

# **Identificación de genes de inmunoglobulinas en el genoma bovino**

Estudiante            Daniela Megrian

Orientadores        Hugo Naya, Otto Pritsch

Unidad de Bioinformática - Institut Pasteur de Montevideo

Licenciatura en Bioquímica – Facultad de Ciencias – Universidad de la República

Junio de 2014

# Identificación de genes de inmunoglobulinas en el genoma bovino

Palabras claves: bovinos, sistema inmune, genes de inmunoglobulinas.

## Resumen

Dada la relevancia del estado de salud del ganado bovino en la productividad es importante conocer los diferentes componentes de su sistema inmunitario, tanto innato como adaptativo. En particular resulta muy importante conocer el tipo de respuesta de anticuerpos o respuesta humoral, tanto a nivel de las respuestas frente a diferentes agentes patógenos como para analizar la respuesta individual frente a vacunas.

En bovinos, el locus que contiene los segmentos génicos que codifican para las cadenas pesadas de las inmunoglobulinas (IGH) no se conoce completamente dado que se encuentra ausente en uno de los ensamblajes reportados del genoma bovino, y parcialmente en el otro. Las regiones constantes correspondientes a todas las clases humanas y murinas se mapearon en el cromosoma 21, pero la información sobre los segmentos VH, DH y JH no está sistematizada, es parcial y en algunos casos inconsistente.

Este trabajo propone sistematizar la información disponible sobre los genes de inmunoglobulina bovinos.

Se lograron identificar en los genomas bovinos reportados, un conjunto pequeño de segmentos VH en comparación con otras especies estudiadas. A este repertorio limitado se le agrega que solamente un tercio presenta las señales necesarias para su correcta recombinación, y que de aquellos segmentos potencialmente funcionales solamente 9 fueron identificados en un conjunto de bibliotecas de expresión (ADNc). El conjunto de segmentos génicos VH bovinos funcionales y expresados pertenecen a una sola familia filogenética homóloga a la familia VH2 humana.

Los segmentos DH identificados concuerdan con lo reportado en análisis previos donde se identificaron CDR3H de longitud variable, destacándose un conjunto particular que presenta longitudes excepcionalmente largas.

La cantidad de segmentos JH se condice con lo conocido para la mayoría de las

especies, y el patrón de expresión parece asemejarse a lo conocido para humanos.

Finalmente, este trabajo propone una nueva nomenclatura sistematizada para los segmentos VH, DH y JH identificados en el genoma bovino de manera análoga a la existente para humanos. Dado que la ubicación cromosómica del locus de cadena pesada se presenta fraccionado en varios cromosomas, se discute el posible origen de esta fragmentación y se proponen estrategias experimentales específicas para resolver este problema mediante la secuenciación completa de esta región.

## Índice

<b>A.</b>	Introducción	5
<b>A.1.</b>	El sistema inmune innato	6
<b>A.2.</b>	El sistema inmune adaptativo	8
<b>A.2.1.</b>	Mecanismos de la respuesta inmunitaria adaptativa	9
<b>A.2.2.</b>	Receptores del sistema inmune adaptativo	10
<b>A.2.3.</b>	El receptor de linfocitos T	10
<b>A.2.4.</b>	El receptor de linfocitos B	11
<b>A.2.5.</b>	El dominio constante de las Inmunoglobulinas	12
<b>A.2.6.</b>	El dominio variable de las inmunoglobulinas	14
<b>A.2.7.</b>	Formación de una inmunoglobulina	14
<b>A.2.8.</b>	Organización de los segmentos génicos de la cadena pesada de las inmunoglobulinas	16
<b>A.2.9.</b>	Organización de los segmentos génicos de la cadena liviana de las inmunoglobulinas	18
<b>A.2.10.</b>	Aspectos evolutivos de los genes VH	18
<b>A.2.11.</b>	Repertorio y generación de la diversidad de las Inmunoglobulinas	19
<b>A.2.12.</b>	Herramientas informáticas de análisis de los genes de Ig	22
<b>A.3.</b>	Análisis derepertorios de anticuerpos en bovinos	22
<b>A.3.1.</b>	Proyecto genoma bovino	22
<b>A.3.2.</b>	Segmentos génicos de inmunoglobulinas en el genoma bovino	23
<b>A.4.</b>	Problema Identificado	26
<b>A.5.</b>	Objetivo	26
<b>B.</b>	Materiales y métodos	27
<b>B.1.</b>	Obtención de genes de referencia	27
<b>B.2.</b>	Identificación de genes de inmunoglobulina en el genoma bovino	27
<b>B.3.</b>	Identificación de patrones necesarios para la expresión de los segmentos identificados	27
<b>B.4.</b>	Análisis filogenético de los segmentos VH identificados	28
<b>B.5.</b>	Identificación de mRNA correspondientes a inmunoglobulinas	28
<b>C.</b>	Resultados	29

<b>C.1.</b>	Obtención de genes de referencia	29
<b>C.2.</b>	Identificación de genes de inmunoglobulina en el genoma bovino	29
<b>C.2.1.</b>	Segmentos VH	29
<b>C.2.2.</b>	Segmentos DH	31
<b>C.2.3.</b>	Segmentos JH	31
<b>C.3.</b>	Identificación de patrones necesarios para la expresión de los segmentos identificados	31
<b>C.3.1.</b>	Segmentos VH	32
<b>C.3.2.</b>	Segmentos DH	34
<b>C.3.3.</b>	Segmentos JH	38
<b>C.4.</b>	Análisis filogenético de los segmentos identificados	38
<b>C.4.1.</b>	Segmentos VH	38
<b>C.4.2.</b>	Segmentos DH	40
<b>C.4.3.</b>	Segmentos JH	40
<b>C.5.</b>	Identificación de mRNA correspondientes a inmunoglobulinas	41
<b>D.</b>	Discusión y conclusiones	43
<b>E.</b>	Bibliografía	49

## **A. Introducción**

Los datos para el año 2012 demuestran la existencia de 11,4 millones de vacunos en el Uruguay. En el mismo año se faenaron más de 2 millones de ejemplares. La producción de carne bovina en Uruguay ocupa un 62% de la superficie del área productiva del país. La actividad pecuaria en Uruguay emplea una importante cantidad de mano de obra calificada y no calificada, directa o indirectamente, desde criadores hasta distribuidores.

Dadas las características de la producción de carne en Uruguay, el país ha entrado en mercados internacionales exigentes de calidad. Históricamente, el principal producto exportado por Uruguay ha sido la carne bovina congelada, habiendo sido recientemente desplazada por las habas de soja. De todas formas, en el año 2013 dos de los principales productos exportados fueron carne bovina y lácteos. Según el Anuario Estadístico Agropecuario 2013 del Ministerio de Ganadería, Agricultura y Pesca, y el Informe de Comercio Exterior de Uruguaya 2013 de Uruguay XXI, durante el período Enero-Diciembre de 2013 la exportación de carne bovina congelada reportó US\$ 936 millones, la carne bovina fresca US\$ 363 millones y la leche US\$ 365 millones. Estos datos implicaron una participación del 10%, 4% y 5% respectivamente de las exportaciones totales del país.

La producción pecuaria tiene importantes pérdidas económicas debido a los problemas de salud de los animales, como enfermedades respiratorias, digestivas o metabólicas. Estas pérdidas no están solamente determinadas por la muerte del animal, sino también por el gasto de recursos en alimentación, productos veterinarios y mejoramiento. De esta manera, la salud del ganado es determinante en la rentabilidad de la actividad pecuaria.

Los animales suelen estar sometidos a manejos, transportes, temperaturas extremas o presiones reproductivas que pueden tener efecto negativo sobre la salud del mismo. En el caso de la producción de leche, para obtener una performance rentable de la lactación, es un prerrequisito producir un ternero a intervalos regulares, siendo el intervalo parto-concepción el factor limitante. Desde la década del '60 a la actualidad se ha visto una disminución del 20 al 30% en la tasa de preñez, ya que la presión por selección para producción de leche ha afectado los índices reproductivos. El conflicto entre el desempeño productivo y el desempeño reproductivo está condicionado en gran parte por el manejo diferencial en el que se puede incidir especialmente durante el período de transición. Este período es en el que se concentran las enfermedades

metabólicas o tecnopatías de la producción de leche actual. Del equilibrio con que la vaca resuelva este proceso dependerá la capacidad de maximizar la producción de leche, evitar enfermedades metabólicas y asegurar la siguiente preñez. Se ha descrito una disminución transitoria de la respuesta inmune luego del parto, dada por el pasaje de neutrófilos al tracto reproductivo, y por cambios en la respuesta de linfocitos, afectando la producción de anticuerpos.

Además de los cambios metabólicos comandados por señales hormonales durante y luego de la preñez, la calidad de los caminos, las distancias recorridas para acceder al alimento y la disponibilidad de alimento con valor nutricional pueden representar un gran potencial traumático para la homeostasis del animal. La clínica individual carece de efecto preventivo: cuando las vacas se enferman, la leche se pierde, por lo que la intervención no es capaz de controlar las pérdidas económicas que representa una patología. Dado que el sistema inmunitario es el responsable del reconocimiento, resistencia y eliminación de agentes patógenos, este es fundamental para la performance del animal (Meikle, Cavestany et al. 2013).

El sistema inmunitario de un individuo está formado por un conjunto de órganos, tejidos, células y moléculas que actúan en forma colectiva y coordinada en respuesta a la presencia de estructuras extrañas, agentes patógenos causantes de enfermedades o estructuras propias transformadas, dañadas o envejecidas. El sistema inmune tiene como característica principal la capacidad de discriminar “lo propio” de “lo extraño”, generando potentes respuestas inmunes contra “lo extraño”.

El sistema inmune puede dividirse para su estudio en el sistema inmune innato y el sistema inmune adaptativo. El sistema inmune innato responde básicamente de la misma manera cada vez que se enfrenta a una infección. Los mecanismos de respuesta son generales para aquellas estructuras comunes presentes en agentes patógenos. La respuesta inmunitaria innata está siempre inmediatamente disponible para actuar, pero no lleva a una inmunidad duradera. Por el contrario, una respuesta inmunitaria adaptativa es específica, tiene la propiedad de “recordar” exposiciones repetidas al mismo agente patógeno, y aparece durante un lapso de vida de un individuo como una adaptación a la infección por ese agente patógeno.

### **A.1. El sistema inmune innato**

El reconocimiento y respuesta inicial frente a un agente patógeno es llevado a cabo por el sistema inmune innato. Este sistema responde de manera inmediata y repetida ante cualquier infección dado que los mecanismos moleculares efectores se

encuentran en el lugar previamente a la infección. Así la mayoría de los agentes infecciosos a los que un individuo sano se enfrenta normalmente no suelen causar una enfermedad perceptible ya que son reconocidos y eliminados rápidamente (de 0 a 96 horas en humanos).

El sistema inmunitario innato es la primera barrera contra la infección; bloquea la entrada de agentes infecciosos y elimina o limita el crecimiento de aquellos que tienen la capacidad de colonizar los tejidos del organismo.

El sistema inmunitario innato responde a estructuras comunes presentes en muchos microorganismos, por lo cual no discrimina con alta especificidad a los distintos agentes patógenos que ingresan al organismo. Estas estructuras comunes corresponden a moléculas simples conocidas bajo el nombre de Patrones Moleculares Asociados a Patógenos (del inglés, PAMPs), presentes en diversos microorganismos pero no en células propias. Las células del sistema inmune innato (macrófagos, neutrófilos y células dendríticas) expresan Receptores de Reconocimiento de Patrones (del inglés, PRRs), los cuales son inespecíficos o con especificidad amplia, pero discriminan eficazmente entre las células del hospedador y los patógenos. Los receptores del sistema inmune innato son codificados por la línea germinal, por lo tanto son restringidos a un repertorio limitado e invariable. Dado que los microorganismos pueden evolucionar con mayor rapidez que sus hospedadores, es lógico que los PRRs solo reconozcan estructuras moleculares que han permanecido sin cambios durante la evolución.

Los principales componentes del sistema inmune innato son las barreras físicas y químicas en los epitelios, las células efectoras (fagocitos mononucleares, neutrófilos, células natural killer (NK)) y proteínas como las citoquinas o las que forman parte del sistema de complemento. Estos componentes proveen al organismo de diversos mecanismos para responder a la infección.

Los agentes infecciosos vivos se pueden agrupar en virus, bacterias, hongos, protozoarios y helmintos. Estos pueden crecer en todos los compartimientos del organismo, donde los diferentes mecanismos de defensa del sistema inmunitario deberán combatirlos.

Los sitios de interacción entre los individuos y el ambiente externo son los epitelios (piel, mucosa gastrointestinal y respiratoria). Estos epitelios actúan como barrera eficaz para evitar la entrada de la mayoría de los agentes infecciosos, y tienen la capacidad de repararse rápidamente en caso de herida. Además, estos epitelios, así como algunos leucocitos producen péptidos con propiedades antimicrobianas, enzimas

degradantes potentes y sustancias químicas tóxicas. Si un agente infeccioso supera estas barreras y accede al tejido subyacente o a la circulación, se replicará y en general será reconocido por las células efectoras.

La gran mayoría de las células efectoras del sistema inmune innato provienen de un linaje mielóide. Estas se originan en la médula ósea y migran a los tejidos por la circulación sanguínea. Las células NK provienen de un progenitor linfóide.

Dentro de las células que provienen de un progenitor mielóide común se encuentran los macrófagos, neutrófilos y células dendríticas. Los macrófagos son células fagocíticas que se encuentran en los tejidos subepiteliales cuya función intrínseca es identificar, fagocitar y destruir agentes infecciosos. Los neutrófilos también son células fagocíticas pero migran al sitio de infección poco tiempo después de la entrada del agente infeccioso al organismo.

Las células dendríticas, al igual que los macrófagos y neutrófilos, captan materia por medio de fagocitosis, pero su función principal no es la eliminación de agentes infecciosos. Estas células tienen un rol importante en el vínculo entre el sistema inmunitario innato y el adaptativo.

## **A.2. El sistema inmune adaptativo**

El sistema inmune innato defiende de manera eficaz al individuo de agentes patógenos que presentan determinados patrones conservados en su composición, como oligosacáridos, peptidoglucanos, lipopolisacáridos de la membrana externa de las bacterias gram negativas, ARN de doble cadena de algunos virus o islas CpG desmetiladas. Estos PAMPs son comunes a muchos agentes patógenos y se han conservado durante la evolución. El sistema inmune adaptativo se superpone con el sistema inmune innato de forma tardía, pero tiene la capacidad de distinguir el agente infeccioso particular y dar una respuesta específica para combatirlo.

Los linfocitos específicos para antígeno son células que, al igual que las células NK, provienen de un progenitor linfóide común. Se diferencian en la capacidad de dar una respuesta inmunitaria contra una gran variedad de agentes infecciosos, dada la variabilidad en los receptores de antígeno que presentan en su superficie.

Los linfocitos se clasifican en linfocitos B y linfocitos T. Ambos se originan en la médula ósea, pero mientras los precursores de linfocitos B maduran allí mismo, los precursores de linfocitos T migran al timo. Dada su función, la médula ósea y el timo son denominados órganos linfoides centrales o primarios. Una vez alcanzada la maduración,

ambos tipos de linfocitos vírgenes pasan a la circulación sanguínea y se distribuyen en los órganos linfoides periféricos o secundarios (ganglios linfáticos, tejidos linfoides asociados a las mucosas, bazo, etc.), donde se inician las respuestas inmunitarias adaptativas. Los linfocitos recirculan entre la sangre y estos órganos hasta que encuentran su antígeno específico y son activados.

**A.2.1. Mecanismos de la respuesta inmunitaria adaptativa:** Para que se genere una respuesta inmunitaria adaptativa es necesaria la previa activación de las células dendríticas, que actúan como células presentadoras de antígeno. Sus receptores de reconocimiento de patrones, al unirse al antígeno, promueven la fagocitosis y degradación del agente patógeno. Además, las células dendríticas captan material extracelular no reconocido por sus receptores mediante macropinocitosis independiente de receptor.

Una vez internalizado el agente patógeno, las células dendríticas migran a los órganos linfoides periféricos. Allí maduran en células presentadoras de antígeno, presentando fragmentos del antígeno en su superficie y expresando moléculas coestimuladoras. También secretan citocinas, que tienen implicancias tanto en la inmunidad adaptativa como en la innata. Así las células dendríticas serán capaces de presentar el antígeno a los linfocitos T vírgenes, estos últimos luego proliferarán y se diferenciarán a linfocitos activados. Esta activación es necesaria para la posterior activación de linfocitos B.

Los linfocitos maduros indiferenciados que no reconocen su antígeno recirculan por los tejidos linfoides periféricos, accediendo a los mismos desde la sangre, y salen a través de los vasos linfáticos eferentes, regresando a la sangre mediante el conducto torácico. Este proceso se repite hasta que reconocen su antígeno específico o mueren. Los antígenos de agentes patógenos son transportados por las células presentadoras de antígeno desde los sitios de infección hacia el tejido linfoide periférico a través de los vasos linfáticos aferentes. Los órganos linfoides periféricos están compuestos de un almacén de células del estroma que emiten señales de supervivencia para los linfocitos. Este tejido también está especializado para atrapar las células presentadoras de antígeno y facilitar el inicio de respuestas inmunitarias adaptativas.

Cuando un linfocito es activado por el encuentro de su receptor específico con su antígeno, comenzará un período de proliferación y diferenciación para finalmente abandonar el órgano linfoide periférico como linfocito efector. Este linfocito regresa al torrente sanguíneo a través de los vasos linfáticos eferentes, y es transportado hacia los

tejidos donde actuará. Este es un proceso que se extiende en el tiempo, por lo cual una respuesta inmunitaria adaptativa a un antígeno al cual el individuo no se enfrentó antes recién se hará efectiva aproximadamente una semana después de la infección.

**A.2.2. Receptores del sistema inmune adaptativo:** A diferencia de los PRRs que reconocen estructuras diferentes compartidas por varios agentes patógenos, cada linfocito porta receptores de antígeno específicos para una estructura molecular única que difiere de los receptores de los otros linfocitos. Las estructuras reconocidas por PRRs son bastantes distintos de aquellos reconocidos por linfocitos. Los mecanismos del sistema inmunitario adaptativo permiten vencer las restricciones inherentes a la respuesta inmunitaria innata. La posibilidad de reconocer una diversidad de antígenos teóricamente infinita le permite al sistema inmunitario adaptativo una respuesta específica a cada agente patógeno.

A partir de los mecanismos genéticos de generación de la diversidad que suceden durante el desarrollo de los linfocitos es posible para un individuo generar un repertorio de receptores de linfocitos suficientemente extenso para asegurar que pueda reconocer casi cualquier estructura. Aquellos linfocitos que presentan receptores específicos para moléculas propias “auto-reactivas” se eliminan en una etapa temprana del desarrollo linfocitario. A partir del repertorio de linfocitos maduros, solamente aquellos linfocitos que se encuentren con el antígeno que se une a su receptor con alta afinidad proliferarán y se diferenciarán en células efectoras. Así es que un individuo solamente producirá anticuerpos contra los antígenos a los cuales haya estado expuesto. El proceso de proliferación se denomina expansión clonal, en ese momento la célula se divide y produce muchas descendientes idénticas que son capaces de secretar anticuerpos clonotípicos con una especificidad idéntica. Todo el mecanismo explicado anteriormente se denomina teoría de la selección clonal.

**A.2.3. El receptor de linfocitos T:** El receptor de linfocitos T (TCR) está compuesto de dos cadenas de similar tamaño, denominadas cadena  $\alpha$  y cadena  $\beta$ . El TCR solamente existe en forma de receptor de membrana. Cada una de estas cadenas está compuesta por una región variable y una región constante, siendo la combinación de las regiones variables de la cadena  $\alpha$  y cadena  $\beta$  la responsable de la generación de un sitio único para la unión de antígeno. Estos receptores no son capaces de unir directamente el antígeno, sino que solamente podrán reconocer fragmentos peptídicos cortos de antígenos proteicos unidos a glicoproteínas llamadas moléculas del complejo

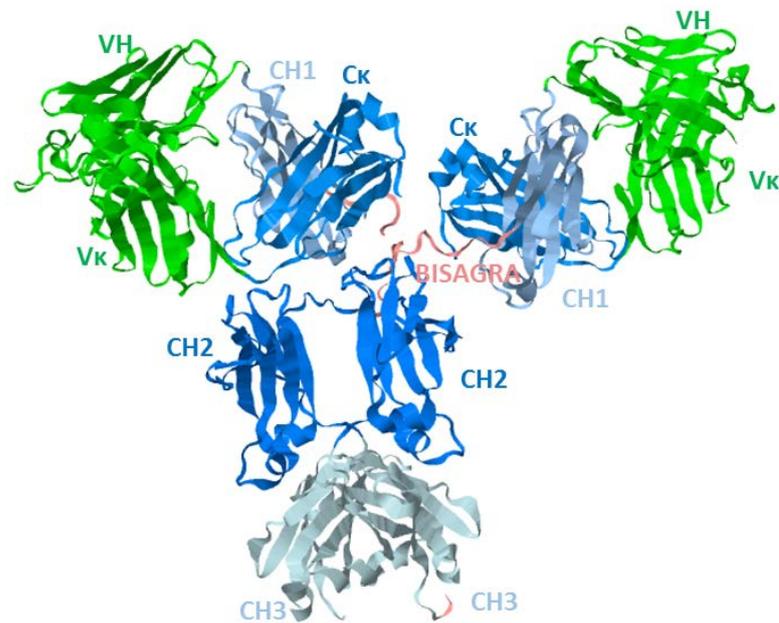
principal de histocompatibilidad (MHC) presentes en la superficie de las células presentadoras de antígeno. Este reconocimiento hará que se emita una señal al linfocito para activarse.

El MHC es un locus con un alto contenido en genes sumamente polimorfos, es decir que dentro de una población existen muchos alelos. Esto aumenta la probabilidad de que un mismo individuo exprese dos alelos distintos de la molécula, permitiendo que pueda unir una mayor variabilidad de péptidos antigénicos. Existen dos tipos principales de moléculas de MHC, las de clase I y las de clase II. Las primeras entran en contacto con antígenos citosólicos, mientras que las de clase II con antígenos extracelulares que hayan sido endocitados. Estructuralmente las moléculas de MHC presentan una hendidura a través de su superficie exterior, lugar donde unen los fragmentos peptídicos de los antígenos.

**A.2.4. El receptor de linfocitos B:** Las moléculas de reconocimiento de antígeno de los linfocitos B se denominan inmunoglobulinas (Ig). Estas moléculas se producen como receptores de membrana en la superficie de los linfocitos B (BCRs) o en forma de inmunoglobulinas o anticuerpos secretados al medio extracelular con la misma especificidad de antígeno para una misma célula. La función de los receptores de membrana de linfocitos B es reconocer y unirse a su antígeno, lo cual transmite una señal que causa la activación del linfocito, produciéndose la expansión clonal y secreción de anticuerpos específicos para el antígeno. Esta respuesta tiene como fin neutralizar la actividad del antígeno o facilitar su eliminación.

Las inmunoglobulinas son la forma secretada del receptor de antígeno de células B terminales diferenciadas, llamadas células plasmáticas. A diferencia de los receptores de linfocitos T, las inmunoglobulinas pueden unirse al antígeno de forma directa, reconociendo epítopes desplegados sobre la superficie del antígeno. Las inmunoglobulinas se unen a agentes patógenos o sus productos tóxicos en los espacios extracelulares del organismo. La secreción de anticuerpos es la principal función efectora de las células plasmáticas.

Las inmunoglobulinas se pueden esquematizar como una molécula en forma de Y (Figura 1). Cada molécula de anticuerpo está formada por dos cadenas pesadas idénticas y dos cadenas livianas idénticas, estas se ubican en la inmunoglobulina de manera simétrica formando la Y (Figura 2A). Los “brazos” de la Y corresponden a las dos regiones variables idénticas (dominio V) que son los sitios de unión a antígeno. La “pata” de la Y corresponde al dominio constante (dominio C) del anticuerpo, que es la región que determina la función efectora del anticuerpo.



**Figura 1. Estructura cristalográfica de una inmunoglobulina IgG.** La molécula de inmunoglobulina tiene un eje de simetría y está compuesta de dos cadenas pesadas idénticas (formada por los segmentos  $V_H$ - $CH_1$ - $CH_2$ - $CH_3$ ) y dos cadenas ligeras iguales (formada en este caso por los arreglos  $V_k$ - $C_k$ ). La combinación de la región variable de la cadena pesada y de la cadena liviana determina la especificidad de la inmunoglobulina (*verde*). La región constante determina la función efectora del anticuerpo y sólo adopta unas pocas formas distintas (isotipo IgG en este caso) (*azul y celeste*). Figura generada a partir de PDB ID 1HZH. (Saphire, Parren et al. 2001)

**A.2.5. El dominio constante de las Inmunoglobulinas:** El dominio constante solamente adopta unas pocas formas distintas determinadas por la clase. Las clases de inmunoglobulinas están definidas por los isotipos del dominio C de las cadenas pesadas. En mamíferos se conocen cinco clases, IgM, IgA, IgD, IgG e IgE codificadas por los genes  $\mu$ ,  $\alpha$ ,  $\delta$ ,  $\gamma$  y  $\epsilon$  respectivamente. Se han identificado isotipos adicionales en vertebrados inferiores mandibulados, como aves, reptiles, anfibios, peces óseos y peces cartilaginosos (Litman, Anderson et al. 1999). La clase IgM se encuentra estructuralmente conservada a lo largo de la evolución y se expresa en todos los vertebrados mandibulados. La clase IgD, a pesar de ser primitiva como la IgM, se encuentra ausente en aves y varios mamíferos como conejo, zarigüella y elefante. Sí se ha detectado en otros mamíferos, reptiles, anfibios y peces tanto óseos como cartilaginosos (Wilson, Bengten et al. 1997, Zhao, Kacs Kovics et al. 2002, Ohta and

Flajnik 2006, Zhao, Pan-Hammarstrom et al. 2006, Wei, Wu et al. 2009). Además se han identificado otras clases en distintos vertebrados como IgY en tetrápodos inferiores (Zhao, Rabbani et al. 2000, Lundqvist, Middleton et al. 2001, Wei, Wu et al. 2009), IgNAR e IgNARC en peces cartilaginosos (Greenberg, Avila et al. 1995, Dooley and Flajnik 2006), IgT/IgZ en trucha y zebrafish (Danilova, Bussmann et al. 2005, Hansen, Landis et al. 2005), IgX (mismo nombre dado para las IgNARC e IgW según la especie) e IgF en anfibios (Hsu, Flajnik et al. 1985, Zhao, Pan-Hammarstrom et al. 2006), e IgO en ornitorrinco (Zhao, Cui et al. 2009). Las aves además de IgD tampoco expresan las clases IgG e IgE pero sí la clase IgY (Lundqvist, Middleton et al. 2006, Choi, Kim et al. 2010).

La clase IgY, se cree que deriva de la duplicación de IgM. La clase IgY parece haber tenido un rol central en la evolución de las Ig en el linaje de los tetrápodos, dando lugar a la clase IgA (presuntamente a partir de la recombinación con IgM) y a IgG e IgE a partir de eventos de duplicación génica (Warr, Magor et al. 1995, Bengten, Wilson et al. 2000, Deza, Espinel et al. 2007). Estas características son compartidas para todas las especies de aves estudiadas, incluyendo el avestruz. Dado que el avestruz es una de las aves más primitivas vivientes, tiene sentido que estas características hayan surgido muy temprano durante la divergencia de las especies de aves. En algunas especies, como en algunas aves, se han encontrado inversiones en los genes que codifican estas regiones. En particular la orientación de transcripción del segmento génico que codifica IgA es opuesta a la del resto de las especies estudiadas (Zhao, Naude et al. 1996).

Cada uno de los isotipos se especializa en activar diferentes mecanismos efectores. La región que codifica las clases se denomina región constante ya que no está sometida a los mecanismos de recombinación generadores de diversidad de la región variable. Así es que las funciones de reconocimiento y efectoras están separadas de forma estructural en la molécula de anticuerpo. De la misma manera una molécula de anticuerpo tiene dos funciones, una es unirse de manera específica al antígeno del agente patógeno que desencadenó la respuesta inmunitaria, y una vez unido al antígeno, la otra función es reclutar células y moléculas para deshacerse del agente patógeno. Estas funciones efectoras son características de los anticuerpos ya que la región C del receptor de linfocitos B de membrana permanece insertada en la membrana de la célula.

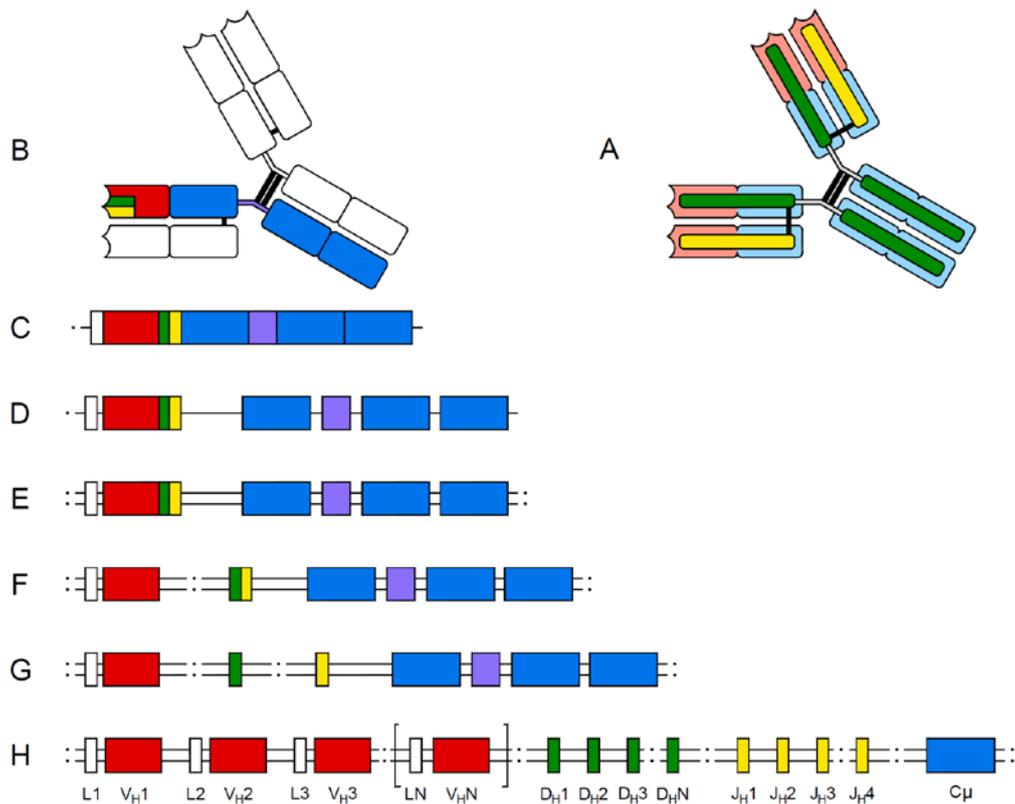
Una vez que se alcanza la especificidad por un antígeno, la función de la Ig puede ser modulada por cambio de clase de la región constante. El cambio de clase es una fase secundaria de diversificación que ocurre en los linfocitos B activados, que solamente involucra a la región C de la inmunoglobulina. En este proceso se reemplaza

la región C de la cadena pesada  $C_{\mu}$  original por una región C alternativa, aumentando la diversidad funcional del repertorio de inmunoglobulinas. A diferencia de los segmentos  $V_H$ ,  $D_H$  y  $J_H$  que se recombinan tempranamente en la diferenciación de linfocitos B, la región C puede expresarse en la progenie conforme éstas maduran y proliferan en el transcurso de una respuesta inmunitaria. El primer receptor de antígeno expresado por una célula B en una respuesta inmunitaria es el IgM. Más tarde durante la respuesta inmunitaria, la misma región V ensamblada puede expresarse en otras clases de anticuerpo de manera irreversible. El cambio de clase ocurre solamente después de que el linfocito B haya sido estimulado por antígeno.

**A.2.6. El dominio variable de las inmunoglobulinas:** Tanto las cadenas pesadas como las livianas están formadas por una región variable y una constante. Las regiones variables de una cadena pesada y de una cadena liviana se combinan para formar el sitio de unión a antígeno del anticuerpo (dominio variable o dominio V) (Figura 2A). La secuencia aminoacídica de esta región de unión varía mucho entre diferentes moléculas de anticuerpos, y es por ello que se denomina “región variable”. La estructura en forma de dominios V y C y el mecanismo de generación de diversidad para los receptores de linfocitos B es igual al empleado para los TCR. Esta estructura básica puede presentar una gran variabilidad de estructuras tridimensionales con capacidad de unirse en forma altamente específica a los diferentes tipos de antígenos.

**A.2.7. Formación de una inmunoglobulina:** Prácticamente cualquier sustancia puede ser el desencadenante de una respuesta de anticuerpos. Además, la respuesta a un solo antígeno puede comprender la producción de muchos anticuerpos diferentes cada uno con especificidades y afinidades diferentes (respuesta policlonal).

Cada variante de receptor no se puede codificar por completo en el genoma ya que requeriría un número de genes codificantes mayor al número de genes que hay en todo el genoma. Los genes que codifican las regiones variables de los receptores de linfocitos se heredan como sets de segmentos génicos (Hozumi and Tonegawa 1976).



**Figura 2.** A. Una molécula de anticuerpo está compuesta por dos cadenas pesadas idénticas (verde), y dos ligeras idénticas (amarillo). Cada una de las cadenas consta de una región constante (celeste) y una región variable (rosado). Cada una de las regiones variables está formada por una cadena ligera y una pesada, que conforman el sitio de unión a antígeno. B. La región variable de una cadena pesada se construye a partir de un segmento génico  $V_H$  (rojo), uno  $D_H$  (verde) y uno  $J_H$  (amarillo). Con azul se representa la región constante. G-C. En primer lugar, los segmentos génicos  $D_H$  y  $J_H$  se unen, después el segmento génico  $V_H$  se une a la secuencia  $DJ$  combinada. Un gen de la región C de la cadena pesada es codificado por varios exones y una región bisagra (violeta). Los exones de la región C, junto con la secuencia líder (blanco) son empalmados a la secuencia del dominio V durante el procesamiento del transcrito de ARN de la cadena pesada. H. El locus de la cadena pesada tiene una agrupación de segmentos génicos  $V_H$ , otra de segmentos  $D_H$  seguida de una agrupación de segmentos  $J_H$ . A continuación se ubica una agrupación de segmentos C, que comienza con el segmento  $C_\mu$ . Un segmento de cada agrupación se selecciona para formar la cadena pesada de las moléculas de inmunoglobulina. Figura adaptada de Fig. 1-13, Fig. 4-2 y Fig. 4-4 (Murphy, Travers et al. 2008).

Durante el desarrollo de estas células, los segmentos génicos de la región variable se seleccionan y ensamblan de manera irreversible mediante el mecanismo de recombinación de ADN. Este mecanismo se conoce como reordenamiento génico. Una secuencia de la región V entonces estará formada por dos o tres tipos de segmentos génicos, cada uno de los cuales está presente en múltiples copias en el genoma de la línea germinal. Dado que la selección de un segmento génico de cada tipo ocurre de manera aleatoria es posible un gran número de posibles combinaciones y por ende la gran diversidad del repertorio de receptores. Se estima que en seres humanos este repertorio puede llegar a más de  $10^{11}$  moléculas de inmunoglobulinas con estructuras diferentes.

La diversidad que se puede generar en la región V de la molécula de inmunoglobulina tiene una dependencia básica en el número de segmentos génicos y la distinción en sus secuencias. Dada la multiplicidad de segmentos génicos en el ADN de la línea germinal, ninguno de estos es esencial. Esto reduce la presión evolutiva sobre cada segmento génico para permanecer intacto, acumulando mutaciones. Así es que no todos los segmentos génicos identificados son funcionales, cierta proporción ha acumulado mutaciones que no les permiten codificar proteínas funcionales. Estos se denominan pseudogenes. Por ejemplo, en humanos existen 87 genes  $V_H$  (de los cuales 46 son funcionales) (Berman, Mellis et al. 1988, Cook, Tomlinson et al. 1994).

Dada la existencia de diversos segmentos génicos en cada set, dicha selección generará para cada célula un gen singular para la región variable de las cadenas pesada y liviana (Figura 2H). Así es que a partir del reordenamiento de un número limitado de segmentos génicos somáticos es posible generar un gran número de proteínas diferentes. Si una combinación de segmentos es funcional, el reordenamiento se detiene y ese linfocito solo expresará esa especificidad de receptor. Como consecuencia de esto, toda la progenie de esa célula heredará genes que codifican la misma especificidad de receptor.

**A.2.8. Organización de los segmentos génicos de la cadena pesada de las inmunoglobulinas:** El dominio V de una inmunoglobulina es codificada por más de un segmento génico (Figura 2H). El primero de estos se denomina segmento génico  $V_H$  o variable y codifica los primeros 95 a 101 aminoácidos. Este segmento forma la mayor parte del dominio V. Dado que las inmunoglobulinas son proteínas extracelulares, cada segmento  $V_H$  es precedido por un exón que codifica un péptido líder (L) que dirige la proteína hacia las vías secretoras de la célula. El péptido líder presenta un intrón en su secuencia.

En el caso de la cadena ligera, el segundo segmento génico se llama segmento génico J o de unión, y codifica el resto del dominio. Así es que la unión de un segmento génico V con uno J crea un exón que codifica la región V de la cadena liviana completa.

Para producir un ARNm de cadena ligera dicho exón se une a la secuencia de la región C mediante corte y empalme de ARN luego de la transcripción; el segmento génico J<sub>L</sub> (segmento génico J de cadena liviana) del dominio V reordenado está separado del dominio C por un intrón corto.

En la cadena pesada existe además un tercer tipo de segmento génico llamado segmento génico D<sub>H</sub> o de diversidad, que se encuentra entre los segmentos V<sub>H</sub> y J<sub>H</sub>. Es decir que el dominio V de la cadena pesada se codifica en tres segmentos génicos. Entonces para formar un exón del dominio V de la cadena pesada completo, primero un segmento génico D<sub>H</sub> se recombina con un segmento génico J<sub>H</sub>, y luego un segmento génico V<sub>H</sub> se reordena con el DJ<sub>H</sub> (Figura 2G-B). Los segmentos génicos V<sub>H</sub>, D<sub>H</sub> y J<sub>H</sub> de la cadena pesada, están flanqueados por secuencias señalizadoras de la recombinación (RSSs) muy conservadas. Al comienzo de la recombinación, un par de RSSs son acercadas por los componentes RAG-1 y RAG-2 de la recombinasa, que introducen cortes doble-hebra entre las RSSs y los segmentos génicos codificantes. Luego, la vía “non homologous end joining” (NHEJ) se emplea para unir los extremos de los segmentos génicos para formar la unión codificante y la unión señal respectivamente (Bassing, Swat et al. 2002, Jung, Giallourakis et al. 2006). El ARN se procesa de la misma manera que para la cadena ligera.

El dominio de la inmunoglobulina codificado por los segmentos génicos V<sub>H</sub>, D<sub>H</sub> y J<sub>H</sub> contiene tres regiones determinantes de la complementariedad (CDRs). Estas regiones de la proteína son las que unen al antígeno. Los CDRs están separados por regiones framework (FRs) conservadas que le dan soporte estructural a la proteína (Kirkham, Mortari et al. 1992). El segmento génico V<sub>H</sub> codifica las regiones denominadas CDR1H, CDR2H, FR1, FR2 y FR3. La región CDR3H está codificada por el extremo 3' del segmento V<sub>H</sub> y la unión de los segmentos D<sub>H</sub> y J<sub>H</sub> (Tonegawa 1983).

Los segmentos génicos de inmunoglobulinas de mamíferos están organizados en tres loci, el IGH que codifica la cadena pesada (o cadena H) y los loci IGK e IGL que codifican dos tipos equivalentes de cadena liviana (cadenas κ y cadenas λ respectivamente). Cada uno de estos loci puede ensamblar una secuencia de dominio V completa. En algunos casos excepcionales, como en algunas subclases de IgGs de camélidos, o las IgNARs de tiburón solamente se utilizan cadenas pesadas (Hamers-Casterman, Atarhouch et al. 1993, Greenberg, Avila et al. 1995).

El loci IGH en mamíferos, anfibios y reptiles está organizado en forma de translocón, en el cual los segmentos génicos están ordenados como  $(V_H)_n-(D_H)_n-(J_H)_n-(C_H)_n$ . En algunas especies se ha identificado un patrón inusual en esta organización. En peces óseos los segmentos génicos que codifican la región IgZ/T (gen  $\zeta/\tau$ ) se encuentran ubicados entre los genes  $V_H$  y  $D_H$ , algo similar a lo que ocurre en el locus de los TCR $\alpha$  y TCR $\beta$  de ratón.

En peces cartilaginosos (vertebrados mandibulados filogenéticamente más primitivos), el loci IGH presenta una organización multicluster, donde múltiples clusters se agrupan en tándem. Cada cluster está ordenado como  $V_H-D_H-D_H-J_H-C_H$ .

**A.2.9. Organización de los segmentos génicos de la cadena liviana de las inmunoglobulinas:** Las cadenas livianas (o cadenas L) tienen una contribución considerable a la diversidad combinatoria de los anticuerpos al asociarse con las cadenas pesadas. Tanto en mamíferos como en reptiles solamente existen los tipos de cadena  $\kappa$  ( $\rho$  en anfibios, L1/L3 en teleósteos o tipo 3 en elasmobranquios) y  $\lambda$  (tipo III en anfibios o tipo 2 en elasmobranquios). Sin embargo, en anfibios y peces cartilaginosos y teleósteos se identificó un tercer tipo llamado cadena  $\sigma$  (L2 en teleósteos) (Haire, Rast et al. 2000, Criscitiello and Flajnik 2007, Qin, Ren et al. 2008). Un cuarto tipo,  $\sigma$ -cart, se encontró en elasmobranquios (Fleurant, Changchien et al. 2004).

Para el locus IGL se han identificado tres formas principales de organización de sus segmentos: multicluster (en elasmobranquios), minimalista (en aves) y en translocón (en mamíferos) (Zhang, Nie et al. 2003). Sin embargo no todos los tipos de cadena son funcionales en todas las especies, por ejemplo las aves solamente expresan cadenas  $\lambda$  (Sanders and Travis 1975, Magor, Higgins et al. 1994). Además, en pollo y pinzón cebrado se identificó solamente un segmento génico funcional  $\kappa$  y uno  $\lambda$ . Sin embargo, se encontraron varios pseudogenes, dejando la generación de diversidad en manos de la conversión génica intracromosómica (Parvari, Ziv et al. 1987, Das, Mohamedy et al. 2010).

**A.2.10. Aspectos evolutivos de los genes  $V_H$ :** Los genes  $V_H$  de mamíferos pueden ser clasificados en tres grandes clanes (denominados I, II y III) que han coexistido en el genoma por más de 350 Myr (Tutter and Riblet 1989, Kirkham, Mortari et al. 1992, Ota and Nei 1994). Una familia se define como tal cuando sus miembros comparten más de 80% de identidad nucleotídica.

Se ha estudiado que los segmentos génicos del clan II están presentes en todas

las especies de mamíferos estudiadas (Tutter and Riblet 1989, Johansson, Aveskog et al. 2002, De Genst, Saerens et al. 2006). En caballo se han identificado miembros de los tres clanes: familia 1 y 5 del clan I, familias 2,4, 6 y 7 del clan II, y familia 3 del clan III (Sun, Wang et al. 2010). En oveja, los genes  $V_H$  expresados pertenecen al clan II principalmente (Sun, Kacskovics et al. 1994, Dufour, Malinge et al. 1996). Además, los genes  $V_H$  en conejo, ornitorrinco, zarigüeya, cerdo y pollo se ha identificado que solo expresan una familia o familias muy relacionados pertenecientes al clan III (Currier, Gallarda et al. 1988, Reynaud, Dahan et al. 1989, Ota and Nei 1994, Sun, Kacskovics et al. 1994, Miller, Grabe et al. 1998, Johansson, Aveskog et al. 2002). En humanos y ratón, los segmentos  $V_H$  del clan II también son preferencialmente expresados durante el estadio fetal, sugiriendo que estos segmentos  $V_H$  tienen un rol esencial en el desarrollo de la respuesta inmunitaria (Alt, Blackwell et al. 1987, Tutter and Riblet 1989, Schroeder and Wang 1990).

#### **A.2.11. Repertorio y generación de la diversidad de las inmunoglobulinas:**

Durante la recombinación la diversidad del repertorio del receptor de linfocitos B se incrementa por la enzima desoxinucleotidiltransferasa terminal (TdT) que añade nucleótidos no codificados por una plantilla (nucleótidos N) en las uniones entre segmentos génicos reordenados. Durante la recombinación, la acción de las proteínas del complejo RAG crea una horquilla de ADN en los extremos de los segmentos seleccionados. Estas horquillas son con frecuencia escindidas asimétricamente por un complejo formado por la proteína Artemis. La cadena más corta se debe prolongar con nucleótidos complementarios a la cadena más largas antes de la ligadura de los dos segmentos. Estos nucleótido añadidos según la plantilla se denominan nucleótidos P (palindrómicos). Otro mecanismo de diversidad de la unión es la adición aleatoria de hasta 20 nucleótidos no codificados por una plantilla denominados nucleótidos N. La adición de nucleótidos P y N en los sitios de recombinación puede introducir desplazamiento del marco de lectura, lo que podrían generar codones de terminación. La acción de la enzima TdT contribuye especialmente con la diversidad de la denominada región CDR3H (Alt and Baltimore 1982, Tonegawa 1983). Los otros mecanismos de generación de diversidad son la asociación combinatorial al azar entre una cadena H y una cadena L en la formación de la molécula de anticuerpo (de Wildt, Hoet et al. 1999), la hipermutación somática y la conversión génica (Weill and Reynaud 1996, Flajnik 2002).

La **hipermutación somática** ocurre luego de que los linfocitos B han sido activados por su antígeno, o sea que opera sobre los genes de inmunoglobulina ya

ensamblados. Este proceso provoca mutaciones puntuales por transición (T por C o A por G) mediadas por la enzima desaminasa de citidina inducida por activación (AID). Este proceso altera las secuencias de aminoácidos en la inmunoglobulina, siendo favorecida la supervivencia de aquellas clonas que produzcan mayor afinidad por el antígeno. Pruebas actuales sugieren que la AID actúa como desaminasa de citidina de ADN, transformando los residuos de citidina de los genes de inmunoglobulina en uridina. El residuo uracilo producido puede ser el sustrato para la enzima de reparación por escisión de bases de uracilo-DNA-glucosilasa (UNG) que elimina la base pirimidina para formar un sitio abásico. La replicación sin molde a través del sitio puede generar mutaciones por transición o por transversión. Al incrementar considerablemente el daño en el ADN, aumenta en gran medida la probabilidad de que este daño sea reparado de modo incorrecto, provocando una mutación. Estas mutaciones están distribuidas en toda la región variable, pero no de manera totalmente aleatoria, ya que se conocen “puntos calientes” en los que se concentran las mutaciones. La hipermutación somática en genes funcionales tiende a alterar la estructura básica de las inmunoglobulinas. Estas son luego seleccionadas de forma negativa porque el proceso ocurre en el centro germinal donde clonas de linfocitos B compiten entre sí por la interacción con el antígeno. Sin embargo, algunas mutaciones favorecen la unión de la inmunoglobulina al antígeno. Los linfocitos que expresan estos receptores son seleccionados para madurar y convertirse en linfocitos secretores de anticuerpos.

La **conversión génica** es un proceso por el cual secuencias cortas del gen de la región V reordenado son reemplazadas por secuencias provenientes de un pseudogen. Este proceso también estaría mediado por la acción de las proteínas AID y UNG. Una vez eliminada la base pirimidina, la endonucleasa apúrica/apirimídica 1 (APE1) puede escindir el resto del residuo, produciendo una mella en una sola cadena del ADN, en el sitio de la citosina original. Se cree que estas mellas son las señales que inician un proceso de reparación dirigido por homología en el cual un segmento génico homólogo se usa como molde para la replicación de ADN que repara al segmento. Eventos repetidos de conversión génica puede continuar diversificando el repertorio. De forma análoga a lo que sucede durante la hipermutación somática, algunas de estas conversiones harán que el segmento ya no sea funcional, mientras que otras favorecerán la interacción de la inmunoglobulina con el antígeno.

En la mayoría de la especies, el locus  $V_H$  contiene múltiples segmentos génicos  $V_H$  funcionales, múltiples segmentos  $V_H$  pseudogénicos, así como también múltiples segmentos  $V_H$  con secuencia codificante pero carentes de sitios de splicing o señales de recombinación (RSSs). El número total de segmentos  $V_H$  varía considerablemente

de una especie a otra; se han identificado alrededor de 160 en roedores, pero menos de 20 en ungulados como vacas y ovejas. Mientras tanto, la cantidad de segmentos D y J es bastante menor. Así es que el repertorio de VDJ en rumiantes, cerdo, oveja y pollo es menor que en otras especies estudiadas. En las especies como estas, donde el repertorio de genes de inmunoglobulinas es limitado, la conversión génica e hipermutación somática tienen un papel importante en la generación de diversidad post-rearreglo génico (Zhao, Jackson et al. 2006, Verma and Aitken 2012). Además, en estos casos, generalmente todos los segmentos V<sub>H</sub> pertenecen a una sola familia génica, lo cual no promueve la diversidad. En pollos la conversión génica se encarga de diversificar el repertorio primario de inmunoglobulinas. Esto sería necesario dado que sólo 1 segmento V<sub>H</sub> y uno V<sub>L</sub> pueden rearrreglarse. La conversión génica somática tiene lugar en la Bursa de Fabricius de pollos jóvenes, donde se reemplazan regiones de los exones de cadena pesada y cadena liviana con secuencias derivadas de pseudogenes ( $\psi$ ) (Thompson and Neiman 1987, Reynaud, Dahan et al. 1989, Carlson, McCormack et al. 1990, Thompson 1992). Un proceso similar sucede en conejos, donde la mayoría de los linfocitos B rearreglan solamente el segmento V<sub>H</sub> ubicado en 3', el cual luego es modificado por conversión génica (Becker and Knight 1990, Knight and Winstead 1997).

En la oveja, el cerdo y la zarigüeya, el repertorio de segmentos V<sub>H</sub> funcionales está formado por algunos pocos segmentos V<sub>H</sub> similares y pertenecientes al mismo clan. Es decir que el repertorio está limitado tanto en número como en diversidad. En cerdo se han identificado alrededor de 30 segmentos V<sub>H</sub> pertenecientes a la misma familia (Sun, Kacs Kovics et al. 1994), cinco segmentos D<sub>H</sub> y cinco J<sub>H</sub>. Solamente dos segmentos D<sub>H</sub> y uno J<sub>H</sub> son funcionales (Butler, Sun et al. 1996, Eguchi-Ogawa, Wertz et al. 2010).

El análisis estructural de los genes de inmunoglobulinas indica que los genes V<sub>H</sub> son el resultado de la duplicación de genes en tandem en el contexto de la conversión génica. Dado que solo algunos de estos segmentos V<sub>H</sub> contribuyen sustancialmente al repertorio de anticuerpos, se mantiene que la poligenia podría ser un vestigio de tiempos anteriores a cuando los procesos somáticos se hicieron lo suficientemente evolucionados para generar un repertorio de anticuerpos (Butler and Wertz 2012).

#### **A.2.12. Herramientas informáticas de análisis de los genes de Ig.**

El sitio web IMGT (the international ImMunoGeneTics information system, <http://imgt.cines.fr>) recopila y permite el acceso a información estandarizada de proteínas que pertenecen al sistema inmune. En particular provee de secuencias nucleotídicas, aminoacídicas; mapas génicos; polimorfismos genéticos; especificidades

y estructuras 2D y 3D (Lefranc, Giudicelli et al. 2005).

Dentro de IMGT es posible trabajar con la base de datos LIGM-DB. Esta contiene todas las secuencias nucleotídicas de inmunoglobulinas y receptores de células T reportadas para varias especies de vertebrados, pero incluye secuencias de DNA genómico de la línea germinal (no rearrreglado), rearrreglado (gDNA) y cDNA (Giudicelli, Duroux et al. 2006). Es muy complejo trabajar con este tipo de información dados los mecanismos involucrados en la generación de la diversidad de inmunoglobulinas, como los rearrreglos de cientos de bases de ADN, recombinaciones, deleciones, inserciones e hipermutación somática.

La base de datos GENE-DB fue desarrollada para manejar la complejidad de las IG y TR, a partir del análisis de la información existente en LIGM-DB. Comprende todas las secuencias nucleotídicas de inmunoglobulinas y receptores de células T reportadas principalmente para humano y ratón. Define una secuencia de referencia para cada gen y alelo individual, y cada una de estas secuencias está identificada según su funcionalidad. Además cada gen está definido como V, D o J, y asociado a la especie a la cual pertenece. Dada la estandarización y anotación de la información, este trabajo se realizó basándose en los datos aportados por esta base de datos. Actualmente GENE-DB se encuentra en desarrollo la expansión a otras especies de vertebrados. Al día de hoy, sobre bovinos solamente se encuentran disponibles genes de TRs (Giudicelli, Chaume et al. 2005).

### **A.3. Análisis de repertorios de anticuerpos en bovinos**

**A.3.1. Proyecto genoma bovino:** Al día de hoy existen depositados en bases de datos públicas solamente dos proyectos de secuenciación del genoma de la vaca. Ambos proyectos están basados en la misma “*raw data*” pero utilizan distintas herramientas bioinformáticas. Uno de estos, el denominado UMD\_3.1, fue llevado a cabo por el Center for Bioinformatics and Computational Biology (University of Maryland). El otro, Btau\_4.6.1, por el Cattle Genome Sequencing and Analysis Consortium.

Ambos proyectos del genoma bovino fueron secuenciados a partir de la combinación de los métodos “*Hierarchical Shotgun sequencing*” y “*Whole-genome shotgun sequencing*” (WGS). La primer estrategia se basa en la secuenciación de BACS, clones de insertos extensos de entre 100 y 150 Kbp. La segunda ensambla lecturas cortas sin mapeo previo.

La biblioteca de ADN de BACs se generó a partir de una muestra de sangre de un toro Hereford, denominado L1 Domino 99375. El ADN a partir del cual se obtuvieron las secuencias de WGS se extrajo de células blancas de una vaca hija de Domino (L1 Dominette 01449). El coeficiente de “*inbreeding*” es del 31%.

El proyecto Btau\_4.6.1 fue ensamblado con el sistema Atlas por el Medicine Human Genome Sequencing Center del Baylor College. Se ensamblaron 13.387 scaffolds, en un total de 2,940 Gb. El 90% de scaffolds se ubicaron en 31 cromosomas (29 autosomas, cromosoma X y cromosoma Y)(Liu, Qin et al. 2009).

El proyecto UMD\_3.1 consta de 36.82 millones de lecturas en 2,670 Gb, de los cuales 2 Gb (99%) han sido ubicados en 30 cromosomas (29 autosomas y cromosoma X). El 97% de las lecturas fueron secuenciadas por el Baylor College of Medicine. El genoma se ensambló con Celera Assembler 5.2 a partir de todas las lecturas disponibles en el NCBI Trace Archive. Las lecturas fueron ensambladas en 66.141 scaffolds de 40.86 Kbp de largo promedio (Zimin, Delcher et al. 2009).

A pesar de que en ambos proyectos los scaffolds no ubicados en cromosomas son una pequeña fracción del total (10% de los scaffolds para Btau\_4.6.1, 1% de las lecturas para UMD\_3.1) es probable que los loci de genes de inmunoglobulinas hayan presentado dificultad para el ensamblaje, dado que las muestras a partir de las cuales se secuenció el genoma fueron de sangre periférica donde se incluyen los linfocitos que presentan reordenamiento génico.

**A.3.2. Segmentos génicos de inmunoglobulinas en el genoma bovino:** El locus de la cadena  $\lambda$  fue ubicado en el cromosoma 17. Se identificaron 63 segmentos  $V\lambda$  (25 de los cuales serían potencialmente funcionales), 3 segmentos  $J\lambda$  y 5 segmentos  $C\lambda$ . Los segmentos  $V\lambda$  se pueden organizar en 8 familias filogenéticas, las cuales también agruparían la mayoría de los segmentos  $V\lambda$  identificados en ovinos. La mitad de los segmentos  $V\lambda$  identificados no pudieron ser asignados a una posición cromosomal específica. Dado que fueron identificados en contigs genómicos sin ubicar en el genoma, varios podrían corresponder a segmentos huérfanos. Por su parte, el locus de la cadena  $\kappa$  se ubicó en el cromosoma 11, identificándose 22 segmentos  $V\kappa$  (siendo 8 de ellos potencialmente funcionales), 3 segmentos  $J\kappa$  y 1 segmento  $C\kappa$ . Los segmentos  $V\kappa$  se pueden agrupar en 4 familias filogenéticas, que también conservarían el agrupamiento de los segmentos  $V\kappa$  identificados en ovinos. Este análisis de identificación de segmentos génicos de cadena liviana se basó en la versión Btau\_3.1 del genoma bovino, generada por el Cattle Genome Sequencing and Analysis

Consortium. En humanos y en ratón se identificaron significativamente más segmentos génicos funcionales de cadena liviana que en bovinos (el doble y el triple respectivamente). (Ekman, Niku et al. 2009).

El locus IGH no se conoce completamente dado que se encuentra ausente del ensamblaje Btau\_4.2, y parcialmente en el UMD\_3.1. Las regiones constantes correspondientes a todas las clases humanas y murinas han sido mapeadas en BTA21q23-24 (Tobin-Janzen and Womack 1992, Chowdhary, Fronicke et al. 1996, Zhao, Kacs Kovics et al. 2002, Zhao, Kacs Kovics et al. 2003). Un estudio identificó que el locus JH se localiza en BTA21 y contiene 6 genes de los cuales 2 son funcionales. Sin embargo analizando datos de transcriptómica solamente uno de ellos sería utilizado. Los otros 4 segmentos presentan secuencias divergentes del consenso en las RSSs y los sitios de splicing (Zhao, Kacs Kovics et al. 2003). Otro estudio identifica también 6 segmentos ordenados de manera similar a lo observado en oveja. De los 6 segmentos también clasifica como pseudogenes a 4 de estos. Encuentran un segmento en el cromosoma 11 que se rearregla a baja frecuencia y otro en el cromosoma 21 el cual se rearregla con alta frecuencia. Esta información difiere de lo esperado, ya que no se suelen expresar cadenas pesadas procedentes de ubicaciones genómicas distintas (Hosseini, Campbell et al. 2004). Se identificaron probablemente 7 segmentos  $D_H$  distintos (Koti, Kataeva et al. 2008). Otro elemento interesante es que se identificaron algunos segmentos  $D_H$  de longitud extremadamente larga, produciendo regiones CDR3H de hasta 62 aminoácidos (Saini, Allore et al. 1999, Saini and Kaushik 2002, Shojaei, Saini et al. 2003, Koti, Kataeva et al. 2008, Kaushik, Kehrl et al. 2009, Koti, Kataeva et al. 2010). También se identificó una secuencia  $\mu$ -like aparentemente funcional y un locus JH en BTA11q23 (Hayes and Petit 1993, Hosseini, Campbell et al. 2004, Kuroiwa, Kasinathan et al. 2009).

El conocimiento de segmentos  $V_H$  en rumiantes es muy limitado. Las primeras evidencias identificaron una sola familia, homóloga a la VH2 humana y a la Q52 murina, con aproximadamente 10 segmentos únicos, pero que presentan baja diversidad entre sí. Por Southern blotting se identificaron algunas familias más pero probablemente no funcionales ya que no se identifican en transcritos. También observaron que el patrón de sustitución nucleotídica en transcritos de bazo es consistente con el de diversificación por hipermutación somática, y que este proceso comenzaría más temprano que en otras especies estudiadas (Berens, Wylie et al. 1997, Sinclair, Gilchrist et al. 1997).

Más recientemente se identificaron 36 genes  $V_H$  en el genoma bovino, la mayoría de estos a partir del genoma UMD\_3.1, pocos de estos incluidos en Btau\_4.2. Siete de

estos genes fueron solamente identificados en los datos originales o high throughput data de un BAC clone. Los genes fueron identificados en los cromosomas 7 y 21. De los 36 genes que se encontraron 10 serían funcionales, con promotor conservado y exones intactos. Además todos los segmentos  $V_H$  identificados en el cDNA del íleon fetal alinearon con los segmentos  $V_H$  identificados. Estos 10 segmentos pertenecen a la familia descrita como VH1. Los segmentos pertenecientes a las familias VH2 e VH3 consisten solamente de pseudogenes. (Niku, Liljavirta et al. 2012). Otro estudio identifica también 36 genes  $V_H$  de los cuales 13 serían potencialmente funcionales. El análisis filogenético reveló 2 familias, a las que denominaron boVH1 y boVH2, donde boVH1 comprende a todos los segmentos funcionales (Walther, Czerny et al. 2013).

Si la cantidad y diversidad de segmentos  $V_H$  es así de limitada, el rearreglo sería incapaz de crear un repertorio con variabilidad significativa. Además, existe escasa contribución a la diversidad por la cadena liviana, ya que dichos loci también tienen un tamaño reducido (Ekman, Niku et al. 2009, Pasman, Saini et al. 2010). Inicialmente se propuso que el proceso de conversión génica podría ser el mecanismo principal de generación de la diversidad en el repertorio de Ig en bovinos, pero varios estudios han sugerido a la hipermutación somática como proceso de diversificación. Dado que en la conversión génica las secuencias que causan diversificación son tomadas de donantes pseudogénicos surgidos de duplicaciones o fusiones de segmentos  $V_H$ , se esperan modificaciones confinadas en el marco de lectura de la Ig. Sin embargo se han identificado sustituciones nucleotídicas tanto en el exón VDJ como en las regiones downstream no codificantes, sugiriendo que el mecanismo de hipermutación somática sería el predominante en el proceso de generación de la diversidad de los anticuerpos bovinos (Berens, Wylie et al. 1997, Saini, Hein et al. 1997, Sinclair, Gilchrist et al. 1997, Lopez, Perez et al. 1998, Kaushik, Shojaei et al. 2002, Das, Nozawa et al. 2008, Kaushik, Kehrl et al. 2009, Verma and Aitken 2012).

#### **A.4. Problema Identificado**

Dada la relevancia del estado de salud del animal en la productividad es importante conocer los diferentes componentes del sistema inmune bovino, tanto innato como adaptativo. En particular resulta muy importante conocer en el bovino el tipo de respuesta de anticuerpos o respuesta humoral, tanto a nivel de las respuestas frente a diferentes agentes patógenos como para analizar la respuesta individual frente a vacunas.

El gran repertorio de inmunoglobulinas con especificidades únicas ha hecho

complejo el estudio de la diversidad real generada a partir del potencial genómico. Para vaca se estima un potencial de  $\sim 1 \times 10^7$  especificidades únicas, no siendo posible el análisis exhaustivo con métodos tradicionales de secuenciación. Los avances recientes en secuenciación de alto rendimiento permitirán la secuenciación de estos repertorios en profundidad. Estos estudios a gran escala contribuyen al entendimiento sobre cómo el sistema inmunitario se enfrenta a infecciones, desarrollo de vacunas y determinar marcadores de riesgo, diagnóstico y pronóstico de enfermedades. Aunque los métodos de secuenciación han tenido un rol importante en la biología desde los '80, la secuenciación de repertorios de inmunoglobulinas es reciente, propiciado por el rápido desarrollo de técnicas y la disminución de costos.

Para poder analizar correctamente los resultados del análisis de repertorios por secuenciado masivo es imprescindible contar con una información sistematizada sobre el conjunto de segmentos génicos germinales que codifican para las dierentes cadenas de inmunoglobulinas.

#### **A.5. Objetivo**

El objetivo general de este trabajo es sistematizar la información disponible sobre los genes de inmunoglobulina bovinos. Primero recopilando la información relativa a los genes de inmunoglobulinas conocidos y el genoma bovino, y posteriormente analizando la información con herramientas disponibles, para finalmente reportar una lista de potenciales genes de inmunoglobulina bovinos.

## **B. Materiales y métodos**

### **B.1. Obtención de genes de referencia**

Con el objetivo de identificar los genes de inmunoglobulina en bovinos se obtuvieron todas las secuencias codificantes en formato FASTA para genes homólogos de otras especies presentes en la base de datos GENE-DB de IMGT depositadas en el sitio ftp de EBI (European Bioinformatics Institute, <ftp://ftp.ebi.ac.uk>). También se obtuvieron las secuencias de los segmentos VH reportadas por Niku, Liljavirta et al. 2012 y Walther, Czerny et al. 2013, y las de los segmentos JH reportadas por Zhao,

Kacskovics et al. 2003 y Hosseini, Campbell et al. 2004.

## **B.2. Identificación de genes de inmunoglobulina en el genoma bovino**

Para identificar potenciales genes de inmunoglobulina en el genoma bovino se realizó una búsqueda de homología con la herramienta blastn del paquete BLAST a partir segmentos génicos de inmunoglobulinas de otras especies (Tabla 1) contra los scaffolds ubicados y no ubicados en cromosomas de los dos proyectos de secuenciación del genoma bovino existentes. Las secuencias de los scaffolds se descargaron de NCBI (AC GJ060423:GJ063645 y GJ057137:GJ060422 para UMD\_3.1 y AC JH119571:JH126398 para Btau\_4.6.1) (Liu, Qin et al. 2009, Zimin, Delcher et al. 2009).

## **B.3. Identificación de patrones necesarios para la expresión de los segmentos identificados**

Los segmentos de inmunoglobulinas para ser funcionales deben tener adyacentes a ellos diversas señales, siendo particularmente importantes los sitios de unión a factores de transcripción en el promotor, sitios de splicing o secuencias de señal de recombinación.

A los efectos de identificar los RSS adyacentes a los segmentos génicos de inmunoglobulinas se utilizó la herramienta FUZZNUC de EMBOSS, permitiendo 5 diferencias nucleotídicas (mismatches) como máximo. Se utilizaron como patrón CACAGTGN(23)TCAGAAACC para segmentos  $V_H$ , GGTTTTTGAN(11-13)CACTGTGN(200)CACAGTGN(11-13)TCAAAAACC para segmentos  $D_H$ , y GGTTTTTGTN(23)CAATGTG para  $J_H$ . Estos patrones fueron determinados a partir del consenso de los RSS identificados en humanos depositados en IMGT. Para los segmentos  $V_H$  se buscó el patrón RSS en la región 3' de la secuencia, y para los segmentos  $J_H$  en la región 5' correspondiente. También se utilizó la herramienta FUZZNUC para identificar el patrón octamérico "ATGCAAAT" en el promotor, permitiendo 1 mismatch como máximo.

Se utilizó la herramienta FEX (<http://linux1.softberry.com/>) para identificar el codón de inicio (ATG) y los sitios de splicing en el péptido líder de los segmentos  $V_H$ .

#### **B.4. Análisis filogenético de los segmentos V<sub>H</sub> identificados**

Para el análisis filogenético de los segmentos V<sub>H</sub> se realizó un alineamiento múltiple con Clustal Ω a partir de un miembro de cada familia de segmentos génicos V<sub>H</sub> de humano y ratón en la base de datos GENE-DB de IMGT y los segmentos V<sub>H</sub> bovinos identificados. Como outgroup se utilizó la secuencia VH8-1 de *Oncorhynchus mykiss*. Se utilizó la herramienta JModelTest (<https://code.google.com/p/jmodeltest2/>) para la selección del modelo de sustitución nucleotídica (GTR+G+I). Se analizaron las secuencias con BEAUTi y BEAST ([http://beast.bio.ed.ac.uk/Main\\_Page](http://beast.bio.ed.ac.uk/Main_Page)) para la reconstrucción de la filogenia. Se analizaron los datos generados por BEAST con Tracer (<http://tree.bio.ed.ac.uk/software/tracer/>). Finalmente se generó y editó el árbol con la herramienta treeannotator de BEAST y FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

#### **B.5. Identificación de mRNA correspondientes a inmunoglobulinas**

Para analizar la expresión, y por ende la funcionalidad de los segmentos de inmunoglobulina identificados en el genoma bovino, se hizo una búsqueda de mRNAs en sangre y tejidos linfoides donde se encuentran grandes cantidades de inmunoglobulinas. Para esto se realizó un análisis de blastn utilizando como query los segmentos V<sub>H</sub> identificados en bovinos contra las secuencias de las librerías bovinas de sangre (Lib.19002), intestino (Lib.15490, Lib.17884, Lib.15436, Lib.17251, Lib.18995, Lib.22657, Lib.10204, Lib.10222, Lib.22658, Lib.10203, Lib.18988, Lib.22656, Lib.17250) y tejido linforeticular (Lib.19006, Lib.10111, Lib.18999, Lib.17578, Lib.22653, Lib.22654, Lib.17575, Lib.18994, Lib.19012, Lib.19020, Lib.22655) depositadas en la base de datos UniGene de NCBI.

### **C. Resultados**

#### **C.1. Obtención de genes de referencia**

A enero de 2013 se encontraban depositados en GENE-DB de IMGT 1127 segmentos V<sub>H</sub>, 162 D<sub>H</sub> y 59 J<sub>H</sub> de las especies: *Danio rerio*, *Homo sapiens*, *Macaca mulatta*, *Mus musculus*, *Oncorhynchus mykiss*, *Oryctolagus cuniculus* y *Rattus norvegicus* (Tabla 1).

**Tabla 1.** Cantidad de segmentos VH, DH y JH depositados en GENE-DB a enero de 2013 desglosados por especie.

	VH	DH	JH
<b>Danio rerio</b>	41	7	7
<b>Homo sapiens</b>	341	44	13
<b>Macaca mulatta</b>	23	32	8
<b>Mus musculus</b>	406	39	9
<b>Oncorhynchus mykiss</b>	58	9	7
<b>Oryctolagus cuniculus</b>	49	11	11
<b>Rattus norvegicus</b>	209	20	4

## **C.2. Identificación de genes de inmunoglobulina en el genoma bovino**

### **C.2.1. Segmentos VH**

Se identificaron 57 segmentos VH repartidos en 21 scaffolds de ambos genomas, de los cuales 35 corresponden a secuencias distintas (Tabla 2). Al trabajar con 2 genomas distintos esperamos que la información esté duplicada, por lo tanto, para identificar a 2 segmentos como el mismo se buscó una identidad de más de 98%, para descartar polimorfismos. De los 35 segmentos distintos, 33 se encuentran ubicados en cromosomas (VH1-1 al VH2-33, la nomenclatura se comentará en el apartado C.4.1.). Todos los segmentos ubicados en cromosomas pertenecen al proyecto UMD\_3.1, y 22 de estos también fueron obtenidos en scaffolds del proyecto Btau\_4.6.1. En el proyecto UMD\_3.1 también se pudo identificar un segmento sin ubicar en cromosoma (VH2-33). En el proyecto Btau\_4.6.1 se identificaron solamente 2 segmentos distintos a los identificados en el proyecto UMD\_3.1 (VH1-34 y VH1-35). Dentro de un mismo scaffold se identificaron segmentos tanto en la hebra positiva (VH2-17 al VH1-21) como en la negativa (VH2-22 al VH1-28).

**Tabla 2.** Segmentos VH identificados en el genoma bovino por homología con segmentos VH de otras especies.

ID	UBICACIÓN	PROYECTO	SCAFFOLD	HEBRA	
● VH1-1	Chr7	UMD_3.1	NW_003064413.1	-	
● VH1-2			NW_003064411.1	-	
● VH2-3			NW_003064299.1	-	
● VH3-4				-	
● VH2-5			NW_003064298.1	-	
● VH1-6				-	
● VH2-7			NW_003064297.1	-	
● VH2-8				-	
● VH3-9			NW_003064296.1	-	
● VH1-10				-	
● VH3-11			NW_003064295.1	+	
● VH2-12				+	
● VH1-13			NW_003064291.1	+	
● VH1-14				-	
● VH3-15			NW_003064290.1	+	
● VH3-16				-	
● VH2-17	Chr21		NW_003104538.1	-	
● VH3-18				-	
● VH3-19				-	
● VH3-20				-	
● VH1-21				-	
● VH2-22				+	
● VH1-23				+	
● VH1-24				+	
● VH3-25				+	
● VH3-26				+	
● VH2-27				+	
● VH1-28				+	
● VH1-29				-	
● VH2-30				NW_003104530.1	-
● VH3-31					-
● VH3-32				-	
● VH2-33	ChrUn	NW_003100762.1	+		
● VH1-2*	ChrUn	Btau_4.6.1	NW_001503306.1	+	
● VH2-7*			NT_183109.1	+	
● VH2-30*				+	
● VH3-31*				+	
● VH3-32*				+	
● VH3-16*				+	
● VH1-34				+	
● VH2-33*				+	
● VH3-4*				+	
● VH2-5*				+	
● VH3-9*				-	
● VH2-8*			NT_182449.1	-	
● VH1-13*				NW_001504477.2	-
● VH2-17*			NT_185907.1	-	
● VH2-12*			NT_182448.1	+	
● VH3-11*				+	
● VH3-20*			+		
● VH1-24*			+		
● VH1-35			-		
● VH3-25*			NT_185036.1	-	
● VH2-27*				-	
● VH1-28*			-		
● VH3-18*			NT_186922.1	+	
● VH3-19*				+	

Se indica en qué genoma se identificó, en qué scaffold y en cuál hebra del scaffold. En la columna “UBICACIÓN” se indica en qué cromosoma está ubicado el scaffold. En caso de los scaffolds no ubicados en cromosomas se indica “ChrUn”. Los segmentos indicados con “\*” corresponden a segmentos identificados en el genoma Btau\_4.6.1 repetidos en UMD\_3.1. La nomenclatura y colores se comentarán en el apartado C.4.1..

### **C.2.2. Segmentos DH**

Debido a la gran diversidad intrínseca en los segmentos DH, no fue posible identificar dichos segmentos a partir de búsquedas de homología con segmentos DH de otras especies. Por esta razón la estrategia de búsqueda se basó en la identificación de secuencias RSS lo cual se describe en la sección C3.2.

### **C.2.3. Segmentos JH**

Se identificaron 15 segmentos JH repartidos en 3 scaffolds de ambos genomas, de los cuales 9 corresponden a secuencias distintas (Tabla 3). Fueron identificados 5 segmentos previamente identificados por Zhao, Kacskovics et al. 2003 (JH1, JH2, JH-PS1, JH-PS2 y JH-PS4) y 5 previamente identificados por Hosseini, Campbell et al. 2004 (JH2, JH3, JH4, JH5 y JH6), de los cuales solamente uno es común. Todos los segmentos fueron ubicados en el cromosoma 8 a excepción del JH-7, JH-8 y JH-9 que no pudieron ser ubicados.

## **C.3. Identificación de patrones necesarios para la expresión de los segmentos identificados**

En este trabajo se definió como ORF a aquel segmento génico que tiene un codón ATG, sitios de splicing adecuados y al menos un marco abierto de lectura, y que no contiene codones de terminación o frameshifts en su secuencia, por lo que potencialmente podría traducirse en una proteína. Un segmento clasificado como ORF no se pudo considerar potencialmente funcional dado que no contiene una o más de las secuencias necesarias para su expresión. Para considerar a un segmento como potencialmente funcional, se tomaron en cuenta la existencia de RSS de longitud, identidad y ubicación adecuadas, y la presencia del octámero “ATGCAAAT” en el promotor. Las secuencias clasificadas como pseudogenes carecen de un ORF.

**Tabla 3.** Segmentos JH identificados en el genoma bovino por homología con segmentos identificados por Zhao, Kacs Kovics et al. 2003 y Hosseini, Campbell et al. 2004.

ID	UBICACIÓN	PROYECTO	SCAFFOLD	HEBRA	HIT
JH-1	Chr8	UMD_3.1	NW_003066919.1	-	JH-PS1 (Zhao)
JH-2				-	JH2 (Hosseini) / JH-PS2 (Zhao)
JH-3				-	JH3 (Hosseini)
JH-4				-	JH4 (Hosseini)
JH-5				-	JH5 (Hosseini)
JH-6				-	JH6 (Hosseini)
JH-1*	ChrUn	Btau_4.6.1	NW_001494075.1	-	JH-PS1 (Zhao)
JH-2*				-	JH2 (Hosseini) / JH-PS2 (Zhao)
JH-3*				-	JH3 (Hosseini)
JH-4*				-	JH4 (Hosseini)
JH-5*				-	JH5 (Hosseini)
JH-6*				-	JH6 (Hosseini)
JH-7			NT_186572.1	+	JH1 (Zhao)
JH-8				+	JH-PS4 (Zhao)
JH-9				+	JH2 (Zhao)

Se indica en qué genoma se identificó, en qué scaffold y en cuál hebra del scaffold. En la columna “UBICACIÓN” se indica en qué cromosoma está ubicado el scaffold. En caso de los scaffolds no ubicados en cromosomas de indica “ChrUn”. Los segmentos indicados con “\*” corresponden a segmentos identificados en el genoma Btau\_4.6.1 repetidos en UMD\_3.1. La columna HIT indica a qué secuencia es homólogo cada segmento. Los segmentos indicados como “Zhao” corresponden a los identificados por Zhao, Kacs Kovics et al. 2003 y los indicados como “Hosseini” corresponden a los identificados por Hosseini, Campbell et al. 2004.

### C.3.1. Segmentos VH

De los 57 segmentos identificados, solamente 11 no contienen la secuencia RSS adyacente en 3'. Estos 11 segmentos corresponden a 6 segmentos distintos (VH2-12, VH3-19, VH1-24, VH3-25, VH1-29 y VH3-32). Las secuencias RSS identificadas presentan un heptámero idéntico al buscado, las secuencias espaciadoras tienen un largo constante de 23 bp y parecen presentar poca variabilidad a nivel de secuencia. Se visualiza cierta variabilidad en la secuencia del nonúmero (Tabla 4).

Tabla 4. Secuencias RSS ubicadas adyacentes al extremo 3' de los segmentos VH identificados.

ID	MIS-MATCHES	HEPTÁMERO	ESPACIADOR	LARGO DE ESPACIADOR (bp)	NONÁMERO
● VH1-1	.	CACAGTG	TGAAAACCCACATCCTGAGGGTG	23	TCAGAAACC
● VH1-2	.	CACAGTG	TGAAAACCCACATCCTGAGGGTG	23	TCAGAAACC
● VH1-2*	.	CACAGTG	TGAAAACCCACATCCTGAGGGTG	23	TCAGAAACC
● VH2-3	2	CACAGTG	AGGGGAAGTCAGTGTGAGCCAG	23	ACAAAAACC
● VH3-4	5	CACAGTG	AGGGGAAGGCAGTGTGAGCCAG	23	ACACCACAC
● VH3-4*	5	CACAGTG	AGGGGAAGGCAGTGTGAGCCAG	23	ACACCACAC
● VH2-5	2	CACAGTG	AGGGGAAATCAGTGTGAGCCAG	23	ACAAAAACC
● VH2-5*	2	CACAGTG	AGGGGAAATCAGTGTGAGCCAG	23	ACAAAAACC
● VH1-6	3	CACAGTG	TGAAAACCCACATCCTGAGGGTG	23	TCAGACCCT
● VH2-7	2	CACAGTG	AGGGGAAGTCAGTGTGAGCCAG	23	ACAAAAACC
● VH2-7*	2	CACAGTG	AGGGGAAGTCAGTGTGAGCCAG	23	ACAAAAACC
● VH2-8	2	CACAGTG	AGGGGAAGTCAGTGTGAGCCAG	23	ACAAAAACC
● VH2-8*	2	CACAGTG	AGGGGAAGTCAGTGTGAGCCAG	23	ACAAAAACC
● VH3-9	4	CACAGTG	AGGGGAAGTCAGTGTGAGCCAG	23	ACACCAGCC
● VH3-9*	4	CACAGTG	AGGGGAAGTCAGTGTGAGCCAG	23	ACACCAGCC
● VH1-10	.	CACAGTG	TGAAAACCCACATCCTGAGGGTG	23	TCAGAAACC
● VH3-11	4	CACAGTG	AGGGGAAGTCAGTGTGAGCCAG	23	ACACCAGCC
● VH3-11*	4	CACAGTG	AGGGGAAGTCAGTGTGAGCCAG	23	ACACCAGCC
● VH2-12					
● VH2-12*					
● VH1-13	.	CACAGTG	TGAAAACCCACATCCTGAGGGTG	23	TCAGAAACC
● VH1-13*	.	CGCAGTG	CATGTAGTCGGTGAAGGTGTAT	22	CCAGAAGCC
● VH1-14	.	CACAGTG	TGAAAACCCACATCCTGAGGGTG	23	TCAGAAACC
● VH3-15	4	CACAGTG	AGGGGAAGGCAGTGTGAGCCAG	23	ACACCAGCC
● VH3-16	4	CACAGTG	AGGGGAAGGCAGTGTGAGCCAG	23	ACACCAGCC
● VH3-16*	4	CACAGTG	AGGGGAAGGCAGTGTGAGCCAG	23	ACACCAGCC
● VH2-17	2	CACAGTG	AGGGGAAATCAGTGTGAGCCAG	23	ACAAAAACC
● VH2-17*	2	CACAGTG	AGGGGAAATCAGTGTGAGCCAG	23	ACAAAAACC
● VH3-18	4	CACACTG	AGGGGACCTCAGTGTGAGCCAG	23	CCACAAGCC
● VH3-18*	4	CACACTG	AGGGGACCTCAGTGTGAGCCAG	23	CCACAAGCC
● VH3-19					
● VH3-19*					
● VH3-20	4	CACAGTG	AGGGGAAGGCAGTGTGAGCCAG	23	ACACCAGCC
● VH3-20*	4	CACAGTG	AGGGGAAGGCAGTGTGAGCCAG	23	ACACCAGCC
● VH1-21	.	CACAGTG	TGAAAACCCACATCCTGAGGGTG	23	TCAGAAACC
● VH2-22	2	CACAGTG	AGGGGAAATCAGTGTGAGCCAG	23	ACAAAAACC
● VH1-23	1	CACAGTG	TAAAACCCACATCCTGAGGGTG	23	TCAGAAACA
● VH1-24					
● VH1-24*					
● VH3-25					
● VH3-25*					
● VH3-26	4	CACAGTG	AGGGGACGTCAGTGTGAGCCAG	23	ACACAGGCC
● VH2-27	2	CACAGTG	AGGGGAAGTCAGTGTGAGCCAG	23	ACAAAAACC
● VH2-27*	2	CACAGTG	AGGGGAAGTCAGTGTGAGCCAG	23	ACAAAAACC
● VH1-28	.	CACAGTG	TGAAAACCCACATCCTGAGGGTG	23	TCAGAAACC
● VH1-28*	.	CACAGTG	TGAAAACCCACATCCTGAGGGTG	23	TCAGAAACC
● VH1-29					
● VH2-30	2	CACAGTG	AGGGGAAGTCAGTGTGAGCCAG	23	ACAAAAACC
● VH2-30*	2	CACAGTG	AGGGGAAGTCAGTGTGAGCCAG	23	ACAAAAACC
● VH3-31	3	CACAGTG	AGGGGATGGCAGTGTGAGCCAG	23	CCACAAGCC
● VH3-31*	3	CACAGTG	AGGGGATGGCAGTGTGAGCCAG	23	CCACAAGCC
● VH3-32					
● VH3-32*					
● VH2-33	3	CACAGTG	AGGGGACGTCAGTGTGAGCCAG	23	ACAAAAACC
● VH2-33*	2	CACAGTG	AGGGGACGTCAGTGTGAGCCAG	23	ACAAAAACC
● VH1-34	3	CACACCG	TGAAAACCCACATCCTGAGGGTG	23	TCAGAAACC
● VH1-35	4	CACAGTG	AGGGGAAGTCAGTGTGAGCCAG	23	ACACCAGCC

El punto (“.”) en la columna “MISMATCHES” indica que la secuencia encontrada es idéntica a la canónica.

El patrón octamérico en el promotor fue identificado adyacente a 27 segmentos, correspondientes a 18 segmentos distintos. En la mayoría de los casos la secuencia del octámero es idéntica a la canónica (Tabla 5). Cada segmento VH es precedido por un segmento líder, el cual presenta un intrón en su interior. El análisis con la herramienta FEX permitió identificar si cada segmento consta de un codón inicio y los sitios necesarios para el corte y empalme de los exones. Esta información junto al largo de los exones y la distancia del codón inicio a la secuencia RSS permitiría identificar si dicho segmento codifica un ORF. Solamente 18 segmentos cumplieron con todas las condiciones establecidas, de los cuales 11 corresponden a segmentos distintos (VH2-3, VH2-5, VH2-7, VH2-8, VH3-9, VH3-11, VH2-17, VH3-20, VH2-27, VH2-30 y VH2-33). Para estos 11 la longitud del exón 1 es de 44 o 45 bp, con la excepción de 3 casos donde la longitud es mayor. La longitud desde el inicio al RSS es más variable, rondando las 440 bp a excepción de 1 segmento (Tabla 5). A partir de la información recabada por estos análisis se le asignó a cada segmento la etiqueta de “funcional”, “ORF” o “pseudogen”. En cualquier caso, estas etiquetas son predicciones a base de unos pocos parámetros y pretenden discriminar aquellos segmentos que serían potencialmente funcionales de aquellos que no. De los 57 segmentos 17 serían potencialmente funcionales, los cuales corresponden a 12 segmentos distintos (Tabla 6). En todas las tablas, a la derecha de cada identificador se indicó con un círculo de color si el segmento se clasificó como funcional (verde), ORF (amarillo) o pseudogen (rojo).

### **C.3.2. Segmentos DH**

Dado que la secuencia de los segmentos D<sub>H</sub> es muy variable, la búsqueda se basó en la identificación de los patrones RSS flanqueando la secuencia codificante. Dado que se ha reportado el largo particularmente extenso de algunos segmentos D<sub>H</sub> bovinos se permitió la búsqueda de segmentos de hasta 200 bp (~66 aas). Se identificaron 15 segmentos DH, de los cuales 8 son distintos. Todos los segmentos presentan nonámeros y heptámeros muy cercanos al consenso, y secuencias espaciadoras del largo esperado. Todos los segmentos presentan al menos un ORF, aunque no todos los segmentos tienen un ORF partiendo del primer nucleótido (DH-1, DH-2 y DH-4). El largo de los exones es muy variable, y codifican desde 10 hasta 51 aminoácidos. Se ubicaron segmentos en los cromosomas 7, 8 y 21 (Tabla 7).

**Tabla 5.** Localización de secuencia octamérica en el promotor e identificación de señales necesarias para codificar un ORF.

ID	OCTÁMERO	MIS-MATCHES	LARGO DE EXÓN 1	EXÓN 2	LARGO DE INICIO A RSS
● VH1-1				SÍ	
● VH1-2				SÍ	
● VH1-2*				SÍ	
● VH2-3	ATGCAAAT	.	44	SÍ	432
● VH3-4	ATGCAAAT	.	51		433
● VH3-4*	ATGCAAAT	.	51		433
● VH2-5	ATGCAAAT	.	44	SÍ	432
● VH2-5*				SÍ	
● VH1-6			119		301
● VH2-7	GTGCAAAT	1	44	SÍ	432
● VH2-7*	GTGCAAAT	1	44	SÍ	432
● VH2-8	ATGCAAAT	.	59	SÍ	451
● VH2-8*	ATGCAAAT	.	59	SÍ	451
● VH3-9	ATGCAAAT	.	58	SÍ	437
● VH3-9*	ATGCAAAT	.	58	SÍ	437
● VH1-10					
● VH3-11	ATTCAAAT	1	45	SÍ	414
● VH3-11*	ATTCAAAT	1	45	SÍ	414
● VH2-12				SÍ	
● VH2-12*				SÍ	
● VH1-13			119	SÍ	301
● VH1-13*			119	SÍ	301
● VH1-14				SÍ	
● VH3-15	ATTCAAAT	1	45		427
● VH3-16			45		427
● VH3-16*	ATTCAAAT	1	45		427
● VH2-17	ATGCAAAT	.	44	SÍ	441
● VH2-17*	ATGCAAAT	.	44	SÍ	441
● VH3-18	ATGCAAAT	.			
● VH3-18*					
● VH3-19					
● VH3-19*					
● VH3-20	ATTCAAAT	1	89	SÍ	585
● VH3-20*	ATTCAAAT	1	89	SÍ	585
● VH1-21			119	SÍ	302
● VH2-22	ATGCAAAT	.	44	SÍ	431
● VH1-23					
● VH1-24			119	SÍ	
● VH1-24*			119	SÍ	
● VH3-25					
● VH3-25*					
● VH3-26	ATGCAAAT	.			
● VH2-27	ATGCAAAT	.	44	SÍ	431
● VH2-27*	ATGCAAAT	.	44	SÍ	431
● VH1-28			119		301
● VH1-28*			119		301
● VH1-29					
● VH2-30	ATGCAAAT	.	44	SÍ	432
● VH2-30*	ATGCAAAT	.	44		432
● VH3-31					
● VH3-31*					
● VH3-32					
● VH3-32*					
● VH2-33	ATGCAAAT	.	44	SÍ	432
● VH2-33*			44	SÍ	432
● VH1-34			119	SÍ	301
● VH1-35	ATGCAAAT	.			

**Tabla 6.** Resumen de análisis de búsqueda de patrones necesarios para la expresión de los segmentos identificados.

ID	RESUMEN	RSS	ORF	OCTÁMERO
● VH1-1	PSEUDOGEN	0		
● VH1-2	PSEUDOGEN	0		
● VH1-2*	PSEUDOGEN	0		
● VH2-3	FUNCIONAL	0	0	0
● VH3-4	PSEUDOGEN	0		0
● VH3-4*	PSEUDOGEN	0		0
● VH2-5	FUNCIONAL	0	0	0
● VH2-5*	PSEUDOGEN	0		
● VH1-6	PSEUDOGEN	0		
● VH2-7	FUNCIONAL	0	0	0
● VH2-7*	FUNCIONAL	0	0	0
● VH2-8	FUNCIONAL	0	0	0
● VH2-8*	FUNCIONAL	0	0	0
● VH3-9	FUNCIONAL	0	0	0
● VH3-9*	FUNCIONAL	0	0	0
● VH1-10	PSEUDOGEN	0		
● VH3-11	FUNCIONAL	0	0	0
● VH3-11*	FUNCIONAL	0	0	0
● VH2-12	PSEUDOGEN			
● VH2-12*	PSEUDOGEN			
● VH1-13	ORF	0	0	
● VH1-13*	ORF	0	0	
● VH1-14	PSEUDOGEN	0		
● VH3-15	PSEUDOGEN	0		0
● VH3-16	PSEUDOGEN	0		
● VH3-16*	PSEUDOGEN	0		0
● VH2-17	FUNCIONAL	0	0	0
● VH2-17*	FUNCIONAL	0	0	0
● VH3-18	PSEUDOGEN	0		0
● VH3-18*	PSEUDOGEN	0		
● VH3-19	PSEUDOGEN			
● VH3-19*	PSEUDOGEN			
● VH3-20	FUNCIONAL	0	0	0
● VH3-20*	FUNCIONAL	0	0	0
● VH1-21	ORF	0	0	
● VH2-22	FUNCIONAL	0	0	0
● VH1-23	PSEUDOGEN	0		
● VH1-24	PSEUDOGEN			
● VH1-24*	PSEUDOGEN			
● VH3-25	PSEUDOGEN			
● VH3-25*	PSEUDOGEN			
● VH3-26	PSEUDOGEN	0		0
● VH2-27	FUNCIONAL	0	0	0
● VH2-27*	FUNCIONAL	0	0	0
● VH1-28	PSEUDOGEN	0		
● VH1-28*	PSEUDOGEN	0		
● VH1-29	PSEUDOGEN			
● VH2-30	FUNCIONAL	0	0	0
● VH2-30*	PSEUDOGEN	0		0
● VH3-31	PSEUDOGEN	0		
● VH3-31*	PSEUDOGEN	0		
● VH3-32	PSEUDOGEN			
● VH3-32*	PSEUDOGEN			
● VH2-33	FUNCIONAL	0	0	0
● VH2-33*	ORF	0	0	
● VH1-34	ORF	0	0	
● VH1-35	PSEUDOGEN	0		0

Tabla 7. Segmentos DH identificados a partir de la búsqueda de RSSs.

ID	UBICACIÓN	PROYECTO	SCAFFOLD	HEBRA	MIS-MATCHES	5'	ESPACIADOR (bp)	5'	EXÓN		3'	ESPACIADOR (bp)	3'	ORF
						NONÁMERO		HEPTÁMERO	LARGO (bp)	LARGO (aa)	HEPTÁMERO		NONÁMERO	
						SECUENCIA			SECUENCIA			SECUENCIA		
DH-8	Chr7	UMD_3.1	NW_003064411.1	-	3	GGTTTCTGA	12	CACGGTG	57	19	CACAGTG	12	ACAAAAACC	
DH-7				-	3	GGTTTCTGA	12	CACGGTG	153	51	CACAGTG	12	ACAAAAACC	
DH-6	Chr8		NW_003066918.1	-	3	GGTTTCTGA	12	CACGGTG	57	19	CACAGTG	12	ACAAAAACC	
DH-5				-	2	GGTTTTTGA	12	CACGGTG	66	22	CACAGTG	12	ACAAAAACC	
DH-4				-	4	GGTTTCTGA	12	CACGGTG	35	11	CACAGTG	12	ACAGAAACC	+1
DH-3	Chr21		NW_003104538.1	-	3	GGTTTCTGA	12	CACGGTG	42	14	CACAGTG	12	ACAAAAACC	
DH-2				-	4	GGTTTCTGA	12	CACGGTG	35	11	CACAGTG	12	ACAGAAACC	+1
DH-1				-	4	GGATTTTGA	12	CACTGTG	30	10	CACAGTG	12	ACATAAAGC	+1
DH-7*				ChrUn	NT_186153.1	-	3	GGTTTCTGA	12	CACGGTG	153	51	CACAGTG	12
DH-4*	ChrUn		Btau_4.6.1	NW_001503306.1	+	4	GGTTTCTGA	12	CACGGTG	35	11	CACAGTG	12	ACAGAAACC
DH-5*		+			2	GGTTTTTGA	12	CACGGTG	66	22	CACAGTG	12	ACAAAAACC	
DH-6*		+			3	GGTTTCTGA	12	CACGGTG	57	19	CACAGTG	12	ACAAAAACC	
DH-3*	ChrUn	NW_001504477.2	-	3	GGTTTCTGA	12	CACGGTG	42	14	CACAGTG	12	ACAAAAACC		
DH-2*			-	4	GGTTTCTGA	12	CACGGTG	35	11	CACAGTG	12	ACAGAAACC	+1	
DH-1*			-	4	GGATTTTGA	12	CACTGTG	30	10	CACAGTG	12	ACATAAAGC	+1	

### C.3.3. Segmentos JH

Se identificaron secuencias RSS ubicadas adyacentes al extremo 5' de los segmentos JH en 9 de los 15 segmentos identificados. Estos 9 segmentos corresponden a 6 segmentos distintos (JH-1, JH-2, JH-4, JH-6, JH-7 y JH-9). Se observa mayor variabilidad en las regiones nonaméricas y heptaméricas, así como también en la longitud de la secuencia espaciadora, en comparación con lo visto para los segmentos VH (Tabla 8).

**Tabla 8.** Secuencias RSS ubicadas adyacentes al extremos 3' de los segmentos JH identificados.

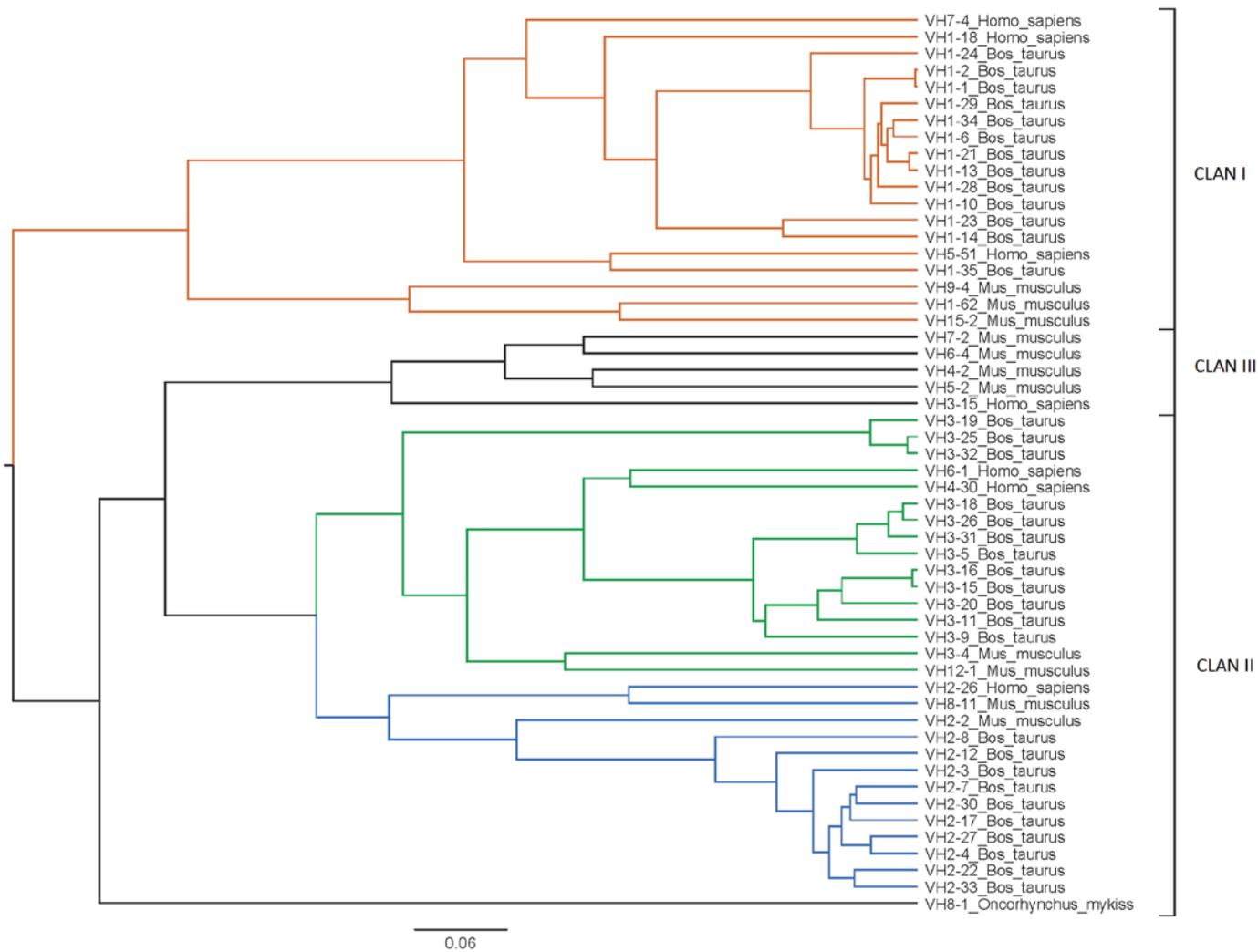
ID	MM	NONÁMERO	ESPACIADOR	LARGO ESPACIADOR (bp)	HEPTÁMERO
JH-1	5	GGGTTTCCG	TGCTCCCAGCACGGGGCCAGCC	22	CACTGTG
JH-1*	5	GGGTTTCCG	TGCTCCCAGCACGGGGCCAGCC	22	CACTGTG
JH-2					
JH-2*	4	TGTTTTTGT	GGAAAAGAAGCAACAGAATGGAAG	24	CCATGCT
JH-3					
JH-3*					
JH-4	2	GGTTTTTGC	ACAGCCCCTAGCGGGGCCCATGG	23	CACTGTG
JH-4*	2	GGTTTTTGC	ACAGCCCCTAGCGGGGCCCATGG	23	CACTGTG
JH-5					
JH-5*					
JH-6	1	GGTTTTTGT	TGGGCGAGGCTGGAGATATCAC	22	CACTGTG
JH-6*	1	GGTTTTTGT	TGGGCGAGGCTGGAGATATCAC	22	CACTGTG
JH-7	5	GGTTTTGGC	CAAGCCAGGAACCTCAGGGGAA	22	CTCGGTG
JH-8					
JH-9	1	GGTTTTTGT	TGGGCGAGGCTGGAGATATCAC	22	CACTGTG

La columna MM indica la cantidad de mismatches que presenta la RSS respecto del consenso.

## C.4. Análisis filogenético de los segmentos identificados

### C.4.1. Segmentos VH

En el análisis filogenético de los segmentos V<sub>H</sub> se observa que se agruparon en 3 grandes grupos. Uno de estos grupos está comprendido dentro del clan I humano, que agrupa a las familias VH1, VH5 y VH7 humanas. Dentro de este clan aparecen las familias VH1, VH9 y VH15 de ratón. Los segmentos bovinos agrupados allí corresponden a segmentos que fueron clasificados como ORF o pseudogenes. A este grupo se le denominó VH1 (Figura 3).



**Figura 3.** Análisis filogenético de los segmentos VH. Con marrón se identificó la familia a la que se denominó VH1, con azul la familia VH2 y con verde la familia VH3.

Ningún segmento bovino fue agrupado dentro del clan III humano, el cual comprende a la familia VH3 humana, y agrupó a las familias VH4, VH5, VH6 y VH7 murinas. El clan II humano comprende las familias VH2, VH4 y VH6 humanas, y agrupa las familias VH2, VH3, VH8 y VH12 murinas. Dentro de este clan los segmentos VH bovinos se ubicaron en 2 grandes grupos a los cuales se denominó VH2 y VH3. El grupo VH2 se agrupó más cercano a la familia VH2 humana, mientras que el grupo VH3 se observa más cercano a las familias VH4 y VH6 humanas. La mayoría de los segmentos identificados como potencialmente funcionales se encuentran comprendidos en el grupo VH2 (VH2-3, VH2-5, VH2-7, VH2-8, VH2-17, VH2-22, VH2-27, VH2-30 y VH2-33), mientras que unos pocos en el grupo VH3 (VH3-9, VH3-11 y VH3-20) (Figura 3).

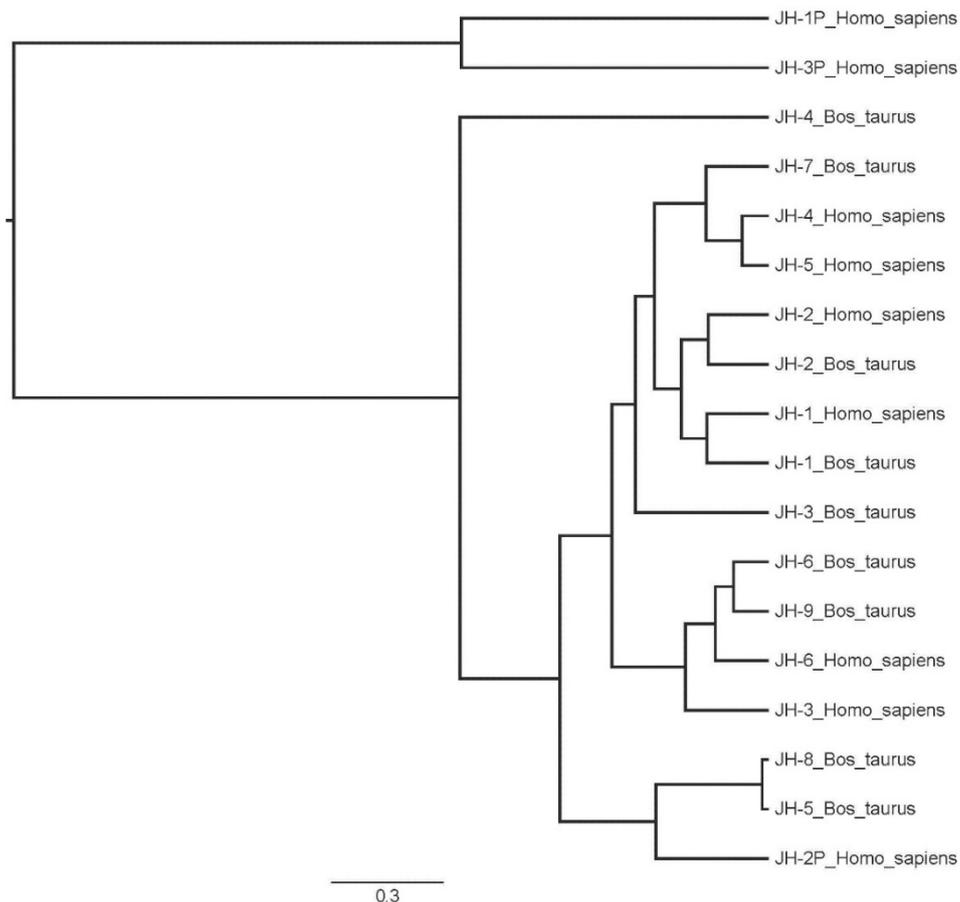
La nomenclatura de los segmentos se basó en esta agrupación de los segmentos y las reglas de nomenclatura que define IMGT para humano. Luego del prefijo VH se indica a qué grupo pertenece, y al final un número que se definió de manera más o menos arbitraria. Para los segmentos ubicados en cromosomas se numeró de menor a mayor leyéndolos de 3' a 5', partiendo del segmento VH inmediatamente adyacente en 5' del primer segmento DH. En humano, así como en la mayoría de las especies descritas, no se conocen segmentos funcionales en más de un cromosoma. Dado que para bovinos en la bibliografía se había identificado el locus VH en el cromosoma 21 se comenzaron a numerar a partir de este cromosoma. Para aquellos segmentos no ubicados en cromosomas la numeración fue arbitraria.

#### **C.4.2. Segmentos DH**

Dada la variabilidad en longitud y diversidad en secuencia de los segmentos DH identificados con bovinos no se realizó un análisis filogenético. La naturaleza de los segmentos DH implica una gran diversidad, por lo que este tipo de análisis sería complejo de interpretar.

#### **C.4.3. Segmentos JH**

Los segmentos JH identificados en bovinos parecen encontrarse alejados filogenéticamente de 2 de los 3 pseudogenes identificados en humanos (JH-1P y JH-3P). Sin embargo, el segmento pseudogénico JH-2P humano se encuentra cercano a los segmentos JH-5 y JH-8 bovinos. El segmento JH-4 bovino no se encuentra agrupado con ningún segmento humano. El resto de los segmentos bovinos se encuentran agrupados con los segmentos humanos funcionales (Figura 4). En particular, destacamos que el segmento JH7 bovino se agrupa junto con el JH4 y JH5 humanos.



**Figura 4.** Análisis filogenético de los segmentos JH. Para los segmentos JH humanos la letra “P” adyacente al nombre indica que se trata de un pseudogen.

### **C.5. Identificación de mRNA correspondientes a Iq**

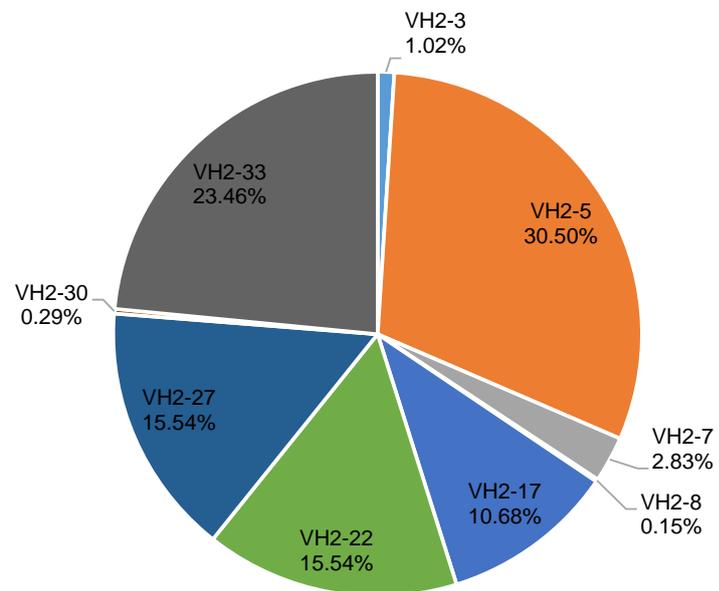
De los 35 segmentos VH distintos identificados solamente 9 se encontraron en las librerías bovinas analizadas (VH2-3, VH2-5, VH2-7, VH2-8, VH2-17, VH2-22, VH2-27, VH2-30, VH2-33). Estos 9 segmentos corresponden a todos los segmentos del grupo VH2 clasificados como potencialmente funcionales y se reparten entre el cromosoma 7 y el 21. Los 3 segmentos del grupo VH3 clasificados como potencialmente funcionales no fueron identificados en dichas librerías. Ningún segmento clasificado como ORF o pseudogen se encontró en las librerías (Figura 5A).

La gran mayoría de los hits contra los transcritos se repartió entre 5 segmentos. En particular los segmentos VH2-5 y VH2-33 se reparten más de la mitad de los transcritos identificados (Figura 5B).

De los segmentos JH solamente se identificaron 3 en estas librerías. En particular se detectó 426 veces el segmento JH-7, 15 veces el JH-4 y 11 veces el JH-6.

**A**

ID	CANTIDAD mRNAs
● VH1-1	
● VH1-2	
● VH2-3	14
● VH3-4	
● VH2-5	420
● VH1-6	
● VH2-7	39
● VH2-8	2
● VH3-9	
● VH1-10	
● VH3-11	
● VH2-12	
● VH1-13	
● VH1-14	
● VH3-15	
● VH3-16	
● VH2-17	147
● VH3-18	
● VH3-19	
● VH3-20	
● VH1-21	
● VH2-22	214
● VH1-23	
● VH1-24	
● VH3-25	
● VH3-26	
● VH2-27	214
● VH1-28	
● VH1-29	
● VH2-30	4
● VH3-31	
● VH3-32	
● VH2-33	323
● VH1-34	
● VH1-35	

**B**

**Figura 5.** Distribución de segmentos VH identificados en transcritos.

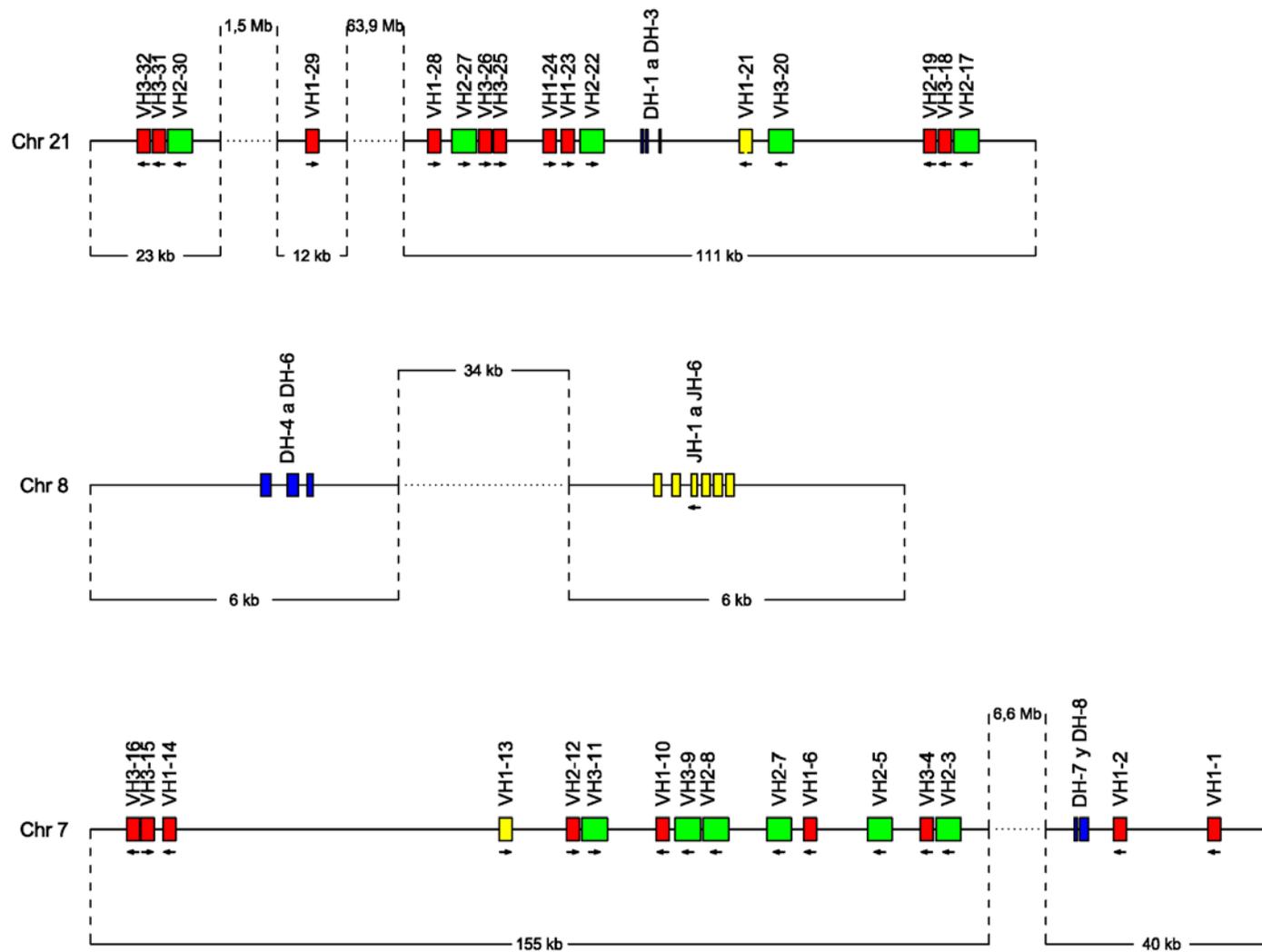
## D. Discusión y conclusiones

Este trabajo tuvo como objetivo sistematizar la información disponible sobre los genes de inmunoglobulina bovinos, identificando y analizando los segmentos VH, DH y JH existentes en los 2 genomas bovinos reportados y disponibles en las bases de datos públicas.

Se pudieron identificar 57 **segmentos VH germinales** repartidos en 21 scaffolds en ambos genomas. Treinta y cinco de estos segmentos fueron identificados como únicos y se ubicaron en los cromosomas 7 y 21. Uno de ellos se encuentra en un scaffold que no pudo ser ubicado en ningún cromosoma. Es notorio que en bovinos existe un repertorio de segmentos VH germinales muy limitado en número y también llama la atención la distribución en dos cromosomas de dichos segmentos (Figura 6). Esto no ha sido observado en otras especies estudiadas, pero sí para bovinos. Un estudio anterior examinó e identificó dos BACs de regiones potencialmente contenedoras de segmentos génicos y se pudo ver que un segmento de BTA21 conteniendo segmentos VH se ubicó incorrectamente en BTA7 en UMD\_3.1 (Niku, Liljavirta et al. 2012).

El análisis filogenético de los segmentos VH los clasifica en 3 grandes grupos, uno presente similitud con el clan I humano (al que se denominó VH1) y 2 grupos similares al clan II humano (VH2 y VH3 bovinos). Esta agrupación permitió proponer una nomenclatura sistemática aproximada a la que existe para humano (Lefranc 2001). Los segmentos se numeran de forma creciente desde 3' hacia 5', partiendo del segmento adyacente al primer segmento DH (leído de 5' a 3'). Previo al número de segmento se indica a qué grupo filogenético pertenece (Figura 6). Esta nomenclatura no pretende ser definitiva dadas las inconsistencias a nivel cromosómico que se identificaron, pero propone un acercamiento a la organización de la información disponible.

A la baja cantidad de segmentos VH identificados se le suma la pequeña fracción de estos que se encontró en el análisis de transcriptómica: solamente 9 fueron encontrados en las librerías bovinas analizadas. Esto se aproxima a los datos sugeridos por la búsqueda de patrones adyacentes a los segmentos necesarios para su expresión. Según estos datos se clasificaron 12 segmentos distintos como potencialmente funcionales. De estos 12, los 9 identificados en transcritos corresponden a segmentos del grupo llamado VH2. Parece un número insuficiente para generar la diversidad que caracteriza a las moléculas de inmunoglobulinas.



**Figura 6.** Organización cromosómica de los segmentos VH, DH y JH. Aunque se respeta la distribución, el tamaño de los segmentos fue magnificado.

**Tabla 9.** Equivalencias de las denominaciones dadas por los distintos autores para el mismo segmento. Niku et al. 2012 no reporta repetidos.

ID	Walther et al. 2013	Niku et al. 2012
● VH1-1		IGHV3S7
● VH1-2		IGHV3S8
● VH2-3	IGHV19	IGHV1S5
● VH3-4	IGHV18	IGHV2S6
● VH2-5	IGHV17	IGHV1S6
● VH1-6		IGHV3S4
● VH2-7	IGHV16	IGHV1S7
● VH2-8	IGHV15	IGHV1S8
● VH3-9	IGHV14	IGHV2S9
● VH1-10		IGHV3S5
● VH3-11	IGHV21	IGHV2S10
● VH2-12	IGHV13	IGHV1S9
● VH1-13		IGHV3S6
● VH1-14		
● VH3-15	IGHV12	IGHV2S11
● VH3-16	IGHV11	IGHV2S12
● VH2-17	IGHV10	
● VH3-18	IGHV9	IGHV2S1
● VH3-19	IGHV8	
● VH3-20	IGHV7	IGHV2S2
● VH1-21		IGHV3S1
● VH2-22	IGHV6	IGHV1S2
● VH1-23		
● VH1-24		IGHV3S2
● VH3-25	IGHV5	
● VH3-26	IGHV4	IGHV2S4
● VH2-27	IGHV3	IGHV1S3
● VH1-28		IGHV3S3
● VH1-29		
● VH2-30	IGHV2	IGHV1S4
● VH3-31	IGHV1	IGHV2S5
● VH3-32		
● VH2-33	IGHV36	IGHV1S10
● VH1-34		
● VH1-35		
● VH1-2*		
● VH3-4*	IGHV30	
● VH2-5*	IGHV31	
● VH2-7*	IGHV25	
● VH2-8*	IGHV24	
● VH3-9*	IGHV23	
● VH3-11*		
● VH2-12*	IGHV20	
● VH1-13*		
● VH3-16*	IGHV28	
● VH2-17*	IGHV34	
● VH3-18*	IGHV35	
● VH-19*		
● VH3-20*	IGHV22	
● VH1-24*		
● VH-25*		
● VH2-27*	IGHV33	
● VH1-28*		
● VH2-29*	IGHV26	
● VH2-30*	IGHV27	
● VH3-31*		
● VH2-33*	IGHV29	

No se identificaron segmentos VH que se alejaran significativamente de lo previamente reportado por Niku, Liljavirta et al. 2012 y Walther, Czerny et al. 2013. Algunos de los segmentos reportados en estos trabajos no fueron identificados en esta tesina, ya que no se pudieron recuperar a partir de las ubicaciones indicadas. Los árboles filogenéticos reportados por ambos trabajos coinciden con el obtenido aquí. Walther, Czerny et al. (2013) no identificaron los segmentos que en este trabajo fueron clasificados como pertenecientes al grupo VH1, mientras que Niku, Liljavirta et al. (2012) los clasificó como VH3. Walther, Czerny et al. (2013) sí identificó los del grupo VH2, a los que clasificó como boVH1, y los del grupo VH3 a los que clasificó como boVH2. Niku, Liljavirta et al. (2012) los clasificó como VH1 e VH2 respectivamente. Los estudios de expresión de ambos autores encuentran también que los únicos identificados son los del VH2 (Tabla 9). Esto se contradice con lo observado en humanos, donde la familia VH2 está subrepresentada en el repertorio expresado.

En relación a los **segmentos DH germinales** sucede algo similar a lo observado para los segmentos VH, donde se identificaron segmentos en los cromosomas 7, 8 y 21. Dada la variabilidad de secuencia de los segmentos DH es complejo determinar si la identidad de estos es la misma de los reportados previamente. Sí se observan las mismas características particulares en la extremadamente larga longitud de los mismos, en comparación con lo conocido para otras especies.

La cantidad e identidad de **segmentos JH germinales** condice parcialmente lo previamente descrito para bovinos. Dentro de los 9 segmentos reportados se identificaron todos menos 2 segmentos, el denominado JH1 previamente por Hosseini, Campbell et al. 2004 y el denominado JH-PS3 por Zhao, Kacskovics et al. 2003. Seis de estos segmentos fueron ubicados en el cromosoma 8 y 3 no pudieron ser ubicados. Los estudios previos localizaron de manera no informática los segmentos en los cromosomas 11 y 21, y plantean la incógnita de si es posible el rearrreglo trans-cromosómico. A diferencia de lo observado para los segmentos VH la cantidad de segmentos JH no parece tan reducida, ya que, por ejemplo, en ratón se conocen 4 y en humanos 6.

Los datos de expresión también se corresponden con lo previamente reportado. El segmento JH-7, correspondiente al JH-1 reportado por Zhao, Kacskovics et al. (2003), parece expresarse en gran cantidad, casi exclusivamente. El segmento JH-7 se encuentra filogenéticamente cercano a los segmentos JH-4 y JH-5 bovinos. Los segmentos JH-4 y JH-6 correspondientes a los mismos reportados por Hosseini, Campbell et al. (2004) se expresarían a un nivel mucho menor. El segmento JH-6, junto con el JH-9 se encuentra filogenéticamente cercano a los segmentos JH-6 y JH-3

humanos. En humanos se ha encontrado que el segmento JH-4 es el más expresado en cualquier tipo de linfocitos B, dato que coincide con el segmento JH-7 bovino, lo que sugiere que serían segmentos homólogos (Wu, Kipling et al. 2010). En humanos le siguen en nivel de expresión de mayor a menor los segmentos JH-6, JH-5, JH-3, JH-2 y JH-1. Parece también existir homología entre el segmento JH-6 humano y el bovino. La cercanía de los segmentos JH-5 y JH-8 bovinos al pseudogen JH-2P humano podría sugerir que se trata también de segmentos pseudogénicos, aunque son necesarios más análisis para poder determinarlo.

Aunque los datos reportados para los segmentos VH, DH y JH se aproximan a lo previamente reportado en la bibliografía y a lo conocido para la mayoría de las especies se han identificado algunas inconsistencias. Una es la cantidad acotada de segmentos VH, y la otra la localización de segmentos funcionales en distintos cromosomas. Además, previamente un grupo comparó ambos genomas mediante el software GenAlyzer e identificaron tres tipos de inconsistencias en las secuencias: inversiones (coincide el mapeo en un cromosoma pero difiere la orientación, INV), mapeo en distintas posiciones de un cromosoma (MDP), y presencia en un cromosoma en un ensamblaje y ausencia en el otro (ODS). Trabajaron con aquellas mayores a 100 kpb (Partipilo, D'Addabbo et al. 2011). Se identificaron 137 inversiones y 16 mapeos en distintas posiciones. Se encontraron 198 secuencias en cromosomas de UMD\_3.1 que no mapean en los correspondientes de Btau\_4.6, contra 37 secuencias en Btau\_4.6 que no mapean en los correspondientes de UMD\_3.1. Esto coincide con la mayor cantidad de scaffolds sin mapear en un cromosoma en el ensamblaje Btau\_4.6. De las ODSs identificadas en UMD\_3.1 se encontraron 17 secuencias que no se pudieron identificar en ningún scaffold de Btau\_4.6. En dos de estos casos, localizados en el cromosoma 21, se identificaron genes VH. En diversos experimentos realizados para esclarecer las ambigüedades se encontró que generalmente el ensamblaje UMD\_3.1 tenía la razón. De todas formas sostienen que sería importante disponer información sobre la evolución del cariotipo para mejorar el ensamblaje.

Intentando buscar una causa que explique estas inconsistencias reportadas para los loci de inmunoglobulinas, encontramos que los 2 genomas reportados en la bibliografía fueron realizados a partir de ADN genómico obtenido de muestras de sangre periférica. El ADN obtenido a partir de estas muestras proviene del conjunto de células circulantes en la sangre, entre las que se encuentran en una proporción importante a los linfocitos T y linfocitos B maduros. En el proceso de linfopoyesis que sucede en los órganos linfoides primarios (médula ósea y timo) las células progenitoras de los linfocitos sufren el rearreglo génico de los segmentos que codifican para los receptores T y para

las inmunoglobulinas. Estos rearrreglos implican procesos de recombinación a distancia a nivel del genoma, el cual es modificado irreversiblemente. Por otro lado, en la sangre periférica también podemos encontrar linfocitos que sufrieron el proceso de hipermutación somática y eventualmente de conversión génica así como de cambio de clase. Por tanto es totalmente esperable encontrar inconsistencias en el ensamblaje de los loci de los segmentos de inmunoglobulinas.

Así es que la arquitectura a gran escala del locus no es confiable, aunque a nivel de contigs se podría utilizar para estudiar la estructura de genes individuales. Dada la complejidad del locus IGH son necesarios esfuerzos específicos en la secuenciación de esta región. Parece ser imprescindible resolver las inconsistencias con nuevos análisis de secuenciación de muestras biológicas que no presenten rearrreglos en los loci de segmentos de inmunoglobulinas, ni hipermutación somática ni cambio de clase. Por esta razón, una propuesta es secuenciar a partir de muestras de hígado, ya que el tejido hepático no presenta rearrreglo en dicho locus y es barato y fácil de obtener en grandes cantidades.

## E. Bibliografía

- Alt, F. W. and D. Baltimore (1982). "Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions." Proc Natl Acad Sci U S A**79**(13): 4118-4122.
- Alt, F. W., T. K. Blackwell and G. D. Yancopoulos (1987). "Development of the primary antibody repertoire." Science**238**(4830): 1079-1087.
- Bassing, C. H., W. Swat and F. W. Alt (2002). "The mechanism and regulation of chromosomal V(D)J recombination." Cell**109** **Suppl**: S45-55.
- Becker, R. S. and K. L. Knight (1990). "Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits." Cell**63**(5): 987-997.
- Bengtén, E., M. Wilson, N. Miller, L. W. Clem, L. Pilstrom and G. W. Warr (2000). "Immunoglobulin isotypes: structure, function, and genetics." Curr Top Microbiol Immunol**248**: 189-219.
- Berens, S. J., D. E. Wylie and O. J. Lopez (1997). "Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains." Int Immunol**9**(1): 189-199.
- Berman, J. E., S. J. Mellis, R. Pollock, C. L. Smith, H. Suh, B. Heinke, C. Kowal, U. Surti, L. Chess, C. R. Cantor and et al. (1988). "Content and organization of the human Ig VH locus: definition of three new VH families and linkage to the Ig CH locus." EMBO J**7**(3): 727-738.
- Butler, J. E., J. Sun and P. Navarro (1996). "The swine Ig heavy chain locus has a single JH and no identifiable IgD." Int Immunol**8**(12): 1897-1904.
- Butler, J. E. and N. Wertz (2012). "The porcine antibody repertoire: variations on the textbook theme." Front Immunol**3**: 153.
- Carlson, L. M., W. T. McCormack, C. E. Postema, E. H. Humphries and C. B. Thompson (1990). "Templated insertions in the rearranged chicken IgL V gene segment arise by intrachromosomal gene conversion." Genes Dev**4**(4): 536-547.
- Choi, J. W., J. K. Kim, H. W. Seo, B. W. Cho, G. Song and J. Y. Han (2010). "Molecular cloning and comparative analysis of immunoglobulin heavy chain genes from *Phasianus colchicus*, *Meleagris gallopavo*, and *Coturnix japonica*." Vet Immunol Immunopathol**136**(3-4): 248-256.
- Chowdhary, B. P., L. Fronicke, I. Gustavsson and H. Scherthan (1996). "Comparative analysis of the cattle and human genomes: detection of ZOO-FISH and gene mapping-based chromosomal homologies." Mamm Genome**7**(4): 297-302.
- Cook, G. P., I. M. Tomlinson, G. Walter, H. Riethman, N. P. Carter, L. Buluwela, G. Winter and T. H. Rabbitts (1994). "A map of the human immunoglobulin VH locus completed by analysis of the telomeric region of chromosome 14q." Nat Genet**7**(2): 162-168.
- Criscitello, M. F. and M. F. Flajnik (2007). "Four primordial immunoglobulin light chain isotypes, including lambda and kappa, identified in the most primitive living jawed vertebrates." Eur J Immunol**37**(10): 2683-2694.
- Currier, S. J., J. L. Gallarda and K. L. Knight (1988). "Partial molecular genetic map of the rabbit VH chromosomal region." J Immunol**140**(5): 1651-1659.
- Danilova, N., J. Bussmann, K. Jekosch and L. A. Steiner (2005). "The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z." Nat Immunol**6**(3): 295-302.
- Das, S., U. Mohamedy, M. Hirano, M. Nei and N. Nikolaidis (2010). "Analysis of the immunoglobulin light chain genes in zebra finch: evolutionary implications." Mol Biol Evol**27**(1): 113-120.
- Das, S., M. Nozawa, J. Klein and M. Nei (2008). "Evolutionary dynamics of the immunoglobulin heavy chain variable region genes in vertebrates." Immunogenetics**60**(1): 47-55.
- De Genst, E., D. Saerens, S. Muyldermans and K. Conrath (2006). "Antibody repertoire development in camelids." Dev Comp Immunol**30**(1-2): 187-198.
- de Wildt, R. M., R. M. Hoet, W. J. van Venrooij, I. M. Tomlinson and G. Winter (1999). "Analysis of heavy and light chain pairings indicates that receptor editing shapes the human antibody repertoire." J Mol Biol**285**(3): 895-901.

- Deza, F. G., C. S. Espinel and J. V. Beneitez (2007). "A novel IgA-like immunoglobulin in the reptile *Eublepharis macularius*." Dev Comp Immunol**31**(6): 596-605.
- Dooley, H. and M. F. Flajnik (2006). "Antibody repertoire development in cartilaginous fish." Dev Comp Immunol**30**(1-2): 43-56.
- Dufour, V., S. Malinge and F. Nau (1996). "The sheep Ig variable region repertoire consists of a single VH family." J Immunol**156**(6): 2163-2170.
- Eguchi-Ogawa, T., N. Wertz, X. Z. Sun, F. Piumi, H. Uenishi, K. Wells, P. Chardon, G. J. Tobin and J. E. Butler (2010). "Antibody repertoire development in fetal and neonatal piglets. XI. The relationship of variable heavy chain gene usage and the genomic organization of the variable heavy chain locus." J Immunol**184**(7): 3734-3742.
- Ekman, A., M. Niku, J. Liljavirta and A. Iivanainen (2009). "Bos taurus genome sequence reveals the assortment of immunoglobulin and surrogate light chain genes in domestic cattle." BMC Immunol**10**: 22.
- Flajnik, M. F. (2002). "Comparative analyses of immunoglobulin genes: surprises and portents." Nat Rev Immunol**2**(9): 688-698.
- Fleurant, M., L. Changchien, C. T. Chen, M. F. Flajnik and E. Hsu (2004). "Shark Ig light chain junctions are as diverse as in heavy chains." J Immunol**173**(9): 5574-5582.
- Giudicelli, V., D. Chaume and M. P. Lefranc (2005). "IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes." Nucleic Acids Res**33**(Database issue): D256-261.
- Giudicelli, V., P. Duroux, C. Ginestoux, G. Folch, J. Jabado-Michaloud, D. Chaume and M. P. Lefranc (2006). "IMGT/LIGM-DB, the IMGT comprehensive database of immunoglobulin and T cell receptor nucleotide sequences." Nucleic Acids Res**34**(Database issue): D781-784.
- Greenberg, A. S., D. Avila, M. Hughes, A. Hughes, E. C. McKinney and M. F. Flajnik (1995). "A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks." Nature**374**(6518): 168-173.
- Haire, R. N., J. P. Rast, R. T. Litman and G. W. Litman (2000). "Characterization of three isotypes of immunoglobulin light chains and T-cell antigen receptor alpha in zebrafish." Immunogenetics**51**(11): 915-923.
- Hamers-Casterman, C., T. Atarhouch, S. Muyldermans, G. Robinson, C. Hamers, E. B. Songa, N. Bendahman and R. Hamers (1993). "Naturally occurring antibodies devoid of light chains." Nature**363**(6428): 446-448.
- Hansen, J. D., E. D. Landis and R. B. Phillips (2005). "Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: Implications for a distinctive B cell developmental pathway in teleost fish." Proc Natl Acad Sci U S A**102**(19): 6919-6924.
- Hayes, H. C. and E. J. Petit (1993). "Mapping of the beta-lactoglobulin gene and of an immunoglobulin M heavy chain-like sequence to homoeologous cattle, sheep, and goat chromosomes." Mamm Genome**4**(4): 207-210.
- Hosseini, A., G. Campbell, M. Prorocic and R. Aitken (2004). "Duplicated copies of the bovine JH locus contribute to the Ig repertoire." Int Immunol**16**(6): 843-852.
- Hozumi, N. and S. Tonegawa (1976). "Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions." Proc Natl Acad Sci U S A**73**(10): 3628-3632.
- Hsu, E., M. F. Flajnik and L. Du Pasquier (1985). "A third immunoglobulin class in amphibians." J Immunol**135**(3): 1998-2004.
- Johansson, J., M. Aveskogh, B. Munday and L. Hellman (2002). "Heavy chain V region diversity in the duck-billed platypus (*Ornithorhynchus anatinus*): long and highly variable complementarity-determining region 3 compensates for limited germline diversity." J Immunol**168**(10): 5155-5162.
- Jung, D., C. Giallourakis, R. Mostoslavsky and F. W. Alt (2006). "Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus." Annu Rev Immunol**24**: 541-570.
- Kaushik, A., F. Shojaei and S. S. Saini (2002). "Novel insight into antibody diversification from cattle." Vet Immunol Immunopathol**87**(3-4): 347-350.
- Kaushik, A. K., M. E. Kehrl, Jr., A. Kurtz, S. Ng, M. Koti, F. Shojaei and S. S. Saini (2009). "Somatic hypermutations and isotype restricted exceptionally long CDR3H contribute

- to antibody diversification in cattle." Vet Immunol Immunopathol**127**(1-2): 106-113.
- Kirkham, P. M., F. Mortari, J. A. Newton and H. W. Schroeder, Jr. (1992). "Immunoglobulin VH clan and family identity predicts variable domain structure and may influence antigen binding." EMBO J**11**(2): 603-609.
- Knight, K. L. and C. R. Winstead (1997). "Generation of antibody diversity in rabbits." Curr Opin Immunol**9**(2): 228-232.
- Koti, M., G. Kataeva and A. K. Kaushik (2008). "Organization of D(H)-gene locus is distinct in cattle." Dev Biol (Basel)**132**: 307-313.
- Koti, M., G. Kataeva and A. K. Kaushik (2010). "Novel atypical nucleotide insertions specifically at VH-DH junction generate exceptionally long CDR3H in cattle antibodies." Mol Immunol**47**(11-12): 2119-2128.
- Kuroiwa, Y., P. Kasinathan, T. Sathiyaseelan, J. A. Jiao, H. Matsushita, J. Sathiyaseelan, H. Wu, J. Mellquist, M. Hammitt, J. Koster, S. Kamoda, K. Tachibana, I. Ishida and J. M. Robl (2009). "Antigen-specific human polyclonal antibodies from hyperimmunized cattle." Nat Biotechnol**27**(2): 173-181.
- Lefranc, M. P. (2001). "Nomenclature of the human immunoglobulin genes." Curr Protoc Immunol**Appendix 1**: Appendix 1P.
- Lefranc, M. P., V. Giudicelli, Q. Kaas, E. Duprat, J. Jabado-Michaloud, D. Scaviner, C. Ginestoux, O. Clement, D. Chaume and G. Lefranc (2005). "IMGT, the international ImMunoGeneTics information system." Nucleic Acids Res**33**(Database issue): D593-597.
- Litman, G. W., M. K. Anderson and J. P. Rast (1999). "Evolution of antigen binding receptors." Annu Rev Immunol**17**: 109-147.
- Liu, Y., X. Qin, X. Z. Song, H. Jiang, Y. Shen, K. J. Durbin, S. Lien, M. P. Kent, M. Sodeland, Y. Ren, L. Zhang, E. Sodergren, P. Havlak, K. C. Worley, G. M. Weinstock and R. A. Gibbs (2009). "Bos taurus genome assembly." BMC Genomics**10**: 180.
- Lopez, O., C. Perez and D. Wylie (1998). "A single VH family and long CDR3s are the targets for hypermutation in bovine immunoglobulin heavy chains." Immunol Rev**162**: 55-66.
- Lundqvist, M. L., D. L. Middleton, S. Hazard and G. W. Warr (2001). "The immunoglobulin heavy chain locus of the duck. Genomic organization and expression of D, J, and C region genes." J Biol Chem**276**(50): 46729-46736.
- Lundqvist, M. L., D. L. Middleton, C. Radford, G. W. Warr and K. E. Magor (2006). "Immunoglobulins of the non-galliform birds: antibody expression and repertoire in the duck." Dev Comp Immunol**30**(1-2): 93-100.
- Magor, K. E., D. A. Higgins, D. L. Middleton and G. W. Warr (1994). "cDNA sequence and organization of the immunoglobulin light chain gene of the duck, *Anas platyrhynchos*." Dev Comp Immunol**18**(6): 523-531.
- Meikle, A., D. Cavestany, M. Carriquiry, M. d. L. Adrien, V. Artegoitia, I. Pereira, G. Rupprechter, P. Pessina, G. Rama, A. Fernández, M. Breijo, D. Laborde, O. Pritsch, J. M. Ramos, E. de Torres, P. Nicolini, A. Mendoza, J. Dutour, M. Fajardo, A. L. Astessiano, L. Olazábal, D. Mattiauda and P. Chilibruste (2013). "Avances en el conocimiento de la vaca lechera durante el período de transición en Uruguay: un enfoque multidisciplinario." Agrociencia Uruguay**17**(1): 7.
- Miller, R. D., H. Grabe and G. H. Rosenberg (1998). "V(H) repertoire of a marsupial (*Monodelphis domestica*)." J Immunol**160**(1): 259-265.
- Murphy, K., P. Travers, M. Walport and C. Janeway (2008). Janeway's immunobiology. New York, Garland Science.
- Niku, M., J. Liljavirta, K. Durkin, E. Schroderus and A. Iivanainen (2012). "The bovine genomic DNA sequence data reveal three IGHV subgroups, only one of which is functionally expressed." Dev Comp Immunol**37**(3-4): 457-461.
- Ohta, Y. and M. Flajnik (2006). "IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates." Proc Natl Acad Sci U S A**103**(28): 10723-10728.
- Ota, T. and M. Nei (1994). "Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family." Mol Biol Evol**11**(3): 469-482.
- Partipilo, G., P. D'Addabbo, G. M. Lacalandra, G. E. Liu and M. Rocchi (2011). "Refinement of *Bos taurus* sequence assembly based on BAC-FISH experiments." BMC

Genomics**12**: 639.

- Parvari, R., E. Ziv, F. Lentner, S. Tel-Or, Y. Burstein and I. Schechter (1987). "Analyses of chicken immunoglobulin light chain cDNA clones indicate a few germline V lambda genes and allotypes of the C lambda locus." EMBO J**6**(1): 97-102.
- Pasman, Y., S. S. Saini, E. Smith and A. K. Kaushik (2010). "Organization and genomic complexity of bovine lambda-light chain gene locus." Vet Immunol Immunopathol**135**(3-4): 306-313.
- Qin, T., L. Ren, X. Hu, Y. Guo, J. Fei, Q. Zhu, J. E. Butler, C. Wu, N. Li, L. Hammarstrom and Y. Zhao (2008). "Genomic organization of the immunoglobulin light chain gene loci in *Xenopus tropicalis*: evolutionary implications." Dev Comp Immunol**32**(2): 156-165.
- Reynaud, C. A., A. Dahan, V. Anquez and J. C. Weill (1989). "Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region." Cell**59**(1): 171-183.
- Saini, S. S., B. Allore, R. M. Jacobs and A. Kaushik (1999). "Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies." Eur J Immunol**29**(8): 2420-2426.
- Saini, S. S., W. R. Hein and A. Kaushik (1997). "A single predominantly expressed polymorphic immunoglobulin VH gene family, related to mammalian group, I, clan, II, is identified in cattle." Mol Immunol**34**(8-9): 641-651.
- Saini, S. S. and A. Kaushik (2002). "Extensive CDR3H length heterogeneity exists in bovine foetal VDJ rearrangements." Scand J Immunol**55**(2): 140-148.
- Sanders, B. G. and J. C. Travis (1975). "Evidence for one immunoglobulin light-chain type in chickens: absence of a blocked N-terminal light-chain type." Biochem Genet**13**(11-12): 779-782.
- Saphire, E. O., P. W. Parren, R. Pantophlet, M. B. Zwick, G. M. Morris, P. M. Rudd, R. A. Dwek, R. L. Stanfield, D. R. Burton and I. A. Wilson (2001). "Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design." Science**293**(5532): 1155-1159.
- Schroeder, H. W., Jr. and J. Y. Wang (1990). "Preferential utilization of conserved immunoglobulin heavy chain variable gene segments during human fetal life." Proc Natl Acad Sci U S A**87**(16): 6146-6150.
- Shojaei, F., S. S. Saini and A. K. Kaushik (2003). "Unusually long germline DH genes contribute to large sized CDR3H in bovine antibodies." Mol Immunol**40**(1): 61-67.
- Sinclair, M. C., J. Gilchrist and R. Aitken (1997). "Bovine IgG repertoire is dominated by a single diversified VH gene family." J Immunol**159**(8): 3883-3889.
- Sun, J., I. Kacs Kovics, W. R. Brown and J. E. Butler (1994). "Expressed swine VH genes belong to a small VH gene family homologous to human VHIII." J Immunol**153**(12): 5618-5627.
- Sun, Y., C. Wang, Y. Wang, T. Zhang, L. Ren, X. Hu, R. Zhang, Q. Meng, Y. Guo, J. Fei, N. Li and Y. Zhao (2010). "A comprehensive analysis of germline and expressed immunoglobulin repertoire in the horse." Dev Comp Immunol**34**(9): 1009-1020.
- Thompson, C. B. (1992). "Creation of immunoglobulin diversity by intrachromosomal gene conversion." Trends Genet**8**(12): 416-422.
- Thompson, C. B. and P. E. Neiman (1987). "Somatic diversification of the chicken immunoglobulin light chain gene is limited to the rearranged variable gene segment." Cell**48**(3): 369-378.
- Tobin-Janzen, T. C. and J. E. Womack (1992). "Comparative mapping of IGHG1, IGHM, FES, and FOS in domestic cattle." Immunogenetics**36**(3): 157-165.
- Tonegawa, S. (1983). "Somatic generation of antibody diversity." Nature**302**(5909): 575-581.
- Tutter, A. and R. Riblet (1989). "Conservation of an immunoglobulin variable-region gene family indicates a specific, noncoding function." Proc Natl Acad Sci U S A**86**(19): 7460-7464.
- Verma, S. and R. Aitken (2012). "Somatic hypermutation leads to diversification of the heavy chain immunoglobulin repertoire in cattle." Vet Immunol Immunopathol**145**(1-2): 14-22.
- Walther, S., C. P. Czerny and U. S. Diesterbeck (2013). "Exceptionally long CDR3H are not isotype restricted in bovine immunoglobulins." PLoS One**8**(5): e64234.
- Warr, G. W., K. E. Magor and D. A. Higgins (1995). "IgY: clues to the origins of modern

- antibodies." Immunol Today**16**(8): 392-398.
- Wei, Z., Q. Wu, L. Ren, X. Hu, Y. Guo, G. W. Warr, L. Hammarstrom, N. Li and Y. Zhao (2009). "Expression of IgM, IgD, and IgY in a reptile, *Anolis carolinensis*." J Immunol**183**(6): 3858-3864.
- Weill, J. C. and C. A. Reynaud (1996). "Rearrangement/hypermutation/gene conversion: when, where and why?" Immunol Today**17**(2): 92-97.
- Wilson, M., E. Bengten, N. W. Miller, L. W. Clem, L. Du Pasquier and G. W. Warr (1997). "A novel chimeric Ig heavy chain from a teleost fish shares similarities to IgD." Proc Natl Acad Sci U S A**94**(9): 4593-4597.
- Wu, Y. C., D. Kipling, H. S. Leong, V. Martin, A. A. Ademokun and D. K. Dunn-Walters (2010). "High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations." Blood**116**(7): 1070-1078.
- Zhang, Y. A., P. Nie, H. Y. Luo, Y. P. Wang, Y. H. Sun and Z. Y. Zhu (2003). "Characterization of cDNA encoding immunoglobulin light chain of the mandarin fish (*Siniperca chuatsi*)." Vet Immunol Immunopathol**95**(1-2): 81-90.
- Zhao, M., R. J. Naude, K. Muramoto and W. Oelofsen (1996). "Purification and characterization of ostrich pancreatic secretory trypsin inhibitor." Int J Pept Protein Res**48**(2): 174-181.
- Zhao, Y., H. Cui, C. M. Whittington, Z. Wei, X. Zhang, Z. Zhang, L. Yu, L. Ren, X. Hu, Y. Zhang, L. Hellman, K. Belov, N. Li and L. Hammarstrom (2009). "Ornithorhynchus anatinus (platypus) links the evolution of immunoglobulin genes in eutherian mammals and nonmammalian tetrapods." J Immunol**183**(5): 3285-3293.
- Zhao, Y., S. M. Jackson and R. Aitken (2006). "The bovine antibody repertoire." Dev Comp Immunol**30**(1-2): 175-186.
- Zhao, Y., I. Kacs Kovics, Q. Pan, D. A. Liberles, J. Geli, S. K. Davis, H. Rabbani and L. Hammarstrom (2002). "Artiodactyl IgD: the missing link." J Immunol**169**(8): 4408-4416.
- Zhao, Y., I. Kacs Kovics, H. Rabbani and L. Hammarstrom (2003). "Physical mapping of the bovine immunoglobulin heavy chain constant region gene locus." J Biol Chem**278**(37): 35024-35032.
- Zhao, Y., Q. Pan-Hammarstrom, S. Yu, N. Wertz, X. Zhang, N. Li, J. E. Butler and L. Hammarstrom (2006). "Identification of IgF, a hinge-region-containing Ig class, and IgD in *Xenopus tropicalis*." Proc Natl Acad Sci U S A**103**(32): 12087-12092.
- Zhao, Y., H. Rabbani, A. Shimizu and L. Hammarstrom (2000). "Mapping of the chicken immunoglobulin heavy-chain constant region gene locus reveals an inverted alpha gene upstream of a condensed epsilon gene." Immunology**101**(3): 348-353.
- Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassel, T. S. Sonstegard, G. Marcais, M. Roberts, P. Subramanian, J. A. Yorke and S. L. Salzberg (2009). "A whole-genome assembly of the domestic cow, *Bos taurus*." Genome Biol**10**(4): R42.