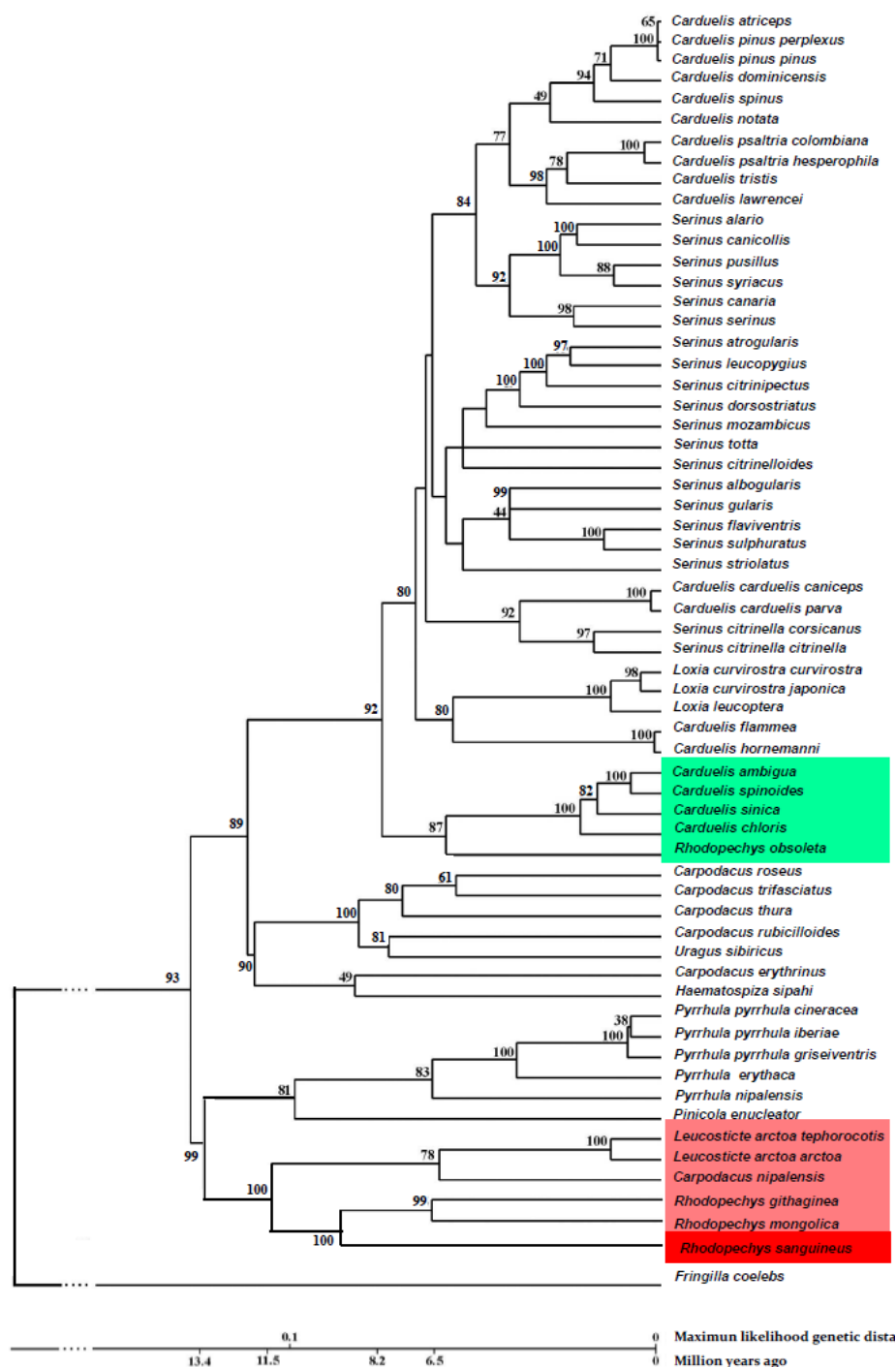


Figura R10. Árbol de máxima verosimilitud (ML) linearizado basado en la secuencia de 924 pb del citocromo b mitocondrial (ln L = 8.612,08). Las frecuencias de nucleótidos obtenidas fueron: A = 0.29280, C = 0,41180, G = 0,13990, T = 0,15550; las tasas de sustitución de nucleótidos: AC = 0.848200, AG = 6.167300, AT = 2.208400, CG = 0.175600, CT = 21.150900, GT = 1.0000). Modelo de Thorne (Thorne *et al.*, 1998) se utilizó para estimar longitudes de rama. Los valores de los Bootstrap se indican para cada nodo. La escala temporal indica la aproximación de los tiempos de divergencia de acuerdo con los resultados anteriores (Arnaiz-Villena *et al.*, 1998; Arnaiz-Villena *et al.*, 1999a; Arnaiz-Villena *et al.*, 2001; Fleischer *et al.*, 1998).

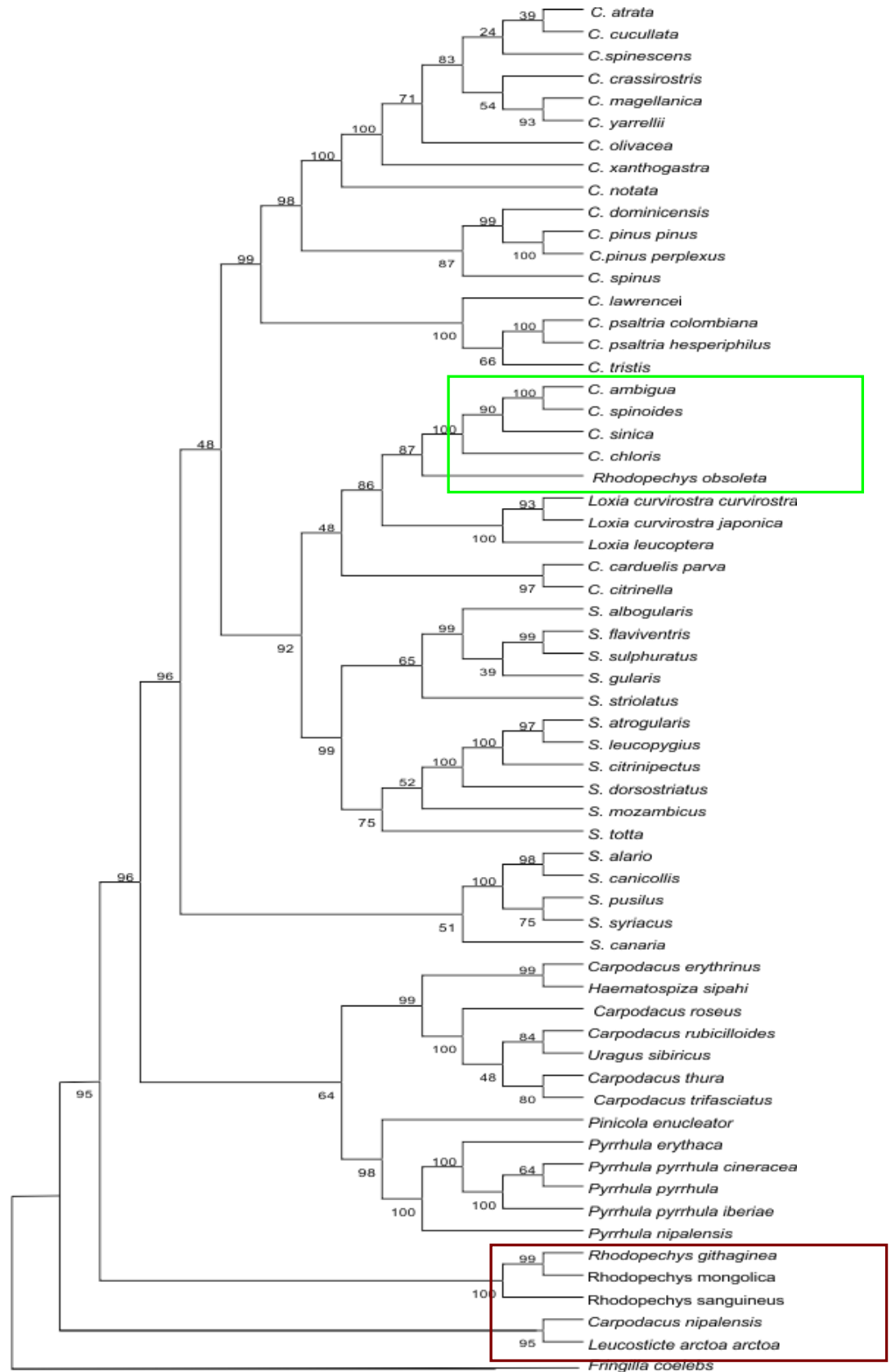


Figura R11. Arbol bayesiano (BI) basado en secuencias de citocromo b mitocondrial. El modelo de evolución utilizado fue GTR + I + G. 5,000,000 generaciones, los valores de probabilidad (x100) están indicados en cada nodo.

DISCUSIÓN

1. Sistema Principal de Histocompatibilidad en aves silvestres de canto

Análisis de la tasa de sustitución

Como se ha comentado varias ocasiones en este trabajo, la mayor parte de la variabilidad de las moléculas de MHC de clase I, está localizada en los exones 2 y 3, que codifican para los dominios alfa 1 y alfa 2, responsables de la característica forma de “valva” de unión peptídica. Numerosos estudios (Hughes y Hughes, 1995a; Hughes y Nei, 1988; Hughes y Nei, 1989b), han comprobado que dentro del sitio de reconocimiento antigénico (ARS), existen un mayor número de cambios nucleotídicos no sinónimos, que en el resto de la molécula.

El alto nivel de este tipo de sustituciones, indica que el polimorfismo está sujeto a un tipo de selección de balance. Este patrón de sustitución, que es el opuesto al visto en la mayoría de los genes que codifican para proteínas, indica que la selección natural favorece en cierta medida la variabilidad en esta región, posiblemente porque la heterocigosis del MHC puede aumentar la capacidad para presentar una mayor gama de péptidos distintos.

Al observar las tasas de sustitución por sitio sinónimo (d_s) y por sitio no sinónimo (d_n) dentro del ARS, y compararlas con los mismos valores dentro del resto de lugares de los exones 2 y 3 (no ARS) en todas las secuencias de aves silvestres de canto utilizadas en este trabajo (tablas R3 y R4), se puede apreciar cómo el número de sustituciones que producen cambio de aminoácido es significativamente mayor en la región de reconocimiento antigénico respecto al resto, no obstante en cualquiera de los casos, siempre d_s es mayor a d_n en las dos regiones.

El cociente entre d_n / d_s nos indica que tipo de evolución siguen las moléculas al variar, si la evolución es neutral (sin fuerzas selectivas) es igual a 1. En nuestro caso, el cociente d_n / d_s es < 1 tanto en la región ARS, como en el resto de posiciones, lo que nos indica que sobre estas moléculas existe un tipo de selección negativa, donde se tiende a eliminar la mayoría de las sustituciones no sinónimas.

Ante estos resultados, no parece que la selección de balance esté actuando sobre los genes MHC de clase I estudiados. Por tanto, el tipo de selección que experimentan las secuencias de histocompatibilidad aquí estudiadas aparentemente no es el característico de las moléculas presentadoras de antígenos de clase I clásicas. Sin embargo, este tipo de análisis requerirían disponer de una muestra más amplia, para evitar posibles errores de muestreo.

Por otro lado, no hay que descartar la posibilidad de que las posiciones aminoacídicas involucradas en sitio de reconocimiento antigénico sean otras distintas a las señaladas en este trabajo por analogía con las moléculas humanas

Análisis de las secuencias proteicas en aves *Passeriformes*

Al comparar las secuencias de las especies de aves silvestres estudiadas, observamos que los cambios obtenidos son poco numerosos, ya que son especies que han radiado en un corto periodo de tiempo y por lo tanto están muy próximas genéticamente. La mayoría de los cambios encontrados en nuestras especies (géneros *Carduelis* y *Serinus* especialmente), se localizan fuera de las zonas de interacción con el péptido a excepción de la posición 9 y la posición 58, la primera implicada en sitio de unión al péptido (tabla R6) y la segunda implicada en el receptor de células T (tabla R7).

Si se observan las figuras R2 y R3, se distinguen dos pequeñas regiones de mayor variabilidad comprendidas en el dominio alfa 1. La primera se localiza entre los residuos 9-11 y la segunda entre los residuos aminoacídicos 49-58.

Por lo tanto, la variabilidad observada está fuera de sitio de unión del péptido (excepto la posición 9) y fuera de las regiones hipervariables y de los puntos de contacto con otras moléculas excepto la posición número 10 que interacciona con las posiciones 56 y 62 de la $\beta 2m$ (Grossberger y Parham, 1992) y la posición 58 que está implicada en la interacción con el receptor de células T.

Esta escasa variabilidad apoya las dudas de si los genes de histocompatibilidad de clase I que estamos estudiando son clásicos o no clásicos, lo que nos hace tener en cuenta los errores de muestreo aunque son menos probables ya que las especies estudiadas presentan las mismas características y sabemos que están separadas en su aparición en la Tierra por millones de años. El análisis de la tasa de sustitución nos acercaba a la idea de que la selección a la que estaban sujetas estas moléculas era una selección negativa, muy poco característica de las moléculas clásicas por otro lado el hecho de no haber encontrado regiones hipervariables nos conduce de nuevo a la sospecha de que estas moléculas podrían ser no clásicas.

Variabilidad aminoacídica en aves silvestres de canto

Como ya se ha dicho anteriormente, tras comparar numerosas secuencias de histocompatibilidad de clase I, se determinó que la mayor variabilidad se concentra precisamente en los dominios alfa 1 y alfa 2, generalmente en las posiciones que forman el ARS (Parham *et al.*, 1988), por lo que nos basamos en estas regiones para analizar la variabilidad aminoacídica de nuestras secuencias.

Para identificarlas, se tomó como referencia la secuencia consenso de las proteínas humanas de clase I clásicas (HLA-Ia). En ellas se localizaron las posiciones implicadas en el reconocimiento antigénico (Bjorkman *et al.*, 1987b), así como las posiciones que fueron definidas como hipervariables en las moléculas HLA de clase I clásicas humanas (Parham *et al.*, 1988).

Hemos observado que ninguna de las posiciones variables de nuestras secuencias presentan más de dos alternativas, por lo que no existe hipervariabilidad en las secuencias proteicas MHC de clase I de estas aves, tal y como fue definida en las moléculas HLA de clase I clásicas humanas (Parham *et al.*, 1988) (tablas R6 y R7).

La mayor parte de las variaciones aminoacídicas se localizan fuera del sitio de reconocimiento antigénico, que se mantiene muy conservado.

Todo esto, nos lleva a pensar que puede que las especies analizadas en este trabajo estén expuestas a un repertorio limitado de patógenos, de ahí la reducida variedad de péptidos que pueden unir estas moléculas. También es posible que las moléculas MHC de clase I estudiadas, unan péptidos muy concretos o que tuvieran una función distinta de la presentación antigénica pudiendo ser homólogos de moléculas MHC no clásicas con funciones inmunosupresoras (Gomez-Prieto *et al.*, 2009; Donadi *et al.*, 2011). También esta poca variabilidad podría ser debida a un error de muestreo.

Cabe destacar, que el nicho aéreo que ocupan estas aves silvestres pequeñas de canto podría influir en el procesamiento y presentación de péptidos, siendo radicalmente diferente a animales que ocupan nichos terrestres, por lo que cabe la posibilidad de que las posiciones implicadas en la unión de péptidos sean distintas.

Finalmente, como no se han estudiado previamente series de animales silvestres emparentados, puede que esta sea la imagen de MHC de especies silvestres y no del hombre, ratón o pollos, especies sujetas a un fuerte incremento de individuos y por tanto a una consanguinidad que favoreciese la variación molecular de MHC por aumento de sobrecruzamiento de cromátidas homólogas.

Unión de péptidos

La capacidad que tienen las moléculas MHC de clase I de unir un amplio repertorio de péptidos radica en la variabilidad de las posiciones del ARS no implicadas en el anclaje del péptido por sus extremos (Janeway *et al.*, 2005) por lo que la diversidad de péptidos que

pueden ser unidos por una molécula de histocompatibilidad de clase I, es debida a la variación de los aminoácidos intermedios.

Esto implica que los 7 aminoácidos reconocidos como invariantes que sirven de anclaje sean siempre los mismos (o muy similares) independientemente del péptido al que se unan.

En las tablas R6 y R7, podemos observar cómo de los 7 residuos del sitio de reconocimiento antigénico (ARS) identificados en humanos por Bjorkman (Bjorkman *et al.*, 1987a) como lugares invariantes y caracterizados por la función de anclar el péptido a través de sus extremos, 5 son compartidos por nuestras aves silvestres de canto estudiadas, lo que nos hace pensar que estas moléculas poseen la capacidad de unir péptidos.

De las posiciones restantes una está dentro sitio de unión al péptido (143) donde se cambia treonina (T) por una asparagina (N) los dos de carácter polar neutro, a excepción de la especie *S. canaria* que continúa teniendo en este lugar una N. La otra posición se encuentra dentro de la región que conforma el receptor de células T del dominio alfa 2 (tabla R7). Este lugar presenta también un aminoácido distinto suponiendo un cambio de lisina (K) por histidina (H), los dos de carácter polar básico. En esta posición se ha encontrado una excepción en la especie *S. canaria* que posee en este lugar una arginina (R), también de tipo polar básico (anexo 4).

Estas posiciones parecen participar en la unión de péptidos así como en la interacción del receptor de células T (TCR) según la dirección de las cadenas laterales (Bjorkman *et al.*, 1987b).

Residuos conservados

La función de las moléculas de histocompatibilidad de clase I es la unión y presentación de un gran número de péptidos al receptor clonotípico de células T, esta función está estrechamente relacionada con la estructura tridimensional de las moléculas MHC.

Debido a esto, una gran variabilidad de la estructura de la proteína permite la unión de muchos péptidos diferentes.

Sin embargo, en la conformación de estas proteínas, no todos los residuos son variables, algunos residuos han de mantenerse invariables para mantener una estructura compatible con esa presentación antigénica.

Dentro de estos residuos, deben encontrarse una serie de aminoácidos que aparezcan conservados a lo largo de la evolución en todas las moléculas de MHC de clase I. Grossberger

y Parham (Grossberger y Parham, 1992) describieron estos residuos conservados entre especies de mamíferos, aves *Galliformes*, reptiles, anfibios y peces en los dominios alfa 1, 2 y 3.

En los dos primeros dominios, estas posiciones están colocadas simétricamente y contribuyen al contacto intra e interdominios, sin afectar a las zonas que interaccionan con el péptido.

A pesar de que en el tercer dominio se encuentra el mayor número de residuos conservados (un total de 11), son los dominios alfa 1 y 2 los que parecen estar implicados en mantener la estructura de valva de éstas moléculas. De hecho se ha demostrado que una molécula es capaz de mantener integridad estructural e incluso su capacidad de unión peptídica careciendo del tercer dominio.

En este trabajo se han comprobado 6 de las 7 posiciones identificadas por Grossberger como conservadas para los dominios alfa 1 y alfa 2.

De esas 6 posiciones, 2 están conservadas pero con aminoácidos distintos. La posición 10, que en el resto de vertebrados analizados presenta una treonina es sustituida en aves *Passeriformes* por una valina a excepción de las especies *S. canaria* y *C. crassirostris* y la posición 96, que en los vertebrados analizados es una glutamina, cambia en aves *Passeriformes* a leucina.

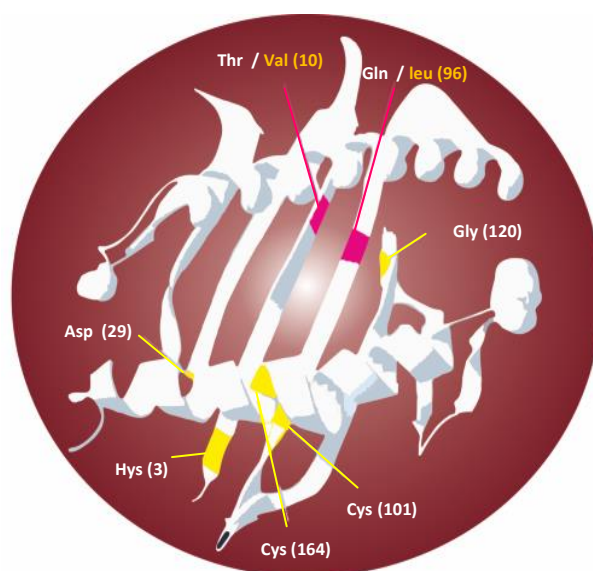


Figura D1. Modelo tridimensional de los dominios α -1 y α -2 de una molécula de histocompatibilidad de clase I en aves *Passeriformes*. En ella se pueden observar coloreados los 7 residuos conservados en vertebrados (Grossberger y Parham, 1992) cuyas posiciones se especifican entre paréntesis. En rosa están destacadas las dos posiciones conservadas (10 y 96) que muestran un diferente aminoácido (en amarillo) en los pájaros pequeños de canto estudiados en comparación con el resto de los vertebrados.

Los dos aminoácidos sustituidos dentro de estos residuos conservados pasan de ser de carácter polar neutro (vertebrados) a carácter apolar (anexo 4), pero poseen un tamaño muy similar, por lo que se cree que debe tener una estructura tridimensional parecida a las moléculas de clase I funcionales.

Estas dos posiciones (10 y 96), están implicadas en la interacción con la β 2-microglobulina. Concretamente la posición 10 interacciona con la 56 ó 62 d la β 2m y la posición 96 de la molécula MHC de clase I con las posiciones 31, 56, 60 y 82 de la β 2m.

El cambio de la leucina (96) se traduce en la pérdida de los puentes de hidrógeno que estabilizan la unión con la β 2m, por otro lado, el cambio de la valina (10) puede compensar esa pérdida ya que se establecen uniones Van der Waals con las posiciones nombradas anteriormente (56 y 62) de la β 2m (Tysoe-Calnon *et al.*, 1991).

Otro residuo que interacciona con el primer residuo de esta proteína y que está conservado en éstas aves además de en el resto de vertebrados analizados es la glicina localizada en la posición 120.

Según algunos estudios la β 2m del pollo y de mamíferos presenta un porcentaje de homología de un 44 % teniendo la misma capacidad de unión a la cadena α (Kaufman *et al.*, 1992; Riegert *et al.*, 1996; Welinder *et al.*, 1991), lo que nos lleva a pensar que esta interacción se produce de manera correcta y funcional en las aves silvestres de canto estudiadas. Es probable que la β 2m y el MHC de clase I hayan seguido un proceso de coevolución.

El resto de los residuos conservados son, las cisteínas localizadas en las posiciones 101 y 164, encargadas de formar el puente disulfuro con el dominio alfa 2; el aspártico y la glicina de las posiciones 29 y 120.

A su vez, estas secuencias de aves silvestres de canto presentan un único punto conservado de N-glucosilación que corresponde a las posiciones 86 (asparagina), 87 (glutamina) y 88 (serina) (Figura R2 y anexo 4) y una serie de residuos encargados de estabilizar la molécula mediante puentes salinos en el dominio alfa 2 localizados en las posiciones 166 (glutamato) y 169 (arginina) (Hashimoto *et al.*, 1999) (Figura R3 y anexo 4).

Por otra parte, de los 3 residuos conservados en humanos que interaccionan con la molécula CD8 (Gln115, Asp122, Glu128) (Gao *et al.*, 1997; Sun *et al.*, 1995), dos (Gln115 y Asp122) están conservados también en vertebrados inferiores (Hashimoto *et al.*, 1999) y sólo uno (Asp122) se mantiene conservado en todas las aves *Passeriformes* estudiadas (Figura R3).

La excepción de *S. canaria*

Como hemos podido observar en los resultados obtenidos en las secuencias MHC de aves *Passeriformes*, la especie *Serinus canaria* ha sido una de las excepciones constantes a lo

largo de este trabajo (Apartados 1.4-1.6 y 1.8-1.9) corroborando estudios anteriores de nuestro grupo (Arnaiz-Villena *et al.*, 1999a).

La variabilidad encontrada en esta molécula, en general respecto al resto de aves *Passeriformes* analizadas y en particular respecto a las especies de su mismo género nos hace pensar que puede ser debido a la gran diferencia de patógenos existentes en las Islas Canarias en comparación con los continentes.

Sin embargo, no se puede excluir que puedan estar actuando sobre la molécula de MHC de clase I de este canario, otro tipo de selección.

En la secuencia aminoacídica de esta especie (Figuras R2 y R3) se han observado cambios dentro del sitio de unión al péptido (tablas R6 y R7) pero también se han encontrado numerosos cambios fuera de estas regiones.

Esto nos lleva a pensar que no sólo la selección mediada por patógenos (dentro de la región ARS) actúa, si no que el ambiente insular influiría también en esta molécula.

Tampoco se puede descartar que no exista, por ejemplo, una selección dependiente de frecuencia (basada en la importancia evolutiva de los alelos raros) (Bodmer, 1972;Kojima, 1971) actuando sobre estas secuencias (Filardi y Moyle, 2005).

Esta divergencia en la molécula de MHC de clase I encontrada en canarios silvestres coincide con la observada en otras aves insulares (Filardi y Moyle, 2005) supuestamente influida por la presión selectiva de estos ambientes, los patógenos y la endogamia característica de poblaciones tan pequeñas (Sato *et al.*, 2001).

La teoría de “mínimo esencial”

Uno de los objetivos de este trabajo ha sido la comparación de los genes de histocompatibilidad en modelos silvestres de aves silvestres de canto, respecto a modelos “artificiales” (hombre, ratón y pollo) para comprobar si el diseño tan compacto del Sistema Principal de Histocompatibilidad del pollo se mantiene en otras aves. Entendiendo como “artificial” la gran consanguinidad que hoy día tienen el hombre el ratón y el pollo, que partiendo de un número pequeño de individuos han dado lugar a grupos relativamente grandes.

Como hemos dicho en la introducción, la peculiaridad de dicho diseño se debe al escaso número de genes de histocompatibilidad, a la gran proximidad existente entre ellos en el

cromosoma y a la corta longitud de los genes debida al pequeño tamaño de sus intrones (Guillemot *et al.*, 1988;Kaufman *et al.*, 1995;Kaufman *et al.*, 1999b).

Hasta ahora no se había obtenido el mapa genético completo del Sistema Principal de Histocompatibilidad en aves de canto excepto el de *Taeniopygia guttata* (Balakrishnan *et al.*, 2010), donde se ha observado que el pinzón cebra, presenta una estructura de MHC mucho más compleja, incluyendo múltiples genes de clase I y clase II, algunos de los cuales parecen ser pseudogenes, extendiéndose por una región genómica mucho mayor que la del pollo.

En nuestro caso, como se muestra en los resultados, apartado 1.15, ha sido secuenciado el intrón 2, en distintos géneros de aves *Passeriformes* (*Carduelis*, *Fringilla*, *Passer*, y *Serinus*) midiéndose la longitud del mismo. En la tabla R15 se observa cómo la longitud media de este intrón en aves silvestres de canto es de 303pb, muy parecida a la del resto de mamíferos comparados (a excepción en *Homo sapiens* cuya longitud es de 241 pb) lo que sugiere que, al menos en este aspecto, no se puede hablar de un “*minimal essential MHC*”. En estudios realizados en otras aves, tanto de canto como de otros órdenes, tampoco se han encontrado evidencias que permitan afirmar que los MHCs analizados sean tan compactos como el del pollo (Balakrishnan *et al.*, 2010;Bonneaud *et al.*, 2004;Karlsson y Westerdahl, 2013;Moon *et al.*, 2005;Shiina *et al.*, 2004;Westerdahl *et al.*, 1999;Xia *et al.*, 2005).

La incongruencia que dentro de la familia *Galliforme* haya intrones cortos en el pollo y no en la codorniz (*Coturnix japonica*) se explicaría por el hecho de que la consanguinidad en los grupos de pollos es mucho más alta que en los grupos de codornices, por motivos de la industria alimentaria.

Tamaño y evolución de los intrones

Como sabemos, un intrón es una fracción de ADN que no codifica proteínas y se encuentra inserta en el interior de un gen codificante. El intrón debe ser eliminado del ARN transcrito para que éste pueda ser traducido en una proteína.

Desde el primer momento se pensó que eran regiones no codificantes de ahí el nombre de “exones” (expressed region) e “intrones” (intrinsic region) (Gilbert, 1978;Gilbert, 1987).

De esta forma, un gen consistiría en una serie de exones entre los que se intercalan uno o varios intrones no codificantes y las proteínas se formarían a partir del ensamblaje de los exones (Blake, 1978).

Hoy sabemos que cerca del 30% del ADN de los eucariotas está formado por intrones, mientras que los procariotas carecen de ellos. En la mayoría de los mamíferos, hay un promedio de más de ocho intrones por gen (Roy y Gilbert 2006; Farlow *et al.*, 2011).

Sin embargo, ahora está bien aceptado que los intrones no son simplemente ADN "basura", ya que son la base de splicing alternativo, que puede generar múltiples proteínas a partir de un único gen; algunos intrones también codifican moléculas de ARN no codificante que regulan la transcripción.

Debido a sus funciones y conservación recién descubiertas en el genoma, ahora se cree que muchos intrones pueden evolucionar en virtud de las restricciones selectivas; de hecho, varios estudios han encontrado evidencia de evolución adaptativa en la forma de segregación de los intrones (Parsch *et al.*, 2010; Hayden *et al.*, 2011; Cagliani *et al.*, 2012), lo que sugiere que tanto el tamaño como la secuencia pueden estar seleccionados por fuerzas no neutrales, como se creía en un principio.

Algunos estudios han encontrado que dentro de las especies el tamaño de los intrones varía sustancialmente entre los diferentes genes (los genes en tejidos o genes implicados en el desarrollo poseen intrones más largos y conservados durante más tiempo en la evolución que el resto) (Castillo-Davis *et al.*, 2002; Eisenberg y Levanon, 2003; Urrutia y Hurst, 2003; Vinogradov, 2004).

Incluso dentro de un solo gen, los intrones son diferentes, los primeros intrones siempre son más largos que los siguientes (Marais *et al.*, 2005; Gaffney y Keightley, 2006; Gazave *et al.*, 2007; Bradnam y Korf, 2008), lo que puede reflejar diferentes propiedades funcionales.

Por otra parte, el tamaño del intrón también varía entre las especies, como se ha comentado varias veces en este trabajo, los intrones de la especie *Gallus gallus* son mucho más reducidos en comparación con los de otros mamíferos excepto el hombre, según algunos estudios esto puede ser debido a la presión selectiva impuesta por comportamientos metabólicamente exigentes tales como el vuelo (Hughes y Hughes, 1995b), donde los intrones pequeños podrían proporcionar una ligera mejora en la eficacia de la transcripción o en la precisión del "splicing" (Lynch, 2002).

Sin embargo, en un estudio posterior, Vinogradov y colaboradores (1999) examinaron 176 intrones de 55 genes homólogos de pollo-humano y no pudieron encontrar diferencias significativas entre los tamaños intrónicos de éstas dos especies.

En trabajos posteriores Waltari y Edwards (2002) estudiando 14 intrones de 19 aves voladoras y no voladoras diferentes, llegaron a la conclusión de que el tamaño de los intrones aviares se asemeja a un movimiento Browniano (movimiento al azar de una molécula de fluido afectada por ruido térmico). Estos resultados sugieren que los cambios en el tamaño de los intrones no se pueden asociar significativamente con los comportamientos metabólicamente costosos tales como el vuelo en las aves.

Sin embargo, el número de intrones de este estudio era bastante pequeño. Por lo tanto, hoy por hoy, no hay ninguna conclusión firme con respecto a si los intrones son más pequeños en las especies de aves que en mamíferos y si el vuelo puede imponer presiones selectivas sobre el tamaño de los mismos.

Se ha propuesto que en organismos amniotas de sangre caliente, el tamaño de su genoma puede estar bajo limitaciones fisiológicas (Waltari y Edwards, 2002), donde se ven favorecidas las células más pequeñas cuya relación superficie-volumen es mayor, lo que conlleva a una mayor capacidad de intercambio de gases para el mantenimiento de una alta tasa metabólica (Szarski, 1983; Hughes y Hughes, 1995b; Organ *et al.*, 2007).

En apoyo de estas teorías, varios estudios han encontrado genomas más pequeños en aves y murciélagos en comparación con otros mamíferos euterios (Hughes y Hughes, 1995b; Van den Bussche *et al.*, 1995). También en colibríes, caracterizados por su famoso vuelo estacionario, de alto consumo energético, se ha encontrado que tienen el genoma más pequeño de todas las aves estudiadas hasta ahora (Gregory *et al.*, 2009).

Sin embargo, los estudios de Organ y colaboradores (2007) sobre el origen del tamaño del genoma aviar mediante la reconstrucción de los genomas ancestrales, sugirieron que la reducción del tamaño del genoma se produjo mucho antes del origen de las aves (Organ *et al.*, 2007).

De acuerdo con este patrón, los análisis de Qu Zhang y Scott V. Edwards (2012) mostraron que dentro de los reptiles y los mamíferos analizados, aquellos que poseen la capacidad de volar tienen unos intrones más pequeños, lo que sugiere una posible correlación entre el tamaño del intrón / tamaño del genoma y la capacidad de vuelo.

También existen teorías como las de Lynch y Conery (2003) que sugieren que los cambios en la complejidad del genoma responden a cambios en el tamaño de la población.

En base a su hipótesis, la contracción de los genomas e intrones que han observado en las aves y murciélagos es el resultado de un mayor tamaño de su población en relación con sus

parientes cercanos no voladores, por lo que, según esta teoría, un genoma más pequeño puede ser más eficiente en poblaciones grandes que en poblaciones pequeñas.

En nuestro trabajo, hemos secuenciado el intrón 2 del MHC de clase I en aves silvestres de canto (*Carduelis*, *Serinus*). Estas secuencias han sido comparadas con especies de los géneros (*Fringilla*, *Passer*) y a su vez con el resto de vertebrados objeto de estudio. En ellas hemos observado cómo en todas las secuencias de aves *Passeriformes* se conservan más allá de las regiones de reconocimiento que intervienen en el “splicing”.

Muchos estudios corroboran la idea de que los intrones largos contienen mayor cantidad de secuencia conservada que los intrones cortos y que la presencia de estas secuencias conservadas explica la relación entre tamaño de los intrones y variación en las proteínas. Existen evidencias de que las secuencias no codificadoras que están conservadas funcionan como reguladoras de la expresión de un gen (Petit, N *et al.*, 2007).

Esto puede sugerir que la presencia de estas secuencias conservadas en los genes MHC de clase I en aves silvestres de canto, está estrechamente relacionada con la complejidad en el patrón de expresión de la proteína.

La conservación del intrón 2 tan marcada durante largos periodos de tiempo encontrada por nosotros, podría sugerir que los intrones son influenciados por factores desconocidos externos, o sea que podrían ser evolutivamente seleccionados.

Respecto al tamaño de nuestro intrón hemos observado cómo en la especie *Gallus gallus* el tamaño es significativamente más pequeño respecto a las aves *Passeriformes* estudiadas, siguiendo la teoría de Kaufman (Kaufman *et al.*, 1995) de MHC “mínimo esencial”, caracterizado por no encontrarse todos los genes contenidos en la región MHC, por encontrarse solo el gen C4 del complemento representando la región clase III, por poseer intrones más cortos y una proximidad en sus genes mucho mayor.

A su vez hemos observado cómo el tamaño del intrón 2 de la molécula MHC de clase I en la especie *Gallus gallus*, es muy parecido al que presenta el intrón de *Homo sapiens* no pudiendo encontrar diferencias significativas entre los mismos apoyando los datos de Vinogradov y colaboradores (1999) citados anteriormente.

En relación a la reducción del tamaño de los intrones para favorecer procesos metabólicos de alto contenido energético, tales como el vuelo, nosotros no hemos encontrado relación alguna con el intrón 2 de la molécula de MHC de clase I. De hecho las dos especies que presentan el intrón más pequeño son el gallo y el humano, las dos especies terrestres. Por

otro lado la otra especie de ave galliforme (*Coturnix japonica*) posee una longitud muy similar al de las aves *Passeriformes* y al del resto de mamíferos y primates comparados en la tabla R15.

No podemos afirmar que nuestras aves silvestres de canto no sigan las teorías anteriormente citadas ya que para ello habría que analizar muchos intrones no sólo uno, pero si podemos decir que, al menos, el intrón 2 de la molécula de MHC de clase I no parece seguir la tendencia de la reducción del tamaño de las secuencias no codificadoras para el mantenimiento de una alta tasa metabólica.

2.Filogenia y evolución de aves Passeriformes

La filogenia de los géneros estudiados en este trabajo, basada en las secuencias de ADN del gen mitocondrial citocromo b de las especies que lo componen, nos ofrece su historia evolutiva de una manera más lineal, facilitando la interpretación de los resultados obtenidos con las secuencias de histocompatibilidad de clase I de algunas de ellas. El ADN del citocromo b mitocondrial es más adecuado para estudiar filogenias.

Estas secuencias de citocromo b comprenden los géneros: *Carduelis*, *Serinus*, *Loxia*, *Carpodacus*, *Uragus*, *Haematospiza*, *Pyrrhula*, *Pinicola*, *Leucosticte*, *Rhodopechys* y *Fringilla*. En el árbol ML linearizado (Figura R10) y en el análisis bayesiano (Figura R11) se muestran los resultados de estos análisis y además se indican los tiempos de divergencia aproximados de los nodos más relevantes.

La mayor parte de las secuencias utilizadas forman parte del archivo del grupo (tabla M2), exceptuando una nueva, secuenciada por primera vez en este trabajo, la especie *Rhodopechys sanguineus* (Arnaiz-Villena *et al.*, 2014a).

La mayoría de las especies de los géneros *Carduelis* y *Serinus* representadas en las figuras R10 y R11 se muestran entremezcladas en el tiempo y parecen no pertenecer a un grupo monofilético (no poseen un ancestro común).

Las especies más tempranas del género *Serinus* aparecieron probablemente en la época del Mioceno hace aproximadamente 9 millones de años ligeramente después de la primera especie del género *Carduelis* (aproximadamente 9,5 millones de años) posiblemente en el continente asiático, coincidiendo con los cambios dramáticos de clima cuando el Mar Mediterráneo oriental (Mar Tethys) se estaba cerrando (Arnaiz-Villena *et al.*, 1999a;Cox y Moore, 1995;Cox y Moore, 2000;Paturi, 1991;Smith y Funnel, 2009).

Por otro lado, dentro de los árboles filogenéticos obtenidos, podemos observar cómo algunas especies de la tribu *Carduelini* se han quedado fuera de 'sus propios' géneros.

Dentro de estos géneros podemos observar como el Camachuelo del desierto (Zamora *et al.*, 2006a) aparece estrechamente relacionado con el linaje de los verderones (*Carduelis ambigua*, *Carduelis spinoides*, *Carduelis sinica* y *Carduelis Chloris*) surgido a finales del Mioceno (hace aproximadamente 6,5 millones de años) (Figura R10 y R11) (Zamora *et al.*, 2006a). Es muy posible que el Camachuelo del desierto, durante periodos de sequía extrema hace 2.6 m.a (Uriarte-Cantolla, 2003) hubiese salido de las zonas semidesérticas a las que estaba acostumbrado hacia hábitats más húmedos evolucionando como un ancestro de los verderones (Zamora *et al.*, 2006a). Sus características fenotípicas, como un plumaje pálido ya mostraban una separación frente al resto de especies de su mismo género cuyo colorido tiende a ser rosado o rojizo (Arnaiz-Villena *et al.*, 1998; Arnaiz-Villena *et al.*, 2008).

Otra de las especies que estaría fuera de nodo del resto de especies de su mismo género sería *Carpodacus nipalensis*, alejado del resto de *Carpodacus* americanos y asiáticos (*C. roseus*, *C. trifasciatus*, *C. thura*, *C. rubicilloides* y *C. erythrinus*). Fenotípicamente ya se observa que este Camachuelo oscuro difiere en su pico, cabeza y características corporales respecto al resto de especies de su mismo género (Arnaiz-Villena *et al.*, 2001; Zamora *et al.*, 2006a). Dentro de este grupo, también podemos ver como la especie *C. erythrinus* posee una estrecha relación con *Haematospiza sipahi*, y cómo *Uragus sibiricus* se presenta como especie hermana de *C. rubicilloides* (Gill, 1999) formando ambos dos clados diferentes.

A su vez, si observamos el grupo de aves de género *Pyrrhula*, vemos que forman un único clado con el Camachuelo picogruoso (*Pinicola enucleator*), el que parece ser la especie más antigua existente dentro de este grupo monofilético.

Por otro lado, en los dos árboles filogenéticos obtenidos (Figura R10 y R11) podemos ver que las especies del género de los piquituertos (*Loxia*) están fuertemente emparentadas con otras especies de *Carduelis* que comparten hábitats fríos del hemisferio norte. Esto nos hace pensar que los piquituertos deberían ser clasificados junto con los pardillos sizerines (*Carduelis flammea*) y los pardillos de Hornemann (*Carduelis hornemanni*) con los que forman un grupo monofilético (Arnaiz-Villena *et al.*, 1998; Arnaiz-Villena *et al.*, 2001).

Otro de los grupos que observamos en nuestros resultados, sería el denominado “pinzones de zonas áridas” que comprenden los géneros (*Leucosticte*, *Carpodacus* y *Rhodopechys*) (Arnaiz-Villena *et al.*, 2008). Según el árbol ML (Figura R10) lo antepasados de este grupo

podrían haber aparecido hace 13,5 ma, junto con género *Pyrrhula* que habrían ocupado un hábitat más húmedo y lleno de vegetación que los “pinzones de zonas áridas” (Clement *et al.*, 1993).

Una de las singularidades de estos resultados es que por primera vez ha sido secuenciada la especie *Rhodopechys sanguineus*. Se ha comprobado, tras los análisis filogenéticos, que queda incluida dentro del grupo de “pinzones de zonas áridas” caracterizados fenotípicamente todos por un grado de plumaje rojizo-rosado (Arnaiz-Villena *et al.*, 2014a). Todas estas aves prosperan en climas extremos, en hábitats como desiertos o montañas con escasa vegetación.

Rhodopechys sanguineus parece ser antepasado del Camachuelo trompetero (*Rhodopechys githaginea*) y del Camachuelo mongol (*Rhodopechys mongólica*). De acuerdo con nuestro dendrograma linealizado (Figura R10) apareció en la Tierra hace aproximadamente 10 ma lo que coincide más o menos con la aparición de zonas áridas en África y Asia.

Esta radiación de “pinzones de zonas áridas” podría haber comenzado alrededor de este tiempo en Asia (o menos probable en África), desde donde algunos de los linajes posteriormente podrían haber sido sometidos a la dispersión durante las condiciones más cálidas.

Las especies *Leucosticte arctoa* y *Carpodacus nipalensis* poseen más melanina en su plumaje favorecido como protección bacteriana para ambientes más húmedos en comparación con los ambientes más áridos donde habitan los *Rhodopechys*. Además este *Carpodacus* posee un alto contenido de carotenoides en su plumaje lo que le da ese característico color intenso tan diferente del resto de especies que habitan en zonas áridas, posiblemente producto de una dieta rica en bayas de las que se alimenta a grandes alturas (Clement *et al.*, 1993).

Todo esto nos hace creer que de acuerdo con nuestro análisis filogenéticos basado en secuencias de ADN mitocondrial citocromo b, los géneros *Rhodopechys* y *Carpodacus* deberían ser revisados, así como una gran parte de la taxonomía de la familia *Carduelini*.

CONCLUSIONES

En este trabajo se han obtenido datos moleculares procedentes de aves de canto silvestres del Orden *Passeriformes*, concretamente han sido secuenciadas 26 especies de muestras recogidas a lo largo de la geografía mundial. Los resultados obtenidos de la secuenciación del exón 2, intrón 2 y exón 3 del gen codificante para la cadena pesada de las moléculas MHC de clase I de estas muestras, han sido comparados con 20 secuencias de distintos vertebrados, obtenidas de la base de datos GenBank.

También se ha realizado un estudio de las relaciones evolutivas existentes entre las especies estudiadas de aves silvestres *Passeriformes* con el fin de interpretar el proceso evolutivo al que han estado sometidas las secuencias de histocompatibilidad obtenidas en este trabajo. Para ello se recurrió a la secuenciación y análisis del gen mitocondrial, citocromo b para el estudio filogenético.

De los resultados obtenidos se puede concluir que:

1. Las moléculas de histocompatibilidad de clase I de las aves *Passeriformes* estudiadas presentan unos cambios aminoacídicos específicos de estas aves que no existen en otras aves (*Gallus gallus*, *Coturnix japonica*), ni en otros vertebrados estudiados. Estos cambios se encuentran en los residuos de las posiciones 10 y 96.
2. La estabilidad de estas moléculas con dos residuos diferentes y su capacidad de unión a la β 2-microglobulina es muy similar a la del resto de vertebrados aquí estudiados.
3. Esto sugiere que las pequeñas aves de canto (*Passeriformes*) cuyo nicho principal es aéreo, han seguido un camino evolutivo diferente al de otras aves más terrestres (*Galliformes*).
4. El tamaño del intrón 2 de las moléculas de histocompatibilidad de clase I, es muy similar en vertebrados y en los *Passeriformes* estudiados por nosotros (media de 304 pb) sin embargo son significativamente más largos que en el pollo (203 pb) pero no que en la codorniz (305 pb) (perteneciendo las dos al orden *Galliformes*) y que en humanos (241 pb).

5. La secuencia nucleotídica del intrón 2 en aves *Passeriformes* se conserva en la familia *Carduelinae* durante al menos 9 millones de años, sugiriendo la presencia de una selección evolutiva positiva sobre este intrón.

6. Por otra parte, se presenta una nueva filogenia y sistemática molecular de numerosas especies de aves *Passeriformes* de los 5 continentes, que varían sustancialmente de la filogenia fenotípica existente y se pueden correlacionar con las variaciones ambientales conocidas en la Tierra desde la época del Mioceno, hace 23 millones de años.

BIBLIOGRAFÍA

Referencias

- ✓ Amiel J L, 1967. Study of the leukocyte phenotypes in Hodgkin's disease. *Histocompatibility Testing* .Munksgaard, Copenhagen, 79-81.
- ✓ Alberts SC, Ober C, 1993. Genetic variability in the major histocompatibility complex: a review of non-pathogen-mediated selective mechanisms. *Physical anthropology* 36, 71-89.
- ✓ Allende LM, Rubio I, Ruiz-del-Valle V, Guillen J, Martinez-Laso J, Lowy E, Varela P, Zamora J, Arnaiz-Villena A, 2001. The Old World sparrows (genus *Passer*) phylogeography and their relative abundance of nuclear mtDNA pseudogenes. *J. Mol. Evol.* 53, 144-154.
- ✓ Anderson S, Bankier AT, Barrell BG, Debruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG, 1981. Sequence and Organization of the Human Mitochondrial Genome. *Nature* 290, 457-465.
- ✓ 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, 1031-1041.
- ✓ 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, 1999a. Rapid radiation of canaries (Genus *Serinus*). *Mol. Biol. Evol.* 16, 2-11.
- ✓ Arnaiz-Villena A, Iliakis P, Gonzalez-Hevilla M, Longas J, Gomez-Casado E, Sfyridaki K, Trapaga J, Silvera-Redondo C, Matsouka C, Martinez-Laso J, 1999b. The origin of Cretan populations as determined by characterization of HLA alleles. *Tissue Antigens* 53, 213-226
- ✓ Arnaiz-Villena A, Vargas-Alarcon G, Granados J, Gomez-Casado E, Longas J, Gonzales-Hevilla M, Zuniga J, Salgado N, Hernandez-Pacheco G, Guillen J, Martinez-Laso J, 2000. HLA genes in Mexican Mazatecos, the peopling of the Americas and the uniqueness of Amerindians. *Tissue Antigens* 56, 405-416.
- ✓ Arnaiz-Villena A, Guillen J, Ruiz-del-Valle V, Lowy E, Zamora J, Varela P, Stefani D, Allende LM, 2001. Phylogeography of crossbills, bullfinches, grosbeaks, and rosefinches. *Cell. Mol. Life. Sci.* 58, 1159-1166.
- ✓ Arnaiz-Villena A, Martinez-Laso J, Moscoso J, Livshits G, Zamora J, Gomez-Casado E, Silvera-Redondo C, Melvin K, Crawford MH, 2003. HLA genes in the Chuvashian population from European Russia: admixture of Central European and Mediterranean populations. *Hum. Biol.* 75, 375-392.
- ✓ Arnaiz-Villena A, Siles N, Moscoso J, Zamora J, Serrano-Vela JI, Gomez-Casado E, Castro MJ, Martinez-Laso J, 2005. Origin of Aymaras from Bolivia and their relationship with other Amerindians according to HLA genes. *Tissue Antigens* 65, 379-390.
- ✓ Arnaiz-Villena A, Lowy E, Ruiz-del-Valle V, Westerdahl H, Moscoso J, Serrano-Vela JI, Witzell H, Zamora J, 2007a. Evolution of the major histocompatibility complex class I genes in *Serinus canaria* from the Canary Islands is different from that of Asian and African continental *Serinus* species. *Journal of Ornithology* 148, S479-S484.
- ✓ Arnaiz-Villena A, Ruiz-del-Valle V, Moscoso J, Serrano-Vela JI, Zamora J, 2007b. mtDNA phylogeny of North American *Carduelis pinus* group. *Ardeola* 54, 1-14.
- ✓ Arnaiz-Villena A, Moscoso J, Ruiz-del-Valle V, Gonzalez J, Reguera R, Wink M, Serrano-Vela JI, 2007c. Bayesian phylogeny of *Fringillinae* birds: status of the singular African Oriole Finch (*Linurgus olivaceus*) and evolution and heterogeneity of genus *Carpodacus*. *Acta Zoologica Sinica* 53, 826-834.

- ✓ Arnaiz-Villena A, Vargas-Alarcon G, Serrano-Vela JI, Reguera R, Martinez-Laso J, Silvera-Redondo C, Granados J, Moscoso J, 2007d. HLA-E polymorphism in Amerindians from Mexico (Mazatecans), Colombia (Wayu) and Chile (Mapuches): evolution of MHC gene. *Tissue Antigens* 69, 132-135.
- ✓ Arnaiz-Villena A, Moscoso J, Granados J, Serrano-Vela JI, de la Peña A, Reguera R, Ferri A, Seclen E, Izaguirre R, Perez-Hernandez N, Vargas-Alarcon G, 2007e. HLA genes in Mayos population from Northeast Mexico. *Mol. Immunol.* 8, 466-475.
- ✓ Arnaiz-Villena A, Moscoso J, Ruiz-del-Valle V, Gonzalez J, Reguera R, Ferri A, Wink M, Serrano-Vela JI, 2008. Mitochondrial DNA Phylogenetic Definition of a Group of 'Arid-Zone' *Carduelini* Finches. *Open Ornithology Journal* 1, 1-7.
- ✓ Arnaiz-Villena A, Ruiz-del-Valle V, Gomez-Prieto P, 2009a Phylogeography of finches and sparrows. Rechi, L. (Ed.), Nova Science, New York, 1-54.
- ✓ Arnaiz-Villena A, Ruiz-del-Valle V, Gomez-Prieto P, Reguera R, Parga-Lozano C, Serrano-Vela JI, 2009b. Estrildinae Finches (*Aves, Passeriformes*) from Africa, South Asia and Australia: a Molecular Phylogeographic Study. *Open Ornithology Journal* 2, 29-36.
- ✓ Arnaiz-Villena A, Reguera R, Parga-Lozano C, Abd-El-Fatah S, Monleon L, Barbolla L, Gomez-Prieto P, Martinez-Laso J, Silvera-Redondo C, 2009c. HLA Genes in Afro-American Colombians (San Basilio de Palenque): The First Free Africans in America. *The Open Immunology Journal* 2, 59-66.
- ✓ Arnaiz-Villena A, Gonzalez-Alcos V, Moscoso J, Reguera R, Barbolla L, Ferri A, Fernandez PC, Abd-El-Fatah S, Serrano-Vela JI, 2009d. HLA genes in Uros from Peru: Origin and relationships with other Amerindians and worldwide populations. *International Journal of Immunogenetics* 36, 159-167.
- ✓ Arnaiz-Villena A, Parga-Lozano C, Moreno E, Areces C, Rey D, Gomez-Prieto P, 2010a. The Origin of Amerindians and the Peopling of the Americas According to HLA Genes: Admixture with Asian and Pacific People. *Curr. Genomics* 11, 103-114.
- ✓ Arnaiz-Villena A, Ruiz-del-Valle V, Reche P, Gomez-Prieto P, Lowry E, Zamora J, Areces C, Rey D, Parga C, Serrano-Vela JI, 2010b. Songbirds conserved sites and introns size of MHC class I molecules reveal a unique evolution in vertebrates. *The Open Ornithology Journal* 3, 156-165.
- ✓ Arnaiz-Villena A, Abd-El-Fatah S, Granados-Silvestre MA, Parga-Lozano C, Gomez-Prieto P, Rey D, Areces C, Penaranda P, Menjivar M, Rodriguez-Perez JM, Granados J, Vargas-Alarcon G, 2011. Human leukocyte antigen-DRB1 class II genes in Mexican Amerindian Mazahuas: genes and languages do not correlate. *Genet Test. Mol. Biomarkers* 15, 97-102.
- ✓ Arnaiz-Villena A, Areces C, Rey D, Enriquez-de-Salamanca M, Alonso-Rubio J, Ruiz-del-Valle, 2012. Three different North American Siskin/Goldfinch evolutionary radiations (Genus *Carduelis*): Pine Siskin green morphs and European Siskin in America. *The Open Ornithology Journal* 5, 73-81.
- ✓ Arnaiz-Villena A, Enriquez-de-Salamanca M, Areces C, Alonso-Rubio J, Abd-El-Fatah-Khalil S, Fernandez-Honrado M, Rey D, 2013. HLA-G(*):01:05N null allele in Mayans (Guatemala) and Uros (Titikaka Lake, Peru): evolution and population genetics. *Hum. Immunol.* 74, 478-482.
- ✓ Arnaiz-Villena A, Ruiz-del-Valle V, Gomez-Prieto P, Rey D, Enriquez-de-Salamanca M, Marco J, Muñiz E, Martín-Villa M, Areces C, 2014a. *Carduelini* New Systematics: Crimson-winged Finch (*Rhodopechys sanguineus*) is Included in "Arid-Zone" *Carduelini* Finches by Mitochondrial DNA Phylogeny. *The Open Ornithology Journal* 7, 55-62.
- ✓ Arnaiz-Villena A, Vargas-Alarcon G, Areces C, Enriquez-de-Salamanca M, Abd-El-Fatah-Khalil S, Fernandez-Honrado M, Marco J, Martín-Villa JM, Rey D, 2014b. Mixtec Mexican Amerindians: an HLA Alleles Study for America Peopling, Pharmacogenomics and Transplantation. *Immunol. Invest.* 43, 738-755

-
- ✓ Balakrishnan CN, Ekblom R, Volker M, Westerdahl H, Godinez R, Kotkiewicz H, Burt DW, Graves T, Griffin DK, Warren WC, Edwards SV, 2010. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biol* 8, 29.
 - ✓ Barber J, 1984. Has the Manganese-Protein of the Water Splitting Reaction of Photosynthesis Been Isolated. *Trends in Biochemical Sciences* 9, 79-80.
 - ✓ Becquemont L, 2010. HLA: a pharmacogenomics success story. *Pharmacogenomics* 11, 277-281.
 - ✓ Bembridge GP, MacHugh ND, McKeever D, Awino E, Sopp P, Collins RA, Gelder KI, Howard CJ, 1995. CD45RO expression on bovine T cells: relation to biological function. *Immunology* 86, 537-544.
 - ✓ Bernardi G, Bernardi G, 1986. Compositional constraints and genome evolution. *J Mol Evol* 24, 1-11.
 - ✓ Bernardi G, Hughes S, Mouchiroud D, 1997. The major compositional transitions in the vertebrate genome. *J Mol Evol.* 44, 44-51.
 - ✓ Bjorkman PJ, Saper M.A, Samraoui B, Bennett WS, 1987a. Structure of the human class I histocompatibility antigen, HLA-A 2. *Nature* 329, 506-512.
 - ✓ Bjorkman PJ, Saper M.A, Samraoui B, Bennett W, Strominger J, Wiley DC, 1987b. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* 329, 512-518.
 - ✓ Black FL, Salzano FM, 1981. Evidence for heterosis in the HLA system. *Am J Hum Genet* 33, 894-899.
 - ✓ Blake CCF, 1978. Do genes-in-pieces imply proteins-in-pieces?. *Nature* 273-267.
 - ✓ Bloom SE, Briles WE, Briles RW, Delany ME, Dietert RR, 1987. Chromosomal localization of the major histocompatibility (B) complex (MHC) and its expression in chickens aneuploid for the major histocompatibility complex/ribosomal deoxyribonucleic acid microchromosome. *Poult. Sci.* 66, 782-789.
 - ✓ Bodmer W, 1972. Evolutionary significance of the HLA system. *Nature* 237, 139-145.
 - ✓ Boles WE, 1995. The Worlds Oldest Songbird. *Nature* 374, 21-22.
 - ✓ 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, 855-865.
 - ✓ Bonneaud C, Chastel O, Federici P, Westerdahl H, Sorci G, 2006. Complex Mhc-based mate choice in a wild passerine. *Proc Biol Sci.* 273, 1111-1116.
 - ✓ Brackenbury J, 1982. The structural basis of voice production and its relationship to sound characteristics. Kroodsma DE, Miller EH (eds) *Acoustic Communication in Birds*. Academic Press, New York Edition.
 - ✓ Bradnam KR, Korf I, 2008. Longer first introns are a general property of eukaryotic gene structure. *PLoS One* 3,e3093.
 - ✓ Briles W, McGibbon W, 1948. Heterozygosity of inbred lines of chickens of two loci effecting cellular antigens. *Genetics* 33, 605.
 - ✓ Briles W, McGibbon W, Irwin DM, 1950. On multiple alleles affecting cellular antigens in the chicken. *Genetics* 35, 633-640.
 - ✓ Buckley P, 1987 Mendelian genes, Cooke F, Buckley P (eds) *Avian Genetics*. Academic Press, New York Edition.

- ✓ Caceres P, 2015. El árbol evolutivo de las aves. *Aves y naturaleza* 17,8-10.
- ✓ Cagliani R, et al. 2012. Variants in SNAP25 are targets of natural selection and influence verbal performances in women. *Cell Mol Life Sci.* 69, 1705–1715.
- ✓ Campbell RD, Trowsdale J, 1993. Map of the human MHC. *Immunol.* 14, 349-352.
- ✓ Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ, 1999. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* 283, 1748-1752.
- ✓ Case JT, Wallace DC, 1981. Maternal Inheritance of Mitochondrial-Dna Polymorphisms in Cultured Human-Fibroblasts. *Somatic Cell Genetics* 7, 103-108.
- ✓ Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA, 2002. Selection for short introns in highly expressed genes. *Nat Genet.* 31, 415–418.
- ✓ Charron D, 2012. [From HLA to cell therapy]. *Rev. Med Interne* 33, 62-63.
- ✓ Clement P, Harris P, Davies J, 1993. *Finches and Sparrows.* croom Helm, London.
- ✓ Cox CB, Moore PD, 1995 *Biogeography*, Blackwell. Humanity Press, Cambridge, England. Edition.
- ✓ Cox B, Moore P, 2000 *Biogeography: an ecological and evolutionary approach*, 6, illustrated Edition. Wiley-Blackwell.
- ✓ Crofts A, Robinson H, Andrews K, Van Doren S, Berry E, 1987. Catalytic sites for reduction and oxidation of quinones. Papa S, Chance B, Ernster L (eds) *Cytochrome Systems: Molecular Biology and Bioenergetics.* Plenum Press, New York Edition.
- ✓ Crone M, Simonsen M, 1987 *Avian major histocompatibility complex: a birds eye view.*, *Avian Immunology: basis and practice.* Edition.
- ✓ Dausset J, Nenna A, 1952. [Presence of leuko-agglutinin in the serum of a case of chronic agranulocytosis]. *C. R. Seances Soc Biol Fil.* 146, 1539-1541.
- ✓ Dausset J, 1958. [Iso-leuko-antibodies]. *Acta Haematol.* 20, 156-166.
- ✓ Distant S, Robson KJ, Graham-Campbell J, Arnaiz-Villena A, Brissot P, Worwood M, 2004. The origin and spread of the HFE-C282Y haemochromatosis mutation. *Hum Genet* 115, 269-279.
- ✓ Doherty PC, Zinkernagel RM, 1975. Enhanced immunologic surveillance in mice heterozygous at the H2 complex. *Nature* 256, 50-52.
- ✓ Donadi EA, Castelli EC, Arnaiz-Villena A, Roger M, Rey D, Moreau P, 2011. Implications of the polymorphism of HLA-G on its function, regulation, evolution and disease association. *Cell. Mol. Life Sci.*;68, 369-95.
- ✓ Edwards SV, Arctander P, Wilson AC, 1991. Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. *Proc. R. Soc. Lond. B-Biol. Sci.* 243, 99-107.
- ✓ Edwards SV, Gasper J, Garrigan D, Martindale D, Koop BF, 2000. A 39-kb sequence around a blackbird Mhc class II gene: ghost of selection past and songbird genome architecture. *Mol. Biol Evol.* 17, 1384-1395.
- ✓ Egid K, Brown JL, 1989. The major histocompatibility complex and female mating preference in mice. *Anim Behav* 38, 548-560.

- ✓ Eisenberg E, Levanon EY, 2003. Human housekeeping genes are compact. *Trends Genet.* 19, 362–365.
- ✓ Encyclopaedia Britannica. 1999. Encyclopaedia Britannica International Ltd, Sutton, Surney.
- ✓ Farlow A, Meduri E, Schlotterer 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, 368-376.
- ✓ Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, Vyse TJ, Rioux JD, 2008. Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet* 4, e1000024.
- ✓ Filardi CE, Moyle RG, 2005. Single origin of a pan-Pacific bird group and upstream colonization of Australasia. *Nature* 438, 216-219.
- ✓ Fitch WM, 1971. Toward defining the course of evolution, minimum change from a specific tree topology. *Syst. Zool.* 20, 406-415.
- ✓ Flajnik MF, Ohta Y, Namikawa-Yamada C, Nonaka M, 1999. Insight into the primordial MHC from studies in ectothermic vertebrates. *Immunol. Rev.* 167, 59-67.
- ✓ Fleischer RC, McIntosh C, Tarr CL, 1998. Evolution on a volcanic conveyor belt: using phylogeographic reconstructions and K-A-based ages of the Hawaiian Islands to estimate molecular evolutionary rates. *Mol. Ecol.* 7, 533-545.
- ✓ Gaffney DJ, Keightley PD, 2006. Genomic selective constraints in murid noncoding DNA. *PLoS Genet.* 2:e204.
- ✓ Gao GF, Tormo J, Gerth UC, Wyer JR, McMichael AJ, Stuart DI, Bell JI, Jones EY, Jakobsen BK, 1997. Crystal structure of the complex between human CD8alpha(alpha) and HLA-A2. *Nature* 387, 630-634.
- ✓ Garcia-Moreno J, Mindell DP, 2000. Rooting a phylogeny with homologous genes on opposite sex chromosomes (gametologs): a case study using avian CHD. *Mol Biol Evol.* 17, 1826-1832.
- ✓ Garcia-Ortiz JE, Sandoval-Ramirez L, Rangel-Villalobos H, Maldonado-Torres H, Cox S, Garcia-Sepulveda CA, Figuera LE, Marsh SG, Little AM, Madrigal JA, Moscoso J, Arnaiz-Villena A, Arguello JR, 2006. High-resolution molecular characterization of the HLA class I and class II in the Tarahumara Amerindian population. *Tissue Antigens* 68, 135-146.
- ✓ Garrigan D, Hedrick PW, 2003. Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution* 57, 1707-1722.
- ✓ Gasper JS, Shiina T, Inoko H, Edwards SV, 2001. Songbird genomics: analysis of 45 kb upstream of a polymorphic Mhc class II gene in red-winged blackbirds (*Agelaius phoeniceus*). *Genomics* 75, 26-34.
- ✓ Gaunt A, Wells M, 1973. Models of syringeal mechanisms. *Am Zool* 13, 1227-1247.
- ✓ Gazave E, Marques-Bonet T, Fernando O, Charlesworth B, Navarro A. 2007. Patterns and rates of intron divergence between humans and chimpanzees. *Genome Biol.* 8:R21.
- ✓ Gilbert W. 1978. Why genes in pieces? *Nature* 271,501.
- ✓ Gilbert W. 1987. The exon theory of genes. *Cold Spring Harbor Symp. Quant Biol.* 52, 901–905.

-
- ✓ Giles RE, Blanc H, Cann HM, Wallace DC, 1980. Maternal Inheritance of Human Mitochondrial-Dna. Proc Natl Acad Sci. 77, 6715-6719.
 - ✓ Gill FB, 1999. Ornithology. W. H. Freeman and Company, New York.
 - ✓ Gillespie JH, 1978. A general model to account for enzyme variation in natural populations. V. The SAS--CFF model. Theor. Popul. Biol 14, 1-45.
 - ✓ Glick B, Schatz G, 1991. Import of Proteins Into Mitochondria. Annual. Annual Review of Genetics 25, 21-44.
 - ✓ Gómez-Prieto P, Reguera R, Parga-Lozano C, Moreno E, Rey D, Areces C, Arnaiz Vilena A, 2009. HLA-G, -F y -E: Polymorphism, Function and Evolution. HLA complex in Biology and Medicine: A resource book. Ed. N. Mehra. Jaypee Brothers Medical Publishers Ltd, New Delih, India.
 - ✓ Gorer PA, 1936. The detection of a hereditary antigenic difference in the blood of mice by means of human group A serum. J Genet 32, 17-31.
 - ✓ Grahovac B, Sukernik RI, O'hUigin C, Zaleska-Rutczynska Z, Blagitko N, Raldugina O, Kosutic T, Satta Y, Figueroa F, Takahata N, Klein J, 1998. Polymorphism of the HLA class II loci in Siberian populations. Hum. Genet. 102, 27-43.
 - ✓ Greenewalt C, 1968. Bird Song: Acoustics and Physiology. Smithsonian Institution Press, Washington DC.
 - ✓ Greenewalt C, 1969. How birds sing. Sci Am 221, 126-139.
 - ✓ Gregory TR, Andrews CB, McGuire JA, Witt CC, 2009. The smallest avian genomes are found in hummingbirds. Proc Biol Sci R Soc. 276, 3753–3757.
 - ✓ Grossberger D, Parham P, 1992. Reptilian class I major histocompatibility complex genes reveal conserved elements in class I structure. Immunogenetics 36, 166-174.
 - ✓ Groth JG, Barrowclough GF, 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol Phyl Evol 12, 115-123.
 - ✓ Guillemot F, Billault A, Pourquie O, Behar G, Chausse AM, Zoorob R, Kreibich G, Auffray C, 1988. A molecular map of the chicken major histocompatibility complex: the class II beta genes are closely linked to the class I genes and the nucleolar organizer. EMBO J. 7, 2775-2785.
 - ✓ Guillemot F, Billault A, Auffray C, 1989a. Physical linkage of a guanine nucleotide-binding protein-related gene to the chicken major histocompatibility complex. Proc Natl. Acad. Sci. USA 86, 4594-4598.
 - ✓ Guillemot F, Kaufman JF, Skjoedt K, Auffray C, 1989b. The major histocompatibility complex in the chicken. Trends Genet 5, 300-304.
 - ✓ Hala K, Plachy J, Schulmannova J, 1981. Role of the B-G-region antigen in the humoral immune response to the B-F-region antigen of chicken MHC. Immunogenetics 14, 393-401.
 - ✓ Härlid A, Janke A, Arnason U, 1998. The complete mitochondrial genome of *Rhea americana* and early avian divergences. J Mol Evol. 46, 669-679.
 - ✓ Härlid A, Arnason U, 1999. Analyses of mitochondrial DNA nest ratite birds within the Neognathae: supporting a neotenus origin of ratite morphological characters. Proc R Soc LondB Biol Sci 266, 305-309.

- ✓ Hashimoto K, Okamura K, Yamaguchi H, Ototake M, Nakanishi T, Kurosawa Y, 1999. Conservation and diversification of MHC class I and its related molecules in vertebrates. *Immunol Rev* 167, 81-100.
- ✓ Hayden EJ, Ferrada E, Wagner A, 2011. Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* 474, 92-95.
- ✓ Hedrick PW, Ginevan M, Ewing E, 1976. Genetic polymorphism in heterogeneous environments. *Annual Rev. Ecol. Syst.* 7, 1-32.
- ✓ Hedrick PW, Black FL, 1997. HLA and mate selection: no evidence in South Amerindians. *Am. J Hum. Genet* 61, 505-511.
- ✓ Hedrick PW, 1998. Balancing selection and MHC. *Genetica* 104, 207-214.
- ✓ Hedrick PW, 2007. Balancing selection. *Curr. Biol* 17, R230-R231.
- ✓ Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, Greenwood BM, 1991. Common west African HLA antigens are associated with protection from severe malaria. *Nature* 352, 595-600.
- ✓ Hill AV, 2001. Immunogenetics and genomics. *Lancet* 357, 2037-2041.
- ✓ Horton R, Wilming L, Rand V, Lovering R.C, Bruford EA, Khodiyar VK, Lush MJ, Povey S, Talbot CC, Wright MW, Wain HM, Trowsdale J, Ziegler A, Beck S, 2004. Gene map of the extended human MHC. *Nat. Rev. Genet* 5, 889-899.
- ✓ Howell N, Gilbert K, 1988. Mutational Analysis of the Mouse Mitochondrial Cytochrome-B Gene. *J Mol Biol* 203, 607-618.
- ✓ Howell N, 1989. Evolutionary conservation of protein regions in the proton-motive cytochrome *b* and their possible roles in redox catalysis. *J. Mol. Evol.* 29, 157-169.
- ✓ Huelsenbeck JP, Ronquist F, 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17, 754-755.
- ✓ Hughes AL, Nei M, 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335, 167-170.
- ✓ Hughes AL, Nei M, 1989a. Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. *Mol. Biol Evol.* 6, 559-579.
- ✓ Hughes AL, Nei M, 1989b. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl. Acad. Sci. U. S. A* 86, 958-962.
- ✓ Hughes AL, Hughes MK, 1995a. Natural selection on the peptide-binding regions of major histocompatibility complex molecules. *Immunogenetics* 42, 233-243.
- ✓ Hughes AL, Hughes MK. 1995b. Small genomes for better flyers. *Nature* 377-391.
- ✓ Hughes JM, 2003. Phylogeny of cooperatively breeding cuckoos (Cuculidae, Crotophaginae) based on mitochondrial gene sequences. *Naturwissenschaften* 90, 231-233.
- ✓ Huxley T, 1867. On the classification of birds and on the taxonomic value of the modifications of certain of the cranial bones observable in that class. *Proc. Zool. Soc.* 415-472.
- ✓ Irestedt M, Johansson US, Parsons TJ, Ericson PGP, 2001. Phylogeny of major lineages of suboscines (Passeriformes) analysed by nuclear DNA sequence data. *J. Avian Biol.* 32, 15-25.

-
- ✓ Irwin DM, Kocher TD, Wilson AC, 1991. Evolution of the cytochrome b gene of mammals. *J. Mol. Evol.* 32, 128-144.
 - ✓ Janeway CA, Travers P, 1996. *Immunobiology. The immune system in health and disease*, second Edition. Blackwell Scientific Publications, Oxford.
 - ✓ Janeway CA, Travers P, Walport, Shlomchik, 2005. *Immunobiology: The Immune System in Health and Disease*. Garland Science, New York. Edition.
 - ✓ Jarvis ED, Mirarab S, Aberer AJ, *et al*, 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346, 1320-1331.
 - ✓ Jensen CO, 1903. Experimentelle untersuchungen iiber krebs bei mausen. *Zentralbl BakteriolParasitenkd Infektkr* 39, 28.
 - ✓ Johnson KP, 2001. Taxon sampling and the phylogenetic position of *Passeriformes*: evidence from 916 avian cytochrome b sequences. *Syst. Biol.* 50, 128-136.
 - ✓ Juul-Madsen HR, Zoorob R, Auffray C, Skjodt K, Hedemand JE, 1997. New chicken Rfp-Y haplotypes on the basis of MHC class II RFLP and MLC analyses. *Immunogenetics* 45, 345-352.
 - ✓ Kapustin S, Lyshchov A, Alexandrova J, Imyanitov E, BlinovM, 1999. HLA class II molecular polymorphisms in healthy Slavic individuals from North-Western Russia. *Tissue Antigens* 54, 517-520.
 - ✓ 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, Skjodt K, Salomonsen J, 1991. The B-G multigene family of the chicken major histocompatibility complex. *Crit Rev. Immunol.* 11, 113-143.
 - ✓ Kaufman J, Andersen R, Avila D, Engberg J, Lambris J, Salomonsen J, Welinder K, Skjodt K, 1992. Different features of the MHC class I heterodimer have evolved at different rates. Chicken B-F and beta 2-microglobulin sequences reveal invariant surface residues. *J. Immunol.* 148, 1532-1546.
 - ✓ 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.
 - ✓ Kaufman J, Jacob J, Shaw I, Walker B, Milne S, Beck S, Salomonsen J, 1999a. Gene organisation determines evolution of function in the chicken MHC. *Immunol. Rev* 167, 101-117.
 - ✓ Kaufman J, Milne S, Gobel T.W, Walker B.A, Jacob J P, Auffray C, Zoorob R, Beck S, 1999b. The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401, 923-925.
 - ✓ Kendall E, Sargent CA, Campbell RD, 1990. Human major histocompatibility complex contains a new cluster of genes between the HLA-D and complement C4 loci. *Nucleic Acids Res* 18: 7251-7257.
 - ✓ Klein J, 1986. *Natural history of the Major Histocompatibility Complex*. J. Willey and Sons, New York, USA.
 - ✓ Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX, Wilson AC, 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. U. S. A* 86, 6196-6200.
 - ✓ Kojima K, 1971. Is there a constant fitness for a given genotype? No! *Evolution* 25, 281-285.

- ✓ Landry C, Garant D, Duchesne P, Bernatchez L, 2001. 'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). *Proc Biol Sci.* 268, 1279-1285.
- ✓ Lazaro AM, Moraes ME, Marcos CY, Moraes JR, Fernandez-Vina MA, Stastny P, 1999. Evolution of HLA-class I compared to HLA-class II polymorphism in Terena, a South-American Indian tribe. *Hum. Immunol.* 60, 1138-1149.
- ✓ Leffell MS, Fallin MD, Hildebrand WH, Cavett JW, Iglehart BA, Zachary AA, 2004. HLA alleles and haplotypes among the Lakota Sioux: report of the ASHI minority workshops, part III. *Hum. Immunol.* 65, 78-89.
- ✓ Lewontin R, Hubby J, 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and the degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 54, 595-609.
- ✓ Little CC, Johnson BW, 1922. The inheritance of susceptibility to implants of splenic tissue in mice. *Proc. Soc. Exp. Biol. Med.* 163.
- ✓ Loeb L, 1908. Ueber Entstehung eines Sarkoms nach Transplantation eines Adenocarcinoms einer japanischen Maus. *J Cancer Res Clin Oncol* 7.
- ✓ Lowy, E. Evolución y sistema principal de histocompatibilidad de canarios (género *Serinus*). 2003. Tesis doctoral Universidad Complutense Madrid, Director: Dr. A. Arnaiz-Villena, Sobresaliente *Cum Laude*.
- ✓ Lynch M, 2002. Intron evolution as a population-genetic process. *Proc Natl Acad Sci U S A.* 99, 6118-6123.
- ✓ Lynch M, Conery JS, 2003. The origins of genome complexity. *Science* 302, 1401-1404
- ✓ Marais G, Nouvellet P, Keightley PD, Charlesworth B, 2005. Intron size and exon evolution in *Drosophila*. *Genetics* 170, 481-485.
- ✓ Martinez-Laso J, Gazit E, Gomez-Casado E, Morales P, Martinez-Quiles N, Alvarez M, Martin-Villa J.M, Fernandez V, Arnaiz-Villena A, 1996. HLA DR and DQ polymorphism in Ashkenazi and non-Ashkenazi Jews: comparison with other Mediterraneans. *Tissue Antigens* 47, 63-71.
- ✓ Martinez-Laso J, Sartakova M, Allende L, Kononkov V, Moscoso J, Silvera-Redondo C, Pacho A, Trapaga J, Gomez-Casado E, Arnaiz-Villena A, 2001. HLA molecular markers in Tuvinians: a population with both Oriental and Caucasoid characteristics. *Ann. Hum. Genet.* 65, 245-261.
- ✓ Martinez-Laso J, Siles N, Moscoso J, Zamora J, Serrano-Vela JI, Ira-Cachafeiro J, Castro MJ, Serrano-Rios M, Arnaiz-Villena A, 2006. Origin of Bolivian Quechua Amerindians: their relationship with other American Indians and Asians according to HLA genes. *Eur. J. Med. Genet.* 49, 169-185.
- ✓ Martinez-Laso J, Montoya F, Areces C, Moscoso J, Silvera C, Rey D, Parga-Lozano C, Gomez-Prieto P, Enriquez de Salamanca M, Arnaiz-Villena A, 2011. HLA in Jaidukama: an Amerindian secluded Colombian population with new haplotypes and Asian and Pacific-shared alleles. *Mol. Biol Rep.* 38, 3689-3701.
- ✓ Miller MM, Goto R, Bernot A, Zoorob R, Auffray C, Bumstead N, Briles WE, 1994. Two Mhc class I and two Mhc class II genes map to the chicken Rfp-Y system outside the B complex. *Proc Natl. Acad. Sci. U. S. A* 91, 4397-4401.
- ✓ Miller MM, Goto RM, Taylor RL, Zoorob R, Auffray C, Briles RW, Briles WE, Bloom SE, 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. Proc Natl. Acad. Sci. U. S. A 93, 3958-3962.
- ✓ Miller HC, Lambert DM, 2004. Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). Immunogenetics 56, 178-191.
 - ✓ Mindell DP, Sorenson MD, Huddleston CJ, Miranda Jr HC, Knight A, Sawchuck SJ, Yuri T, 1997 Phylogenetic relationships among and within select avian orders based on mitochondrial DNA., Mindell DP (ed) Avian molecular evolution and systematics. Academic Press, San Diego Edition.
 - ✓ Mindell DP, Sorenson MD, Dimcheff DE, Hasegawa M, Cast JC, Yuri T, 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. Syst Biol 48, 138-152.
 - ✓ Montoya J, Ojala D, Attardi G, 1981. Distinctive Features of the 5'-Terminal Sequences of the Human Mitochondrial Messenger-Rnas. Nature 290, 465-470.
 - ✓ 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, 6702-6712.
 - ✓ Mourer-Chauvire C, Hugueney M, Jonet P, 1989. Decouverte de *Passeriformes* dans l'Oligocène superieur de France. C R Acad Sci Paris 309, 843-849.
 - ✓ Mungall AJ, Palmer SA, Sims SK, Edwards CA, Ashurst JL, Wilming L, Jones MC, Horton R, *et al.*, 2003. The DNA sequence and analysis of human chromosome 6. Nature 425, 805-811.
 - ✓ Munkhbat B, Sato T, Hagihara M, Sato K, Kimura A, Munkhtuvshin N, Tsuji K, 1997. Molecular analysis of HLA polymorphism in Khoton-Mongolians. Tissue Antigens 50, 124-134.
 - ✓ NABC, 2003. Manual para anillar *Passeriformes* y *CuasiPasseriformes* del anillador de Norte America (excluyendo Colibries y Buhos). Point Reyes Station, California, USA.
 - ✓ Nei M, 1972. Genetic distances between populations. Am. Nat. 106, 283.
 - ✓ Nei M, Gojobori T, 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418-426.
 - ✓ Nei M, 1987 Molecular Evolutionary Genetics. Columbia University Press, New York, NY (USA).
 - ✓ Ober C, Elias S, Kostyu DD, Hauck WW, 1992. Decreased fecundability in Hutterite couples sharing HLA-DR. Am. J Hum. Genet 50, 6-14.
 - ✓ Ojala D, Montoya J, Attardi G, 1981. Transfer-Rna Punctuation Model of Rna Processing in Human Mitochondria. Nature 290, 470-474.
 - ✓ Olivo A, Debaz H, de la Rosa G, Alaez C, Juarez V, Gorodezky C, Guedez Y, Dominguez E, Herrera F, Soto M, Matos M, Scorza J, Layrisse Z, 1996 Alelos y Haplotipos del Complejo Principal de Histocompatibilidad en tres poblaciones indigenas Americanas. In: Municio,A., Barreno,P. (Eds.), Polimorfismo Genico HLA en Poblaciones Hispanoamericanas. Real Academis de Ciencias Exactas, Fisicas y Naturales, Madrid, Spain, pp. 195-213.
 - ✓ Olson SL, 1983. Evidence for a polyphyletic origin of the Piciformes. Auk 100, 126-133.
 - ✓ Organ CL, Shedlock AM, Meade A, Pagel M, Edwards SV, 2007. Origin of avian genome size and structure in non-avian dinosaurs. Nature 446, 180-184.
 - ✓ Parga-Lozano C, Rey-Medrano D, Gomez-Prieto P, Areces C, Moscoso J, Abd-El-Fatah-Khalil S, Moreno E, Arnaiz-Villena A, 2011. HLA genes in Amerindian immigrants to Madrid (Spain):

- epidemiology and a virtual transplantation waiting list: Amerindians in Madrid (Spain). *Mol. Biol Rep.* 38, 2263-2271.
- ✓ Parham P, Lomen CE, Lawlor DA, Ways JP, Holmes N, Coppin H, Salter RD, Wan AM, Ennis PD, 1988. Nature of polymorphism in HLA-A, -B, and -C molecules. *PNAS* 85, 4005-4009.
 - ✓ Parham P, Adams EJ y Arnett KL. 1995. The origins of HLA-A, B, C polymorphism. *Immunol Rev* 143, 141-180.
 - ✓ Parsch J, Novozhilov S, Saminadin-Peter SS, Wong KM, Andolfatto P, 2010. On the utility of short intron sequences as a reference for the detection of positive and negative selection in *Drosophila*. *Mol Biol Evol.* 27,1226–1234.
 - ✓ Paterson S, Pemberton JM, 1997. No evidence for major histocompatibility complex-dependent mating patterns in a free-living ruminant population. *Proc Biol Sci.* 264, 1813-1819.
 - ✓ Paturi FR, 1991. *Die Chronik der Erde.*, Haremburg Kommunikation, Dortmund, Germany.
 - ✓ Payne R, Rolfs MR, 1958. Fetomaternal leukocyte incompatibility. *J Clin Invest* 37, 1756-1763.
 - ✓ Petit N, Casillas S, Ruiz A, Barbadilla A, 2007. Protein Polymorphism Is Negatively Correlated with Conservation of Intronic Sequences and Complexity of Expression Patterns in *Drosophila melanogaster*. *Journal of Molecular Evolution* 64, 511-518.
 - ✓ Pfanner N, Neupert W, 1990. The Mitochondrial Protein Import Apparatus. *Annu Rev Biochem* 59, 331-353.
 - ✓ Pharr GT, Gwynn AV, Bacon LD, 1996. Histocompatibility antigen(s) linked to Rfp-Y (Mhc-like) genes in the chicken. *Immunogenetics* 45, 52-58.
 - ✓ Plachy J, Pink JR, Hala K, 1992a. Biology of the chicken MHC (B complex). *Crit Rev. Immunol.* 12, 47-79.
 - ✓ Plachy J, Chausse AM, Thoraval P, Coudert F, 1992b. Molecular genotyping of recombinant congenic lines provides evidence for crossing-over within the B-G region of the major histocompatibility complex of the chicken. *Immunogenetics* 36, 270-273.
 - ✓ Potts WK, Manning CJ, Wakeland EK, 1991. Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature* 352, 619-621.
 - ✓ Prum RO, 1993. Phylogeny, biogeography, and evolution of the broadbills (Eurylaimidae) and asities (Philepittidae) based on morphology. *Auk* 110, 304-324.
 - ✓ Qu Zhang, Scott V. Edwards, 2012. The Evolution of Intron Size in Amniotes: A Role for Powered Flight?. *Genome Biol. Evol.* 4,1033–1043.
 - ✓ Questiau S, Eybert MC, Gaginskaya AR, Gielly YL, Taberlet P, 1998. Recent divergence between two morphologically differentiated subspecies of bluethroat (Aves: Muscicapidae: *Luscinia svecica*) inferred from mitochondrial DNA sequence variation. *Mol. Eco.* 7, 239-45.
 - ✓ Raikow RJ, 1978. Appendicular myology and relationships of the New World nine-primaried oscines (Aves: *Passeriformes*). *Bull Carnegie Mus. Nat. Hist.* 7, 1-43.
 - ✓ Raikow RJ, 1982. Monophyly of the *Passeriformes*: test of a phylogenetic hypothesis. *Auk* 99, 431-455.
 - ✓ Raikow RJ, 1984. Hindlimb myology and phylogenetic position of the New Zealand wens. *AmZool* 24, 446.

- ✓ Regal PJ, 1977. Ecology and evolution of flowering plant dominance. *Science* 196, 622-629.
- ✓ Renzo Esuperanzi, 2008. Los fringílicos. Jilgueros, pardillos, verderones y otros pájaros silvestres, Editorial Hispano Europea, S.A.; Edition.
- ✓ Reusch TB, Haberli MA, Aeschlimann PB, Milinski M, 2001. Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. *Nature* 414, 300-302.
- ✓ Rey D, Parga-Lozano C, Moscoso J, Areces C, Enriquez-de-Salamanca M, Fernandez-Honrado M, Abd-El-Fatah-Khalil S, Alonso-Rubio J, Arnaiz-Villena A, 2013. HLA genetic profile of Mapuche (Araucanian) Amerindians from Chile. *Mol. Biol Rep.* 40, 4257-4267.
- ✓ Riegert P, Andersen R, Bumstead N, Dohring C, Dominguez-Steglich M, Engberg J, Salomonsen J, Schmid M, Schwager J, Skjodt K, Kaufman J, 1996. The chicken beta 2-microglobulin gene is located on a non-major histocompatibility complex microchromosome: a small, G+C-rich gene with X and Y boxes in the promoter. *Proc Natl Acad Sci U S A* 93, 1243-1248.
- ✓ Roitberg-Tambur A, Witt CS, Friedmann A, Safirman C, Sherman L, Battat S, Nelken D, Brautbar C, 1995. Comparative analysis of HLA polymorphism at the serologic and molecular level in Moroccan and Ashkenazi Jews. *Tissue Antigens* 46, 104-110.
- ✓ Roy SW, Gilbert W, 2006. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat Rev Genet.* 7, 211-221.
- ✓ Saitou N, Nei M, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406-425.
- ✓ Sanz J, 1999. Los dinosaurios voladores: historia evolutiva de las aves primitivas, 1st reprint Edition.
- ✓ Saraste M, 1984. Location of Heme-Binding Sites in the Mitochondrial Cytochrome-B. *FebsLetters* 166, 367-372.
- ✓ Sato A, Mayer WE, Tichy H, Grant PR, Grant BR, Klein J, 2001. Evolution of Mhc class II B genes in Darwin's finches and their closest relatives: birth of a new gene. *Immunogenetics* 53, 792-801.
- ✓ Schat KA, 1987. Marek's disease: a model for protection against herpesvirus-induced tumours. *Cancer Surv.* 6, 1-37.
- ✓ Schatz G, 1996. The protein import system of mitochondria. *J Biol Chem* 271, 31763-31766.
- ✓ Schierman LW, ordskog AW, 1961. Relationship of blood type to histocompatibility in chickens. *Science* 134, 1008-1009.
- ✓ Shiina T, Ando A, Imanishi T, Kawata H, Hanzawa K, Gojobori T, Inoko H, Watanabe S, 1995. Isolation and characterization of cDNA clones for Japanese quail (*Coturnix japonica*) major histocompatibility complex (MhcCoja) class I molecules. *Immunogenetics* 42, 213-216.
- ✓ Shiina T, Shimizu S, Hosomichi K, Kohara S, Watanabe S, Hanzawa K, Beck S, Kulski JK, Inoko H, 2004. Comparative genomic analysis of two avian (quail and chicken) MHC regions. *J Immunol* 172, 6751-6763.
- ✓ Sibley CG, Ahlquist J, 1982. The relationships of the Hawaiian Honeycreepers (*Drepaninini*) as indicated DNA - DNA hybridization. *Auk* 99, 130-140.
- ✓ Sibley CG, Monroe BL, 1990. Distribution and Taxonomy of Birds of the World. Yale University Press, New Haven and London.

- ✓ Sibley CG, Ahlquist J, 1990. *Phylogeny and classification of birds*. Yale University Press, New Haven, Conn.
- ✓ Sibley CG, Monroe BL, 1993. *A World Checklist of Birds.*, Yale University Press, New Haven and London Edition.
- ✓ Sibley CG, 1995. *Birds of the World*. Thayer Birding Software.
- ✓ Sibley CG, 1996. *Birds of the World*, on diskette, Windows version 2.0., Santa Rosa, CA Edition.
- ✓ Silvera C, Vargas-Alarcon G, Areces C, Rey D, Parga-Lozano C, Gomez-Prieto P, Barbolla L, Martinez-Laso J, Arnaiz-Villena A, 2011. HLA genes in Wayu Amerindians from Colombia. *Immunol. Invest* 40, 92-100.
- ✓ Simonsen M, Crone M, Koch C, Hala K, 1982. The MHC haplotypes of the chicken. *Immunogenetics* 16, 513-532.
- ✓ Simonsen M, Arnul M, Sorensen P, 1989. The chicken MHC and its importance. Zijpp A, Sybesma W, *Improving genetic disease resistance in farm animals*. Kluwer.
- ✓ Singh R, Kaul R, Kaul A, Khan K, 2007. A comparative review of HLA associations with hepatitis B and C viral infections across global populations. *World J Gastroenterol.* 13, 1770-1787.
- ✓ Smith D, Funnell B, 2009. *Atlas of Mesozoic and Cenozoic Coastlines*, Illustrated Edition. Cambridge University Press, Cambridge.
- ✓ Sneath PHA, Sokal RR, 1973. *Numerical Taxonomy* San Francisco.
- ✓ Snell GD, 1948. Methods for the study of histocompatibility genes. *J Genet* 49, 87-108.
- ✓ Stewart CA, Horton R, Allcock RJ, Ashurst JL, Atrazhev AM, Coggill P, Dunham I, Forbes S, Halls K, Howson JM, Humphray SJ, Hunt S, Mungall AJ, Osoegawa K, Palmer S, Roberts AN, Rogers J, Sims S, Wang Y, Wilming LG, Elliott JF, de Jong PJ, Sawcer S, Todd JA, Trowsdale J, Beck S, 2004. Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res* 14, 1176-1187.
- ✓ Sun J, Leahy DJ, Kavathas PB, 1995. Interaction between CD8 and major histocompatibility complex (MHC) class I mediated by multiple contact surfaces that include the alpha 2 and alpha 3 domains of MHC class I. *J Exp Med* 182, 1275-1280.
- ✓ Swofford, D.L. PAUP. *Phylogenetic Analysis Using Parsimony (* and other methods)* version 4. 2002. Sunderland, Massachusetts, Sinauer Associates.
- ✓ Szarski H, 1983. Cell size and the concept of wasteful and frugal evolutionary strategies. *J Theor Biol.* 105, 201-209.
- ✓ Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512-526.
- ✓ Tamura K, Nei M, Kumar S, 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A* 101, 11030-11035.
- ✓ Tamura K, Stecher G, Peterson D, Filipski A, Kumar S, 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 30, 2725-2729.
- ✓ The MHC sequencing consortium. 1999. Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401, 921-923.
- ✓ Thorsby E, 2009. A short history of HLA. *Tissue antigens* 74, 101-116.

- ✓ Thomas ML, Harger JH, Wagener DK, Rabin BS, Gill TJ., III, 1985. HLA sharing and spontaneous abortion in humans. *Am. J Obstet. Gynecol.* 151, 1053-1058.
- ✓ Thomsen M, Neugebauer M, Arnaud J, Borot N, Sevin A, Baur M, Cambon-Thomsen A, 1994. Recombination fractions in the HLA system based on the data set 'provinces Francaises': indications of haplotype-specific recombination rates. *Eur. J Immunogenet.* 21, 33-43.
- ✓ Thorne JL, Kishino H, Painter IS, 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* 15, 1647-1657.
- ✓ Trachtenberg E, Erlich H, 2001. A review of the role of the human leukocyte antigen (HLA) system as a host immunogenetic factor influencing HIV transmission and progression to AIDS . *HIV molecular immunology* 43-60.
- ✓ Tysoe-Calnon V, Grundy J, Perkins S, 1991. Molecular comparisons of the B2M-binding site in Class I Major Histocompatibility Complex a-chains and proteins of related sequences. *Biochem J.* 277, 359-369.
- ✓ Uriarte-Cantolla A, 2003 *Historia del Clima de la Tierra*, 1 Edition. Servicio Central de Publicaciones del Gobierno Vasco, Vitoria-Gasteiz, Spain.
- ✓ Urrutia AO, Hurst LD, 2003. The signature of selection mediated by expression on human genes. *Genome Res.* 13, 2260–2264.
- ✓ Van den Bussche RA, Longmire JL, Baker RJ, 1995. How bats achieve a small C-value: frequency of repetitive DNA in *Macrotus*. *Mamm Genome* 6,521–525.
- ✓ van Rood JJ, Eernisse JG, van Leeuwen A, 1958. Leucocyte antibodies in sera from pregnant women. *Nature* 181, 1735-1736.
- ✓ van Tuinen M, Sibley CG, Hedges SB, 2000. The early history of modern birds inferred from DNA sequences of nuclear and mitochondrial ribosomal genes. *Mol Biol. Evol.* 17, 451-457.
- ✓ Vargas-Alarcon G, Moscoso J, Martinez-Laso J, Rodriguez-Perez JM, Flores-Dominguez C, Serrano-Vela JI, Moreno A, Granados J, Arnaiz-Villena A, 2007. Origin of Mexican Nahuas (Aztecs) according to HLA genes and their relationships with worldwide populations. *Mol. Immunol.* 44, 747-755.
- ✓ Vargas-Alarcon G, Granados J, Rodriguez-Perez JM, Parga C, Perez-Hernandez N, Rey D, Zuniga J, Arnaiz-Villena A, 2010. Distribution of HLA class II alleles and haplotypes in Mexican Mestizo population: comparison with other populations. *Immunol. Invest* 39, 268-283.
- ✓ Vargas-Alarcon G, Granados J, Perez-Hernandez N, Rodriguez-Perez J.M, Canto-Cetina T, Coral-Vazquez RM, Areces C, Gomez-Prieto P, Arnaiz-Villena A, 2011. HLA-class II genes in Mexican Amerindian Mayas: relatedness with Guatemalan Mayans and other populations. *Immunol. Invest* 40, 101-111.
- ✓ Vinogradov AE, 1999. Intron-genome size relationship on a large evolutionary scale. *J Mol Evol.* 49,376–384.
- ✓ Vinogradov AE, 2004. Compactness of human housekeeping genes: selection for economy or genomic design? *Trends Genet.* 20, 248–253.
- ✓ Wallace D, Lott M, Torroni A, Brown M, Shoffner J, 1994. Report of the committee on human mitochondrial DNA., Cuticchia AJ, Pearson PL (eds) *Human Gene Mapping*. Johns Hopkins Univ. Press, Baltimore Edition.
- ✓ Waltari E, Edwards SV, 2002. Evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs. *Am Nat.* 160, 539–552.

-
- ✓ Welinder KG, Jespersen HM, Walther-Rasmussen J, Skjodt K, 1991. Amino acid sequences and structures of chicken and turkey beta 2-microglobulin. *Mol Immunol* 28, 177-182.
 - ✓ 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, 158-170.
 - ✓ Widger WR, Cramer WA, Herrmann RG, Trebst A, 1984. Sequence Homology and Structural Similarity Between Cytochrome-B of Mitochondrial Complex-Iii and the Chloroplast-B6-F Complex - Position of the Cytochrome-B Hemes in the Membrane. *Proc Natl Acad Sci USA* 81, 674-678.
 - ✓ Wink M, 1995. Phylogeny of Old and New World vultures (Aves: *Accipitridae* and *Cathartidae*) inferred from nucleotide sequences of the mitochondrial cytochrome b gene. *Z Naturforsch C*. 50, 868-82.
 - ✓ 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, 479-490.
 - ✓ Xia C, Hu T, Yang T, Wang L, Xu G, Lin C, 2005. cDNA cloning, genomic structure and expression analysis of the goose (*Anser cygnoides*) MHC class I gene. *Vet Immunol Immunopathol* 107, 291-302.
 - ✓ Yamazaki K, Boyse EA, Mike V, Thaler HT, Mathieson BJ, Abbott J, Boyse J, Zayas ZA, Thomas L, 1976. Control of mating preferences in mice by genes in the major histocompatibility complex. *J Exp Med* 144, 1324-1335.
 - ✓ Zachary AA, Kopchaliiska D, Jackson AM, Leffell MS, 2010. Immunogenetics and immunology in transplantation. *Immunol. Res* 47, 232-239.
 - ✓ Zamora J, Lowy E, Ruiz-del-Valle V, Moscoso J, Serrano-Vela JI, Rivero-de-Aguilar J, Arnaiz-Villena A, 2006a. *Rhodopechys obsoleta* (desert finch): a pale ancestor of greenfinches (*Carduelis* spp.) according to molecular phylogeny. *J. Ornithol.* 147, 448-456.
 - ✓ Zamora J, Moscoso J, Ruiz-del-Valle V, Lowy E, Serrano-Vela JI, Ira-Cachafeiro J, Arnaiz-Villena A, 2006b. Conjoint mitochondrial phylogenetic trees for canaries (*Serinus* spp.) and goldfinches (*Carduelis* spp.) show several specific polytomies. *Ardeola* 53, 1-17.
 - ✓ Zinkernagel RM, Doherty PC, 1979. MHC-restricted cytotoxic T cells: studies on the biological role of polymorphic major transplantation antigens determining T-cell restriction-specificity, function, and responsiveness. *Adv. Immunol.* 27, 51-177.

ANEXOS

1. Listado de las abreviaturas más empleadas en este trabajo

LISTA DE ABREVIATURAS

ADN	Ácido desoxirribonucleico
mtADN	ADN mitocondrial
ARN	Ácido ribonucleico
ARNm	ARN mensajero
ARNt	ARN transferente
ARS	Sitio de reconocimiento antigénico
BI	Método filogenético Bayesiano
β2m	β2-microglobulina
d_s	Tasa de sustitución sinónima
d_n	Tasa de sustitución no sinónima
dNTPs	Desoxirribonucleótidos trifosfato
H-2	Sistema principal de histocompatibilidad en ratones
HFE	Hemocromatosis
HLA	Antígenos leucocitarios humanos
l	Litro(s)
m	Metro(s)
M	Molar
ma	Millones de años
MHC	Sistema mayor de histocompatibilidad
ML	Maximum likelihood, Máxima probabilidad
ml	Mililitro
mM	Milimolar
μl	Microlitro
MLR	Reacción leucocitaria mixta
MP	Máxima parsimonia
ng	Nanogramo
NJ	Neighbor-joining, Asociación de vecinos
NK	Natural Killer
NOR	Región de organización nucleolar
pb	Unidad de pares de bases
PCR	Reacción en cadena de la polimerasa
RC	Rango de conservación
SEM	Error estandar de la media
Taq	Thermus aquaticus (polimerasa)

TAP	Presentador asociado con la presentación antigénica
TCR	Receptor de linfocitos T
x MHC	MHC extendido
µg	Microgramo
µl	Microlitro

2. Clasificación de las aves (Sibley, 1996)

ORDEN	FAMILIA	ESPECIES
STRUTHIONIFORMES	4	11
TINAMIFORMES	1	48
CRACIFORMES	2	72
GALLIFORMES	3	218
ANSERIFORMES	4	163
TURNICIFORMES	1	17
PICIFORMES	5	357
GALBULIFORMES	2	52
BUCEROTIFORMES	2	56
UPUIFORMES	3	12
TROGONIFORMES	1	39
CORACIORMES	10	154
COLIIFORMES	1	6
CUCULIFORMES	6	82
PSITTACIFORMES	2	360
APODIFORMES	2	106
TROCHILIFORMES	1	139
MUSOPHAGIFORMES	1	23
STRIGIFORMES	9	137
COLUMBIFORMES	2	318
GRUIFORMES	10	199
CICONIIFORMES	30	1058
PASSERIFORMES	48	5875

SUBORDEN TYRANNI → **suboscines**

Familia

Acanthisittidae
Pittidae
Eurylaimidae
Phileptidae
Incertae Sedis
Tyrannidae
Thamnophilidae
Furnariidae
Formicariidae
Conopophagidae
Rhinocryptidae

SUBORDEN PASSERI → **oscines o pájaros de canto**

PARVAORDEN CORVIDA

SUPERFAMILIA MENUROIDEA

Familia Climacteridae
 Familia Menuridae
 Familia Ptilonorhynchidae
 Familia Turnagridae

SUPERFAMILIA MELIPHAGOIDEA

Familia Maluridae
 Familia Meliphagidae
 Familia Pardalotidae

SUPERFAMILIA CORVOIDEA

Familia Petroicidae
 Familia Irenidae
 Familia Orthonychidae
 Familia Pomatostomidae
 Familia Laniidae
 Familia Vireonidae
 Familia Corvidae
 Familia Callaeatidae
 Familia Picathartidae

PARVAORDEN PASSERIDA

SUPERFAMILIA MUSCICAPOIDEA

Familia Bombycillidae
 Familia Cinclidae
 Familia Muscicapidae
 Familia Sturnidae

SUPERFAMILIA SYLVIOIDEA

Familia Sittidae
 Familia Certhiidae
 Familia Paridae
 Familia Aegithalidae
 Familia Hirundinidae
 Familia Regulidae
 Familia Pycnonotidae
 Familia Hypocoliidae
 Familia Priniidae / Cisticolidae
 Familia Zosteropidae
 Familia Sylviidae

SUPERFAMILIA PASSEROIDEA

Familia Alaudidae
 Familia Nectariinidae
 Familia Melanocharitidae
 Familia Paramythiidae
 Familia Passeridae
 Familia Fringillidae

Subfamilia Peucedraminae
Subfamilia Fringillinae

Tribu Fringillini
 Tribu Carduelini
 Tribu Drepanidini

SubFamilia Emberizinae

Tribu Emberizini
 Tribu Parulini
 Tribu Thraupini
 Tribu Cardinalini
 Tribu Icterini

3.Simbología y código genético

Símbolo	Nucleótido	Ambigüedad	Símbolo
A	Adenina	A/C	M
C	Citosina	A/G	R
G	Guanina	A/T	W
T	Timina	C/G	S
		C/T	Y
		G/T	K
		A/C/G	V
		A/C/T	H
		A/G/T	D
		C/G/T	B
		A/C/T/G/	N

Código genético

		Segunda posición					
		U	C	A	G		
Primera posición (5')	U	Phe	Ser	Tyr	Cys	U	Tercera posición (3')
		Leu		STOP	Trp	A	
	C	Leu	Pro	His	Arg	U	
				Gln		A	
	A	Ile	Thr	Asn	Ser	U	
		Met		Lys	Arg	A	
	G	Val	Ala	Asp	Gly	U	
				Glu		A	
						C	
						A	
						G	

4. Información sobre aminoácidos

AMINOÁCIDO	SÍMBOLO	SÍMBOLO	POLARIDAD
Alanina	Ala	A	Apolar
Arginina	Arg	R	Polar básico
Asparagina	Asn	N	Polar neutro
Aspartato	Asp	D	Polar ácido
Cisteína	Cys	C	Polar neutro
Fenilalanina	Phe	F	Apolar
Glicina	Gly	G	Polar neutro
Glutamato	Glu	E	Polar ácido
Glutamina	Gln	Q	Polar neutro
Histidina	His	H	Polar básico
Isoleucina	Ile	I	Apolar
Leucina	Leu	L	Apolar
Lisina	Lys	K	Polar básico
Metionina	Met	M	Apolar
Prolina	Pro	P	Apolar
Serina	Ser	S	Polar neutro
Tirosina	Tyr	Y	Polar neutro
Treonina	Thr	T	Polar neutro
Triptófano	Trp	W	Apolar
Valina	Val	V	Apolar

5. Publicaciones relacionadas

156

The Open Ornithology Journal, 2010, 3, 156-165

Open Access

Songbirds Conserved Sites and Intron Size of MHC Class I Molecules Reveal a Unique Evolution in Vertebrates

A. Arnaiz-Villena*, V. Ruiz-del-Valle, P. Reche, P. Gomez-Prieto, E. Lowry, J. Zamora, C. Areces, D. Rey, C. Parga and J.I. Serrano-Vela

Department of Immunology, University Complutense, the Madrid Regional Blood Center, Madrid, Spain

Abstract: Birds are considered dinosaurs that passed the 65 million years ago bottleneck. Songbirds (Passeriformes) include about half extant bird species (about 5000) and are generally the most air-thriving bird species, concordantly with their small size. Major Histocompatibility complex (MHC) molecules stimulate immune responses against microbes and its class I molecules have seven conserved residues in all vertebrates from jawed-fishes, 300 million years ago, to humans, including chickens.

All wild songbird species tested by us (n=18) and others (n= 2) differ in $\alpha 1$ domain residue 10 and $\alpha 2$ residue 96 from all other vertebrates. Amplification, cloning and sequencing were performed by standard methods. Sequences alignment were done by using PAUP and MEGA programs software. Crystallographic studies were performed by using mammal and bird MHC molecules from MPID database and other sources and showed that these changes did not significantly vary the MHC class I molecule stability in songbirds.

Further $\alpha 1$ and $\alpha 2$ domain comparisons by simple Composition Distances and Bayesian Inference showed that songbirds overall MHC class I molecules are phylogenetically more separated from mammal than other birds molecules. In addition MHC class I introns from Passeriformes (songbirds) were found to be longer than humans, chicken introns being the shortest ones.

These small mainly air-borne dinosaurs (Passeriformes) have undergone a different evolutive pathway, regarding to MHC, than all other tested vertebrates and more terrestrial birds. This may have been originated by an altogether different dinosaurs lineage origin or to adaptation to more aerial than terrestrial environment or other unknown cause. In any case, the specific changes observed in this work for class I molecules in songbirds have reached a entropic, stable solution similar to that reached by other vertebrates.

Keywords: Songbirds, MHC evolution, immunology, passerines, carduelis, serinus.

INTRODUCTION

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 [1]. These proteins evolve rapidly and are quite different in primary sequence of different species [2]. However, certain amino-acid residues are conserved in MHC class I molecules from reptilian to humans for keeping the overall tertiary structure [3].

MHC class I genes have been completely sequenced ($\alpha 1$ and $\alpha 2$ protein domains [1]) and thoroughly studied in *Gallus gallus* (chicken) [4, 5], and mammals [1, 6]. The MHC locus [7] was first defined in chicken [8, 9], in particular the highly consanguineous variety 'Leghorn' [4, 10]. The chicken's MHC genetic region is considerably smaller than that of mammals —remarkably it has shorter introns [4, 11]—, and is organized quite differently [4]. Thus, a noteworthy difference is the existence of short

introns in the chicken; that supported the hypothesis that the chicken's MHC represented a 'minimal essential MHC' [4, 5]. It has been assumed that chicken (order *Galliformes*) MHC was similar to all species included in the whole class *Aves*. However, it was shown later that this was not the case (see below).

Class I MHC genes have been sequenced and studied in four songbird species: *Acrocephalus arundinaceus* (great reed warbler) [12], *Serinus canaria* (wild canary), *Serinus mozambicus* (yellow-fronted canary), and *Serinus thibetanus* (Tibetan serin) [13, 14]. These sequences showed an apparently more variable exon 3 than *G. gallus* [12-14], and that class I MHC evolution in islands was different than in sister continental species —wild canary (*Serinus canaria*) vs. Asian Tibetan serin and African yellow-fronted canary— [14].

On the other hand, mammals and dinosaurs both appeared in Triassic Epoch —about 300 million years ago (MYA)—, and both survived the 65 MYA Cretaceous extinction. Currently there are about 4200 mammal species, and 9600 dinosaur (bird) species [15, 16]. After the extinction, when all terrestrial dinosaurs disappeared, many mammals occupied these terrestrial niches recently left empty.

*Address correspondence to this author at the Departamento de Inmunología, Facultad de Medicina, Universidad Complutense de Madrid, Avenida Complutense s/n, 28040 Madrid, Spain; Tel: +34 913017354; Fax: +34 913017356; E-mail: arnaiz@med.ucm.es

Carduelini New Systematics: Crimson-winged Finch (*Rhodopechys sanguineus*) is Included in “Arid-Zone” *Carduelini* Finches by Mitochondrial DNA Phylogeny

A. Amaiz-Villena*, V. Ruiz-del-Valle, P. Gomez-Prieto, D. Rey, M. Enriquez-de-Salamanca, J. Marco, E. Muñiz, M. Martín-Villa and C. Areces

Department Immunology, University Complutense, The Madrid Regional Blood Center, Madrid, Spain

Abstract: *Rhodopechys sanguineus* phylogeny together with a group of *Carduelini* finches has been analyzed. Mitochondrial cyt b molecule has been used for species comparison and maximum likelihood and Bayesian methods have been employed in order to obtain a solid phylogeny. Compared *Carduelini* finches groups include: Greenfinches, “Arid-Zone” finches and Genera *Rhodopechys* and *Pyrrhula* species. Our results lead to conclude: 1) Genus *Rhodopechys* included species should need a new taxonomic classification; 2) Genus *Pyrrhula* shares a common ancestor with “Arid-Zone” finches group; the latter is phylogenetically a separate clade, including species from *Carpodacus*, *Rhodopechys* and *Leucosticte* Genera, and 3) *Pinicola enucleator* belongs to Genus *Pyrrhula* and seems to be ancestral. Results show that a systematic revision of *Carduelini* tribe bird species is required.

Keywords: “Arid-Zone” finches, *Carduelini*, *Carpodacus*, Greenfinches, *Pinicola*, *Pyrrhula*, *Rhodopechys*.

INTRODUCTION

Songbirds evolution (*Aves*, *Passeriformes*, *Passeri*) have frequently been studied. Their phenotypic and molecular evolutions are not always concordant for all cases [1, 2]. Thus, the similarities in morphology and behavior may originate shared features among genetically unrelated species occurring in similar environments; conversely, different features may be present among genetic sister taxa thriving under different environments. This phenotypic plasticity has already been described for other bird species [3, 4]. Thus, some genetic and/or phenotypic traits may not correlate with the evolutionary histories of the birds. On the other hand, mitochondrial cytochrome b gene (cyt-b) sequencing has been widely used for molecular systematics studies. Cyt-b gene has been proved to be helpful for defining evolutionary relationships among distant or closely related taxa [5, 6].

Crimson-winged Finch has the widest distribution among that of *Rhodopechys* “Arid-Zone” finches [7, 8]. It thrives throughout deserts/ semi-desert marginal areas in North Africa (Sahara), Arabian Peninsula and Central Asia deserts [7] Fig. (1).

Rhodopechys sanguineus thrives discontinuously across all desert strip margins and mountains that go from Africa to China. Its presence in Atlas Mountains may be due to winter migration, but it has been recorded in breeding season [7]. It seems that this species is mostly an altitudinal migrant although in China it moves from breeding area in East China range to North Central China in winter. It is found in Lebanon, Mount Hermon (Israel) and probably Syria; strikingly it is also found in Caucasus Mountains, North Ossetia and Kashmir-Indian area (Chitral and Ladakh) [7].

*Address correspondence to this author at the Department of Immunology, School of Medicine, University Complutense, The Madrid Regional Blood Center, Pabellón 5, planta 4. Ave. Complutense s/n, 28040 Madrid, Spain; Tel: +34606993161; E-mail: amaiz@med.ucm.es

Birds included within the *Carduelini* tribe (Genera *Rhodopechys*, *Carpodacus* and *Leucosticte*) have been shown to belong to the same evolutionary radiation by using molecular phylogenetic analyses [8]. Our phylogenetic analyses based cytochrome b gene (cyt-b) nucleotide sequences show that some of these birds (*Rhodopechys mongolica*, *Rhodopechys githaginea* and *Carpodacus nipalensis*) do not go in dendrograms with their respective phenetically defined allies and/or classical Genus classification [7]. This newly defined group of birds lives in both hot and cold *arid areas* and are phenetically distinct, probably because of their adaptation to diverse extreme environments. Both maximum likelihood and Bayesian inference dendrograms support the existence of this new group which seems to have appeared about 14 million years ago [8].

In the present work, we study the phylogeny of Crimson-winged Finch (*Rhodopechys sanguineus*) in order to complete the split family and new classification of *Rhodopechys* and “Arid-Zone” finches included species [8, 9, 10]. Our suggested taxonomic revisions are detailed at the end of Discussion section.

MATERIALS AND METHODS

Bird Samples, DNA Extraction and PCR Amplifications

Sixty two species of songbirds (order *Passeriformes*, suborder *Passeri*) have been included for study (see Table 1). They belong to the tribe *Carduelini*. *Rhodopechys sanguineus* was DNA-sequenced (four different birds) and analyzed in the present work together with relevant *Carduelini* species. Other sequences used in this analysis were retrieved from the GenBank (see Table 1) [8, 9, 11-13]. DNA extraction was done using a standard protocol [14]. Amplification and sequencing of cyt b gene 924 base pairs (bp) was performed as previously described [9].

CURRICULUM VITAE

Preparación académica

- Licenciada en Biología por la Universidad Autónoma de Madrid curso 2009-2010.
- Especialización: Itinerario biología celular y genética.
- Posgrado en Ciencias Biomédicas por la Universidad Complutense de Madrid.

Participación en proyectos I+D

- Periodo 2009-2011. Estudio genético-epidemiológico de Amerindios inmigrantes en Madrid: genes de obesidad y de histocompatibilidad. Departamento de Microbiología I (Inmunología), Universidad Complutense de Madrid. Becario predoctoral. Investigador Principal: Dr. Antonio Arnaiz Villena.
- Periodo 2011-2014. Genes de trasplante no clásicos (HLA-E, -F, -G, -DMB y MIC), y de obesidad en Amerindios inmigrantes a España: medicina preventiva, trasplante y farmacocinética. Entidad financiadora: Fondo de Investigaciones Sanitarias (FIS), Instituto de Salud Carlos III. Universidad Complutense de Madrid. Becario predoctoral. Investigador Principal: Antonio Arnaiz Villena.
- Periodo 2012-2014. Genes de obesidad e histocompatibilidad no clásicos (HLA-E, -F, -G) MIC y HLA-DMB en grupos Amerindios de América inmigrantes a Madrid: medicina preventiva, trasplante y farmacocinética. Entidad financiadora: Fundación Mutua Madrileña Automovilística. Universidad Complutense de Madrid. Becario predoctoral. Investigador Principal: Antonio Arnaiz Villena.

Publicaciones científicas

- **Gómez-Prieto P, Reguera R, Parga-Lozano C, Moreno E, Rey D, Areces C, Arnaiz Vilena A.** *HLA-G, -F y -E: Polymorphism, Function and Evolution.* Capítulo 9 en *HLA complex in Biology and Medicine: A resource book.* Editor: N. Mehra. Jaypee Brothers Medical Publishers Ltd, New Delih, India. 2009.
- **Carlos Balmori Boticario, Cristina Areces Viña, Alberto Pacheco Castro, Marta San Celestino Carchenilla, Juan Antonio García Velasco.** *Impact of an antioxidant complex supplementation on sperm DNA fragmentation in infertile men.* Rev Int Androl. 2010;8(3):107-13.
- **Arnaiz-Villena A, Parga-Lozano C, Moreno E, Areces C, Rey D, Gomez-Prieto P.** *The Origin of Amerindians and the Peopling of the Americas According to HLA Genes: Admixture with Asian and Pacific People.* Current Genomics, 11 (2): 104-113, 2010.
- **Vargas-Alarcón G, Granados J, Pérez-Hernández N, Rodríguez-Pérez JM, Canto-Cetina T, Coral-Vázquez RM, Areces C, Gómez-Prieto P, Arnaiz-Villena A.** *HLA-class II genes in Mexican Amerindian Mayas: relatedness with Guatemalan Mayans and other populations.* Immunol Invest.;40(1):101-111, 2011

- **Arnaiz-Villena A, Areces C, Gómez-Prieto P, Parga-Lozano C, Moreno E, Abd-El-Fatah-Kalil S, Rey D.** *The peopling of the Americas: a complex issue for Amerindian, Na-Dene, Aleut and Eskimo first inhabitants.* International Journal of Modern Anthropology, 3:65-79, 2010.
- **Diego Rey, Pablo Gomez-Prieto, Carlos Hernando Parga-Lozano, Cristina Areces, Patricia Peñaranda, Luz Barbolla, Antonio Arnaiz-Villena.** Genetic HLA differences between American Amerindians Na Dene, Eskimo and Aleuts Italia, Tissue Antigens ISSN: 0001-2815, 2010 vol:75 fasc: 5 págs: 499 - 499 .
- **Carlos Hernando Parga-Lozano, Pablo Gomez-Prieto, Diego Rey, Cristina Areces, Luz Barbolla, Sedeka Abs-El-Fatah, Patricia Peñaranda, Antonio Arnaiz-Villena.** America first inhabitants (Amerindians) show HLA gene flow with Australian Aborigines, other Pacific Islanders and Asian populations Italia, Tissue Antigens ISSN: 0001-2815, 2010 vol:75 fasc: 5 págs: 499 – 500.
- **Arnaiz-Villena A, Ruiz-del-Valle V, Reche P, Gomez-Prieto P, Lowry E, Zamora J, Areces C, Rey D, Parga C, Serrano-Vela JI.** Songbirds conserved sites and introns size of MHC class I molecules reveal a unique evolution in vertebrates. The Open Ornithology Journal, 3: 156-165, 2010.
- **Rey D, Areces C, Parga-Lozano C, Gomez-Prieto P, Crawford MH, Arnaiz-Villena A.** HLA genes variation in populations of the Aleutian Islands. Human Biology, 82 (5-6): 737-744, 2010.
- **Silvera C, Vargas-Alarcon G, Areces C, Rey D, Parga-Lozano C, Gomez-Prieto P, Barbolla L, Arnaiz-Villena A.** HLA genes in Wayu Amerindian from Colombia. Immunological Investigations, 40: 92-100, 2011.
- **Parga-Lozano C, Rey D, Gomez-Prieto P, Areces C, Moscoso J, Abd-El-Fatah-Khalil S, Arnaiz-Villena A.** HLA genes in Amerindian immigrants to Madrid: epidemiology and a virtual transplantation waiting list. Molecular Biology Reports, 38 (4): 2263-2271, 2011.
- **Martinez-Laso J, Montoya F, Areces C, Moscoso J, Silvera C, Rey D, Parga-Lozano C, Gomez-Prieto P, Enriquez de Salamanca M, Arnaiz-Villena A.** HLA in Jaidukama: Amerindian secluded Colombian population with new haplotypes and Pacific-shared alleles. Molecular Biology Reports, 38(6): 3689-3701, 2011.
- **Arnaiz-Villena A, Abd-El-Fatah S, Granados-Silvestre MA, Parga-Lozano C, Gomez-Prieto P, Rey D, Areces C, Peñaranda P, Menjivar M, Rodriguez-Perez JM, Granados J, Vargas-Alarcon G.** Human Leukocyte Antigen-DRB1 class II genes in Mexican Amerindian Mazahuas: genes and language do not correlate. Genetic Testing and Molecular Biomarkers, 15 (1-2): 97-102, 2011.
- **Arnaiz-Villena A, Areces C, Ruiz-del-Valle V.** *El Origen de los Canarios.* Ornitología Práctica, 53:3-11, 2012.
- **Longás J, Martínez-Laso J, Rey D, Areces C, Casado EG, Parga-Lozano C, Luna F, de Salamanca ME, Moral P, Arnaiz-Villena A.** Las Alpujarras region (South East Spain) HLA genes study: evidence of a probable success of 17th century repopulation from North Spain. Molecular Biology Reports, 39:1387-1394, 2012.
- **Arnaiz-Villena A, Fernández-Honrado M, Areces C, Arribas I, Coca C, Enriquez-de-Salamanca M, Parga-Lozano C, Abd-El-Fatah S, Rey D.** Amerindians normalized waist

circumference and obesity diagnosis standardized by biochemical and HLA data. *Molecular Biology Reports*, 39(4): 4875-4878, 2012.

- **Rey D, Fernandez-Honrado M, Areces C, Algora M, Abd-El-Fatah S, Enriquez-de-Salamanca M, Coca C, Arribas I, Arnaiz-Villena A.** Ameridians show no association of PC-1 gene *Gln121* allele and obesity: a thrifty gene population genetics. *Molecular Biology Reports*, 39(7): 7687-7693, 2012.
- **Rey D, Areces C, Enríquez-de-Salamanca M, Parga-Lozano C, Abd-El-Fatah S, Fernandez M, Arnaiz-Villena A.** Los primeros pobladores de América y sus relaciones con las poblaciones del Océano Pacífico según los genes HLA. *Inmunología*, 31:83-91, 2012.
- **Arnaiz-Villena A, Areces C, Rey D, Enriquez-de-Salamanca M, Alonso-Rubio J, Ruiz-del-Valle V.** Three different North American Siskin/Goldfinch evolutionary radiations: Pine Siskin green morphs and European Siskin in America. *The Open Ornithology Journal*, 5:73-81, 2012.
- **Arnaiz-Villena A, Areces C, Enríquez-de-Salamanca M, Abd-El-Fatah Khalil S, Arribas I, Coca C, Fernández-Honrado M, Algora M, Rey D.** Amerindians show no association of *PPAR-γ2* gene *Ala12* allele and obesity: an "unthrifty" variant population genetics. *Molecular Biology Reports*, 40(2): 1767-1774, 2013.
- **Arnaiz-Villena A, Rey D, Enríquez-de-Salamanca M, Abd-El-Fatah Khalil S, Arribas I, Coca C, Fernández-Honrado M, Algora M, Areces C.** Amerindians show association of obesity with *adiponectin* gene *SNP45* and *SNP276*: population genetics of a food intake and "thrifty" gene. *Molecular Biology Reports*, 40(2): 1819-1826, 2013.
- **Arnaiz-Villena A, Enríquez-de-Salamanca M, Areces C, Abd-El-Fatah Khalil S, Fernández-Honrado M, Rey D.** *HLA-G*01:05N* null allele in Mayans (Guatemala) and Uros: evolution and population genetics. *Human Immunology*, 74(4): 478-482, 2013.
- **Rey D, Parga-Lozano C, Moscoso J, Areces C, Enríquez-de-Salamanca M, Fernández-Honrado M, Abd-El-Fatah-Khalil S, Alonso-Rubio J, Arnaiz-Villena A.** *HLA genetic profile Mapuche*) Amerindians from Chile. *Molecular Biology Reports*, 40(7): 4257-4267, 2013.
- **Rey D, Areces C, Alonso-Rubio J, Enriquez-de-Salamanca M, Abd-El-Fatah-Khalil S, Bendikuze N, Fernandez-Honrado M, Barbolla L, Martin-Villa JM, Arnaiz-Villena A.** *HLA in Georgians (Caucasus) and their relationship with Eastern Mediterraneans.* *Molecular Biology Reports*, 40(10):5523-5530, 2013.
- **Rey D, Amirzargar A, Areces C, Enriquez-de-Salamanca E, Marco J, Abd-El-Fatah-Khalil S, Fernandez-Honrado M, Muñiz E, Martin-Villa JM, Arnaiz-Villena A.** *Gorgan (Turkmen in Iran) HLA genetics: transplantation, pharmacogenomics and anthropology.* *Immunological Investigations*, doi: 10.3109/08820139.2014.936938, 2014.
- **Arnaiz-Villena A, Vargas-Alarcón G, Areces C, Enriquez-de-Salamanca E, Abd-El-Fatah-Khalil S, Fernandez-Honrado M, Marco J, Martin-Villa JM, Rey D.** *Mixtec Mexican Amerindians: an HLA alleles study for America peopling, pharmacogenomics and transplantation.* *Immunological Investigations*, doi: 10.3109/08820139.2014.926369, 2014.
- **Areces C, Ruiz-del-Valle V, Rey D, Enriquez-de-Salamanca M, Arnaiz-Villena A.** *Crimson-winged Finch*) is included in "Arid-Zone" *Carduelini* Finches by mitochondrial DNA and Bayesian phylogeny. *In press*, 2014.

Comunicaciones a congresos

- V Congreso Europeo de Andrología. Roma, 2008.
- European Society of Human Reproduction and Embryology (ESRHE). Amsterdam, 2009.
- Congreso AETEL. Palma de Mallorca, 2009.
- 2º Congreso Internacional de Instituto Valenciano de Infertilidad (IVI). Madrid, 2009.
- II Congreso de la Sociedad Española de Biología Evolutiva (SESBE). Valencia, 2009.
- 25th International Ornithological Committee. Campos de Jordão, Brasil, 2009.
- “COS/AOU/SCO 2010 Joint Meeting, San Diego, EE.UU.2010.
- 24th European Immunogenetics and Histocompatibility Conference (EFI). Italia, Florencia, 2010.
- XXXV Congreso Nacional de la Sociedad Española de Inmunología (SEI). San Sebastián, España, 2010.
- XX Congreso Español de Ornitología TREMP. Lleida, 2010.
- Convención internacional de antropología. Anthropos 2011. La Habana, Cuba, 2011.
- 25th European Immunogenetics and Histocompatibility Conference (EFI). Prague, Czech Republic, 2011.
- XXXVI Congreso Nacional de la Sociedad Española de Inmunología (SEI). Pamplona, Spain, 2011.
- 129th Stated Meeting of the American Ornithologists' Union. Jacksonville, Florida, USA, 2011.
- III Congreso de la Sociedad Española de Biología Evolutiva (SESBE). Madrid, España, 2011.
- I Congreso Ibérico de Sistemática Animal. Madrid, España, 2012.
- 16th International HLA and Immunogenetics Workshop Report Conference, 26th European Immunogenetics and Histocompatibility Conference, 23rd British Society for Histocompatibility and Immunogenetics Conference. Liverpool, United Kingdom, 2012.
- 5th North American Ornithological Conference (NAOC-V). Vancouver, Canada, 2012.
- XXI Congreso Español y V Ibérico de Ornitología. Vitoria, España, 2012.
- 27th European Immunogenetics and Histocompatibility Conference. Maastricht, the Netherlands, 2013.
- XXXVII Congreso Nacional de la Sociedad Española de Inmunología. Salamanca, Spain, 2013.
- XXXVIII Congreso Nacional de la Sociedad Española de Inmunología. Badajoz, Spain, 2014.
- 56 Congreso de la Sociedad Española de Endocrinología y Nutrición. Valencia, Spain, 2014.

- 28th European Immunogenetics and Histocompatibility Conference. Stockholm, Sweden, 2014. *Tissue Antigens*, 85:5-164, 2014.