

Capítulo 1

Introducción

La teoría de la *Evolución* (Darwin, 1859) expresa que cada especie se derivó a partir de otra más simple. Esto ocurre gracias a un proceso llamado *Selección Natural*, donde el medio ambiente escoge de entre los individuos que compiten entre sí a los individuos genéticamente más aptos para sobrevivir y reproducirse. Darwin hizo una construcción de un árbol evolutivo de acuerdo con las características físicas de las especies. Después, conforme los científicos avanzaron en sus investigaciones, se fueron dando cuenta de que la construcción hecha por Darwin era valiosa pero un tanto "áspera"; pues solo comparaba las especies por su apariencia externa, zona geográfica y alimentación, entre otros. Esta información no podía responder con objetividad cuál era el orden consecutivo en que fueron surgiendo las especies, ni el tiempo que había transcurrido desde su divergencia. Además, los científicos no estaban totalmente de acuerdo en que la apariencia externa pudiera ser el factor determinante. Poco a poco fueron surgiendo otras ramas de la ciencia que también han aportado datos valiosos: 1) la *Anatomía Comparada*, que hace comparaciones de las estructuras óseas, identificando patrones de formación de los huesos; 2) la *Paleontología*, que usa fósiles para calcular el tiempo en que existieron las especies incluso muchas de ellas ya desaparecidas, además, proporciona datos de la estructura ósea, apariencia y en algunos casos hasta de las posibles formas de alimentación, movimiento y comportamiento; 3) la *Embriología Comparada*, que reconstruye la historia evolutiva comparando patrones de desarrollo del embrión; y 4) la *Inmunología y Bioquímica*, que proporcionan datos en el ámbito molecular acerca de reacciones enzimáticas, y estimación de similitud proteica y nucleotídica. Entonces, se ha reconstruido la historia evolutiva comparando desde rasgos burdos como la apariencia hasta datos moleculares. Recientemente con el desarrollo explosivo de la Biología Molecular se pueden ahora obtener las secuencias nucleotídicas de los genes (y también deducir las aminoácidos correspondientes) y con éstas, hacer comparaciones hasta el nivel nucleotídico, el nivel *elemental* del *programa genético* responsable del desarrollo de los organismos. Esta última información ha ido modificando, clarificando e incluso corrigiendo lo que se sabía anteriormente (O'Brien, 1993), creando el camino para una rama de la ciencia, relativamente nueva: La *Evolución Molecular*.

1.1 Evolución Molecular

La Evolución Molecular comprende tres áreas de estudio: (1) *La evolución de macromoléculas*, que se refiere a las velocidades y patrones de cambio que ocurren en el material genético y sus productos durante la evolución, así como a los posibles mecanismos de dichos cambios; (2) *la filogenia molecular* que estudia la reconstrucción de la historia evolutiva de macromoléculas y organismos deducidos de datos moleculares, y (3) *el origen*

de la vida, que estudia los posibles eventos y mecanismos que pudieron originar la vida en nuestro planeta (Li y Graur, 1991).

Es bien conocido que el DNA de todos los organismos está expuesto a cambios en su secuencia por diversos mecanismos llamados *recombinaciones*, *deleciones*, *inserciones* y *mutaciones puntuales*. Estos cambios ocurren a lo largo de todo el genoma y con relativa frecuencia. Sin embargo, dada la relevancia de la región mutada y el tipo de mutación, los cambios no *persisten* uniformemente en todo el genoma a través de las generaciones. Se ha demostrado en diversos estudios que las velocidades de dichos cambios (número de cambios en un sitio por unidad de tiempo) varían entre las diferentes regiones genéticas (regiones codificantes reguladoras e intergénicas); y aún entre ellas mismas (Tabla 1.1). Por estas razones, es deseable analizar la secuencia más completa del gen. Comparando las secuencias nucleotídicas o aminoácidas de un mismo gen entre diferentes especies, se pueden calcular en función del número de cambios y sus patrones, la cronología y las velocidades a las que ocurrieron estos cambios. Estos datos sirven para construir árboles filogenéticos que relacionan las proteínas o las especies entre sí. Al hacer reconstrucciones de árboles filogenéticos, se necesita de secuencias que tengan al menos algunas variaciones, ya que sin éstas no se podría hacer la reconstrucción. Al mismo tiempo, si existen muchas variaciones entre las secuencias, la reconstrucción puede ser difícil o poco confiable. Generalmente se usan secuencias que tengan variación moderada y que estén distribuidas entre las especies en estudio. También se usan secuencias de baja relevancia como pseudogenes para normalizar las estimaciones en tiempo de divergencia en las reconstrucciones.

Secuencia [‡]	Longitud (aa)	Velocidad [†]
Histona 3 (mutaciones no sinónimas [‡])	135	0.00 ± 0.00
Histona 3 (mutaciones sinónimas)	135	6.38 ± 1.19
Hormona de Crecimiento (mutaciones no sinónimas)	189	1.23 ± 0.15
Hormona de Crecimiento (mutaciones sinónimas)	189	4.95 ± 0.77
Interferón γ (mutaciones no sinónimas)	136	2.79 ± 0.31
Interferón γ (mutaciones sinónimas)	136	8.59 ± 2.56
5' UTR Hormona de Crecimiento	21	1.79 ± 0.85
3' UTR Hormona de Crecimiento	91	1.83 ± 0.41
5' UTR Apolipoproteína A-IV	83	3.06 ± 0.68
3' UTR Apolipoproteína A-IV	134	1.73 ± 0.33
α y β Globina pseudogenes (vaca y cabra)		9.10 ± 0.90

[‡] Varían con la relevancia de la región y el tipo de mutación. Los cálculos fueron hechos a partir de secuencias de humano, ratón y rata, excepto donde se indica.

[†] El tiempo de divergencia se fijó en 80 millones de años. Las velocidades son mostradas en unidades de número de sustituciones por sitio por 10^9 años (extraído de Li y Graur, 1991).

[‡] Mutaciones no sinónimas son las que ocurren en las secuencias codificantes y cuya mutación produce también un cambio de aminoácido.

Se pueden construir árboles filogenéticos a partir de secuencias aminoácidas o nucleotídicas. Sin embargo, la mayor resolución se alcanza con las últimas, ya que cada residuo de aminoácido es codificado por 3 nucleótidos, dando 61 tripletes o codones para los 20 aminoácidos. En la actualidad es más fácil obtener la secuencia nucleotídica que la

aminoacídica; además, de la primera se puede deducir la segunda. De igual manera se pueden construir árboles filogenéticos de secuencias codificantes, no codificantes o ambas, y dado que las velocidades de cambio dependen de la cantidad de diferencias por nucleótido, entre más extensa sea la secuencia por analizar, mayor será la resolución en el cálculo de las velocidades.

Conforme el número de secuencias entre especies evolutivamente "lejanas" aumente, se incrementa la ramificación del árbol filogenético de las especies y se eleva la confiabilidad de sus bifurcaciones; y conforme aumente el número de secuencias entre especies evolutivamente "cercanas", se van resolviendo las ramillas o las hojas de su rama. En la actualidad existen programas computacionales que ayudan en la realización de estos cálculos, y en la reconstrucción de árboles filogenéticos basándose en ciertos criterios y algoritmos. Entre otros, se encuentran el PHYLIP y CLUSTALW (Higgins, 1994) que realizan un análisis comparativo de las secuencias y proporcionan un árbol, matriz de cambios y/o alineamiento de las secuencias.

En síntesis, para construir un árbol filogenético, se requiere: (1) escoger el gen a estudiar como modelo evolutivo; (2) aumentar el número de secuencias conocidas de este mismo gen entre las diferentes especies, requiriéndose una metodología que facilite la obtención de dichas secuencias; y (3) utilizar modelos matemáticos para analizar las secuencias y construir el árbol evolutivo.

1.2 Hormona del Crecimiento como Modelo Evolutivo

Las hormonas del crecimiento (GHs) pertenecen a una familia de hormonas con similitud estructural y que incluye también a las prolactinas y lactógenos placentarios (PLs) (Sherwood, 1971; Niall y cols, 1971). Se cree que en los pre-vertebrados los genes de prolactina y GH se originaron de un gen ancestral común por duplicación génica seguida por la divergencia de los productos y que el producto polipeptídico de este gen ancestral debió parecerse más a la prolactina (Niall y cols., 1971; Seeburg, 1982; Miller & Eberhardt, 1983). En primates, una segunda duplicación ahora en el gen de la GH, dio origen a los PLs. Tanto la prolactina como la GH están presentes en todos los vertebrados estudiados hasta la fecha (Niall, 1971), entonces el estudio de la GH como modelo evolutivo nos daría datos relevantes para la reconstrucción de árboles filogenéticos exclusivamente de vertebrados. A pesar de que hay muchas secuencias reportadas, hay una tendencia hacia las especies de interés comercial y no tanto a las de relevancia evolutiva.

Tabla 1.2 Total de secuencias de la GH reportadas en GenBank.

Clases	Secuencias Nucleotídicas Reportadas
Peces	41
Anfibios	3
Reptiles	0
Aves	1
Mamíferos	17
Marsupiales	1

1.3 Hormonas del Crecimiento

Las GHs de mamíferos en su forma madura son péptidos de alrededor de 190 aminoácidos con un peso molecular de 22,000Daltones (22KDa), iniciando generalmente con un residuo fenilalanina. La forma precursora tiene alrededor de 215 aminoácidos incluyendo en su extremo amino terminal un péptido señal de 25 aminoácidos. Estas hormonas son producidas y secretadas por los somatotrofos de la hipófisis anterior y estimulan el crecimiento lineal en los animales. Las GHs contienen dos enlaces disulfuro entre cisteínas que estabilizan la estructura tridimensional activa, no contienen glucosilaciones (Niall y cols., 1973; Santomé y cols., 1973; Sato y cols., 1988) y sus características las hacen muy factibles de secuenciar (Ascacio-Martínez, 1994; Castro-Peralta, 1995) y producir en microorganismos (Escamilla, 1996; Gallardo-Blanco, 1998).

Se ha observado que todas las GHs presentan efectos fisiológicos muy similares en los vertebrados estudiados, pero a veces no solo pueden ser activas en la especie de origen sino también en especies que se sitúan abajo en la escala evolutiva. Por ejemplo, la GH bovina, se usa para estimular el crecimiento en peces, pero la GH de peces no ha mostrado actividad en bovinos (Juskevich, 1988).

La estructura de los genes de las GHs está muy conservada, formada por 5 exones separados por 4 intrones, con ciertas excepciones en algunos peces Salmoniformes que presentan 6 exones y 5 intrones. Las estructuras terciarias se caracterizan por presentar 4 hélices α que están conservadas (Figura 1.1) en mamíferos como lo muestran las estructuras cristalográficas de las GHs del cerdo (Abdel-Mequid y cols., 1987) y del humano (De Voss y cols., 1992), así como la predicha para la vaca (Carlacci y cols., 1991).

En la actualidad se cuenta con las secuencias nucleotídicas y/o aminoacídicas de la GH de varias especies particularmente en las de interés agropecuario (Tabla 1.2). Mayoritariamente solo está disponible la secuencia de aminoácidos. En cuanto a la secuencia nucleotídica, frecuentemente solo se conocen las secuencias que codifican para las proteínas (RNAm o DNA complementario - DNAc -) y éstas son todavía parciales en algunos casos. No obstante, existen muchas especies y órdenes en los cuales no se ha reportado secuencia alguna hasta ahora.

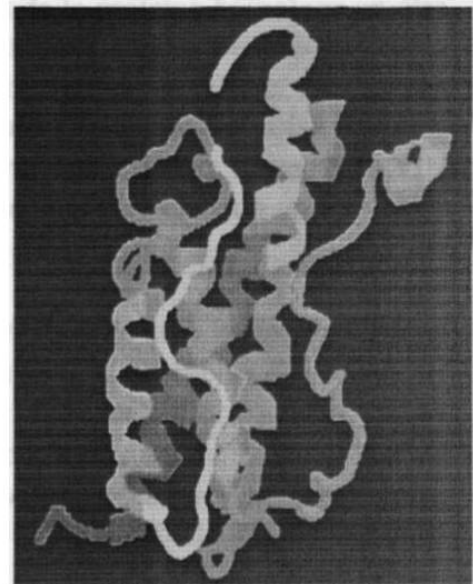


Figura 1.1 Estructura tridimensional (3D) de la GH porcina. En diferentes colores se muestran las 4 hélices α (recreada en RasMol v2.6).

1.4 Obtención de las Secuencias Nucleotídicas

El material genético se puede obtener tanto del DNA de células nucleadas e indirectamente del RNAm de células que expresan el gen a estudiar, una vez que en el laboratorio se retrotranscribe a DNA por la acción enzimática de una transcriptasa reversa. El DNA resultado de transcripción reversa es llamado DNA complementario (DNAc) y no incluye intrones, ya que proviene del RNAm maduro.

Originalmente las secuencias nucleotídicas se obtuvieron al tamizar bancos de genes o genotecas de DNA genómico (DNAg) y bancos de DNAc. Para construirlos, se purifican a partir del tejido de interés, el DNAg o los RNAm. El material se fragmenta cortando con enzimas de restricción el DNAg o se prepara retro-transcribiendo el RNAm en DNAc. La construcción del banco se logra insertando los fragmentos de DNA en un vector derivado de un fago lítico por ejemplo, que se usa para infectar bacterias. La selección de las clonas que tienen el fragmento de interés se puede hacer por hibridación del DNA con sondas homólogas o heterólogas, o por actividad inmunológica o enzimática del producto. Ya seleccionada la colonia fágica, se extrae el DNA, se corta y se obtiene la secuencia del fragmento que contiene. Así, se obtuvo de hipófisis de la trucha arco iris *Oncorhynchus mykiss*, el DNAc de la GH (Agellon y cols., 1986). Cuando no se tiene información sobre la proteína en estudio de una especie dada, pero se conoce la de la proteína homóloga en otra especie relativamente cercana, se pueden utilizar en el tamizaje sondas heterólogas de la especie conocida. De esta manera, por ejemplo, se obtuvieron el DNAc de *Caranx delicatissimus* al usar una sonda de *Seriola quinqueradiata* (Yamakawa y cols., 1988) y para el DNAc de cabra, se utilizó el DNAc de vaca (Yamano y cols. 1988). El gen de la GH bovina se obtuvo a partir de DNAg de hígado cortado con *EcoRI* y clonado en un vector derivado del bacteriófago λ (Gordon y cols. 1982).

Con el incremento de secuencias reportadas y con la aparición de la Reacción en Cadena de la Polimerasa conocida como PCR (por sus siglas del inglés), la obtención de secuencias nucleotídicas se ha facilitado enormemente. Al tener los oligonucleótidos adecuados, se puede amplificar el fragmento de interés y evitar la construcción del banco. Así se obtuvieron la secuencia de DNAc de perro *Canis familiaris* (Ascacio-Martinez y cols., 1995), de gato *Felis catus* (Castro-Peralta y cols., 1995), y de caballo *Equus caballus* (Ascacio-Martinez y cols., 1995).

1.5 Justificación

Diseñando oligonucleótidos consenso de un grupo de especies, se puede amplificar específicamente el gen de la GH por PCR sin la necesidad de construir un banco. Para los casos en los que no se logra obtener un producto de PCR de la GH usando los oligonucleótidos diseñados y para aprovechar los bancos construidos, se puede diseñar un oligonucleótido consenso que sirva para sondear un banco.

En la actualidad, se conocen al menos 4060 especies vivientes de mamíferos, de las cuales 171 (4%) corresponden a Artiodáctilos y a pesar de que es el 5° orden con más especies, es el orden con más secuencias conocidas de DNA de la GH (7 especies, 5%) gracias a su interés agropecuario. Sin embargo de la familia de *Giraffa camelopardalis*, todavía no existe una secuencia de GH reportada. Los Artiodáctilos, constituyen entonces un buen modelo para tratar de diseñar oligonucleótidos consenso que permitan amplificar sus genes de la GHs. Al mismo tiempo, se podría diseñar un oligonucleótido consenso que al usarse como sonda y facilitar el aislamiento de los genes de las GHs de cualquier vertebrado.

Capítulo 2

Objetivos

Contribuir al conocimiento de la filogenia molecular, facilitando la obtención de nuevas secuencias de la hormona del crecimiento mediante el desarrollo de oligonucleótidos específicos para diferentes órdenes de vertebrados.

2.1 Objetivo General

Diseñar un método que facilite la detección de genes de las hormonas del crecimiento de vertebrados.

2.2 Objetivos Particulares

1. Recuperar las secuencias nucleotídicas y/o aminoacídicas de las hormonas del crecimiento (GH) reportadas en los bancos de información existente.
2. En función de la clasificación de las especies, agrupar y alinear dichas secuencias de GHs, seleccionando de cada especie el informe de la secuencia más completa o más actualizada.
3. Seleccionar una región altamente conservada en la escala evolutiva que permita diseñar un(os) oligonucleótido(s) como sonda para tamizar bancos de DNAg o DNAc.
4. Comprobar la utilidad los oligonucleótidos tanto en el laboratorio (por hibridación con DNA de varias especies) y en computador (por comparación con las secuencias reportadas).
5. En los órdenes donde sea factible, diseñar oligonucleótidos consenso que permitan obtener el gen de la GH, amplificando por PCR.
6. Confirmar la utilidad del juego de oligonucleótidos diseñado para Artiodáctilos, amplificando a partir del DNA genómico de jirafa, su gen de la GH.

Capítulo 3

Material y Métodos

3.1 Origen de los Reactivos

Las enzimas de restricción fueron adquiridas de Gibco-Bethesda Research Laboratories, Inc. (Gibco-BRL, Gaithersburg, MD, EUA) y de New England Biolabs, Inc. (NEB, Beverly, MA, EUA).

La enzima *Taq* DNA polimerasa, dNTP's y reactivos necesarios para realizar la PCR fueron adquiridos de Promega (Madison, WI, USA).

La enzima T4 polinucleótido cinasa y su solución amortiguadora fue adquirida de New England Biolabs, Inc. (United Kingdom).

El juego de hexanucleótidos fue adquirido de Gibco BRL (Gaithersburg, MD, EUA).

Los nucleótidos radioactivos [³²P]-αdATP y [³²P]-γdATP NEG502Z se adquirieron de las compañías Amersham International (Buckinghamshire, UK) y New England Nuclear (Boston, MA, EUA) respectivamente.

El DNA del fago λ usado como control negativo marcador de peso molecular en los geles fue adquirido en Gibco BRL (Gaithersburg, MD, EUA).

El Kit de secuenciación AmpliCycle fue de Perkin-Elmer Cetus (Norwalk, CT, EUA).

El oligonucleótido (*GH-Explorer I*) usado como sonda general de la GH en ensayos de hibridación fue sintetizado por Life Technologies (Gibco-BRL, Gaithersburg, MD, EUA).

El juego de oligonucleótidos para Artiodáctilos fue sintetizado en el laboratorio de Química Biomolecular de la ULIEG.

3.2 Material Biológico

El vector utilizado para la clonación del producto amplificado de la hormona del crecimiento de jirafa, pBS(II)-SK+, se obtuvo de la plasmidoteca de la ULIEG.

El plásmido pPic9kBGH (Gallardo y cols., 1998) usado para obtener un fragmento de 494pb utilizado como sonda fue obtenido de la plasmidoteca de la ULIEG, construido y proporcionado amablemente por el Biol. Hugo Gallardo y el Dr. Hugo Barrera.

El plásmido pBR322 usado como control negativo y de peso molecular en los experimentos de hibridación fueron obtenidos de la plasmidoteca de la ULIEG y proporcionados por J. Claudio Moreno Rocha.

El plásmido pBS-hPL-3 usado como marcador de peso molecular fue obtenido de la plasmidoteca de la ULIEG, proporcionado y construido por el Biólogo y M. C. J. Claudio Moreno Rocha.

El plásmido pSV-hPL-1 usado como problema en una hibridación fue obtenido de la plasmidoteca de la ULIEG.

El tejido placentario y sangre de jirafa así como la sangre de gamo, cebrá y avestruz utilizados para la extracción de DNA, amplificación y clonación de la GH, fueron

proporcionados por la Ing. Rocío García Rojas del Departamento de Veterinaria de Africam Safari de Puebla, México.

Las muestras obtenidas de sangre de puerco, vaca y cabra utilizados para la extracción de DNA, fueron obtenidas del Rastro Municipal de Monterrey, Nuevo León, México.

Los tejidos de tiburón, catán, guppy, carpa, tilapia, ameca y serpiente fueron donados por el Dr. Roberto Mendoza y por el Dr. Carlos Aguirre del Laboratorio de Fisiología y del Museo de Historia Natural respectivamente, ambos de la Facultad de Ciencias Biológicas de la UANL.

El tejido de pollo fue donado por Vicant Representantes Internacionales S. de R.L.C.V., en Monterrey, Nuevo León, México.

El tejido de lagartija y tortuga fue donado por Lucia Elizabeth Cuéllar Chávez.

El tejido placentario de gato fue donado por el Departamento de Fisiología de la Facultad de Medicina de la UANL y proporcionado por M.C. y Q.F.B. Ana María Sifuentes Rincón.

Una fracción del DNA humano fue extraída de sangre del M.C. y Biol. J. Claudio Moreno Rocha del laboratorio de Biología Molecular de la ULIEG.

El DNA de gamo, búfalo americano y una fracción del DNA humano fueron proporcionados por la Q.C.B. Dolores Esquivel del laboratorio de Biología Molecular de la ULIEG de la Facultad de Medicina de la UANL.

El tejido de la cepa de ratón CD1 fue donado por el Departamento de Inmunología de la Facultad de Medicina de la UANL, y proporcionado por M.C. y Q.F.B. Irma Martínez Ávila.

Todo el material biológico, aunque no representa algún riesgo para la salud, fue manejado siguiendo los requerimientos impuestos por la Secretaría de Salud. El manejo de productos radiológicos se realizó siguiendo los lineamientos de la Comisión Nacional de Seguridad Nuclear y Salvaguarda de la Secretaría de Energía.

3.3 Equipo de Laboratorio

Se utilizaron microcentrifugas *Eppendorf* modelos 5402 y 5415, fuentes de poder *EC600* e *ISCO494*, un agitador de tubos (vortex) modelo *Maxi Mix*, una placa agitadora con control de temperatura modelo *PC320* marca *Corning*, un ultracongelador marca *So-Low* de *Environmental Equipment*, una campana de flujo laminar marca *Labconco Corporation*, dos termocicladores modelo *PT150* y *PT100* marca *MJ Research*, una báscula digital marca *Sartorius* modelo *1206MP*, un equipo de secuenciación *Genomyx* de *Genomyx Corporation*, un horno incubador *Hybrid Incubator*, así como material clásico de laboratorio.

3.4 Paquetes Computacionales

Se utilizaron los paquetes incluidos en *Microsoft Office '97*, *Word* como procesador de texto, *PowerPoint* como editor de diapositivas y *Excel* como creador de hojas de cálculo. Un sistema *UMAX Scan* (©1990-1991 *UMAX Scanner Driver*, *Impact Research Inc*) de

adquisición de imágenes y como manejador Adobe Photoshop Limited Edition 2.5.1. Un sistema modelo Gel-Doc 1000 de BioRad (U. K.) como digitalizador de imágenes de geles y/o películas y el *Molecular Analyst* incluido con este sistema como cuantificador.

Para análisis en Biología Molecular, se usó *Amplify* v1.2b (Bill Engels © 1992 University of Wisconsin Genetics M. Madison, WI, EUA), *DNA Strider™* v1.1 (© Chemical Merck and C.E.A. 1989 Service de Biochimie-Départament de Biologie-Institut de Recherche Fondamentale-CEA-FRANCE), *OLIGO* v4.0 (© 1992, Plymouth, Mn, EUA).

Para los alineamientos se utilizó *CLUSTALW* versión PC.

Para los análisis densitométricos de las hibridaciones se utilizó *ImageTool* v1.28 (© 1995-1997 The University of Texas Health Science Center en San Antonio, Texas, EUA).

Para el almacenamiento en base de datos y desarrollo de programas se utilizó *Visual Basic* v5.0 (© Microsoft Corp.) y *Visual C++* v5.0 (© Microsoft Corp.) respectivamente.

3.5 Estrategia General

Para conseguir los objetivos planteados en este proyecto, se propuso la estrategia general que se muestra en la Figura 3.1.

3.6 Obtención y Análisis de las Secuencias Nucleotídicas y Aminoacídicas de GH

3.6.1 Obtención de las Secuencias

Para obtener las secuencias, se usaron las direcciones de Internet del GenBank:

www.ncbi.nlm.nih.gov/Entrez/nucleotide.html

www.ncbi.nlm.nih.gov/Entrez/protein.html

Se introdujo la consulta:

<"growth hormone"[Protein Name] | "somatotropin"[Protein Name]>*

* El símbolo "|" en computación significa "OR" que quiere decir "O" (alternativa) y se usa para no provocar confusión con la palabra "OR". Los símbolos "<" ">" se utilizan para delimitar una expresión literal.

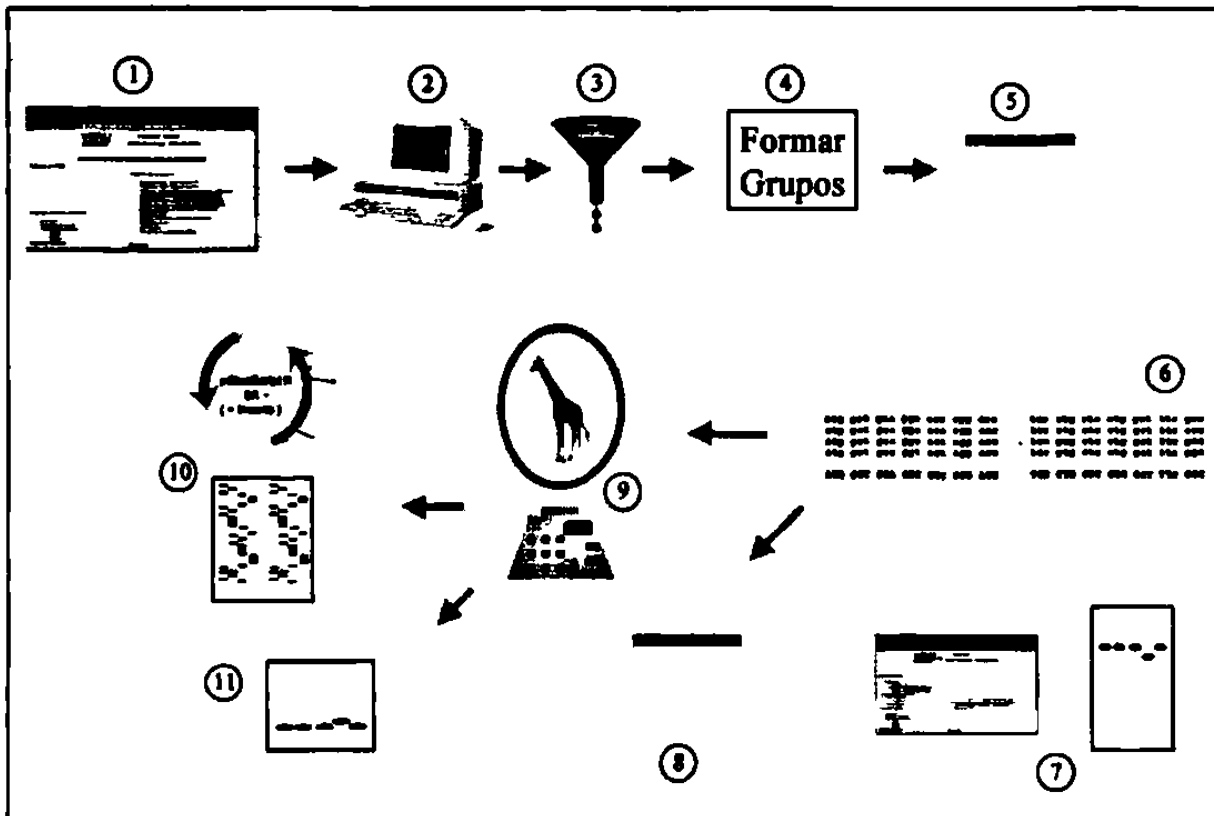


Figura 3.1 Estrategia General. Se ilustran los pasos seguidos en el trabajo y que son: 1) Obtención de las Secuencias. 2) Conversión y almacenamiento local. 3) Filtrado. 4) Agrupación. 5) Alineamiento por grupos. 6) Diseño de iniciadores consenso y sonda(s). 7) Comprobación teórica-virtual y/o práctica de la(s) sonda(s). 8) Comprobación teórica-virtual de los iniciadores. 9) Amplificación de la GH de jirafa usando iniciadores de Artiodáctilos. 10) y 11) Comprobación del producto amplificado por clonación y secuenciación y/o por hibridación con sonda conocida.

Es conocido que existen versiones diferentes de la hormona y, además, variaciones al nombrarlas teniendo nombres como “growth hormone a”, “b”, “c”, “1”, “2”, “i”, “ii”, algunos otros como “growth hormone precursor”, “pregrowth hormone” y combinaciones. Para resolver esto, primero se obtuvieron todas las variaciones en cada uno de los dos términos usando el valor “List Terms” en la opción “Mode” en la misma página. Los términos usados fueron los siguientes:

growth hormone
 growth hormone 12
 growth hormone 9d
 growth hormone alt
 growth hormone factor i
 growth hormone gh precursor
 growth hormone ii
 growth hormone precursor aa
 17 to
 growth hormone precursor aa
 25 to 191
 growth hormone prohormone
 growth hormone wt

 somatotropin
 somatotropin 2 splice frame 2
 precursor

growth hormone 1
 growth hormone 2
 growth hormone a
 growth hormone b
 growth hormone gh 1
 growth hormone i
 growth hormone ii precursor
 growth hormone precursor aa
 17 to 187
 growth hormone precursor aa
 26 to 190
 growth hormone type 2
 growth prehormone aa 12 to
 173
 somatotropin 1 precursor
 somatotropin 20k

growth hormone 1 precursor
 growth hormone 2 precursor
 growth hormone aa 1 217
 growth hormone c variant
 growth hormone gh 2
 growth hormone i precursor
 growth hormone precursor
 growth hormone precursor aa
 23 to 188
 growth hormone prepeptide

 growth hormone variant
 pregrowth hormone

 somatotropin 2 precursor
 somatotropin 22k

somatotropin a
 somatotropin i
 somatotropin ii precursor
 somatotropin, variant

somatotropin alpha chain
 somatotropin i precursor
 somatotropin precursor i

somatotropin b
 somatotropin ii
 somatotropin precursor ii

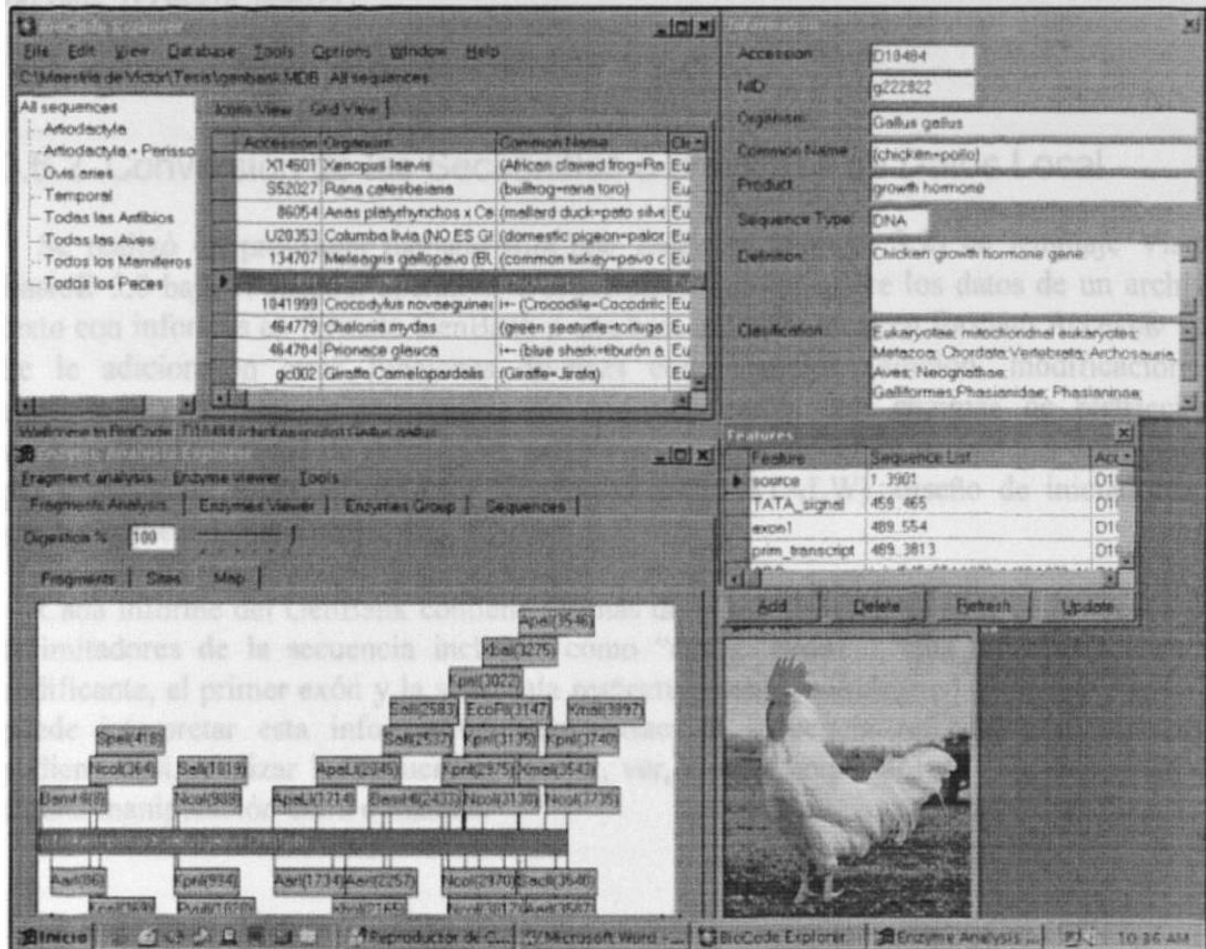


Figura 3.2 Vista de *BioCode Explorer*. Se muestra información de la secuencia seleccionada y un mapa de restricción

Por lo tanto, la consulta definitiva fue:

```

((((((((((((((((((((((((((((((((((((((((("growth hormone"[Protein Name] OR "growth
hormone 1"[Protein Name]) OR "growth hormone 1 precursor"[Protein Name]) OR
"growth hormone 12"[Protein Name]) OR "growth hormone 2"[Protein Name]) OR
"growth hormone 2 precursor"[Protein Name]) OR "growth hormone 9d"[Protein Name])
OR "growth hormone a"[Protein Name]) OR "growth hormone aa 1 217"[Protein Name])
OR "growth hormone alt"[Protein Name]) OR "growth hormone b"[Protein Name]) OR
"growth hormone c variant"[Protein Name]) OR "growth hormone factor i"[Protein
Name]) OR "growth hormone gh 1"[Protein Name]) OR "growth hormone gh 2"[Protein
Name]) OR "growth hormone gh precursor"[Protein Name]) OR "growth hormone
i"[Protein Name]) OR "growth hormone i precursor"[Protein Name]) OR "growth
hormone ii"[Protein Name]) OR "growth hormone ii precursor"[Protein Name]) OR
"growth hormone precursor"[Protein Name]) OR "growth hormone precursor aa 17
to"[Protein Name]) OR "growth hormone precursor aa 17 to 187"[Protein Name]) OR
"growth hormone precursor aa 23 to 188"[Protein Name]) OR "growth hormone
precursor aa 25 to 191"[Protein Name]) OR "growth hormone precursor aa 26 to
190"[Protein Name]) OR "growth hormone prepeptide"[Protein Name]) OR "growth
hormone prohormone"[Protein Name]) OR "growth hormone type 2"[Protein Name]) OR
"growth hormone variant"[Protein Name]) OR "growth hormone wt"[Protein Name]) OR
("growth prehormone aa 12 to 173"[Protein Name] OR "somatotropin"[Protein Name])
    
```

```
OR (((((((((((("somatotropin 1 precursor"[Protein Name] OR "somatotropin 2
precursor"[Protein Name]) OR "somatotropin 2 splice form 2 precursor"[Protein
Name]) OR "somatotropin 20k"[Protein Name]) OR "somatotropin 22k"[Protein Name])
OR "somatotropin a"[Protein Name]) OR "somatotropin alpha chain"[Protein Name])
OR "somatotropin b"[Protein Name]) OR "somatotropin i"[Protein Name]) OR
"somatotropin i precursor"[Protein Name]) OR "somatotropin ii"[Protein Name]) OR
"somatotropin ii precursor"[Protein Name]) OR "somatotropin precursor i"[Protein
Name]) OR "somatotropin precursor ii"[Protein Name]) OR "somatotropin,
variant"[Protein Name]))
```

3.6.2 Conversión de las Secuencias a una Base de Datos Local

Se realizó un programa llamado *BioCode Explorer* (Figura 3.2) en lenguaje Visual Basic® 5.0 bajo el sistema operativo Windows®95 que transfiere los datos de un archivo texto con informes en formato GenBank a una base de datos local en formato Access® 1.1. Se le adicionaron otras funciones diversas como edición (adición, modificación y eliminación) de secuencias, análisis de codones, cortes con enzimas de restricción, visualización del mapa de restricción, extracción y exportación de secuencias delimitadas por marcadores, alineamiento (por ligación con CLUSTALW), diseño de iniciadores y sondas consenso, entre otros.

Cada informe del GenBank contiene además de la información general de la secuencia, delimitadores de la secuencia incluida, como "cds", "exon1", "tata_signal" (secuencia codificante, el primer exón y la señal tata respectivamente, entre otros). *BioCode Explorer* puede interpretar esta información para extraer la secuencia referida y manipularla, pudiendo así, analizar las secuencias (cortar, ver, alinear, exportar, etc.) sin necesidad de alguna manipulación extra o manual.

3.6.3 Selección de la Mejor Secuencia de Cada Especie

Con *BioCode Explorer* se ordenaron las secuencias por especie (organismo). Se activaron las "ventanas" de información acerca de las secuencias. En cada especie, se seleccionaron aquellos informes obsoletos o incompletos y se eliminaron activando la opción "delete sequences". Éste proceso se repitió hasta tener una sola secuencia de cada especie.

3.6.4 Agrupación de las Secuencias

Usando *BioCode Explorer* se ordenaron las secuencias en base a la clasificación y se exportaron usando la opción "Export Information" y se analizaron en *Excel*. El nivel de la clasificación para la agrupación fue variable según el número de secuencias "cercanas" reportadas. Se efectuó una segunda agrupación por clases. Las clases fueron: Peces, Anfibios, Reptiles, Aves, Mamíferos y Marsupiales.

3.6.5 Alineamiento de las Secuencias

Se construyó una matriz de alineamiento (% similitud) de todas las secuencias utilizando *BioCode Explorer* para ordenar por clasificación y para extraer las secuencias aminoacídicas y *CLUSTALW* para calcular el porcentaje de similitud. Cada par de secuencias fue usado como entrada para *CLUSTALW*. Los resultados fueron graficados utilizando *Excel*.

3.7 Diseño de la Sonda “consenso” para las GH

3.7.1 Selección de la Región Conservada como Sonda

Para buscar los posibles oligonucleótidos, primero se elaboró un programa (en lenguaje C++) que suma el número total de secuencias en que *aparea totalmente* un oligonucleótido y se ejecutó con todos los posibles oligonucleótidos de 16 bases de longitud derivados de las todas secuencias codificantes; posteriormente se elaboró otro programa (*MaxProbe*) que calcula el consenso de un oligonucleótido computando el mejor *alineamiento* con cada una de las secuencias y se ejecutó con todos los posibles oligonucleótidos de 16 y 32 bases de longitud de las todas secuencias codificantes. Los resultados de ambos programas se compararon con oligonucleótidos de 16 bases de longitud. El alineamiento utilizado en *MaxProbe* se consiguió utilizando el algoritmo de selección de la mejor diagonal (White, 1984) sin considerar inserciones ni deleciones (sin “gaps”). No se calcularon aquellos oligonucleótidos que formaran parte de dos exones siempre y cuando ésta información estuviera disponible en el informe original. Para seleccionar el mejor oligonucleótido entre “los mejores”, se diseñó una función (f_{sonda} descrita en el capítulo 4) que considera: a) Alineamiento, b) Apareamiento en los extremos, c) Energías intramoleculares e intermoleculares, d) Temperatura de fusión media, y, adicionalmente consideramos un parámetro al que denominamos “*encadenamiento*”, que está en función del número máximo de nucleótidos consecutivos que aparezcan con el oligonucleótido en cuestión. Para determinar el peso aproximado de cada parámetro, se ejecutó el programa variando en cada ejecución un solo parámetro hasta obtener resultados satisfactorios. Para determinar los oligonucleótidos que podrían ser “los mejores”, antes de evaluarlos con f_{sonda} , se formaron dos grupos. El primer grupo se ordenó por apareamiento promedio y el segundo por apareamiento promedio menos dos desviaciones estándar; se analizaron con f_{sonda} solo el 5% de los más altos de cada grupo. Para determinar la suficiencia del 5%, se ordenaron todos los valores de apareamiento promedio y apareamiento promedio menos dos desviaciones estándar, se determinó el punto de inflexión del crecimiento de la curva resultante y se elaboró un gráfico para mostrar los resultados.

3.7.2 Aumento del Desempeño de la Sonda

Para ampliar la longitud de la sonda, a la secuencia obtenida se le agregaron nucleótidos consenso en su respectiva posición en cada extremo hasta que la evaluación (f_{sonda}) fuera menor que la inmediata anterior o rebasara el límite de 50 nucleótidos. También se eliminaron nucleótidos en los extremos mientras la evaluación fuera mayor que la inmediata anterior. Esto con el fin de corregir si al extender un extremo, pueda resultar mejor que se recorte el otro y viceversa, dado que la asignación inicial esta basada únicamente en alineamiento sin tomar en cuenta los demás parámetros.

Idealmente se deseaba que la sonda apareara en iguales circunstancias (de temperatura, apareamiento y energías) a todas las secuencias con el fin de homogeneizar los experimentos. Dado que las secuencias no son exactamente iguales, en general no es posible obtener una sonda con los mismos parámetros para todas las secuencias blanco; entonces, para minimizar la desviación de los parámetros, se diseñó un algoritmo para que funcione de la siguiente manera:

Considerando $Nt(a,b) = 1$ si el nucleótido a es igual al nucleótido b y $Nt(a,b) = 0$ en caso contrario; $Seq(c,m,d) = 1$ si el nucleótido m de la secuencia c es igual al nucleótido d , $Seq(c,m,d) = 0$ en caso contrario; y kS como el nucleótido en la posición k de la secuencia S ...

- 1) Para cada secuencia i se calcula el número de nucleótidos que aparean con el oligonucleótido en cuestión ($\#ap_i = \sum Nt(^kSec_i, ^kOligo)$).
- 2) Para cada posición j en el oligonucleótido y para cada uno de los nucleótidos se calcula el número de secuencias que tienen en la posición j ese nucleótido ($\#Sa_j = \sum Seq(^jSec_i, A)$, $\#Sg_j = \sum Seq(^jSec_i, G)$, $\#Sc_j = \sum Seq(^jSec_i, C)$, $\#St_j = \sum Seq(^jSec_i, T)$) y se calcula $\#Sx_j = \sum Seq(^jSec_i, Oligo)$ como el número de secuencias que tienen el mismo nucleótido que el oligonucleótido en cada posición j .
- 3) Se calcula $G_{j(a,g,c,i)} = \#ap_i / \#S_{(a,g,c,i)j} - \#ap_i / \#S_{xj}$ que representa la "ganancia" de cambiar la posición j por algún nucleótido.
- 4) Si para alguna $G_{j(a,g,c,i)}$ hay un valor positivo, se hace el cambio en el oligonucleótido en la posición j por el nuevo nucleótido. Se evalúa el nuevo oligonucleótido $O_{j(a,g,c,i)} = f_{sonda}$. Se vuelve a ejecutar el algoritmo desde el paso 1 tomando como entrada el nuevo oligonucleótido (solo hasta un nivel de recursividad[†] de 3, o sea, solo se pueden cambiar 3 nucleótidos en cada oligonucleótido).
- 5) De todos los valores $O_{j(a,g,c,i)}$ calculados se toma el mayor siempre y cuando rebase el valor de f_{sonda} del oligonucleótido original.

[†] En computación, "recursividad" se conoce como un proceso que puede necesitar usar su mismo proceso y que usa una condición de salida para evitar quedarse en un ciclo infinito.

3.7.3 Ajustes Manuales de la Sonda

Una vez escogida una secuencia y usando el programa diseñado, se extendió la secuencia en el extremo 3' para incrementar el desempeño en animales superiores. Además, se introdujo en una posición una degeneración de 2 nucleótidos para incrementar el desempeño exclusivamente en primates, sin afectar a las demás especies.

3.8 Comprobación de la Sonda (*GH-Explorer I*)

3.8.1 Comprobación Teórico – Virtual de la Sonda

En la página de Internet de BLAST 2.0.5 (Altschul, 1997) www.ncbi.nlm.nih.gov/cgi-bin/BLAST/mph-newblast?iform=1 se introdujo en el cuadro de opciones <-W 8> y en el cuadro de consulta se introdujo la secuencia:

```
gcctgcttcargaaggacatgcacaaggtggagacctacctgagggtcatgaagtg
```

3.8.2 Comprobación Práctica – Experimental de la Sonda

Se hicieron dos ensayos de hibridación para probar el desempeño de la sonda *GH-Explorer I*. El primer ensayo fue una hibridación tipo slot-blot en membrana de nylon con dispositivo a vacío (Ausubel et. al. 1992) usando en su mayoría 8µg de DNA genómico de varias especies (Tabla 3.1). Las especies a probar en estos ensayos fueron escogidas procurando abarcar a todos los vertebrados, seleccionando especies representativas de cada clase. La temperatura de prehibridación, hibridación y último lavado fueron 42°C, 48°C y 50°C respectivamente. Las concentraciones de sales no fueron cambiadas con respecto del protocolo original. A la solución de prehibridación del protocolo original se omitió el DNA de esperma de Salmón (ya que hibridaría con la *GH-Explorer I*). Se omitieron el paso de la membrana al filtro húmedo con solución de desnaturalización y el paso de la membrana al filtro húmedo con solución de neutralización (Sambrook et. al. 1989). La sonda fue marcada inmediatamente antes de hacer la incubación para hibridación. La exposición fue hecha en película ultrasensible con pantalla amplificadora durante 20h. Se hizo un segundo ensayo para probar la sensibilidad (Tabla 3.2). Se siguió el mismo protocolo propuesto por Ausubel et. al.; esta vez solo se omitió el DNA de esperma de Salmón en la solución de prehibridación. Se usó la misma solución de hibridación que en el ensayo anterior cuya radiactividad tenía una vida media transcurrida. La exposición fue de 72h. Ambas películas fueron sometidas a análisis densitométrico, integrándose únicamente la curva de cada banda. Para calcular la fuerza relativa de apareamiento de *GH-Explorer I* con los diferentes DNAs; se utilizaron las áreas de la titulación de DNAs de gallo para calcular una curva de calibración lineal, transformándose a logaritmos tanto la concentración de DNAs relativa a 8µg de DNAs de gallo como el área densitométrica de las muestras; con el área de todas las demás muestras se calculó un aproximado de la fracción de 8µg de DNAs de gallo que

equivaldrían a hibridar y dar una señal densitométrica aproximadamente igual; la fracción se dividió por el número de genomas relativos a gallo según la especie. Los tamaños de los genomas (Tabla 3.3) fueron extraídos de la página de Internet www.cbs.dtu.dk/databases/DOGS/index.html.

Todos los ensayos involucraron el marcaje radiactivo de la misma. La reacción para el marcaje se llevó a cabo con un volumen final de 10µL con 10U de T4 Polynucleotide Kinase, 20µCi de γ -ATP (6000Ci/mmol), 20pmoles *GH-Explorer I* y 1x del amortiguador apropiado.

Tabla 3.1 DNA de especies ensayadas en hibridación con la sonda *GH-Explorer I*.

DNA*	Pes	DNA*	Pes	DNA*	Pes	DNA*	Pes	DNA*	Pes
Humano	A1	Gallo	B1	Tilapia	C1	PBSjcGH-J5 (12ng)	D1	Gallo (800ng)	E1
Gato	A2	Avestruz	B2	Ameca (7.5µg)	C2	" (1.2ng)	D2	Gallo (80ng)	E2
Vaca	A3	Tortuga	B3	Catán	C3	" (120pg)	D3		
Jirafa	A4	Víbora	B4	Tiburón	C4	" (12pg)	D4		
Cerdo	A5	Gecko (7.5µg)	B5	Guppy (3µg)	C5	" (1.2pg)	D5		
Ratón	A6	Carpa	B6	C(-) PBR322	C6	" (0.12pg)	D6		

*Todas las muestras con 8µg de DNA, excepto donde se indica.

Tabla 3.2 Muestras de DNA usadas en el ensayo de sensibilidad en hibridación con la sonda *GH-Explorer I*.

Columna	Línea A	DNA	Línea B	DNAg	Línea C	DNA	Línea D	DNA
1	pBSjcGH-J1-Eco RI	3.2ng	Gallo	8µg	PSVhpl1-EcoRI	3.2ng	C(-) λ	2 µg
2	PBSjcGH-J1-Eco RI	320pg	Gallo	800ng	PSVhpl1-EcoRI	320pg	C(-) λ	200ng
3	PBSjcGH-J1-Eco RI	32pg	Gallo	80ng	PSVhpl1-EcoRI	32pg	C(-) λ	20ng
4	PBSjcGH-J1-Eco RI	3.2pg	Gallo	8ng	PSVhpl1-EcoRI	3.2pg	C(-) λ	2ng
5	pBSjcGH-J1-Eco RI	320fg			PSVhpl1-EcoRI	320fg	C(-) λ	200pg
6	pBSjcGH-J1-Eco RI	32fg			pSVhpl1-EcoRI	32fg	C(-) λ	20pg

Tabla 3.3 Tamaños de los genomas usados para normalizar las señales de hibridación.

Especie incluida en Ensayo		Especie Fuente	Tamaño Haploide del Genoma (Pares de bases)
Nombre Común	Nombre Científico	Nombre Científico	
Humano	<i>Homo sapiens</i>	<i>Homo sapiens</i>	3.40x10 ⁹
Gato	<i>Felis catus</i>	<i>Canis familiaris</i> [†]	3.36x10 ⁹
Vaca	<i>Bos taurus</i>	<i>Bos taurus</i>	3.65x10 ⁹
Jirafa	<i>Giraffe camelopardalis</i>	<i>Bos taurus</i> [*]	3.65x10 ⁹
Cerdo	<i>Sus scrofa</i>	<i>Sus scrofa</i>	3.11x10 ⁹
Ratón	<i>Mus musculus</i>	<i>Mus musculus</i>	3.45x10 ⁹
Gallo	<i>Gallus gallus</i>	<i>Gallus gallus</i>	1.20x10 ⁹
Avestruz	<i>Struthio camelus</i>	<i>Gallus gallus</i> [†]	1.20x10 ⁹
Tortuga	<i>Trichemys scripta</i>	<i>Caiman crocodylus</i> [*]	2.60x10 ⁹
Víbora	<i>Crotalus atrox</i>	<i>Boa constrictor</i> [*]	2.10x10 ⁹
Gecko	<i>Hemidactylos turcicus</i>	<i>Caiman crocodylus</i> [*]	2.60x10 ⁹
Carpa	<i>Cyprinus carpio</i>	<i>Cyprinus carpio</i>	1.70x10 ⁹
Tilapia	<i>Tilapia sp.</i>	<i>Cyprinus carpio</i> [*]	1.70x10 ⁹
Ameca		<i>Cyprinus carpio</i> [*]	
Catán	<i>Leptoosteus osseus</i>		1.70x10 ⁹
Tiburón [‡]	<i>Squalus sp.</i>	<i>Carcarias obscurus</i>	2.70x10 ⁹
Guppy	<i>Poecilia reticulata</i>	<i>Cyprinus carpio</i> [*]	1.70x10 ⁹

* Especie cercana cuyo tamaño del genoma si está disponible. [‡] Cazón "dogfish".

3.9 Diseño de Iniciadores para PCR

El diseño de iniciadores para PCR se cubrió en dos etapas. La primera, en artiodáctilos, se usó el programa CLUSTALW para hacer los alineamientos, manualmente se detectaron dos zonas conservadas, se utilizó OLIGO para diseñar el par de oligonucleótidos y se verificó su desempeño usando el programa Amplify. En la segunda etapa, que incluye todas las especies, se diseñó un programa, *MaxPrimer* que usa la misma metodología utilizada en *MaxProbe* para buscar las regiones consenso dentro de las secuencias y se hicieron modificaciones en la evaluación para considerar los parámetros propios del proceso de PCR como iniciabilidad¹. Se utilizó *BioCode Explorer* para extraer cuando mucho 500pb de la región 5' justo antes del codón de inicio y 1000pb de la región 3' justo después del codón de terminación de cada secuencia. Se eliminaron las secuencias de longitud menor a 22pb. Se ejecutó *MaxPrimer* con el juego de secuencias 5' y con el juego de secuencias 3'. Se comprobó el desempeño de los iniciadores de 22nt reportados por *MaxPrimer* en *Amplify*. Se comprobó la especificidad de los oligonucleótidos haciendo búsqueda en *GenBank* usando *BLAST*.

¹ Se usó "iniciabilidad" como la capacidad de potenciar la iniciación de polimerización y traduciendo del inglés "primability".

3.10 Comprobación del Juego de Oligonucleótidos Diseñados para Artiodáctilos

3.10.1 Amplificación del Gen de la GH a partir del DNAG de Artiodáctilos

Se hicieron ensayos de PCR usando el juego para amplificar la GH de cabra, vaca, oveja, jirafa, búfalo americano, cerdo y cebrá usando DNAG como templado. El DNAG de cabra, vaca y cerdo se obtuvieron a partir de sangre periférica usando la técnica de TSNT. El DNAG de búfalo americano, oveja y cebrá se obtuvieron de un stock de nuestro laboratorio. La integridad de los DNAG usados se probó en gel de agarosa. Las condiciones y concentraciones de los reactivos fueron las siguientes:

Tabla 3.4 Condiciones y pasos seguidos en los ensayos de PCR.

Reactivo	Concentración	Paso	Tiempo @ Temp
DNAG	75ng	1) Desnaturalización Inicial	3min a 94.0°C
MgCl ₂	1.5mM	2) Desnaturalización	1min a 94.0°C
dNTP's	0.2mM	3) Apareamiento	1min a 58°C~59.5°C
pArt5'-140	0.5µM	4) Extensión	1min a 72.0°C
pArt3'-139	0.5µM	30 veces: Pasos 2~4	
Amortiguador	1.0x	5) Extensión Final	10min a 72.0°C
Taq Polimerasa	1.25U		
H ₂ O	c.b.p. 25µL		

Se hicieron ensayos con temperaturas de apareamiento de 58°C y 59.5°C. En las especies de difícil amplificación se hicieron reacciones adicionales titulando MgCl₂ y variando temperaturas. Los amplicones se observaron en geles de agarosa al 0.8% usando generalmente 7µL de la reacción. La electroforesis se efectuó a temperatura ambiente y a 5V/cm, el amortiguador de corrida fue TBE 0.5x (1x: 89mM Tris base, 89mM ácido bórico y 2mM EDTA) y los geles fueron preparados con TBE 1x.

3.10.2 Comprobación de los Amplicones

Para comprobar que los productos amplificados corresponden a GH, se hizo un ensayo de hibridación en gel (Lueders, 1994)²⁴ de los productos de PCR con una sonda marcada con dATP- α^{32} P y Taq Polimerasa extraída de una región codificante de 494pb de la hormona del crecimiento Bovina clonada en el plásmido pPic9kbGH cuyo fragmento es liberado cortando con la endonucleasa *PvuII*. 60ng del fragmento se purificaron con fibra de vidrio y 25ng con Elutrap (Schleicher & Schuell, Inc). El fragmento fue marcado usando una variación de la técnica de Vector T: 85ng del fragmento de 494pb, 1x de amortiguador

Taq DNA polimerasa, 3.3 μ M (10 μ Ci) de [α -³²P]dATP, 2.5mM de MgCl₂, 2.5U Taq DNA polimerasa en 50 μ L de reacción y 1 gota de aceite mineral e incubando 2h a 72°C.

3.10.3 Clonación del Amplicón de jirafa y Análisis de las Clonas Recombinantes

El producto amplificado obtenido de jirafa se clonó en pBS-SK(+) utilizando la técnica de Vector T (Harrison, 1994) como se muestra en la Figura 3.3. El plásmido pBS-SK(+) se linearizó utilizando la endonucleasa *EcoRV* para producir extremos romos. La reacción para generar el Vector T fue: 2 μ g de plásmido, 1x amortiguador Taq DNA polimerasa, 1mM dTTP, 3mM MgCl₂, 3U de Taq DNA polimerasa y 1 gota de aceite mineral e incubando la reacción 2h a 72°C. La reacción de ligación fue: 150ng de Vector T recién preparado, 1x de amortiguador T4 DNA ligasa, 90ng de producto de PCR y 0.75U de T4 DNA ligasa incubando toda la noche a 16°C. Toda la reacción fue utilizada para la transformación de células *XLI-Blue* calcio - competentes. Tres clonas se caracterizaron utilizando las enzimas de restricción *NciI*, *PvuII*, *PstI*, *SacI*, *EcoRI* y *XhoI*.

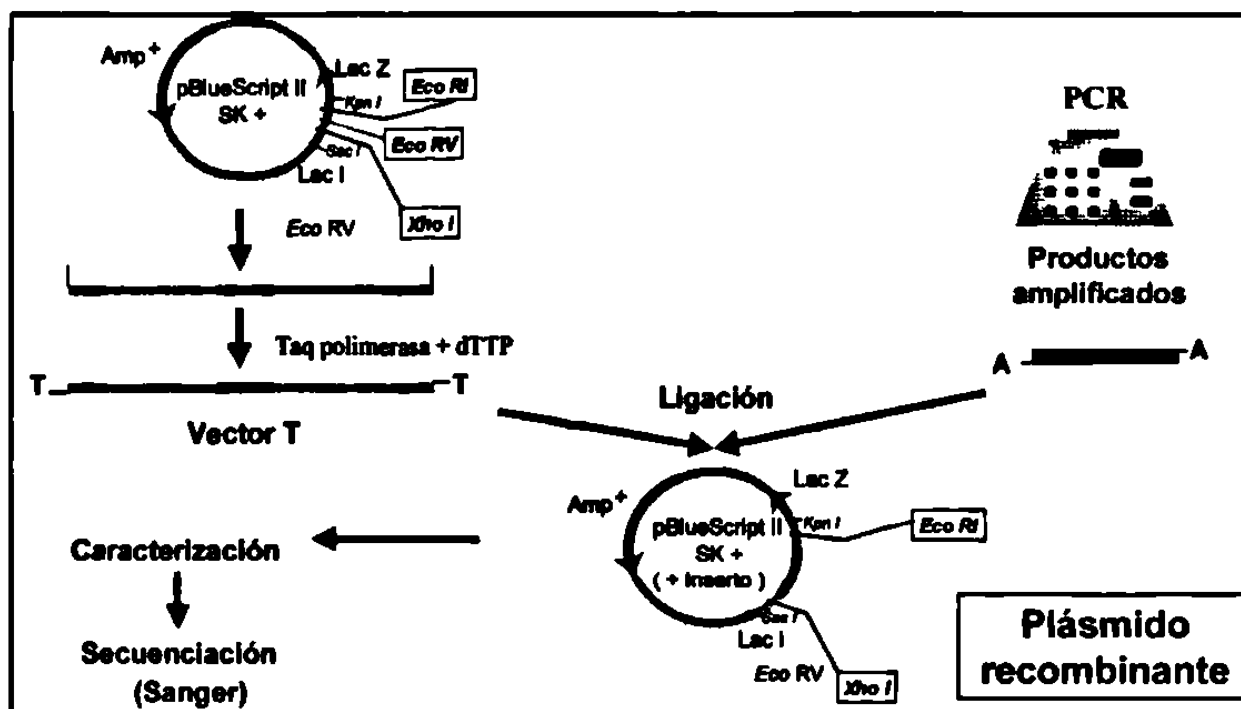
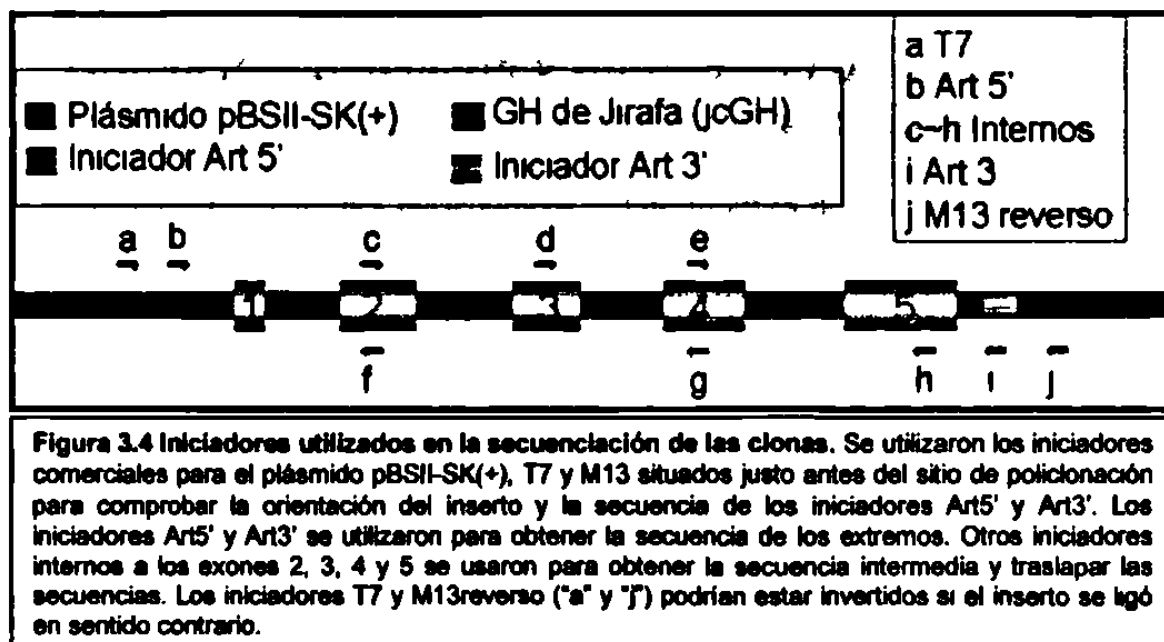


Figura 3.3 Estrategia de Clonación del amplicón. El plásmido pBSII-SK+ es linearizado con *EcoRV* al que se le adicional dT a los extremos 3', por su parte los amplicones ya cuentan con una dA en sus extremos 3'. Los extremos de cada fragmento son complementarios unos con otros favoreciendo la reacción de ligación. El plásmido recombinante obtenido es usado para transformar bacterias calcio competentes para su propegación. El plásmido aislado y multiplicado es utilizado como templatado para caracterizarlo y secuenciarlo.

3.10.4 Secuenciación de las Clonas Recombinantes de Jirafa

Tres clonas fueron sometidas a secuenciación cíclica utilizando un conjunto de iniciadores como se muestra en la Figura 3.4. Se utilizaron, además de los iniciadores de artiodáctilos, Art5' y Art3' y de los iniciadores comerciales T7 y M13 reverso para pBSII-SK(+), 6 iniciadores adicionales que ya se encontraban disponibles en nuestro laboratorio.



3.11 Extracción y Semicuantificación del DNA Genómico

Se utilizó la técnica TSNT para hacer la extracción de DNAG de sangre periférica de vaca, cerdo y cabra. Se usó un stock de nuestro laboratorio de DNAG de placenta humana extraída usando la técnica de Gross-Bellard (Ausubel et. al. 1992), esta misma técnica se utilizó para hacer la extracción de DNAG de sangre de gallo, avestruz, tortuga y serpiente aumentando 3 veces el volumen final sin cambiar la concentración de Proteinasa K. Se usó la técnica de Goelz (Goelz, et. al. 1985) para extraer DNAG de hígado de gato, ratón y lagartija; de músculo de carpa, tilapia, ameca y catán; de aleta de tiburón y de placenta de jirafa. La integridad de los DNAG fue verificada en un gel de agarosa al 0.8%. La semicuantificación fue realizada integrando el área densitométrica de un estándar de DNAG de humano y aproximando proporcionalmente con el área densitométrica de los demás DNA, utilizando el paquete *Molecular Analyst*.

Capítulo 4

Resultados

4.1 Obtención, Selección y Agrupación de las Secuencias Nucleotídicas y Aminoacídicas

S obtuvieron 173 informes nucleotídicos y 330 informes de aminoacídicos en formato GenBank, el informe condensado se muestra en los Anexo 1 y 2.

Tabla 4.1 Especies ordenadas por clasificación.

Accession	Sec	Organismo	Grupo	Accession	Sec	Organismo	Grupo
478615	AA	Acipenser gueldenstaedti	Peces	1174395	As	Esch lucius	Peces
873969	ARN	Amia calva	Peces	X77245	DNAG	Coregonus autumnalis	Peces
882528	ARN	Lepicosteus osseus	Peces	L04688	DNAG	Oncorhynchus keta	Peces
M24066	ARN	Anguilla japonica	Peces	M24768	ARN	Oncorhynchus kisutch	Peces
U63805	DNAG	Poecilia reticulata	Peces	850867	ARN	Oncorhynchus tshawytscha	Peces
U63804	DNAG	Colisa lalia	Peces	Y07741	DNAG	Oncorhynchus mykiss	Peces
482753	AA	Tilapia mossambica	Peces	X61938	DNAG	Salmo salar	Peces
D50368		Seriola quinqueradiata	Peces	AB001865	DNAG	Coregonus lavaretus	Peces
878253	ARN	Morone saxatilis	Peces	464784	AA	Prionace glauca	Peces
X55176	DNAG	Caranx delicatissimus	Peces	X14601	DNAG	Xenopus laevis	Anfibios
L43628	ARN	Seriola dumerilii	Peces	X14602	DNAG	Xenopus laevis	Anfibios
M35627	ARN	Seriola quinqueradiata	Peces	852827	ARN	Rana catesbeiana	Anfibios
U16816	DNAG	Lates calcarifer	Peces	1041999	AA	Crocodylus novaeguineae	Reptiles
A08993	DNAG	Tilapia sp.	Peces	464779	AA	Chelonia mydas	Reptiles
W84774	DNAG	Tilapia nilotica	Peces	86054	AA	Anas platyrhynchos	Av s
X79002	DNAG	Coryphaena hippurus	Peces	134707	AA	Meleagris gallopavo	A es
X65716	DNAG	Dicentrarchus labrax	Peces	D10484	DNAG	Gallus gallus	A e
L43629	ARN	Lateolabrax japonicus	Peces	D00476	DNAG	Capra hircus	Mamíferos
X59377	DNAG	Acanthopagrus butcheri	Peces	M57764	DNAG	Bos taurus	Mamíferos
U48221	DNAG	Sparus aurata	Peces	X72947	DNAG	Bubalus arnee bubalis	Mamíferos
1083902	AA	Acanthopagrus latus	Peces	D16298	DNAG	Capricornis crispus	Mamíferos
134713	AA	Chrysophrys major	Peces	M37310	DNAG	Ovis aries	Mamíferos
85649	AA	Euthynnus pelamis	Peces	259870	AA	Cervidae	Mamíferos
88653	AA	Thunnus albacares	Peces	D16299	DNAG	Cervus nippon	Mamíferos
134721	AA	Thunnus thynnus	Peces	Y12574	DNAG	Cervus elaphus	Mamíferos
D29737	DNAG	Paralichthys olivaceus	Peces	M17704	DNAG	Sus scrofa	Mamíferos
482759	AA	Hippoglossoides platessoides	Peces	1070554	AA	Lama pacos	Mamíferos
103957	AA	Pseudopleuronectes americanus	Peces	823067	ARN	Canis familiaris	Mamíferos
U01143	ARN	Solea senegalensis	Peces	134722	AA	Vulpes vulpes	Mamíferos
U89917	ARN	Sebastes schlegeli	Peces	89010	AA	Alopex lagopus	Mamíferos
L43627	ARN	Sebasticus marmoratus	Peces	U13390	ARN	Felis catus	Mamíferos
U63803	DNAG	Arothron maculatus	Peces	X56120	ARN	Mustela vison	Mamíferos
U63807	DNAG	Pogon rubripes	Peces	284549	AA	Balaenoptera borealis	Mamíferos
Y09421	DNAG	Alburnus alburnus	Peces	838127	DNAG	Oryctolagus cuniculus	Mamíferos
U28388	ARN	Carassius auratus	Peces	U02929	ARN	Equus caballus	Mamíferos
X60419	DNAG	Ctenopharyngodon idella	Peces	U02293	DNAG	Macaca mulatta	Mamíferos
X51969	DNAG	Cyprinus carpio	Peces	M13438	DNAG	Homo sapiens	Mamíferos
U21920	DNAG	Abramis brama	Peces	134706	AA	Loxodonta africana	Mamíferos
M94348	DNAG	Hypophthalmichthys molitrix	Peces	320148	AA	Cavia porcellus	Mamíferos
X60473	DNAG	Hypophthalmichthys nobilis	Peces	866299	ARN	Mesocricetus auratus	Mamíferos
1836078	AA	Clarias gariepinus	Peces	246663	DNAG	Mus musculus	Mamíferos
869215	DNAG	Ictalurus punctatus	Peces	J00739	DNAG	Rattus norvegicus	Mamíferos
L27835	ARN	Pangasianodon gigas	Peces	264607	AA	Macropus rufogriseus	Mamíferos
M63713	ARN	Pangasius pangasius	Peces	AF052192		Trurus	up
482758	AA	Gadus morhua	Peces				
483102	AA	Gadus morhua	Peces				

Las secuencias de cada especie que formaron parte de la base de datos final se encuentran resaltadas en el Anexo 1 y 2. Obtuvimos 90 secuencias correspondientes a 89 especies (Tabla 4.1).

El espacio físico de la base de datos es de alrededor de 5MB sin incluir las imágenes de los organismos. Al hacer el desglose (Tabla 4.2) por clases y sus subclases con más secuencias reportadas, se observa que algunas clases han sido estudiadas preferencialmente, 62% del total de secuencias reportadas son de peces, seguido por los mamíferos con 28% dejando únicamente el 10% a todas las demás clases de vertebrados. Vemos también que el 29% (26/89) de las especies no tiene una secuencia nucleotídica reportada, sin embargo, en mamíferos esta cifra sube a 47% (8/17).

Tabla 4.2 Total de secuencias por clase. Se detallan dos de las clases más estudiadas, peces y mamíferos.

#	Clase	Nt	%	aa	%	Tot Sec	%Tot Sec	#
55	Peces	41	65%	14	54%	55	62%	7 Salmoniformes 5 Siluriformes 7 Cypriniformes 19 Perciformes 4 Pleuriformes
2	Anfibios	3	5%	0	0%	3	3%	
2	Reptiles	0	0%	2	8%	2	2%	
3	Aves	1	2%	2	8%	3	3%	
26	Mamíferos	17	27%	8	31%	25	28%	10 Artiodactyla 5 Carnivora 2 Primates 4 Rodentia
1	Marsupial	1	2%	0	0%	1	1%	
89	Total	63	100%	26	100%	89	100%	

4.2 Alineamiento de las Secuencias

En la Figura 4.1 se muestra un gráfico de la matriz de alineamiento aminoacídico (ordenado según Tabla 4.1) de todas las secuencias seleccionadas de las GH reportadas. Podemos apreciar que entre los peces, el parecido puede ir mayoritariamente desde 100% a 60% y solo entre unos pocos hasta de 40%; conforme hacemos la comparación en la dirección de los mamíferos, el parecido va disminuyendo llegando hasta aproximadamente 20%. *Xenopus Laevis*, un anfibio, se parece más a mamíferos que a peces. En mamíferos, la GH se aprecia más conservada.

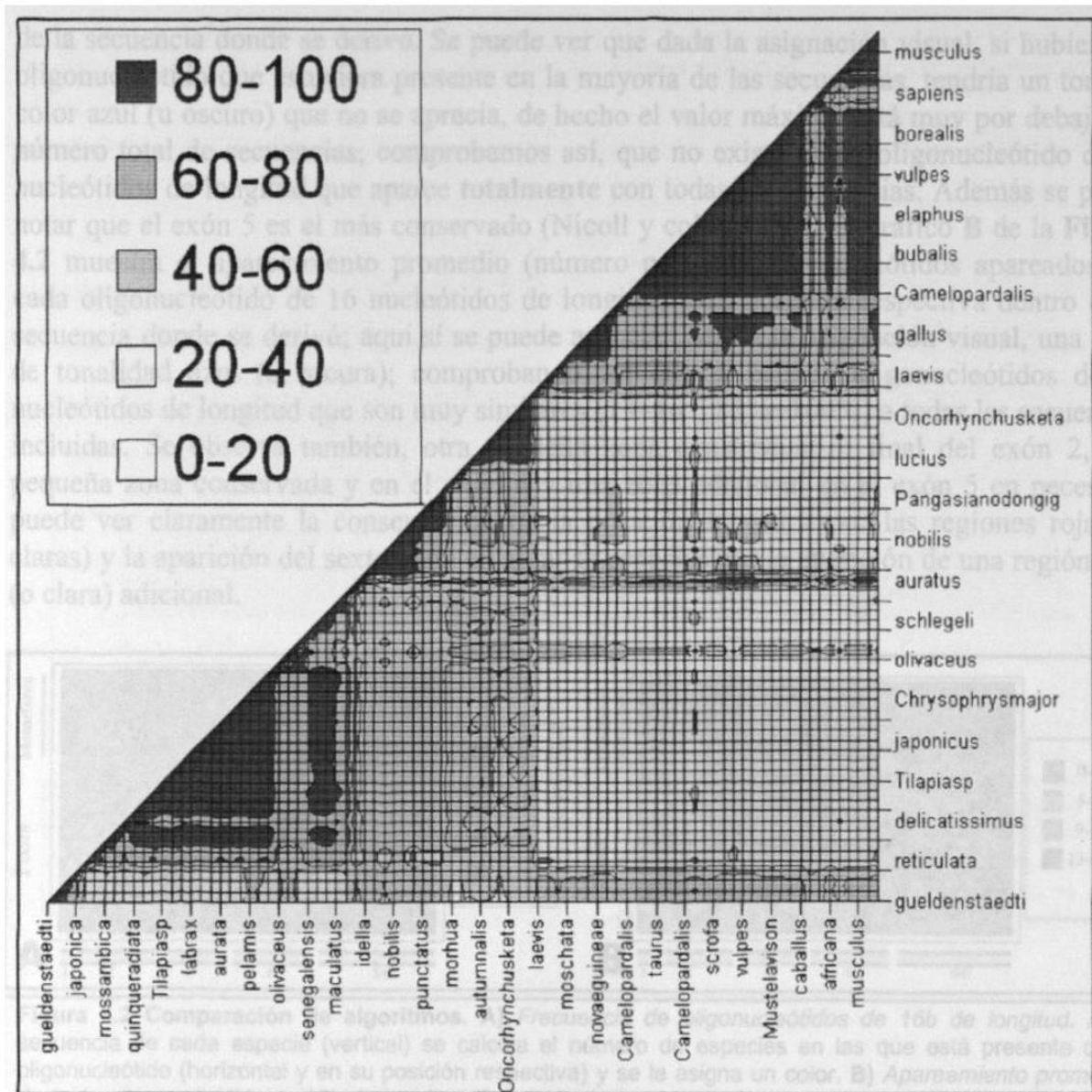


Figura 4.1 Matriz de alineamiento aminoacídico (% similitud). Todas las especies ordenadas por clasificación (se nombraron únicamente 1 de cada 3). En horizontal de izquierda a derecha peces, anfibios, reptiles, aves, mamíferos y marsupiales. De la misma manera en vertical de abajo hacia arriba.

4.3 Selección de la Sonda

Dado que deseábamos obtener un oligonucleótido conservado en todas las especies, pensamos básicamente en dos estrategias de búsqueda y comparamos los resultados. Se utilizaron oligonucleótidos de 16 bases de longitud para efectos de la comparación y 32 bases de longitud como base para el diseño de la sonda.

4.3.1 Comparación de Algoritmos

En el gráfico A de la Figura 4.2, se muestra el número de secuencias en que está presente cada oligonucleótido de 16 nucleótidos de longitud y su posición respectiva dentro

de la secuencia donde se derivó. Se puede ver que dada la asignación visual, si hubiese un oligonucleótido que estuviera presente en la mayoría de las secuencias, tendría un tono de color azul (u oscuro) que no se aprecia, de hecho el valor máximo está muy por debajo del número total de secuencias; comprobamos así, que no existe algún oligonucleótido de 16 nucleótidos de longitud que aparece **totalmente** con todas las secuencias. Además se puede notar que el exón 5 es el más conservado (Nicoll y cols., 1986). El gráfico B de la Figura 4.2 muestra el apareamiento promedio (número promedio de nucleótidos apareados) de cada oligonucleótido de 16 nucleótidos de longitud y su posición respectiva dentro de la secuencia donde se derivó; aquí sí se puede apreciar, según la asignación visual, una zona de tonalidad azul (u oscura); comprobando así que sí existen oligonucleótidos de 16 nucleótidos de longitud que son muy similares (13~16 ó 81%~100%) a todas las secuencias incluidas. Se observa también, otra pequeña zona conservada al final del exón 2, una pequeña zona conservada y en el exón 4 y una zona adicional en el exón 5 en peces. Se puede ver claramente la conservación de la posición intrónica por las regiones rojas (o claras) y la aparición del sexto exón en algunas especies por la aparición de una región roja (o clara) adicional.

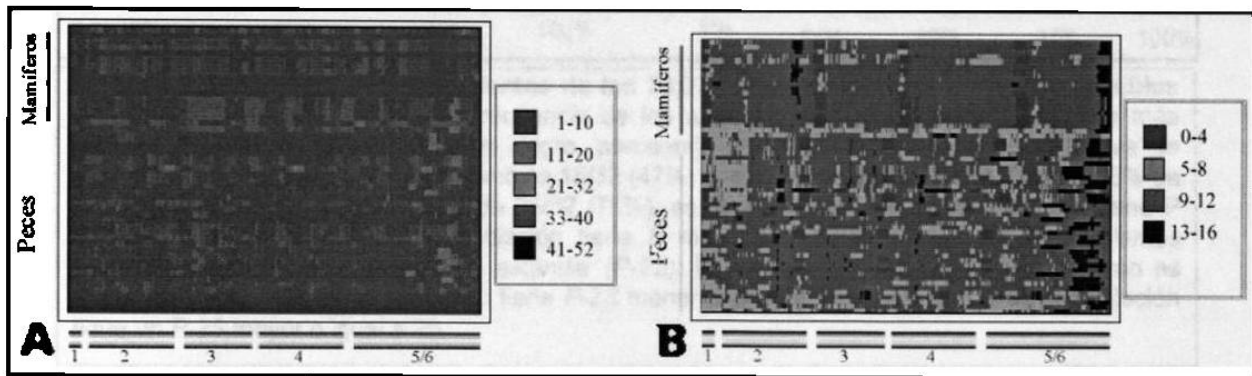


Figura 4.2 Comparación de algoritmos. A) Frecuencia de oligonucleótidos de 16b de longitud. A la secuencia de cada especie (vertical) se calcula el número de especies en las que está presente cada oligonucleótido (horizontal y en su posición respectiva) y se le asigna un color. **B) Apareamiento promedio de cada oligonucleótido de 16b de longitud.** De cada oligonucleótido (horizontal) de cada especie (vertical), se obtiene la suma de nucleótidos con los que tiene el mejor apareamiento en cada secuencia y se divide entre el número de especies (52). La posición aproximada de cada oligonucleótido con respecto a cada exón se muestra en la parte inferior. La asignación de colores se muestra a la derecha de cada gráfico. Las zonas rojas (o claras) indican la posición de los intrones debido a que no se hizo el cálculo y se usó 0 (solo en las secuencias donde sí se tenía tal información y en gráfico B).

4.3.2 Determinación de la Población de Oligonucleótidos por Evaluar

La población total de oligonucleótidos de 32 bases de longitud fue de 28,038 para las 52 secuencias nucleotídicas completas de la GH. El evaluar toda la población resultaría de un esfuerzo inútil dado que las secuencias de bajo apareamiento lógicamente resultarían en una evaluación baja. Para determinar la fracción de la población por evaluar, se determinaron las frecuencias de los alineamientos promedio y alineamientos promedio menos dos desviaciones estándar redondeados al entero más cercano y se analizaron los resultados gráficamente. Se puede ver en la curva A de la Figura 4.3 que el 96% de los oligonucleótidos tiene un apareamiento promedio menor a 78%, el 83% de los oligonucleótidos tienen 68% que es muy bajo contra un máximo de 91%. En la curva B de la Figura 4.3, el 95% de los oligonucleótidos, tienen un valor de alineamiento menos dos

desviaciones estándar menor a 18 que corresponde al 72% con respecto del valor máximo de 25. Por lo tanto, evaluar el 5% superior en ambos criterios resulta ser suficientemente sobrado.

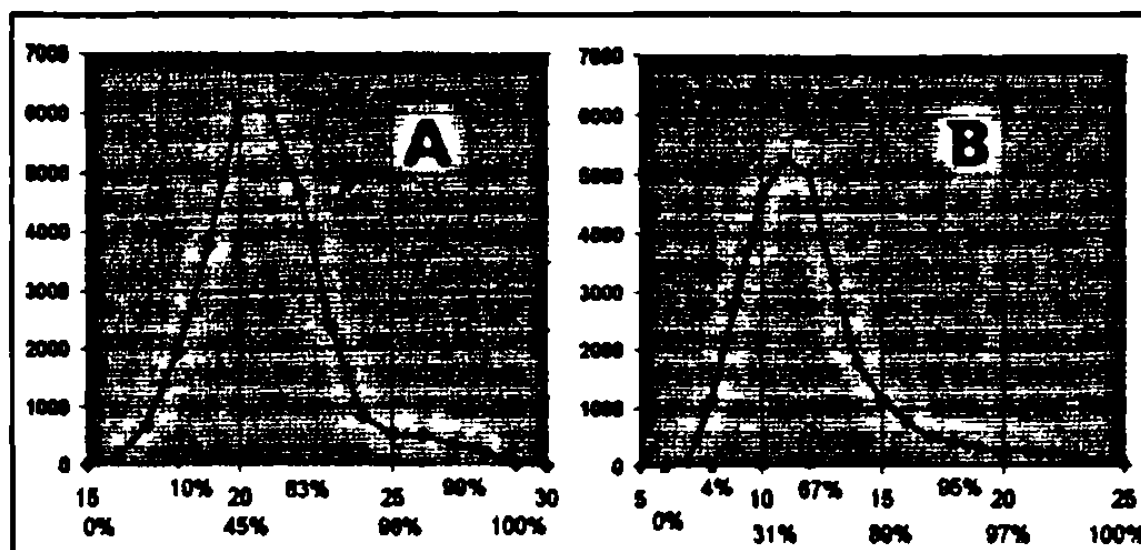


Figura 4.3 Frecuencia de alineamientos de los 28,038 oligonucleótidos consenso posibles de 32nt de longitud. Vertical, la frecuencia de los alineamientos redondeados al entero más cercano. Horizontal, alineamiento en negro, porcentaje de la población en rojo. A) Curva de alineamientos promedio (P): El mínimo es 15/32 (47%) y el máximo 29/32 (91%). Casi el 96% de los oligonucleótidos alinea menos de 25/32 (78%), es decir, solo el 4% de la población tiene P superior a 77%. El 83% de la población tiene P muy bajo, 22/32 (68%). B) Alineamientos promedio menos dos desviaciones estándar ($P-2S$): El mínimo es 7/32 (22%) y el máximo es 26/32 (81%). El 89% de la población tiene $P-2S$ menor o igual a 15 y solo el 3% de la población tiene un $P-2S$ mayor o igual a 20.

4.3.3 Evaluación de los Oligonucleótidos

Se determinó anteriormente que evaluar el 5% de los oligonucleótidos con más alto apareamiento es suficiente para encontrar un buen oligonucleótido como sonda, pero para determinar dentro de esta sub-población los mejores oligonucleótidos, fue necesario diseñar una función que compare los oligonucleótidos entre sí. Las ideas básicas que nos llevaron a diseñar la función a la que llamamos f_{sonda} fueron las siguientes:

- El éxito de una buena hibridación depende directamente del apareamiento entre cada nucleótido de la sonda con la secuencia blanco: Decidimos incluir el número de nucleótidos iguales con cada secuencia blanco en forma de promedio.
- El éxito de una buena hibridación, depende también de la ausencia o minimización del apareamiento de la sonda con ella misma: Decidimos incluir la energía de formación de estructuras de lazos u horquilla y dimerizaciones.
- Dado que la sonda no aparearía perfectamente con todas las secuencias blanco, por lo tanto habrían desapareamientos a lo largo de la sonda, es indeseable que los desapareamientos ocurran en alguno de los extremos de la sonda: Si los extremos no están apareados, se podría facilitar el desapareo de la misma. Decidimos incluir el apareamiento de los últimos tres nucleótidos en cada extremo de la sonda, en forma de promedio y peso a cada uno de ellos.

Asimismo, el correcto apareamiento de alguno de los extremos de la sonda con la secuencia blanco, podría favorecer una posible reacción de PCR.

- Dado que las secuencias blanco no son exactamente las mismas, la temperatura de apareamiento sería variable. *Decidimos homogenizar las temperaturas:* La temperatura óptima depende básicamente del apareamiento con cada secuencia blanco, del contenido de GC y del tamaño del oligonucleótido, así como la concentración de sales. Incluimos la temperatura de fusión media en forma promediada.
- *Dado que se consideran desapareamientos en la secuencia, pero no la distribución de ellas:* Decidimos incluir el tamaño máximo de apareamiento contiguo con cada secuencia blanco, de una forma promediada.
- Es posible que dos oligonucleótidos tengan promedio de apareamiento y promedio de temperatura similares. Para poder evaluar cual de los dos resultaría ser el mejor, se decidió *que el mejor de los oligonucleótidos que tienen promedios muy similares sería el oligonucleótido cuya desviación estándar del apareamiento y/o temperatura sea mínima:* Esto evitaría que algunas secuencias tengan muy poco apareamiento y/o muy baja temperatura, pero compensado con otras secuencias de apareamiento y/o temperatura altos. Decidimos reducir el promedio en dos desviaciones estándar a cada término, bajo la suposición de que la distribución del apareamiento y de las temperaturas se aproxima a una curva estadística normal. Dada la suposición de una distribución normal, la media menos dos desviaciones representan el 96% de la población y resulta en un estadístico tal, que sólo el 2% de la población está por debajo de este valor; es decir, solo una de las 52 secuencias podrían estar por debajo del estadístico escogido. Ahora, la desviación estándar de la población se ve afectada por secuencias blanco que aparezcan muy poco con la sonda o a secuencias blanco que aparezcan mucho con la sonda o a ambos casos en la misma población. Entonces es *preferible que la desviación estándar de un oligonucleótido sea provocada por apareamientos altos que por apareamientos bajos.* Para "premiar" a aquellos oligonucleótidos cuya desviación es provocada por secuencias blanco con muy alto apareamiento y "castigar" aquellos oligonucleótidos cuya desviación es provocada por secuencias blanco con muy bajo apareamiento: decidimos introducir como "corrección", el número de secuencias de apareamiento superior al 89% menos el número de secuencias con apareamiento por debajo de 61%.

Para determinar el peso aproximado de cada término, se formaron 5 grandes grupos:

Término	Unidades	Rango
Apareamiento	Nucleótidos	0~L*
Temperatura media de fusión	°C	0~100
Apareamiento en los extremos (5' y 3')	Nucleótidos	0~N [§]
Encadenamiento	Nucleótidos	0~L*
Energías intermoleculares e intramoleculares	Kcal/mol	-95~0 [†]

*L es la longitud del oligonucleótido. § N es el número de secuencias.

† Calculado para 50 nucleótidos aunque en ejecución de determina en función de la longitud real del oligonucleótido en curso.

Para evaluar todos los parámetros en una sola función, a cada término se le asoció una constante de "peso" relativo. El peso relativo de cada constante fue determinado variando solo una de las constantes, hasta obtener resultados satisfactorios. Para poder sumar cada uno de los parámetros, dado que son unidades completamente diferentes y a escalas diferentes, se normalizaron al valor máximo y mínimo permisible resultando en valores de 0 y 1 antes de multiplicar por el "peso" relativo de cada uno. Para comparar oligonucleótidos de diferente tamaño, los términos se multiplicaron por el inverso de la longitud del oligonucleótido.

La función para evaluar los oligonucleótidos resultó ser la siguiente :

$$f_{sonda} = sP + sE + sG + sJ + sM$$

donde

$$sP = (P_m - 2P_{SD}) * 0.18 * (1/50 + 1/Ol) + 0.333 * (N_{Ov} - N_W) / N_{sec}$$

$$sE = 0.6 N_{m5+0} + 0.3 N_{m5+1} + 0.1 N_{m5+2} + 0.6 N_{m3+0} + 0.3 N_{m3-1} + 0.1 N_{m3-2}$$

$$sG = sGH + sGD$$

$$sGH = 0.04 * (\Delta GH_N + 12.6) / 12.6$$

$$sGD = 0.04 * (\Delta GD_N + 19.8) / 19.8$$

$$sJ = 0.5 * (J_m - J_n) * 0.05 * (1/50 + 1/Ol)$$

$$sM = (Tm_m - 2Tm_{DS}) * 0.18 * (1/50 + 1/Ol) + 0.333 * (N_{Ov} - N_W) / N_{sec}$$

$$Tm_m = Tm_{opt} \text{ promedio } (\Sigma(Tm_{opt}) / N_{sec})$$

$$Tm_{DS} = \text{Desviación estándar de } Tm_{opt}$$

$$Tm_{opt} = Tm_{Primer} * 0.3 + Tm_{Product} * 0.7$$

$$Tm_{Product} = (100.5 + 16.6 * \log_{10}(0.050) + 0.41 * \%GC - (820 / Ol) - \%mismatch$$

$$Tm_{Primer} = (1000.0 * (\Delta h - 3.4) / (\Delta s + 1.987 * \log(1 / 50 * 10^9)) - 273.15 + 16.6 * \log_{10}(50 * 10^3))$$

$$P_m = \text{Apareamiento promedio}$$

$$P_{SD} = \text{Desviación estándar del apareamiento promedio}$$

$$Ol = \text{longitud del oligonucleótido.}$$

$$N_{sec} = \text{Número de secuencias}$$

$$N_{Ov} = \text{Número de secuencias que rebasan 89\% de apareamiento}$$

$$N_W = \text{Número de secuencias que no llegan a 61\% de apareamiento}$$

$$N_{m5+0} = \text{Número de nucleótidos que aparean en el extremo 5'}$$

$$N_{m5+1} = \text{Número de nucleótidos que aparean 1 nucleótido después del extremo 5'}$$

$$N_{m5+2} = \text{Número de nucleótidos que aparean 2 nucleótidos después del extremo 5'}$$

$$N_{m3-2} = \text{Número de nucleótidos que aparean 2 nucleótidos antes del extremo 3'}$$

$$N_{m3-1} = \text{Número de nucleótidos que aparean 1 nucleótido antes del extremo 3'}$$

$$N_{m3-0} = \text{Número de nucleótidos que aparean en el extremo 3'}$$

$$J_m = \text{Promedio del encadenamiento (numero máximo de nucleótidos que aparean y estan consecutivos)}$$

$$J_n = \text{Mínimo de cada uno de los encadenamientos de cada secuencia}$$

$$\Delta GH_N = \text{Energía libre máxima de formación de un complejo de horquilla calculada con el método Neighbor (Breslauer y cols. 1986)}$$

$$\Delta GD_N = \text{Energía libre máxima de formación de un complejo intermolecular calculada con el método Neighbor (Breslauer y cols. 1986)}$$

Los 5 grupos se aproximan a los pesos relativos de 0.36, 0.12, 0.08, 0.10 y 0.36 de peso para sP , sE , sG , sJ y sM respectivamente.

4.4 Diseño de la Sonda (GH-Explorer I)

Las secuencias iniciales de 32 nucleótidos ordenadas por alineamiento promedio reportadas según *MaxProbe* fueron:

- 1) ctggcctgcttcaagaaggacatgcacaaggt
- 2) aaggtggagacctacctgacggtggccaaatg
- 3) tcctatcgactgattgagtcttgggagttccc
- 4) ctacgaactgctggcctgcttcaagaaggaca
- 5) ggctctctccagaggccaactgcaccctgtag
- 6) ttgagtcttgggagttccccagtcggtctctg
- 7) tgaggagcagcgacagctcaacaaaatcttcc
- 8) gaactgctggcctgcttcaagaaggacatgca
- 9) ttcaagaaggacatgcacaaggtggagaccta
- 10) gtggagacctacctgacggtggccaaatgca
- 11) cacaaggtggagacctacctgacggtggccaa

Sin embargo, después de que *MaxProbe* realizó la extensión y cortes automáticos en ambos extremos, 7 de las secuencias, las secuencias 1, 2, 4, 8, 9, 10 y 11 convergieron a la secuencia siguiente:

gcctgcttcaagaaggacatgcacaaggtggagacctacctg (42nt)

La convergencia se explica "alineando" y comparando las secuencias entre si...

```

gcctgcttcaagaaggacatgcacaaggtggagacctacctg (42nt)
ctacgaactgctggcctgcttcaagaaggaca (4)
gaactgctggcctgcttcaagaaggacatgca (8)
ctggcctgcttcaagaaggacatgcacaaggt (1)
ttcaagaaggacatgcacaaggtggagaccta (9)
cacaaggtggagacctacctgacggtggccaa (11)
aaggtggagacctacctgacggtggccaaatg (2)
(10) gtggagacctacctgacggtggccaaatgca

```

Podemos ver que en realidad es la misma zona y que cada oligonucleótido abarca una buena parte de ésta, pero la única que esta totalmente inmersa es el oligonucleótido 9.

En la Figura 4.4 muestra el informe después de indicar a *MaxProbe* que extendiera el extremo 5' de la sonda. Se ve que el apareamiento promedio de la sonda con todas las secuencias es de 48 de 56 nucleótidos (86%); en el peor de los casos (primates), 42 de 56 (75%) y 51 de 56 (91%) en el mejor de los casos (gallo y roedores). La posición 11 se degeneró incluyendo A y G para aumentar el desempeño en primates sin afectar a las demás secuencias (elevando principalmente el encadenamiento -join-). La secuencia definitiva de la sonda a la que llamamos *GH-Explorer I* es la siguiente:

↓

5' gcctgcttcaRgaaggacatgcacaaggtggagacctacctgaggggtcatgaagtg3'

↑

La posición del 5º intrón encontrada en algunos peces se marca con flechas y está situada cerca de la mitad de la sonda. Se especifica también la degeneración introducida.

Obtención de nuevos genes de la hormona del crecimiento (GH) en vertebrados

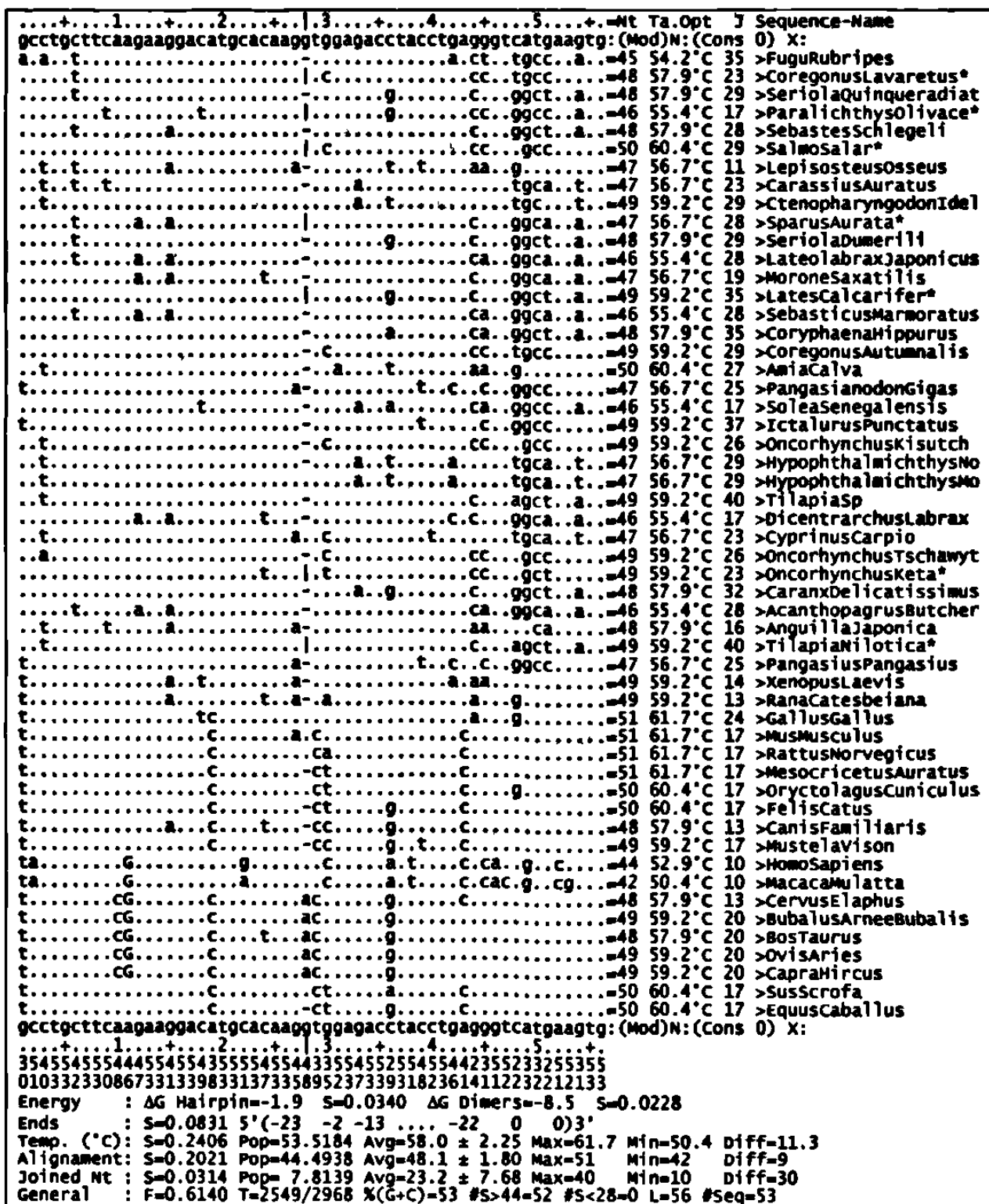


Figura 4.4 Informe del alineamiento de la sonda según MaxProbe. Se muestra el alineamiento. Se resalta con mayúsculas la posición de la degeneración, incrementando el valor de encadenamiento -join- en primates de 10 hasta 19. Se muestra con una barra vertical (|) la posición del intrón 5 solo en las especies que se sabe que lo contienen, un punto (.) en las que se sabe que no lo contienen y un guión (-) en las que no hay evidencia experimental al respecto dado que son secuencias de DNAc. La barra vertical, el punto y el guión, indican igualdad con el nucleótido de la secuencia en cuestión (arriba y abajo).

4.5 Comprobación del Desempeño de *GH-Explorer I*

4.5.1 Comprobación Virtual de *GH-Explorer I*

Para comprobar la especificidad de la sonda por la GH de vertebrados, se efectuó una consulta en la página de Internet del *GenBank* con la secuencia de *GH-Explorer I*, como se indicó en Material y Métodos. En el informe completo (Anexo 3) se obtuvieron 189 secuencias, de los cuales 173 corresponden a todos los informes de GH obtenidos (Anexo 1), por lo tanto, 16 son falsos positivos. Unos fragmentos representativos por analizar se muestran en la Figura 4.4. Casualmente, las secuencias de GH de pato (Figura 4.6) y atún, las que mejor aparean con la sonda *GH-Explorer I*, no habían sido incluidas en el grupo de secuencias en estudio, seguramente por errores en el filtrado. Las secuencias nucleotídicas de *Bufo marinus* (sapo), *Trichosurus vulpecula* (possum) y algunas otras especies, fueron reportadas en el lapso mientras se hacía el estudio y este escrito. Por el alto parecido con Pato, Atún, Sapo y Possum entre otros, queda demostrado que aún con secuencias desconocidas de la hormona del crecimiento, la sonda *GH-Explorer I* aparece positivamente, así sean de aves, peces, anfibios o vertebrados superiores. Indicando que la zona y secuencia escogida tiene una alta importancia evolutiva y/o funcional.

Gallo (*Gallus domesticus*, *Gallus Gallus*) es uno de los que aparean más alto (Figura 4.6) con el grupo de secuencias en estudio (Figura 4.3) y que en el informe también sale evaluado muy alto (por alto "score" y bajo "E"), coincidiendo con el informe; así mismo sucede con macaco que es de los que menos aparean (Figura 4.7). De los falsos positivos, basta con ver el apareamiento* que resulta ser muy pobre, dan un resultado "positivo" debido a que se usó un tamaño mínimo de apareamiento de 8 nucleótidos en el informe, o sea, que todas las secuencias que al menos aparezcan 8 nucleótidos consecutivos serían posibles positivos, sin embargo el apareamiento es realmente bajo.

* El informe de los apareamientos no se muestra completo debido al volumen de información. Para analizarlos, debe reproducir la consulta como se indica en material y métodos e introducir el número máximo posible en el parámetro "Alignments" de la página de Internet correspondiente.

Sequences producing significant alignments:	Score (bits)	E Value
<u>emb X07079 APGH</u> Duck mRNA for growth hormone	<u>83</u>	4e-15
<u>emb X06735 TTGHR</u> Tuna mRNA for growth hormone	<u>83</u>	4e-15
... (omisión) ...		
<u>emb V01238 RMGROW2</u> Rat gene for growth hormone (presomatotropin...)	<u>75</u>	1e-12
<u>emb X17618 GDGHSOM</u> G.domesticus growth hormone somatotropin mRNA	<u>75</u>	1e-12
... (omisión) ...		
<u>gb L25053 RATGHRHR</u> Rattus norvegicus Gonadotropin releasing hor...	<u>75</u>	1e-12
... (omisión) ...		
<u>gb AF062746 AF062746</u> Bufo marinus growth hormone precursor, mRNA...	<u>67</u>	3e-10
... (omisión) ...		
<u>gb U02293 MMU02293</u> Macaca mulatta growth hormone-variant gene, ...	<u>47</u>	3e-04
... (omisión) ...		
<u>gb 882528 882528</u> growth hormone [Lepisosteus osseus-long-nosed ...	<u>41</u>	0.019
<u>emb Z32691 ECT7DN1</u> E.coli (HM174) pT7T7/T7H DNA, 11271 bp	<u>39</u>	0.075
<u>emb Z32693 ECT7DN3</u> E.coli pT7hGH_p1 DNA, 6160bp	<u>39</u>	0.075
<u>gb AF075648 AF075648</u> Equus caballus microsatellite LEX046	<u>37</u>	0.30
<u>gb L16556 MACGH</u> Monkey growth hormone mRNA, complete cds and fl...	<u>37</u>	0.30
... (omisión) ...		
<u>gb D10640 ONHPSOL</u> Chum salmon mRNA for somatolactin, complete cds	<u>35</u>	1.2
... (omisión) ...		
<u>gb U63803 AMU63803</u> Arothron maculatus growth hormone (GH) gene, ...	<u>33</u>	4.8
... (omisión) ...		
<u>gb U90307 HSU90307</u> Human iroquois-class homeodomain protein IRX...	<u>33</u>	4.8

Figura 4.5 Fragmentos representativos del informe del apareamiento de la sonda *GH-Explorer I* con las secuencias reportadas en GenBank.

<u>emb X07079 APGH</u> Duck mRNA for growth hormone	Length = 794
Score = 82.7 bits (41), Expect = 4e-15	
Identities = 51/55 (92%), Positives = 51/55 (92%)	
Query: 2 cctgcttcargaaggacatgcacaagggtggagacctacctgagggtcatgaagtg 56	
Sbjct: 601 cctgcttcaagaaggacctgcacaagggtggagacctacctgaagggtgatgaagtg 655	
<u>emb X17618 GDGHSOM</u> G.domesticus growth hormone somatotropin mRNA	Length = 782
Score = 74.7 bits (37), Expect = 1e-12	
Identities = 50/55 (90%), Positives = 50/55 (90%)	
Query: 2 cctgcttcargaaggacatgcacaagggtggagacctacctgagggtcatgaagtg 56	
Sbjct: 597 cctgcttcaagaaggatctgcacaagggtggagacctacctgaagggtgatgaagtg 651	

Figura 4.6 Mejores apareamientos.

<u>gb U02293 MMU02293</u> Macaca mulatta growth hormone-variant gene, complete cds.	Length = 2594
Score = 46.6 bits (23), Expect = 3e-04	
Identities = 30/33 (90%), Positives = 30/33 (90%)	
Query: 3 ctgcttcargaaggacatgcacaagggtggagac 35	
Sbjct: 1889 ctgcttcaggaaggacatgaacaagggtcgagac 1921	

Figura 4.7 Bajo apareamiento.

Existe un falso positivo que resulta en un apareamiento muy alto, "Rattus norvegicus Gonadotropin releasing hormone" (Figura 4.8), el cual analizando las referencias del informe, descubrimos que la secuencia es derivada de una línea celular de tumor pituitario. Las otras secuencias reportadas son muy poco parecidas o bacterianas; lo que nos sugiere

que pudo haber recombinación específica en esta línea celular que pudo dar lugar a la obtención de esa secuencia.

```
gb|L25053|RATGNRHR Rattus norvegicus Gonadotropin releasing hormone receptor
(GnRH-R) mRNA, complete cds. Length = 2938
Score = 74.7 bits (37), Expect = 1e-12
Identities = 50/55 (90%), Positives = 50/55 (90%)
Query: 2 cctgcttcargaaggacatgcacaagggtggagacctacctgagggtcatgaagtg 56
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 149 cctgcttcaagaaggacctgcacaaggcagagacctacctgagggtcatgaagtg 95
```

Figura 4.8 Falso positivo. Vease texto.

4.5.2 Comprobación Experimental de *GH-Explorer I*

4.5.2.1 Extracción y Semicuantificación del DNAg

En la Figura 4.A se muestra la imagen del gel con los DNAg de humano, carpa, ratón, catán, tortuga (1), tiburón, avestruz, jirafa, gato, ameca, gallo, lagartija, guppy, tilapia, tortuga (2) y serpiente. Los resultados de la cuantificación son 450ng, 350ng, 250ng, 250ng, 200ng, 200ng, 700ng, 300ng, 185ng (en Figura 4.B), 50ng, 300ng (en Figura 4.B), 150ng, 20ng, 350ng, 400ng y 300ng respectivamente. Había duda en los carriles de gato, víbora y gallo que se resolvieron haciendo una extracción Fenol:Sevag[†] (1:1), diluyendo con H₂O y calentando 2min a 60°C. Los resultados se muestran en la Figura 4.B.

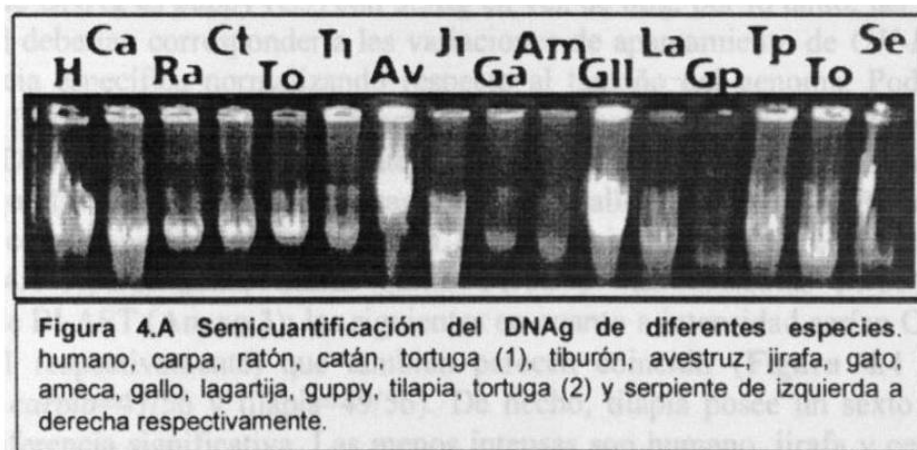
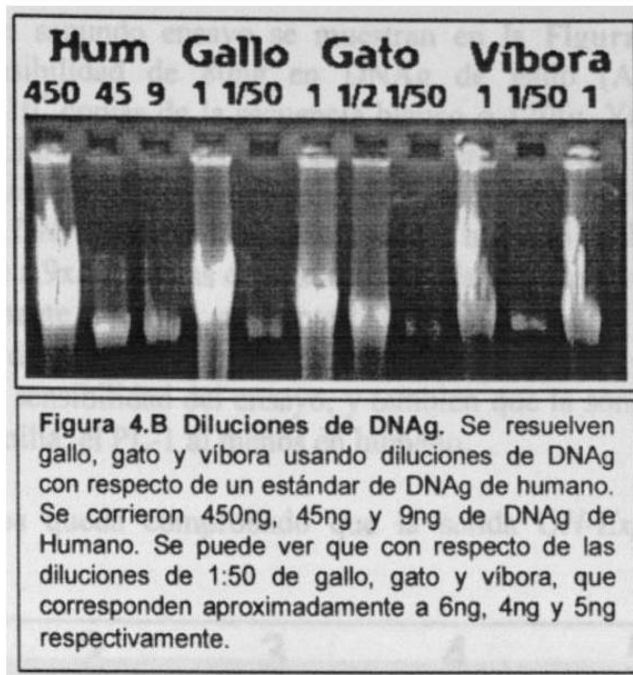


Figura 4.A Semicuantificación del DNAg de diferentes especies. humano, carpa, ratón, catán, tortuga (1), tiburón, avestruz, jirafa, gato, ameca, gallo, lagartija, guppy, tilapia, tortuga (2) y serpiente de izquierda a derecha respectivamente.

[†] Sevag esta compuesto de Cloroformo:Alcohol isoamílico (24:1).



4.5.2.2 Ensayo de Hibridación

Para comprobar el funcionamiento de *GH-Explorer I*, se hicieron dos ensayos de hibridación tipo slot-blot en membrana de nylon usando DNAG de varias especies como secuencias blanco. El primer ensayo demuestra que la sonda *GH-Explorer I* hibridó con todos los DNAG de las especies incluidas y el resultado se muestra en Figura 4.9. Se observan diferencias en la intensidad de detección, las cuales pueden ser debidas a la variación en la cantidad y calidad del DNAG, diferencias en los tamaños de los genomas y al apareamiento de la sonda según la secuencia. La única variación significativa en la cantidad de DNAG es guppy (C5) con 3.5µg en vez de 8µg, por lo tanto, las variaciones en intensidad deberían corresponder a las variaciones de apareamiento de *GH-Explorer I* con la secuencia específica normalizando respecto al tamaño del genoma. Podemos ver que 800ng de DNAG de gallo (E1) se detectan perfectamente, con dificultad se logran apreciar 80ng de DNAG de gallo (E2). La integración densitométrica que se ve en la Tabla 4.3 muestra que las dos líneas más intensas son las de gallo y ratón (A6 y B1) que parecieran coincidir con los resultados obtenidos en el apareamiento tanto en el diseño de la sonda (Figura 4.4 Apareamiento *Gallus gallus*=51/56 y *Mus musculus*=51/56) como en el informe de BLAST (Anexo 3); las siguientes en cuanto a intensidad serían Carpa y Tilapia (B6 y C1 respectivamente) que también parecen coincidir (Figura 4.4 Apareamiento *Cyprinus carpio*=47/56 y tilapia=49/56). De hecho, tilapia posee un sexto exón y no se aprecia diferencia significativa. Las menos intensas son humano, jirafa y cerdo (A1, A4 y A5), en humano también es consistente con el informe de BLAST. En catán y tiburón (C3 y C4) se aprecia una señal aceptable, considerando que éstos peces son muy primitivos y que pudieran tener más variaciones en la secuencia nucleotídica que todavía no se conoce. Toda la línea D es de DNA plasmídico usado como control positivo y de sensibilidad, y no se aprecia señal alguna; dado que el control negativo también es un plásmido, se tuvo duda de la integridad del ensayo y para resolverlo se hizo el segundo ensayo.

Los resultados del segundo ensayo se muestran en la Figura 4.10. En este ensayo confirmamos la sensibilidad de 80ng en DNAG de gallo (A3) lo que equivale a aproximadamente 6×10^4 copias de la secuencia blanco o 120fg. Ya no se aprecian 8ng de DNAG de gallo (A4). En DNA de GH de jirafa se pudieron detectar hasta 12pg (B3) (32pg / (4833/1875)) que equivalen a 6.13×10^6 copias de la secuencia blanco que es 2 órdenes de magnitud menos. En DNA de lactógeno placentario 1 humano (hpl-1) se lograron detectar 3.2ng lo que equivale a 9×10^8 copias de la secuencia blanco que son 4 órdenes de magnitud menos (aproximadamente tomando en cuenta las intensidades y la cantidad de DNA). Por densitometría no se detecta nada en los controles negativos (D1~D6). Comprobamos la reproducibilidad y la sensibilidad del ensayo, y también que la sonda tiene menos afinidad por otro gen de la familia, el PL-1 al menos en humano.

Con estos ensayos queda comprobado que la sonda *GH-Explorer I* funciona para vertebrados.

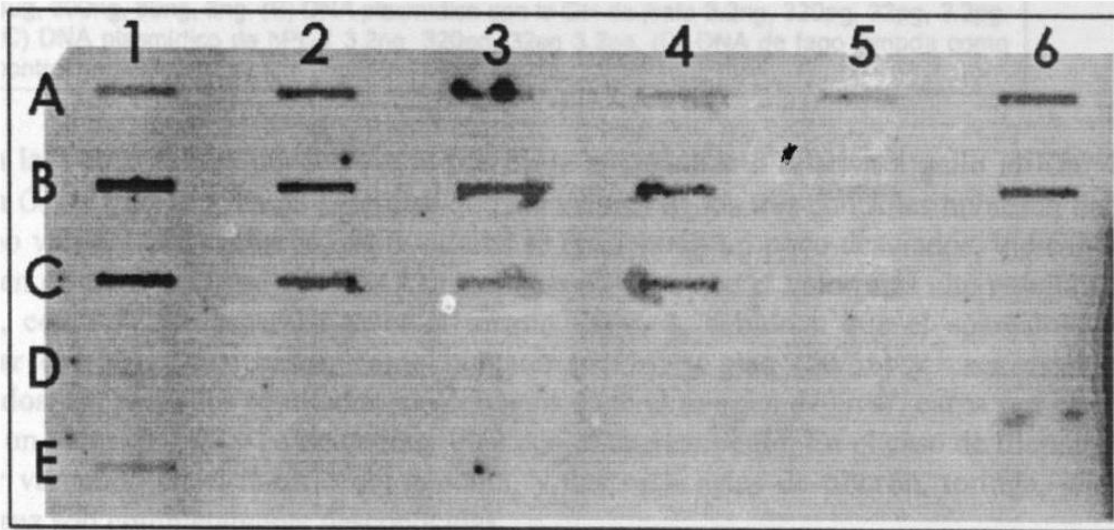


Figura 4.9 Hibridación de *GH-Explorer I* con DNAG de varias especies. Línea A: Humano, gato, vaca, jirafa, cerdo y ratón (de 1 a 6 respectivamente). Línea B: gallo, avestruz, tortuga, víbora, gecko, carpa. Línea C: Tilapia, ameca, catán, tiburón, guppy y pBR322 (Control negativo). Línea D: pBSjcGH-J5 (titulación). Línea E: gallo 800ng y gallo 80ng.

Tabla 4.3 Apareamiento de la sonda *GH-Explorer I* a DNA de diferentes especies ⁵.

	1	2	3	4	5	6		1	2	3	4	5	6
A	0.62	0.77	0.47	0.30	0.45	0.90	A	1.83	2.28	1.45	0.87	1.17	2.77
B	1.00	0.73	0.68	0.64	0.33	0.72	B	1.09	0.77	1.55	1.16	0.75	1.07
C	0.89	0.82	0.56	0.63	0.27		C	1.35	1.33	0.82	0.93	0.84	
D							D						
E	0.098	0.015					E	0.85	1.09				

⁵ (Izquierda) Integración densitométrica de las señales de la Figura 4.10 relativas a gallo. Se subrayan los que participaron en la curva de calibración. (Derecha) Fuerza de apareamiento relativo a gallo ($\{ \text{Antilog} [(\log(\text{Área_Relativa_a_gallo}) - b) / m] / (\text{Tamaño_Genoma_gallo} / \text{Tamaño_Genoma_Especie}) \}$); $m = 0.91374$, $b = -0.03238$, es decir, los que tienen un apareamiento "aparente similar" a gallo, tendrán un valor cercano a 1, menor para los que aparean menos y mayor para los que aparean más. En gecko, ameca y guppy se hicieron correcciones por la cantidad disminuida de DNAG. Los tamaños utilizados en cada genoma se muestran en Tabla 3.4.

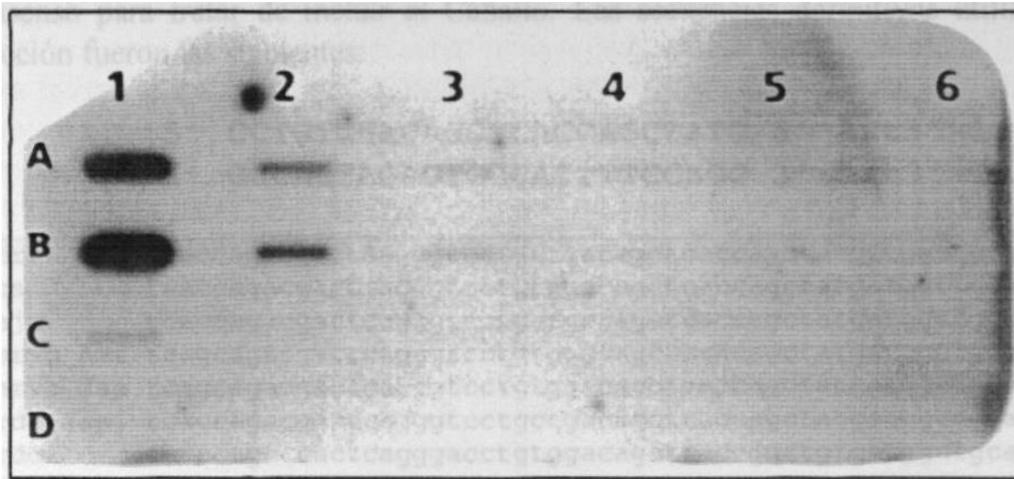


Figura 4.10 Hibridación de *GH-Explorer I* con titulaciones de DNA. (A) DNAg de gallo 8 μ g, 800ng, 80ng, 8ng. (B) DNA plasmídico con la GH de jirafa 3.2ng, 320pg, 32pg, 3.2pg. (C) DNA plasmídico de hPL-1 3.2ng, 320pg, 32pg, 3.2pg. (D) DNA de fago lambda como control negativo.

En la Tabla 4.3, se pueden ver la fuerza de apareamiento relativa a gallo utilizando la sonda *GH-Explorer I*. Es de esperarse que los valores de los tres controles hubiesen dado el mismo valor: 1, sin embargo, estos valores se encuentran un poco desviados, indicando un margen de error cercano al 15%. Asimismo se observa que el valor más alto resulta ser en ratón, con 2.77; se esperaba tener el mismo valor, 1, debido a que el apareamiento es similar (51/56). Otros casos, como humano (44/56) y gato (50/56) y vaca están muy elevados. En peces los resultados son cercanos dado el margen de error, carpa por ejemplo, tiene un valor de 1.07 y coincide muy bien con el apareamiento. En el caso de tilapia puede haber variación en el tamaño del genoma, y, las secuencias de tiburón, tortuga, gecko y avestruz son completamente desconocidas.

4.6 Diseño de Oligonucleótidos Iniciadores para PCR

Para el diseño de iniciadores para PCR se usaron dos metodologías. La primera y la más común fue el alineamiento de las secuencias de los artiodáctilos para buscar un consenso. La segunda metodología fue usando la estrategia de búsqueda de consensos como en el diseño de la sonda *GH-Explorer I*.

4.6.1 Diseño de Iniciadores para Artiodáctilos

El alineamiento realizado en *CLUSTALW* de las secuencias 5' y 3' de Artiodáctilos se muestra en la Figura 4.11 y Figura 4.12 respectivamente. Ambos oligonucleótidos tienen una longitud de 24nt y la temperatura de fusión media (T_m) resulta ser de 69°C y 72°C para el 5' y 3' respectivamente. El nucleótido en la posición 6 de Art5'-140 fue cambiado del consenso para romper una estructura de horquilla (apareamiento intramolecular)