



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Instituto de Investigaciones Biomédicas
Facultad de Medicina

**Secuenciación de ADN por shotgun y
análisis genómico de ~ 500,000 pb
contenidos en once fósidos que
traslapan el locus de Fusión /
Histocompatibilidad (Fu/HC) en el
protocordado *Botryllus schlosseri***

T E S I S
PARA OBTENER EL TÍTULO DE:
Licenciado en Investigación Biomédica
Básica P R E S E N
T A **Ricardo Godínez
Moreno**

Director de Tesis: Dr. Gabriel Moreno Hagelsieb
Co-director honorario: Dr. Bruce A. Roe.



Ciudad Universitaria, México, DF. Junio de 2004



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

El presente trabajo fue realizado en colaboración con el Advanced Center of Genome Technology (ACGT) en la Universidad de Oklahoma y el Centro de Investigación en Fijación del Nitrógeno en la Universidad Nacional Autónoma de México. Agradezco al **Dr. Bruce Roe** por darme la oportunidad de experimentar todos los pasos a seguir en la secuenciación de un genoma, a **Gabriel Moreno Hagelsieb** por su ayuda en la interpretación de resultados, a **Fares Najar** y a todo el equipo del ACGT por su amistad y por enseñarme como trabajar bajo presión utilizando al máximo las tecnologías de secuenciación a gran escala. Al **Dr Laclette** por su gran apoyo y confianza. A las **Doctoras Patricia Ostrosky, Yolanda López Vidal y Edda Sciutto** por sus invaluable consejos durante mi carrera y al **Dr. Edmundo Lamoyi** por su ayuda durante el proceso de titulación y por darme el ejemplo de lo que no quiero hacer de mi carrera como hombre de ciencia, de academia y de negocios.

Finalmente quiero agradecer muy especialmente a mis padres y hermanos por todo su apoyo y por estar siempre cuando y donde los necesito. Y a Nathalie Planteligne por su alegría y gran apoyo intelectual, crítico, físico y moral durante el tiempo que me tomo realizar este trabajo de investigación.

¿Que necesitamos para emprender un verdadero cambio en México?

En 1943 Winston Churchill presagió que;

"Todos los grandes imperios del futuro serán imperios de la mente."

" Y el futuro de tu familia y de tu país depende, en gran parte, de tu habilidad, la de tus vecinos y de tus hijos, para entender y crecer en un ámbito cada vez más tecnológico".

Generar riqueza ya no requiere muchas manos, sino poner a trabajar unas cuantas mentes.

En 1950 Singapur era una isla marginada... Con tales problemas que le pidió a su vecino, Malasia . Que por favor le permitiera... volverse parte de su territorio.

Malasia lo pensó... vio mucha gente pobre... y decidió que no le convenía...Que incorporar a Singapur para ampliar su territorio...Bajaría la productividad y le costaría mucho a Malasia.

A Singapur no le quedó más que EDUCAR a su gente, reformar su gobierno,

Atraer conocimientos y ponerse a trabajar.

Y hoy su gente tiene un nivel de vida ocho veces mejor que el de los mexicanos...

Y casi tres veces más que Malasia.

Ahora, sólo los que producen conocimientos...Y permiten que en su población pueda utilizar estos conocimientos...

Se vuelven más ricos.

Y por esto, los grandes imperios del siglo XXI son... y serán.....

Imperios de la mente.

Churchill tenía razón!

Juan Enriquez Cabot

La mayor parte de los mexicanos seguimos preocupados y ocupados en tratar de establecer o reestablecer las normas

básicas de convivencia y operación de una sociedad. Nos preocupa el creciente crimen en las calles, la corrupción, la falta de empleo y los salarios paupérrimos.

En la medida que la supervivencia diaria nos abrumba, perdemos la habilidad para entender el largo plazo. Hay poco interés y tiempo para dedicarnos a la investigación científica... Y no vemos con buenos ojos que el emprender negocios con la ciencia y crear corporaciones a partir de ideas creadas en Universidades Públicas, es una estrategia a largo plazo que nos permitiría ser más competitivos internacionalmente, no solo como universidad sino como nación. Y nos falta tiempo para estudiar los **CAMBIOS** que están ocurriendo en otros países. Mientras que otros países empiezan a pensar que lo único que tenemos que aportar los Mexicanos... Es cada vez menos relevante... y nos usan sólo para MAQUILAR y no para INVESTIGAR y No para CONSTRUIR. Y nos volvemos empleados... No socios, No equipo, No competitivos, No emprendedores, pese a sus excepciones...

Ricardo GodínezMoreno¹

-Continuación-

Así es que **para que ocurra un verdadero CAMBIO...**

Y que Una nación como México...

Una Compañía...

Un individuo...

Un estudiante...

Un científico...

Lleguen a triunfar...

Depende de su habilidad para trabajar en equipo

Y entender...

Y aplicar...

Cambios en tecnología.

Y de apoyar a aquellos que estén aprendiendo a generar

un **Cambio** y quieran triunfar al lado y en beneficio de su
nación.

No basta...

Controlar la inflación...

La paridad del dólar...

Las políticas demagógicas...

Los micro changarros...

Equilibrar el presupuesto...

Recortar gastos...

Privatizar...

Apoyar a los que ya triunfaron...

Modificadas y complementadas del libro ¹Citas :

AS THE FUTURE CATCHES YOU;
,HOW Genomics & OTHER FORCES ARE CHANGING
YOUR LIFE, WORK, HELATH & WEALTH.

ÍNDICE

| | |
|--|--------------------------------------|
| RESUMEN | ¡ERROR! MARCADOR NO DEFINIDO. |
| ABSTRACT | ¡ERROR! MARCADOR NO DEFINIDO. |
| AGRADECIMIENTOS | ¡ERROR! MARCADOR NO DEFINIDO. |
| ¿QUE NECESITAMOS PARA EMPRENDER UN VERDADERO CAMBIO EN MÉXICO? ¡ERROR! MARCADOR NO DEFINIDO. | |
| -CONTINUACIÓN- | ¡ERROR! MARCADOR NO DEFINIDO. |
| ÍNDICE..... | 1 |
| LISTA DE FIGURAS..... | 3 |
| LISTA DE TABLAS | 4 |
| CAPÍTULO 1 -- INTRODUCCIÓN | ¡ERROR! MARCADOR NO DEFINIDO. |
| 1.1 DEL ADN A LOS GENES A LOS GENOMAS | ¡ERROR! MARCADOR NO DEFINIDO. |
| 1.1.1 ADN | ¡Error! Marcador no definido. |
| 1.1.2 Dogma..... | ¡Error! Marcador no definido. |
| 1.1.3 Genes | ¡Error! Marcador no definido. |
| 1.1.4 Genoma y Modelos Genómicos..... | ¡Error! Marcador no definido. |
| 1.2 EL UROCORDADO COLONIAL <i>BOTRYLLUS SCHLOSSERI</i> | ¡ERROR! MARCADOR NO DEFINIDO. |
| 1.3 SIGNIFICADO, ORIGEN Y FUNCIÓN DEL ALORECONOCIMIENTO. | ¡ERROR! MARCADOR NO DEFINIDO. |
| 1.3.1 El reconocimiento propio/no propio, aloreconocimiento, en <i>Botryllus schlosseri</i> ; | ¡Error! Marcador no definido. |
| 1.4 EL LOCUS DE FUSIÓN/HISTOCOMPATIBILIDAD, FU/HC. | ¡ERROR! MARCADOR NO DEFINIDO. |
| 1.4.1 Importancia Inmunológica y Evolutiva del Fu/HC en <i>Botryllus schlosseri</i> ; | ¡Error! Marcador no definido. |
| 1.5 ¿QUE SE CONOCE SOBRE EL GENOMA DE <i>BOTRYLLUS SCHLOSSERI</i> ? .. | ¡ERROR! MARCADOR NO DEFINIDO. |
| 1.6 ANÁLISIS GENÓMICO..... | ¡ERROR! MARCADOR NO DEFINIDO. |
| CAPÍTULO 2 – OBJETIVOS | ¡ERROR! MARCADOR NO DEFINIDO. |
| OBJETIVO PRINCIPAL..... | ¡ERROR! MARCADOR NO DEFINIDO. |
| OBJETIVOS SECUNDARIOS | ¡ERROR! MARCADOR NO DEFINIDO. |
| CAPÍTULO 3 – MATERIALES Y MÉTODOS... ¡ERROR! MARCADOR NO DEFINIDO. | |
| 3.1 PURIFICACIÓN DE ADN Y SECUENCIACIÓN POR “SHOT GUN” DE LOS FÓSMIDOS QUE TRASLAPAN FU/HC EN <i>BOTRYLLUS SCHLOSSERI</i> | ¡ERROR! MARCADOR NO DEFINIDO. |
| 3.1.1 Purificación de ADN a gran escala | ¡Error! Marcador no definido. |
| 3.1.1.1 Electroforesis en Gel de Agarosa | ¡Error! Marcador no definido. |
| 3.1.2 Construcción de librerías genómicas y secuenciación por shotgun. ... | ¡Error! Marcador no definido. |
| 3.1.3 Shotgun | ¡Error! Marcador no definido. |
| 3.1.3.1 Nebulización | ¡Error! Marcador no definido. |
| 3.1.3.2 Hidro fragmentación “Hydroshear” | ¡Error! Marcador no definido. |
| 3.1.3.2 Reparación de extremos. | ¡Error! Marcador no definido. |
| 3.1.3.4 Selección de insertos por tamaño | ¡Error! Marcador no definido. |
| 3.1.4 Subclonación de insertos y transformación | ¡Error! Marcador no definido. |
| 3.1.5.1 Aislamiento semi-automatizado y automatizado de templados de ADN para secuenciar; | ¡Error! Marcador no definido. |
| 3.1.5.2 Aislamiento semi-automatizado de subclona de ADN | ¡Error! Marcador no definido. |
| 3.1.5.3 Aislamiento automatizado de subclonas usando el robot Zimark..... | ¡Error! Marcador no definido. |
| 3.2 CONDICIONES DE LOS CICLOS DE SECUENCIACIÓN..... | ¡ERROR! MARCADOR NO DEFINIDO. |
| 3.2.1 Eliminación de terminadores no incorporados..... | ¡Error! Marcador no definido. |
| 3.2.2 Secuenciación | ¡Error! Marcador no definido. |
| 3.2.3 Ensamblaje de secuencias..... | ¡Error! Marcador no definido. |
| 3.3 RELLENAMIENTO DE HUECOS “GAP CLOSURE” | ¡ERROR! MARCADOR NO DEFINIDO. |
| 3.3.1 Caminata de iniciadores “Primer- walking” – Clonas de insertos grandes; | ¡Error! Marcador no definido. |
| 3.3.2 Reacción múltiple en cadena de la polimerasa (MPCR) | ¡Error! Marcador no definido. |

3.3.3 *Secuenciación de regiones difíciles* ;**Error! Marcador no definido.**
 3.4 MÉTODOS COMPUTACIONALES Y ANÁLISIS DE DATOS ;**ERROR! MARCADOR NO DEFINIDO.**
 3.4.1 *Análisis de secuencia* ;**Error! Marcador no definido.**

CAPÍTULO 4 – RESULTADOS Y DISCUSIÓN. ;ERROR! MARCADOR NO DEFINIDO.

4.1 SECUENCIACIÓN Y ANOTACIÓN DE 11 FÓSMIDOS QUE TRASLAPAN EL LOCUS FU/HC ;**ERROR! MARCADOR NO DEFINIDO.**
 4.3 NO EXISTE EVIDENCIA DE ELEMENTOS DE INMUNIDAD ADAPTATIVA EN EL FU/HC ;**ERROR! MARCADOR NO DEFINIDO.**
 4.2 REGIONES SINTÉNICAS DEL FU/HC CON EL MHC ;**ERROR! MARCADOR NO DEFINIDO.**
 4.3 ¿EXISTE ALGÚN GENE EN FU/HC ENVUELTO EN REACCIONES INFLAMATORIAS DURANTE EL ALORECONOCIMIENTO? ;**ERROR! MARCADOR NO DEFINIDO.**
 4.5 ELEMENTOS GENÉTICOS MÓVILES EN PLANTAS EN FU/HC ;**ERROR! MARCADOR NO DEFINIDO.**
 4.6 LA PARADOJA DE UN LÓCUS ALTAMENTE POLIMÓRFICO. ;**ERROR! MARCADOR NO DEFINIDO.**
 4.7 DE ASCIDEAS, HONGOS Y PLANTAS A LA UTILIDAD DE LA TEORÍA EVOLUTIVA. CUÁL ES LA RELACIÓN DE TRES PROTEÍNAS CON VARIOS DOMINIOS WD40 Y LA E3 UBIQUITIN LIGASA ENCONTRADA EN EL FU/HC? ;**ERROR! MARCADOR NO DEFINIDO.**

CAPÍTULO 5 –CONCLUSIONES Y PERSPECTIVAS ;ERROR! MARCADOR NO DEFINIDO.

CONCLUSIONES ;**ERROR! MARCADOR NO DEFINIDO.**
 PERSPECTIVAS ;**ERROR! MARCADOR NO DEFINIDO.**

REFERENCIAS ;ERROR! MARCADOR NO DEFINIDO.

FIGURAS ;ERROR! MARCADOR NO DEFINIDO.

Lista de Figuras

| <i>Número</i> | <i>Página</i> |
|--|---------------|
| 1. Filogenia de vertebrados e invertebrados | 57 |
| 2. Evolución del sistema inmune | 58 |
| 3. Anatomía de <i>Botryllus schlosseri</i> | 59 |
| 4. Anatomía de larvas cordadas..... | 59 |
| 5. Ciclo de vida de <i>B. schlosseri</i> | 59 |
| 6. Alorreconocimiento ie. Reconocimiento propio y no propio. | 60 |
| 7. El locus de fusión e Histocompatibilidad Fu/HC..... | 60 |
| 8. Los fenotipos de Fusión y rechazo activados por el Fu/HC | 61 |
| 9. Resumen del Método de Secuenciación por Shot Gun | 61 |
| 10. Resumen detallado del método de secuenciación, descripción de los pasos 1-6..... | 62 |
| 11. Resumen detallado del método de secuenciación, descripción de los pasos 7-10..... | 63 |
| 12. Muestras de ADN de once fósmidos a gran escala | 64 |
| 13. Shot gun de ADN, usando dos métodos de fragmentación aleatoria | 64 |
| 14. Subclonas. | 64 |
| 15. Resultados de Phred visualizados en Consed | 65 |
| 16. Phrap, Ensamblaje de datos visualizados en Consed | 65 |
| 17. Consed es una interfase gráfica, un visualizador para Phred/Phrap | 65 |
| 18. Estrategia para el análisis y la anotación de secuencias genómicas. | 66 |
| 19. Resultados de Blast P visualizados en Artemis. | 66 |
| 20. Predicción de genes y estructuras de exones, visualizadas en Artemis | 67 |
| 21. Resultados de Blastphi y Rps Blast. | 67 |
| 22. Predicción de proteínas transmembranales con SOUSOUI..... | 67 |
| 23. Resultados de la secuenciación y anotación de once fósmidos contenidos en el locus Fu/HC..... | 68 |

Lista de tablas

Número

Página

| | |
|---|----|
| 1. Resultados en GENE BANK relacionados a <i>B.Schlösseri</i> en NCBI | 11 |
| 2. Series de reacciones de ligación | 23 |
| 3. Tablas de Anotación | 37 |

Resumen

Los componentes de la inmunidad adaptativa y del Complejo Mayor de Histocompatibilidad (MHC) son esenciales en cualquier sistema de trasplante basado en la respuesta al reconocimiento propio/no propio que se encuentra en todos los vertebrados. A pesar de no existir una indiscutible correlación filogenética y funcional o una evidencia directa sobre la aparición del MHC o de la inmunidad adaptativa, existe un posible candidato a tener componentes de un MHC ancestral o de vestigios de un sistema inmuno adaptativo primitivo; El locus de Fusión e Histocompatibilidad (Fu/HC) contenido en el genoma del protocordado *Botryllus schlosseri*, quien es capaz de detonar un fenotipo de aceptación o rechazo de tipo MHC entre colonias formadas por múltiples unidades de zooides de *B. schlosseri*. La ubicación filogenética en la evolución de esta ascidea marina, se ubica en la etapa que comprende la Gran Explosión inmunológica “immunological Big Bang”, hace 450-500 millones de años. *B. Schlosseri* es miembro del grupo de los ancestros inmediatos de los vertebrados, los protocordados y también es considerado como un potencial candidato a poseer un MHC ancestral o genes relacionados con la emergencia de la respuesta inmuno-adaptativa. Los datos experimentales y el análisis aquí descrito corresponden a la secuenciación por shotgun de 11 fósidos que cubren aproximadamente 500,000 pares de bases reportados en Genbank, además del análisis y a la anotación de dicha secuencia. Con esta información se estudió el locus Fu/HC, el cual se decodificó parcialmente con el objetivo de elucidar una más detallada organización génica. Así, se reporta el segmento genómico más extenso relacionado con *B. schlosseri*, el cual conforma aproximadamente un 0.07% de su genoma. El ensamblaje de la secuencia fue realizado con Phred/Phrap y Consed. Los resultados preliminares de la anotación de Fu/HC fueron realizados una vez que se alcanzó una cobertura completa de 6X. La predicción preliminar de los modelos génicos fueron realizados con; FGENESH/HMM 2.0 y GENSCAN 1.0 con los cuales se predijeron 45 genes modelo. Además, haciendo uso de Blastx 2.2.6 and Blastp 2.2.6, se detectaron 16 genes modelo homólogos a genes reportados en Genebank con valores $E < 10^{-6}$. En todos ellos, se detectaron dominios conservados mediante el uso de RPS-BLAST 2.2.6. Debido a la importancia de elucidar un receptor potencial de tipo MHC, se utilizó SOSUI, un programa capaz de predecir dominios proteicos transmembranales. Gracias a este análisis, se encontraron 6 proteínas con dominios membranales. Por otra parte, entre los elementos posiblemente relacionados a un sistema inmune ancestral se descubrieron tres genes modelos codificantes de proteínas con varios dominios WD40, una E3 ubiquitin ligasa, una Fosfolipasa A2, una Serin Treonin Kinasa y una proteína transmembranal con una breve homología al factor del complemento C2-Bf. En conjunto, todos estos elementos tienen el potencial de conformar parte de un sistema ancestral de histocompatibilidad, donde Fu/HC juega un papel fundamental. Como el primer estudio de un locus de histocompatibilidad de un metazoano fuera del grupo de los vertebrados, la información obtenida en este trabajo de investigación será de gran valor para futuros análisis de Genómica comparativa y así poder entender los orígenes del sistema inmuno-adaptativo

Abstract

The components of the adaptive immunity and the Major-histocompatibility complex (MHC) are essential in any major transplantation system based on a self/non-self recognition response importantly found in all vertebrates. Even so, no strong phylogenetic or functional specific and direct evidence has been found to the origin of MHC or adaptive immunity. There is a possible candidate to embrace components of an ancestral MHC or vestiges of a primitive immuno-adaptive system. The Fusibility Histocompatibility locus (Fu/HC) contained in the genome of the colonial protochordate *Botryllus schlosseri*, which triggers a rejection or acceptance phenotype resembled by an MHC like response given between colonies formed by multiple units of *B. schlosseri* zooids. Situated in the emerging time that encompass the “immunological big”, 450-500 million years ago, *B. schlosseri* is an immediate vertebrate ancestor and a potential candidate to possess an ancestral MHC or immunoadaptive response related genes. The experimental data and analysis here reported, includes the shotgun sequencing description of eleven fosmids encompassing approximately 500,000 pb of ADN reported by Genbank and preliminary genomic analysis corresponding to the annotation of this sequence. This is the biggest reported genomic information related to the genome of *B. schlosseri*. In this analysis, part of the Fu/HC has been decoded in order to elucidate a more detailed gene organization and analyze 0.07% of the 725bp haploid genome of *B. schlosseri*. When the sequence coverage of every fosmid reached about 6 X coverage, the annotation process started. In a preliminary gene model prediction; FGENESH/HMM 2.0 and GENSCAN 1.0 predicted 45 putative genes. Besides, a consistent homologous gene detection analysis using Blastx 2.2.6 and Blastp 2.2.6, found 18 homologous genes with supporting E values ($E \leq 10^{-6}$) to genes at Genbank. Consistently, functional domains were predicted in 16 genes through a search with RPS-BLAST 2.2.6. Due to the importance of finding a potential receptor MHC like, SOSUI a Transmembranal protein domains predictor program was used to find 6 model genes encoding potential Transmembranal proteins. Regarding an ancestral immune system 7 potential elements were found; three model genes coding for proteins with several WD40 domains, an E3 ubiquitin ligase, a Fosfolipasa A2, a Serin Treonin Kinasa and a Transmembranal protein with a brief homology to the complement system protein C2-Bf. All together they might be possible elements of an ancestral histocompatibility system where Fu/HC plays an essential role. As the first metazoan histocompatibility locus analyzed outside of vertebrates, this study will be very valuable for future comparative genomics analysis to understand the origins of the “adaptive immune system”.

INTRODUCCIÓN

En un amplio sentido esta tesis es sólo un ejemplo del producto de un siglo de investigación con el objetivo de comprender el ADN; la molécula de la vida y la vida misma a través de ella. El punto de inicio se dio a partir del redescubrimiento de las leyes de Mendel a principio del siglo veinte, dando pauta de que la información genética es transmitida de generación en generación, de manera discreta (Stern C y Sherwood E, 1966). Durante el primer cuarto de siglo, los hombres de ciencia descubrieron que las bases celulares de la información genética están en los cromosomas. Durante el segundo cuarto, los oportunistas James D Watson y Francis Crick elucidaron la estructura molecular del ADN (Watson JD y Crick FH, 1953). Durante el tercer cuarto, los biólogos develaron los mecanismos por los cuales las células leen y procesan esta información. Además, desarrollaron herramientas de ADN recombinante, gracias a las cuales los científicos hoy en día, pueden leer, modificar y procesar la misma información. En el comienzo del siglo XXI la tecnología del ADN ha ido incrementando exponencialmente, llevando una tendencia creciente que va desde secuenciar y estudiar genes de manera individual, hasta leerlos y estudiarlos en un contexto global como genomas. Debido a la rápida y cada vez mas eficiente tecnología en secuenciación, hoy en día es posible visualizar un gene no como una sola unidad, sino como parte de un genoma. El principal objetivo de esta tesis es comprender y estudiar globalmente parte de la molécula de la vida situada en un urocordado casi desconocido genómicamente; la ascidea marina colonial *Botryllus schlosseri*. Debido a la secuenciación y anotación de aproximadamente medio millón de pares de bases contenidas en once fósmidos que traslapan el locus de Fusión e Histocompatibilidad (Fu/HC), no sólo la filogenia y la escala del tiempo genómica de la vida será mejor entendida, más aún, genes esenciales posiblemente relacionados a un sistema ancestral de reconocimiento propio y no propio serán revelados. La secuencia de ADN y el análisis global presentado en esta tesis proveerá de información importante del genoma de *B.schlosseri*. Además, en un futuro próximo será de gran utilidad, como información inicial para comprender el origen del sistema immuno-adaptativo de los vertebrados, donde *B.schlosseri* juega un papel central como ancestro inmediato de los vertebrados (Laird et al., 2000).

CAPÍTULO 1

1.1 Del ADN a los Genes a los Genomas

1.1.1 ADN

El ácido desoxiribonucleico (ADN) es una simple molécula esencial para la vida, que resulta un verdadero reto entender (Godínez R., 2001). Tanto en *B. schlosseri* como en todos los organismos vivos, la molécula del ADN consiste de dos hebras que se envuelven entre sí, para formar una estructura parecida a una escalera de caracol compuesta de dos hebras anti paralelas de nucleótidos ligados entre sí (Watson JD y Crick FH., 1953). Un nucleótido está compuesto de cuatro bases nitrogenadas; Timina (T), Citosina (C), Adenina (A) y Guanina (G), el azúcar pentosa 2-desoxiribosa y un grupo fosfato. Las bases nitrogenadas están ligadas por puentes de hidrógeno; Así, Adenina y Timina están ligadas por dos puentes y Citosina y Guanina, por tres. Estas interacciones débiles pero específicas, mantienen un par de bases complementarias apareadas, manteniendo así, las dos hebras de ADN unidas. En el extremo final de cada hebra antiparalela de ADN, un nucleótido 5' tiene un fosfato en la posición cinco prima, mientras que por el contrario, en el extremo final 3' el nucleótido tiene un grupo hidroxilo libre en la posición tres prima y *visceversa*. El orden particular de los nucleótidos adyacentes que permanecen ligados uno a otro a través de una columna de fosfodiesteres, es llamado en conjunto secuencia de ADN. La secuencia de ADN constituye de manera específica las instrucciones genéticas exactas, requeridas para constituir un organismo particular dotado de características propias y únicas.

1.1.2 Dogma

El ADN es el código de la vida. La importancia de su función es reivindicada por rasgos monótonos estructurales pero complejos en su entendimiento. Su papel consiste en preservar y mover información vital en la unidad de la vida: la célula (Micklos D y Freyer G, 1990). El ciclo de vida informacional del ADN es mejor conocido como

“El dogma central de la biología molecular” (Watson JD y Crick FH, 1953). El cual permanece como regla esencial en todos en todos los seres vivos. Este dogma constituye la base de la biología molecular. No obstante, ha sido modificado hasta la fecha, y esta representado por cuatro etapas principales (Crick, 1974). Primero, el ADN es replicado, preservando su información a través de un proceso llamado replicación. Por otro lado, el Ácido Ribonucleico (ARN) es producido usando ADN como plantilla, durante un proceso llamado transcripción. Finalmente mediante un proceso llamado translación, el ARN mensajero acarrea la información codificada por el ADN a los ribosomas, donde es leída y usada para la síntesis de proteínas.

1.1.3 Genes

Cada molécula de ADN contiene decenas de miles de genes, también llamados unidades físicas básicas y funcionales de la herencia. (Micklos D y Freyer G, 1990). Un gene es una secuencia específica de bases nucleotídicas que contiene la información requerida para regular y construir ARN o proteínas, las cuales constituyen los componentes estructurales de las células. Así mismo, son parte de tejidos y enzimas, las cuales llevan a cabo reacciones bioquímicas indispensables para la vida de toda célula. Los genes varían ampliamente en extensión, comúnmente se extienden sobre varias miles de bases. “La mayoría de los genomas eucariontes constan de alrededor del 10% o menos de secuencias codificantes de proteínas, también llamadas exones. (“www.accessexcellence.com”). Dispersadas en muchos genes se encuentran las secuencias intrónicas, las cuales no cuentan con función codificante alguna. Otras regiones no codificantes incluyen secuencias regulatorias, regiones intragénicas y elementos repetitivos. Así, toda secuencia de ADN, incluyendo elementos complejos codificantes y elementos monótonos no codificantes conforman un Genoma.

1.1.4 Genoma y Modelos Genómicos

Un genoma es toda la información contenida en una molécula de ADN. La información genómica provee de muchos de los elementos necesarios para preservar y determinar la vida de todo ser vivo. Durante las últimas décadas, el termino “organismo modelo” ha

sido utilizado de manera estrecha en aquellas especies animales de fácil manejo en el laboratorio. Un ejemplo de ellas son: la mosca de la fruta o *Drosophila melanogaster* y el ratón de casa o *Mus musculus*. Este tipo de organismos modelo cuentan con características óptimas para la experimentación; como lo es un tamaño óptimo para su manejo y un rápido periodo de reproducción (Hedges, 2002). Sin embargo, con el incremento en número de proyectos de secuenciación genómica, esta definición ha sido ampliada. Por ejemplo, hoy en día la atención ha sido dirigida a un mayor número de organismos, tales como la ascidea, el pollo, el tigre, el pez globo, la vaca, el perro, el mono babuino etc. Debido al aspecto único de su posición filogenética en vez de su factibilidad para formar parte como objetos experimentales de estudios científicos. Así, estos organismos son referidos como “modelos genómicos” (Hedges, 2002). Tal y como sucede con el urocordado, *Ciona intestinallis*, cuyo genoma ha sido recientemente secuenciado (Dehal et al., 2002). La importancia genómica de *B. Schlosseri*, un cercano familiar evolutivo de *C.intestinallis*, esta relacionada a su posición como miembro del grupo de los coradados inferiores, los cuales se ubican en el árbol de la vida como la ancestros inmediatos de los vertebrados. (Figura.2) Tanto *B. schlosseri* como *C.intestinallis* son interesantes desde una perspectiva Genómica y evolutiva. Sin embargo, *B.schlosseri* posee un rasgo de importancia inmunológica que *C. intestinallis* no posee: un locus que activa de manera natural, un mecanismo de reconocimiento propio y no propio, que posiblemente se encuentra situado en el periodo de emergencia del sistema inmuno adaptativo (Scofield et al., 1982).

1.2 El urocordado colonial *Botryllus schlosseri*.

B. schlosseri es una ascidea colonial, un organismo secretor de ácido que reside en habitats marinos subtidiales poco profundos. (AW de Tomaso 1998) (Figura. 2). Es un organismo común de los sedimentos marinos que se acumulan generalmente en los barcos. Este animal es en realidad una colonia de individuos que comienzan su vida como larvas individuales e independientes. Una vez establecida la larva, se divide en clonas genéticamente idénticas las cuales forman una colonia, la cual eventualmente

medirá varios centímetros de diámetro (Chadwick-Furman and Weissman,1995).

Como protocolado, *B.schlosseri* es un miembro de la subphylum de los urochordados en la cual ocupa una posición clave como elemento del grupo de organismos que fueron los ancestros inmediatos de los vertebrados. **(Figura. 1)**. (Satoh and Jeffrey, 1995).

El ciclo de vida de *B.schlosseri* oscila entre los ciclos de reproducción sexual y asexual por fisión. Además, sufre una metamorfosis de manera radical entre un plan de desarrollo vertebrado cuando es una larva a uno invertebrado cuando es adulto. Debido a sus características hermafroditas *B.schlosseri* es capaz de producir gametos femeninos y masculinos. El resultado de la reproducción sexual, es una larva de 0.5mm, un vertebrado con todos los rasgos característicos de cualquier cordado que incluyen notocorda, tubo neural dorsal, cola estriada, faringe y una musculatura segmentada (Figura. 3) (Rinkevich B, 2002). Una vez separada de la colonia madre, la larva nada a una superficie cercana y se ancla a un substrato, se establece por varias horas y sufre de una metamorfosis, la cual se caracteriza por una reabsorción de casi todas las características cordadas. A lo largo del periodo de maduración, la larva cambiará todo su plan de desarrollo vertebrado a invertebrado, para convertirse así en un adulto zooide que medirá al menos de 4mm (Figura. 4).

Cada zooide tiene una esperanza de vida de una semana, tiempo durante el cual se reproduce tanto sexualmente como asexualmente al brotar nuevos individuos genéticamente iguales. El tiempo de gestación es de entre 5-7 días, por tanto cada semana un nuevo grupo de larvas es liberado. Cada uno de estos nuevos individuos son hermafroditas y heterogéneos genéticamente. En *B.schlosseri*, la metamorfosis es seguida de un proceso de asexual de reproducción el cual dará origen a una colonia de descendientes genéticamente iguales y unidos por una red vascular común (Weissman, 1982). Cada colonia esta hecha desde 2 hasta >1000 individuos derivados asexualmente (zooides). Conforme una colonia se expande asexualmente, comúnmente interactúan con otras colonias cercanas **(Figura.2)** (Rinkevich, 2002).

1.3 Significado, origen y función del Alorreconocimiento.

Aloreconocimiento es la Habilidad de un organismo individual de distinguir entre células propias y no propias. Este mecanismo ha sido descrito en una gran variedad de organismos multicelulares (F.M Burnet 1971). La habilidad de discriminar entre lo propio y lo no propio, es una estrategia de protección ampliamente conservada. En organismos tales como los hongos, plantas y metazoanos, el alorreconocimiento no solo sirve como un arma en contra de patógenos, mas aún, es una estrategia que sirve para el mantenimiento de la individualidad misma de los organismos (Buss, 1987).

En un contexto de diversidad animal, el alorreconocimiento puede ser encontrado en dos formas distintas; por un lado, mediante la forma de inmunidad a trasplantes que ocurre tanto de manera natural como experimental y en la forma de bloqueo de la auto-fertilización, la cual es bastante común en un gran número de especies hermafroditas.

En muchos organismos, tales como esponjas, cnidarios, brazoanos y tunicados, la forma de trasplatación natural ocurre frecuentemente (DuPasquier 1999). En estos casos, la habilidad del alorreconocimiento es el único sistema indispensable y capaz de proteger los organismos de la fusión somática de individuos distintos genéticamente, permitiendo a los mismos mantener una homogeneidad genética y por tanto una individualidad que los distingue (Khalturin, 2001).

1.3.1 El reconocimiento propio/no propio, alorreconocimiento, en *Botryllus schlosseri*

El alorreconocimiento, es una estrategia para distinguir entre células propias y no propias, esta estrategia de defensa forma parte del fundamento de los orígenes de la respuesta inmunoadaptativa. Es un fenómeno evolutivo ancestral (Cooper, 2001). El subphylum más cercano filogenéticamente a la emergencia de los vertebrados en el cual ocurren reacciones de trasplatación *in vivo*, son los urocordados, también llamados Protochordados o tunicados (**Figura. 1**). El punto de inicio del sistema inmuno adaptativo se sitúa en la emergencia de los vertebrados y posiblemente existan vestigios

de este en los urocordados. A diferencia del origen desconocido de este mecanismo de defensa, es ampliamente conocido que los antecedentes evolutivos del sistema inmune de los vertebrados se origina en los sistemas de reconocimiento propio/no propio de los invertebrados cordados, particularmente en los organismos sedentarios y coloniales como las esponjas, corales, briozoarios y tunicados (Du Pasquier, 1999). El estilo de vida “familiar” y la dispersión restringida de los organismos coloniales, conlleva a una proximidad de con-específicos, lo cual ocasiona frecuentes interacciones tejido tejido. (Rinkevich, 2002) **(Figura. 3)**. *B. schlosseri* es un organismo colonial donde tales contactos resultan en la expresión de un intenso desplegado de respuestas de reconocimiento propio/no propio (Weissman, 1990) **(Figura.5)**. El aloreconocimiento puede ser definido por una expresión jerárquica de alelos de histocompatibilidad ampliamente polimórficos. La evolución y el sostenimiento de esta alta tasa polimórfica y la compleja naturaleza del aloreconocimiento han sido atribuidas a una variedad de fuerzas evolutivas (Magor et al., 1999). Sin embargo, su naturaleza es aún desconocida. La aparición y la respuesta de reconocimiento propio/no propio en *B. schlosseri* permitió la dilucidación de un locus de Fusión/Histocompatibilidad, Fu/HC (Scofield et al., 1982).

1.4 El locus de Fusión/Histocompatibilidad, Fu/HC.

Una amplia variedad de invertebrados marinos pueden destruir trasplantes de manera natural (DuPasquier L et.al 1999). *B. schlosseri*, representa de manera notable el ejemplo mejor caracterizado en estudios de histocompatibilidad y de rechazo (De Tomaso et al., 1998). Estos estudios se han concentrado en investigar las causas de una serie de mecanismos de reconocimiento propio y no propio en *B. Schlosseri*. Gracias a estos estudios ha sido descubierto un grupo especial de alelos que juegan un papel esencial en los eventos de aloreconocimiento en este urocordado (Yund P and Feldgarden M, 1992). Las bases genéticas para estos eventos de reconocimiento propio/no propio residen en un solo locus pero altamente polimórfico, llamado con el término el haplotipo de fusión/histocompatibilidad, Fu/HC (Weissman et al., 1990) **(Figura. 6)**. Basado en varios haplotipos de histocompatibilidad, tanto el fenotipo de rechazo como el de fusión son alternativamente codificados como rasgos simples de carácter Mendeliano y co-

dominante. Ambos fenotipos pueden ser activados en base a los siguientes principios genéticos y fenotípicos (**Figura.7**). Por un lado, el fenotipo de fusión es expresado cuando dos ámpulas de *B.schlosseri* interactúan y se fusionan gracias a que ambas comparten al menos un alelo de Fu/HC. Ambas colonias se fusionan a través de un contacto periférico de las ámpulas, formando sucesivamente colonias quiméricas e individuales de zooides de *B. Schlosseri* unidas mediante un sistema vascular compartido. Por otro lado, el fenotipo de rechazo es activado cuando no se comparte alelo alguno de Fu/HC. Este tipo de respuesta de histocompatibilidad no propia, activa una reacción inflamatoria la cual ocasiona lesiones citotóxicas entre dos ámpulas las cuales son destruidas con el fin de eliminar su interacción (Magor 1999) (**Figura.5**).

El locus Fu/HC es particularmente interesante desde el punto de vista de la genética poblacional. A pesar de ser extremadamente polimórfico, no debería serlo. La razón de este debate radica en que solo los individuos genéticamente relacionados son capaces de fusionarse. Este hecho puede disminuir potencialmente el polimorfismo en el locus (RinKevich et al., 1995). Sin embargo, esto no sucede, ya que Fu/HC codifica la expresión de alelos múltiples y co-dominantes, llegando hasta varios cientos por población (Yund, 1992). A pesar del origen congénico y asexual de gran parte de las colonias de *B.schlosseri*, el nivel polimórfico de Fu/HC asegura eficientemente que la gran mayoría de las colonias sean heterocigotas en este locus. Por estas razones, las poblaciones naturales de *B.Schlosseri* son capaces de mantener una gran variedad de alelos en este locus, aún si las colonias que se fusionan tienen parentesco (Grosberg and Hart , 2000). A la fecha, los intentos por identificar los genes que se encuentran en el Fu/HC o sus proteínas han sido fallidos (RinKevich., 2002). Sin embargo, los esfuerzos para identificar genes prevalentes en librerías de cADN, generadas de tejidos envueltos en el proceso de reconocimiento propio/no propio siguen en el camino. Algunos genes posiblemente involucrados en la respuesta inmune han sido identificados usando esta estrategia. Entre estos se encuentran; una Serin Proteasa, expresada en altos niveles en los sitios de reconocimiento, análogos funcionales a la interleucina-1, al factor de necrosis tumoral alfa (TNF alfa), opsoninas humorales y un receptor bsCD94, homólogo al receptor NK (Natural killer) de los vertebrados. Aunque es posible que estos elementos contribuyan al sistema de reconocimiento propio/no propio como parte de la

maquinaria efectora del reconocimiento, ninguno forma parte del Fu/HC (Margor et al., 1999).

1.4.1 Importancia Inmunológica y Evolutiva del Fu/HC en *Botryllus schlosseri*

Los *Agnathas* son los cordados más primitivos. Hace 450 -500 millones de años, estos vertebrados sin mandíbula y los chondrichthyes (vertebrados con esqueleto de cartílago) divergieron en un momento donde algo sorprendente ocurrió: el “Bing bang” inmunológico, es decir, el periodo de 50 millones de años durante el cual emergió el sistema inmuno-adaptativo (Laird 2000). Hasta la fecha no han sido encontrados elementos concluyentes y representativos de un sistema inmuno adaptativo ni en los invertebrados ni en los *Agnathas*. Además, tampoco ha sido descubierto en ninguno de ellos, vestigio alguno de un sistema inmunidad y transplatación basado en el MHC. Sin embargo, estos organismos resuelven perfectamente el problema del reconocimiento propio y no propio (DuPasquier L.,1999). ¿Que estrategia utilizan para resolver el problema del reconocimiento sin hacer uso de un MHC?. Actualmente no existe evidencia alguna que sustente la respuesta a esta interrogante. Es probable que mecanismos de discriminación propia/no propia, al menos en los deuterostomas, hayan sido conservados en la evolución e incorporados por los vertebrados en sus sistemas de inmunidad adaptativa (DuPasquier L.,2000). Por esta razón, quizás sea posible encontrar en los ancestros cordados invertebrados de los vertebrados, elementos ancestrales o mecanismos relacionados a los sistemas de reconocimiento propio/no propio conservados en vertebrados superiores (DuPasquier L., 2000). La posición filogenética estratégica de *B.schlosseri* se encuentra en la cúspide de la evolución de los cordados invertebrados hace 550 millones de años y el papel del reconocimiento de Fu/HC con sus similitudes funcionales al MHC. Hacen del protocordado *Botryllus schlosseri* un excelente organismo modelo para estudiar los orígenes de la inmunidad adaptativa de los vertebrados (Laird 2000).

1.5 ¿Que se conoce sobre el genoma de *Botryllus schlosseri*?

En la actualidad se cuenta con pocos datos sobre el genoma de *B. schlosseri*. La información existente incluye la determinación por citometría de flujo, “FACS” (Fluorescent Antigen Cell Sorting), de las 725 Mega bases que tiene por extensión el genoma haploide de este protocordado. Además, ha sido mapeada y caracterizada una región de 5.5 cM, la cual contiene el locus de Fusión Histocompatibilidad (Fu/HC) (De Tomaso et al., 1998). Conjuntamente, en ENTREZ, la base de datos de el Centro Nacional de Información Biotecnológica (NCBI), cuenta con 249 vínculos “links” que contienen información nucleotídica referente al genoma de *B.schlosseri*. Esta información consiste de secuencias de ADN complementario (cADN), Etiquetas de Secuencias Expresadas (EST), de ARN mensajero (mARN) además de marcadores genéticos que constan de ADN repetitivo o también llamado satélite. A la fecha, la secuencia contigua de mayor tamaño esta representada por los genes de las proteínas de choque calórico HSP60.1 y HSP60.2. Ambas constan de una extensión de alrededor de 9000 pb. Asimismo, a la fecha existen solo 25 proteínas anotadas y algunas de ellas caracterizadas. Actualmente, más información genómica esta siendo generada. Los resultados preliminares incluidos en esta investigación, sugieren la dilucidación de la secuencia contigua de mayor extensión en el genoma de *B.schlosseri*. La secuencia preliminar de los 11 fósidos secuenciados, analizados y reportados en esta tesis y en NCBI, incluyen una extensión aproximada de 0.5 Mega bases, equivalente al 0.07% de los 725 Mega bases que comprenden el genoma de *Botryllus*. Esta no es solo la mas grande secuencia contigua correspondiente al genoma de *B.Schlosseri* y reportada en NCBI, también tiene el potencial de ser una firma genómica (Hess CM, 2000).

| Registros de <i>Botrylus Schlosseri</i> en Entrez (NCBI) | |
|--|---------------------------|
| Nombre de la Bas de Datos | Vínculos Directos (Links) |
| Nucleótidos | 249 |
| Proteínas | 25 |
| Popset | 1 |
| Central PubMed | 11 |
| Taxonomía | 1 |

1.6 Análisis Genómico

Como estudiante y miembro de un Centro Avanzado de Tecnología Genómica (ACGT), considero de vital importancia el describir en una breve introducción las cuatro

herramientas de Secuenciación. Todas ellas están incluidas en cualquier Análisis Genómico. (Figura. 8 y Figura 9).

Tabla 1. Resultados en GENE BANK relacionados a *B.Schlosseri* en NCBI

La primer herramienta utilizada en cualquier análisis genómico es la Secuenciación. Se realiza mediante la reacción de terminación de cadena de di-desoxinucleótidos (the dideoxy-chain terminating reaction) o reacción de secuenciación de ADN desarrollada por Fred Sanger (Sanger et al., 1977). En este método es posible visualizar al ADN como una mezcla de secuencias llamadas templados, los cuales son en principio empalmados con un oligonucleótido corto, complementario y selectivo, llamado iniciador “primer”. El ADN polimerasa y los deoxi-nucleótidos además de algunos catalizadores e isotopos radioactivos o fluorescentes, son adicionados a la mezcla antes descrita. Así, la enzima ADN polimerasa usa todos los elementos mencionados

anteriormente para catalizar la síntesis de la hebra complementaria de ADN. El iniciador es “el ancla” al punto de inicio de la extensión de la cadena o hebra de ADN, la cual se extiende en dirección 3’.

Sanger incorporó en la mezcla de reacción un nucleótido con el propósito de que “terminara” con la reacción de cadena. Este “terminador” es un deoxinucleótido trifosfatado modificado, donde el grupo hidroxilo 3’ es reemplazado con un átomo de hidrógeno, creando un di-deoxinucleótido. Cuando la reacción de secuenciación procede en presencia de una pequeña cantidad de di-deoxynucleótidos, los productos de la reacción de extensión de cadena resultante comienzan la síntesis en el iniciador extendiéndose en dirección 3’, pero terminando al momento de la incorporación del di-deoxinucleótido, el cual se incorpora de manera opuesta a su nucleótido complementario quien forma parte de la secuencia templado. Este grupo de productos resultantes de la reacción de terminación son llamados secuencia en escalera “sequencing ladder”. Cuatro de estas escaleras se sintetizan usando cuatro di-desoxinucleótidos diferentes (ddA, ddC, ddG, ddT), los cuales terminan cadenas distintas y de manera independiente, formando secuencias en escalera de Adenina, Citosina, Guanina y Timina..

La segunda herramienta requerida en cualquier análisis genómico es una estrategia que tiene como objetivo preparar miles de templados de ADN. Estos templados son requeridos para secuenciar grandes segmentos genómicos de manera más rápida, puesto estos tienen menor tamaño y por tanto son más fáciles de secuenciar. Esta particular herramienta es denominada “Shot Gun” o escopeta (**Figura. 8**) (Bodenteich et al., 1994). En el “Shot Gun”, el ADN genómico se corta aleatoriamente en fragmentos que se separan en distintos tamaños, mismos que se insertan en distintos tipos de vectores, cada uno con una capacidad media diferente (por ejemplo los BACs portan insertos entre 100 y 350 kb y los fósmidos de unas 30 kb). En este trabajo, 11 fósmidos que traslapan el locus Fu-HC fueron secuenciados por la técnica del Shot Gun, ya que se fragmentaron aleatoriamente para crear templados de ADN para ser finalmente secuenciados por el método de Sanger.

La tercera herramienta requerida para la secuenciación genómica es un método automatizado para analizar rápidamente las secuencias en escalera obtenidas mediante el

método de terminación de cadena de Sanger. En el método original empleado por Sanger, un marcador de fósforo radiactivo solía ser incorporado enzimáticamente dentro de los fragmentos de secuenciación, los cuales eran separados a su vez por electroforesis en gel y detectados con una auto radiografía la cual finalmente tenía que ser interpretada. Este método era tedioso y sujeto a errores constantes. Sin embargo en 1986, Leroy Hood ideó una estrategia automatizada de secuenciación de ADN (Lloyd M Smith et al., 1996). Esta estrategia esta basada en la innovación de varios pasos en el proceso de secuenciación, así como en el reemplazamiento de los marcadores o etiquetas radiactivas por etiquetas fluorescentes. El reemplazamiento de la auto-radiografía se lleva a cabo mediante el uso de un detector láser el cual detecta directamente la migración de los fragmentos de ADN a través de una matriz de gel. Mientras que el reemplazamiento de la interpretación manual es realizada a través del uso de un análisis computarizado. Todas estas mejoras en el proceso de secuenciación fueron críticas para su automatización. Aunque este método reemplazaba la interpretación manual y la auto-radiografía, la electroforesis seguía siendo lenta e ineficiente cuando se trataba de secuenciar genomas completos o grandes fragmentos genómicos.

Hoy en día la tecnología ha evolucionado, haciendo posible reemplazar la electroforesis convencional en gel por la electroforesis en capilar, utilizada para la secuenciación a gran escala. Este método emplea un sistema de flujo y un sistema de detección fluorescente inducida por láser que permite secuenciar ADN mediante electroforesis capilar (Swerdlow et al., 1990) Un ejemplo en particular de este método, es el proceso realizado por los 96 capilares en paralelo operados por el secuenciador ABI 3700[™] (**Fig9**).

La cuarta herramienta requerida en cualquier análisis genómico y de secuenciación genómica es el hardware y el software. Es decir, computadoras y programas que permitan ensamblar, anotar y analizar enormes cantidades de datos obtenidos a partir de la secuencia final.

Gracias a estas herramientas de secuenciación rápida y a gran escala; la decodificación de ADN y su organización ha ido incrementando exponencialmente durante la última década. La secuenciación por shot gun, la electroforesis capilar, el software de

ensamblaje y predicción, y las computadoras de última generación, son considerados en conjunto como las innovaciones tecnológicas que revolucionaron la era genética y promovieron la evolución a la era genómica.

CAPÍTULO 2 – OBJETIVOS

Objetivo Principal

El objetivo principal de este trabajo de investigación es el secuenciar, ensamblar y anotar la información de ADN contenida en los once fósidos que traslapan parte del locus de Fusión /Histocompatibilidad, Fu/HC.

Objetivos Secundarios

Los objetivos secundarios tendrán como meta hacer útil la información aquí reportada, con el fin de proveer información indispensable para así desarrollar futuros análisis de genómica comparativa, con el propósito de responder las siguientes hipótesis y preguntas.

1. ¿Existe alguna relación evolutiva entre los genes del Fu/HC y los genes del MHC?. ¿Existe algún ancestro en común para ambos? Si esto es cierto, cabe alguna posibilidad de que las funciones principales de estos genes sean conservadas y tengan el objetivo de unir, procesar y/o presentar péptidos en la superficie celular? Si no existe relación alguna entre el MHC y el Fu/HC, ¿De qué familia de genes forma parte este locus de Fusión/histocompatibilidad?
2. ¿Las similitudes entre el MHC y FuHC son homólogas en secuencia o en función?
3. ¿Existen homólogos vertebrados de los miembros de la familia de genes en Fu/HC, si esto es cierto, ellos tienen alguna función de aloreconocimiento?
4. Si el principal objetivo de Fu/HC es el aloreconocimiento mediante el bloqueamiento de la fusión vascular de organismos no afines y la fusión vascular con organismos afines. ¿Qué funciones biológicas son expresadas mediante el aloreconocimiento? ¿Son estas funciones útiles para mantener la individualidad en organismos solitarios i.e vertebrados, si esto sucede, es posible

que la individualidad y la relación entre organismos afines sea mediada por homólogos de Fu/HC ?

5. ¿Es posible que exista alguna interrelación entre los genes descubiertos en este análisis y los genes que forman parte de otros sistemas inmunológicos y de histocompatibilidad en otros organismos primitivos o contemporáneos?
6. ¿Cómo se relacionan funcionalmente los genes aquí reportados con el complejo de Fusión Histocompatibilidad en *B.schlosseri*?
7. ¿Existe alguna evidencia de la función de Fu/HC en el inicio de la inmunidad adaptativa? o Por qué el sistema de reconocimiento de Fu/HC esta presente o no en vertebrados superiores?
8. ¿Existen rasgos únicos en la secuencia genómica de Fu-HC? Existe algún patrón de homologías respecto a la familia de los vertebrados, invertebrados o las plantas?

CAPÍTULO 3 – MATERIALES Y MÉTODOS

3.1 Purificación de ADN y secuenciación por “shot gun” de los fósmidos que traslapan Fu/HC en *Botryllus schlosseri*.

Las muestras de ADN de *B.schlosseri* fueron recibidas como insertos dentro de once fósmidos: subsf-217g15, subsf-533i14, subsf-555i12, subsf-564g21, subsf-568h21, subsf-532h24, subsf-560b6 subsf-371a15, subsf-557e15, subsf- 53719 and subsf-531d19 . En fósmidos epicentro en células *E. coli* cedidas por el Dr. Anthony W. De Tomaso de la Universidad de Stanford (**Figura.10**).

3.1.1 Purificación de ADN a gran escala

La purificación del ADN a partir de los fósmidos, fue realizada por el método de lisis transparente “clear lysed method” seguido por una doble precipitación de acetato:

El primer paso de la purificación del ADN tiene como meta aumentar la población bacteriana con el fin de obtener una mayor cantidad de ADN.

En este paso, se toma una pequeña población de células a partir del cultivo inicial y se pone en medio LB suplementado con clorafenicol [15mg/ml]. Las células se transfieren cada 8- 10horas de un sub cultivo de 3 ml a otro de 200 ml hasta llegar a un crecimiento celular poblacional óptimo, para así lograr una purificación eficiente de ADN. Cabe señalar que el vector fósrido contiene un gene de resistencia a cloramfenicol, el cual confiere características selectivas a las células sobrevivientes mismas que contendrán el fósrido y por consiguiente, el ADN a extraer. Después del cultivo nocturno, las células fueron cosechadas por centrifugación a 5000 rpm por 30 min en una botella de 500 ml. El pellet obtenido se congeló a una temperatura de -70C por 30 minutos.

Las células del pellet obtenido del cultivo de 200 ml fueron descongeladas y

resuspendidas eficientemente en 8 ml de EDTA 10 mM, pH 8.0.

Esta resuspensión se realizó haciendo uso de una pipeta de 10ml, pipeteando con cuidadosamente hacia arriba y hacia abajo. En este paso las células fueron resuspendidas completamente. Después de mezclar cuidadosamente, la solución fue incubada a temperatura ambiente por un periodo de 5 minutos.

Para separar el ADN de membranas y otras sustancias celulares, se adicionó a las células previamente resuspendidas una mezcla de 16 ml de solución de lisis alcalina (0.2 N NaOH and 1% sodium dodecyl sulphate, SDS). Después, se revolvió lentamente dicha solución junto con las células resuspendidas, hasta volverse homogénea. Posteriormente, se incubó durante 5 minutos en hielo. Todo con el fin de separara el ADN de las membranas celulares. El paso completo debe ser terminado en un periodo no mayor a 10 min.

Nota: NO agite ni incube la mezcla más de 10 min en todo el proceso. Si no se hace caso a esta indicación, la contaminación de ADN genómico de *E. coli* aumentará contaminando el ADN fosmídico, disminuyendo radicalmente la eficiencia del proceso.

Primera precipitación por acetato.

Inmediatamente, 12 ml de 3 M KOAc frío fueron adicionados y mezclados cuidadosamente al mover delicadamente varias veces la botella que contiene la mezcla. Una vez terminado este paso, la solución fue incubada en hielo durante 45 min o almacenada en el ultra congelador a una temperatura de -20 o -80 C toda la noche. El congelar la mezcla en esta etapa, da como resultado un pellet mucho mas compacto y por consiguiente una preparación de ADN final mucho más limpia.

Nota. No agite la solución. Si agita la solución en vez de moverla cuidadosamente, el ADN se fragmentará y la contaminación por *E.coli* aumentará. Promoviendo así, un incremento en la contaminación del vector pUC pues la fragmentación del ADN provocará un aumento en la cantidad insertos cortos . Si no se sigue esta recomendación esto provocará una ineficiencia neta de todo el proceso.

El lisado celular fue limpiado de el complejo precipitado de SDS, proteínas, membranas y ADN cromosomal mediante la una centrifugación a 10,000 rpm por 15 minutos a 4 grados C en la ultra centrifuga RC5-B usando el rotor GSA. Antes de recentrifugar la muestra, el sobre nadante claro, fue filtrado a través de una doble capa de gasa. Esto con el fin de remover cualquier material flotante. Una vez hecho lo anterior, una centrifugación adicional es llevada acabo, con el fin de asegurar que todos los fragmentos insolubles fueron removidos.

El sobre nadante transparente, fue transferido a una botella de 250 ml, esto se hace proporcionalmente en función del cultivo original de 200 ml. Entonces, se adiciona un volumen igualmente proporcional de isopropanol al volumen del sobre nadante. Esto con el fin de precipitar el ADN. Una vez terminado este paso se incuba la mezcla por 5 min y se centrifuga la muestra a 5,000 rpm por 15 minutos haciendo uso de la centrífuga RC5-B usando el rotor GSA nuevamente. Una vez que se precipita la muestra, el sobre nadante es decantado y eliminado, y el pellet conservado.

Una segunda precipitación por acetato fue realizada al disolver rápida pero cuidadosamente el pellet de ADN con una mezcla de 6 ml de 10:50 TE (10 mM Tris-HCl, pH 7.6, 50 mM EDTA, pH 8.0). Una vez que el pellet fue disuelto entonces se adicionaron 1.8 ml de 7.5 M KOAc dentro de cada botella, mismos que fueron transferidos a un tubo Sorval de 50 ml . Nota: La solución de 7.5 M KOAc fue usada sin ajuste de pH).

Después de mezclar, la solución fue congelada hasta solidificarse a -70 C. Esto toma alrededor de 30 min. Pero típicamente se congela toda la noche.

Para purificar el ADN del ARN, se efectuó el siguiente tratamiento con la enzima ribonucleasa. Para reducir la cantidad de RNA presente en la preparación final del ADN perteneciente al fósido. Después de descongelar la solución fue centrifugada a 10K rpm por 10 minutos usando el rotor SS34.

El sobre nadante de cada tubo fue transferido a un tubo Corning de 50 ml(libre de DNAsas). Después de que el sobre nadante fue transferido la RNAsa se adicionó hasta llegar a una concentración final de 100 ug/ml, además también se hizo lo mismo con RNAsaT1 a una concentración final de 40 ul/100 ml, entonces la solución de RNAsa fue

incubada a baño maría de 37 C por una hora.

El ADN fue precipitado al adicionar en cada tubo etanol al 95%. Después de mezclar invirtiendo los tubos estos fueron encubados en un baño de hielo por 15 minutos. El ADN fue entonces compactado en un pellet gracias a la centrifugación que se efectuó a 3000 rpm por 30 min en una centrifuga Beckman GS-R. Entonces cada pellet fue lavado de las sales remanentes con 30ml de etanol al 70%. Mismo que fue secado en una bomba de vacío por 30 min. Una vez secado cada pellet fue disuelto en 210 ul de ddH₂O pipeteando arriba y abajo, se almaceno a 4 C toda la noche con el fin de asegurar una completa disolución de las moléculas de ADN de alto peso molecular. Finalmente, se estimó la concentración de ADN al analizar 10 ul mediante una electroforesis en gel de agarosa. Una preparación de ADN por este método produce aproximadamente 150 ug de ADN por 200ml del cultivo original.

3.1.1.1 Electroforesis en Gel de Agarosa

Un gel de agarosa al 1.5% fue preparado para examinar la calidad y cantidad del ADN. Este gel fue preparado para verificar la calidad y la cantidad obtenida de la preparación de ADN. Para esto se añadieron 1.5 g de agarosa estándar a 1X TAE obteniendo un volumen final de 150 ml dentro de un frasco de 500ml. La solución al 1.5% la solución de agarosa fue calentada en un microondas de 2 a 3 minutos mientras la agarosa se disolvía completamente. Después, 25 ul de 5 mg/ml de bromuro de etidio (EtBr) fueron agregados. Una vez que se mezcló el gel, este se vació en un molde de vidrio donde se dejó de 20 a 30 minutos para solidificarse.

Nota: Debido a las propiedades teratogénicas del EtBr es importante utilizar guantes para evitar el contacto. Una vez que 1.5% de gel de agarosa fue preparado, se cargó cada pozo del gel con 10 ul de ADN y 1-2 ul de colorante de ADN. También, se adicionaron 3 ul de una mezcla de marcadores de peso molecular (Lambda Hind III y HAE). Después, la muestra del ADN fue corrida en la cámara de electroforésis a 20 mA por 30 min. Finalmente, se tomó una fotografía del gel bajo luz UV (**Figura.10**).

3.1.2 Construcción de librerías genómicas y secuenciación por shotgun.

La secuenciación por Shotgun requiere de fragmentar fósmidos en insertos típicamente de entre 2 y 4 kb, y de reparar sus extremos incompletos “blunt ends” para hacerlos accesibles a ser insertados en vectores. Los insertos fueron fosforilados en el extremo 5’ antes de ser ligados al vector. El vector recombinante usado en este proyecto fue pUC18 mismo que fue tratado con la enzima de restricción SmaI con el fin de generar extremos incompletos y de-fosforilados para impedir la recircularización espontánea del vector (Amersham 27-4860-00).

3.1.3 Shotgun

El ADN purificado es literalmente fragmentado por una escopeta... Dos métodos alternativos fueron usados con el objetivo de fragmentar físicamente el ADN **(Figura. 11)**

3.1.3.1 Nebulización

Es un método de fragmentación aleatoria que consiste en aplicar gas nitrógeno en un Nebulizador IPI (Medical Products, Inc., Chicago) a una mezcla de ADN glicerol, y dd H₂O. A una presión de 8.5 p.s.i. por 2.5 minutos. Para obtener como resultado fragmentos de un tamaño promedio de 2 Kb – 4 Kb **(Figura. 11b)**.

La Nebulización fue realizada al disolver entre 100-200 ug de ADN en 2.5 ml de glicerol estéril en agua des-ionizada (*v/v*) al 50% en un nebulizador IPI puesto dentro de un baño de hielo seco y etanol a -15 ° C. El nitrógeno fue introducido por la parte superior de el nebulizador a una presión apropiada por dos y medio minutos. El flujo de gas causa que la solución pase a la cámara superior del tubo lo cual ocasiona su dispersión para después salpicar sobre una pequeña profusión plástica para después generar Pequeñas gotas de ADN fragmentado. La solución de ADN una vez fragmentado, fue dividida en cuatro tubos de 1.5ml (0.5ml de solución /tubo) y los fragmentos de ADN fueron precipitados al adicionar dos volúmenes de una solución de acetato y etanol (95%ethanol/0.12M acetato de sodio). Después de que los fragmentos de ADN fueron precipitados y recolectados por centrifugación, el pellet fue lavado de los residuos de sal

usando etanol al 70%. Después el pellet fue secado y finalmente resuspendido en 30 ul de ddH₂O.

Nota: El tamaño de las gotas y por igual, el tamaño de los fragmentos de ADN son inversamente proporcionales a la presión aplicada. Entonces, un incremento en el tiempo y presión durante la nebulización, ocasiona fragmentos de ADN cada vez más pequeños.

3.1.3.2 Hidro fragmentación “Hydroshear”

Este proceso automatizado de fragmentación aleatoria de ADN es realizado por un aparato llamado Hydroshear, patentado por Gene MachinesTM. Esta máquina fragmenta el ADN dependiendo de la velocidad con que este pasa por un filtro (**Figura. 11a**). La fragmentación del ADN es efectuada al adicionar 100 ul de una solución de 200 ug en la máquina Hydroshear. De esta manera el ADN pasa a través de un filtro de fragmentación por alrededor de 20 veces a una velocidad de nivel 11.

Ambos métodos de fragmentación; Nebulización e Hydroshear haciendo uso de las especificaciones descritas previamente cortan fragmentos de ADN de entre 2-4 kb en promedio.

3.1.3.2 Reparación de extremos.

Los fragmentos generados de la nebulización y/o de la hidro-fragmentación, fueron reparados “end-repaired” y fosforilados al tratarlos con 2ul del fragmento de Klenow de ADN polimerasa y 1ul de un polynucleotido T4 kinasa despues se adicionaron 5ul de un buffer de Kínasa a una concentración 10X (500mM Tris-HCl, pH7.6, 100mM Mg Cl₂, 10mM DTT, y 50ug/ml BSA), 5ul de 10mM rATP, y 7ul of 0.25 mM dNTPs. La reacción fue incubada por 30 minutos a 37°C. Una vez transcurrido este periodo de tiempo, la reacción fue parada al adicionar 10ul de buffer de cargada (0.02% bromophenol blue, 5mM EDTA, pH8.0, 50% glycerol).

3.1.3.4 Selección de insertos por tamaño

Una vez reparados los extremos de los fragmentos de ADN, fueron cargados en un gel de agarosa de bajo punto de fusión al 1% en paralelo con una mezcla de marcadores de peso molecular (*Hind*III-treated λ -ADN mezclado con *Hae*I-treated ϕ X174-ADN de Gibco BRL) (**Figura. 11**). Después de la electroforesis, se cortaron las bandas de entre 2-4kb con navajas estériles. Así, estos fragmentos fueron cortados y puestos en tubos de 1.5 ml para micro centrifugar, después se congelaron a -70°C por al menos 30 min.

Después el ADN se extrajo de los fragmentos en gel por medio de una centrifugación haciendo uso de una micro centrífuga a 13,000 rpm por 5 minutos. Esto con el fin de separar la agarosa del sobre nadante conteniendo los fragmentos de ADN. Los fragmentos del sobre nadante fueron transferidos a un tubo para micro centrífuga de 1.5ml. Esta técnica llamada de congelar y exprimir “freeze and squeeze” fue repetida al menos dos veces mas o hasta que hubiese sido separado el sobre nadante por completo. Una vez separado el sobre nadante y desechada la agarosa , se agregó Etanol y acetato de sodio al (95% ethanol/0.12M Acetato de sodio) al sobre nadante con el fin de precipitar los fragmentos de ADN, a este paso lo siguió una centrifugación en el cuarto frio por 30min a 13,000rpm con el fin de obtener un pellet de ADN. Una vez que se obtuvo un pellet de ADN, este fue lavado dos veces usando 1ml de etanol al 70%, después este fue secado y disuelto en 15ul de agua estéril, destilada y des-ionizada. Posteriormente la solución fue almacenada a -20°C .

3.1.4 Subclonación de insertos y transformación

Después se usan 5ul de la solución de ADN fragmentado y reparado, con el fin de observar su calidad y concentración final. Para tener este conocimiento se corre la muestra en un gel de agarosa al 1% agarose gel. Posteriormente se realizan una serie de reacciones de ligación de acuerdo a la siguiente tabla:

| | 1x | 2x | 4x | 6x |
|----------------------------|-----|-----|-----|-----|
| Fragmentos de ADN | 1ul | 2ul | 4ul | 6ul |
| pUC18 (10ng/ul) | 2ul | 2ul | 2ul | 2ul |
| Ligation buffer | 1ul | 1ul | 1ul | 1ul |
| ddH ₂ O | 5ul | 4ul | 2ul | 0ul |
| T4-ADN Ligasa (400U/ul) | 1ul | 1ul | 1ul | 1ul |

Tabla2. Series de reacciones de ligación.

Las reacciones de ligación fueron incubadas a 4°C por un lapso de 24-48 horas.

Después de la reacción de ligación, se agregaron 2 ul de cada reacción de ligación en 40 ul de células electro competentes de *E.coli* XL1-Blue MRF las cuales se usaron para transformarlas usando un electroporador (Bio-Rad *E.coli* Pulser). La electroporación fue realizada al poner la mezcla de ligación y las células competentes en un recipiente para electroporar (cuvette). Antes de someter las células y el ADN a un pulso eléctrico de 2.5KV por 5 microsegundos a 4°C se debe cerciorar que no haya burbujas en el mismo,. Después de tal pulso eléctrico, las células son inmediatamente diluidas en 1ml de medio YENB frío (7.5 gramos de Bacto Yeast Extract, y 8 gramos de Bacto Nutrient Broth llevado a 1 litro disuelto en agua destilada y esterilizada). Una vez puesto el YENB, la muestra se pone las células se ponen a recuperar en una incubadora a 37°C por 30 minutos, siendo agitada a 250 rpm. Las células transformadas entonces se colectan por centrifugación a 2500 rpm por 5 minutos y se resuspenden en 200ul de medio YENB fresco suplementado con 30 ul de (25mg/ml 5-bromo, 4-chloro, 3-indolyl, Beta D galactoside) (Xgal;Sigma B-4252) y 30ul de 25mg/ml de isopropyl thiogalactoside (IPTG;Sigma I-5502). El IPTG fue incluido como un inductor del operón de lac , un análogo de galactosa el cual no puede ser fragmentado por la enzima Beta-galactosidasa dando como resultado la formación de un color azul intenso cuando el inserto no se une al vector recombinante y por tanto no se obstruye el gene marcador. Posteriormente, los 130 ul de las células resuspendidas fueron plaqueadas en cajas de petri de plástico de 10 cm de diámetro conteniendo 20 ml de LB agar suplementado con 100ug/ml ampicilina. Así, las células fueron incubadas a 37°C por 18 horas. Las colonias blancas se

cosecharon usando el robot cosechador de colonias Flexis (Flexys colony picker™), dentro de los 1.5 ml de TB broth (12 grams of Bacto-tryptone, y 24 gramos de extracto de levadura fueron llevados aun volumen de 900 ml con agua destilada, esterilizado y enfriado. Entonces, 100 ml de 10X TB sales (2.31g KH₂HPO₄ estéril y frío) fueron adicionados hasta producir 1 litro de TB broth adicionado con sales. Entonces se adicionó ampicilina en una concentración final de 100 ug/ml a la solución TB broth con sales, así 200ul de este medio de cultivo a cada uno de los 384 pozos de la placa “384-well flat bottom microtiter plate” (NUNC #242757). Además unas pocas docenas de colonias fueron cosechadas en una placa extra de 96 pozos que contiene 200 ul de medio de cultivo, esto con el fin de reemplazar el contenido de los pozos en los que no hubo crecimiento bacteriano. Después de la inoculación de cada pozo, con colonias individuales, las placas fueron incubadas en la incubadora HiGro “HiGro incubator (Gene Machines, Inc.)”, por un periodo de 22 horas a 37 grados C con agitación a 520 rpm. El flujo de oxígeno fue programado para comenzar 3.5 horas de haber iniciado la agitación, accionando así un mecanismo que permitiera un flujo de oxígeno de 0.5 segundos cada 0.5 minutos. Después de la incubación, las células fueron recolectadas por centrifugación a 2500 rpm durante 10 minutos y congeladas a -80°C.

3.1.5.1 Aislamiento semi-automatizado y automatizado de templados de ADN para secuenciar

Dos métodos fueron desarrollados para purificar templados de ADN i.e. sub clonas, durante el curso de esta investigación. El primero es semi-automatizado y utiliza el robot para manejar líquidos “hydra” así como un agitador manual.

El segundo es automatizado y utiliza el robot Zymark™.

3.1.5.2 Aislamiento semi-automatizado de subclona de ADN

Después de que las células fueron cosechadas, recolectadas y congeladas, en placas de 384 pozos, se adicionaron 60ul de una solución de TE-RNase (50mM Tris-HCl, pH7.6, 0.5 M EDTA-Na, 40ug/ml RNase A and 0.04U/ul RNase T1) usando el Hydra™ (Robbins Scientific, Inc.). Los pellets de las células fueron disueltos y las placas fueron colocadas en un agitador de mesa a una velocidad de 7. Después de 10 minutos de

agitación fueron adicionados 60ul de buffer de lysis buffer (1%SDS, 0.2M NaOH) y las placas fueron incubadas en el agitador a una velocidad de 4, por 30 minutos. Entonces, se adicionaron 60ul de (3M KOAc en un volumen total de 1 L ajustado a un pH de 4.5 con ácido acético) y las placas fueron incubadas a 37°C agitándose a 350rpm por 20 minutos, finalmente se congelaron a -70°C durante toda la noche. Después de haber sido congelados toda la noche, las placas fueron centrifugadas por 30 minutos a 3000rpm. Posteriormente, se transfirieron 60ul del sobre nadante resultante en cada pozo de una nueva placa de 384 pozos. Además, se adicionaron 130ul de etanol al 95% usando el Hydra, esto con el fin de precipitar el ADN. Las placas fueron centrifugadas a 3200rpm por 30 minutos, el sobre nadante fue decantado, y los pellets lavados con 140ul de etanol al 70%. Después de una nueva centrifugación a 3200rpm por 10 minutos, y de haber decantado el exceso de etanol, el ADN fue secado en una cámara de vacío y disuelto en 50ul de dd-H₂O, mismo que fue evaluado por electroforesis en un gel de agarosa el 1%.

3.1.5.3 Aislamiento automatizado de subclonas usando el robot Zimark

La primera etapa del protocolo automatizado de aislamiento de ADN (pasos 14 a 21) fueron hechos usando el mini-Stacatto. Esta maquina contiene dos unidades: el robot Twister II que contiene un brazo robótico el cual mueve las placas de 384 pozos de las columnas de almacenamiento montadas sobre el Twister II. La segunda unidad es el Sciclone la cual es una estación capaz de manejar líquidos haciendo uso de un pipetor de 384 canales y de cuatro agitadores magnéticos.

El aislamiento automatizado de subclonas de ADN fue realizado usando el robot Zimark, mismo que fue programado con las órdenes contenidas en el sitio web (http://www.genome.ou.edu/Zymark_384_well_isolation.html)

3.2 Condiciones de los ciclos de secuenciación

La secuenciación fue realizada usando el kit terminador “BigDye” (PE Applied Biosystems cat.#430151), el cual contiene un buffer de Taq polymerasa (composición

desconocida), AmpliTaq FS ADN polymerasa, los cuatros dNTPs, y los cuatro terminadores dideoxinucleotidos marcados con cuatro fluoróforos. La enzima Ampli Taq contiene dos modificaciones, una deleción en el N-terminal, la cual elimina la actividad de edición 5'-3' de la enzima, lo cual permite una mayor procesividad, y una mutación de fenilalanina a tirosina en el sitio activo, el cual incrementa la afinidad por los nucleótidos terminadores marcados con fluorescencia. Cada terminador esta unido a marcadores fluorescentes, una fluoresceína derivativa (donor dye) conectada a una dRhodamina-derivativa (aceptor dye). La fluoresceína derivativa puede ser excitada por un láser iónico de argón contenido en el secuenciador de ADN de PE Applied Biosystems. La energía emitida del primer fluoróforo en cambio excita los electrones de el aceptor de dRhodamine el cual emite luz que es identificada por un detector PMT en el abi-377 o una cámara ccd en el abi-3700, cuando regresan a su estado original. Los fluoróforos aceptores derivativos de dRhodamina, dR6G,dROX, dTA<ARA, fueron unidos a dR110 por ddATP,ddCTP,ddTTP, y ddTTP respectivamente.

Aproximadamente de 150-200 ng de templado de ADN. 1ul de 13uM de iniciador universal anverso “ primer universal forward” (GACGTTGTAAAACGACGGCC) o iniciador universal reverso “primer universal reverse” (CACAGGAAACAGCTATGACC), dimetilsulfoxido (DMSO) a una concentración final de 5% (V/V) y 2ul del kit de reacción premix Big Dye™ que contiene AmpliTaq FS, pirofosfatasa termoestable, dATP, dCTP, dTTP (100uM cada uno), dITP (500uM), ddATP, ddCTP, ddTTP, y ddGTP (~0.11uM cada uno) a una concentración final de 1/2th, 1/16th , o 1/20th de lo recomendado por el proveedor fueron adicionados a cada pozo de la placa para termocilar de 384 pozos (Robbins Scientific, Inc, cat #1047-00-0). Las reacciones entonces fueron termocicladas por 60 cycles de desnaturalización a 95 °C por 30 segundos, alineamiento “annealing” a 50°C por 20 segundos y extensión a 60°C por 4 minutos.

3.2.1 Eliminación de terminadores no incorporados

Los terminadores no incorporados y las sales de buffer fueron removidos de las reacciones de secuenciación mediante precipitación por etanol. En este caso, se adicionan 10ul de dd-H2O a cada pozo en la placa para termociclar llevándolo a un

volumen final de 20ul, además, fueron adicionados dos volúmenes (40ul) de etanol-acetato a 95% a cada pozo. Entonces, la placa de termociclar fue centrifugada a 3200rpm por 30 minutos. Después, el acetato de etanol fue decantado, los productos precipitados de la reacción fueron lavados con 100ul de etanol al 70% y centrifugados por 10 minutos a 3200rpm. Después de decantar el etanol, las placas fueron secadas a temperatura ambiente y almacenadas a menos -20°C hasta estar listas para ser cargadas en el secuenciador.

3.2.2 Secuenciación

Los productos de secuenciación por shotgun fueron eletroforeticamente analizados ya sea en el secuenciador por gel ABI PRISM 377 o en el secuenciador de capilares ABI PRISM 3700 (**Figura.9**). Para el ABI PRISM 377, se prepare acrilamida al 5.3% (FMC) conteniendo 8 M de urea. La mezcla fue vertida en vidrios de cristal de 48 cm de baja fluorescencia separados con espaciadores de 0.2mm. Después de la polimerización, cada muestra fue disuelta en 1 ul de buffer de cargada (0.02% azul dextrano, 0.2 mM EDTA en formamida desionizada) y cargado en cada pozo. Entonces, la electroforesis fue realizada a 2.5 kV y 52 grados C por 10 horas en buffer 1X TBE (10.8g Tris, 5.5 g ácido bórico, y 8.4g EDTA en un volumen total de 1 litro). Para el ABI PRISM 3700 los capilares fuerin automáticamente rellenos con una mezcla de gel fluido POP-5 (ABI, Inc., Performance Optimized Polymer type 5 cat.# 4313087) y las reacciones fueron electro-cinéticamente inyectadas después de haber sido disueltas en 20ul dd-H₂O. La electroforesis fue realizada a en 6.5 Kv por 2.5 horas. Debido al pequeño diámetro interior de los capilares (50um), el gradiente de temperatura dentro del tubo es minimizado y la disipación del calor es rápida, lo que permite la aplicación de un campo eléctrico elevado. Los datos fueron recolectados automáticamente y analizados usando el llamador de bases ABI, “ABI base caller” en una computadora Macintosh (para el ABI377) o una computadora Dell-PC (para el ABI3700). Los datos entonces fueron transferidos a una estación de trabajo SUN basada en el sistema Unix.

3.2.3 Ensamblaje de secuencias

El ensamblaje de datos y el alineamientos de contigs fueron realizados haciendo uso del

software llamador de bases Phred/Phrap y del software de ensamblaje Consed desarrollado en la Universidad de Washington por Phil Green (Ewing B et al. 1998). Los datos recolectados de los secuenciadores fueron transferidos vía el protocolo de transferencia de archivos (ftp) hacia el directorio de proyectos en la estación SUN. Ahí, los cromatogramas generados por los secuenciadores por cada reacción de los secuenciadores, mejor llamados archivos traza “trace files”, fueron analizados por el software Phred (Ewing B et al. 1998). Phred escribe el formato de la secuencia en forma de bases y asigna valores de calidad a cada base a traza para generar archivos de salida “output files” en el formato PHD que a su vez es transferido al software Phrap para su ensamblaje (Gordon et al. 1998).

Phred analiza los datos para en principio determinar donde estarían centrados los picos de fluorescencia y calidad. Si no hubiese compresiones u otros factores alterando los picos de sus locaciones calculadas teóricamente (**Figura. 13**). Entonces, este examina cada traza hasta encontrar el centro real o en su defecto que los picos observados tengan relación con los picos detectados en el segundo paso con la localización de los picos predichos encontrados en el primer paso (Ewing et al., 1998 7,8). Finalmente, Phred evalúa las trazas que rodean cada base llamada, usando parámetros de valores de calidad para cuantificar la calidad de la traza (Ewing et al. 1998). El valor de calidad (QV) es relacionado a la probabilidad del error de la base llamada (P_e) por la fórmula:

$$QV = -10 * \log_{10} (P_e).$$

Phred usa los datos de un archivo de parámetros químicos llamado phredpar.dat con el objetivo de identificar los datos del iniciador “primer”. Esta información es escrita entonces dentro de archivos PHD individuales llamados de acuerdo a cada archivo traza original y pasados a su vez al programa de ensamblaje Phrap (Ewing et al. 1998). Así, a partir de los archivos PHD, son creados los archivos Fast A, los cuales contienen el formato de la secuencia expresada en bases. Estos son usados por el programa cross_match para filtrar y separar la secuencia del vector. Este programa compara la secuencia obtenida de Phred con las secuencias en “seqs_fasta”. Phrap (Programa de ensamblaje de fragmentos), utiliza los datos generados por Phred para ensamblar las lecturas y generar secuencias consenso contiguas integrando a su vez puntajes de calidad

corroborados por Phred para finalmente construir secuencias consenso o contigs (**Figura. 14**). Esto a su vez puede ser visto y editado por Consed, un programa para visualizar Phrap (**Figura. 15**). (Gordon et al. 1998).

Consed, es un programa utilizado para ver y editar de manera interactiva los ensamblajes generados por Phrap y los archivos resultantes del ensamblaje por Phrap con el fin de crear una representación interactiva de los datos (**Figura. 15**). La interfase gráfica de Consed consiste de la secuencia consenso de cada contig mostrando las lecturas individuales que las componen. Las lecturas individuales son ligadas tanto a los archivos traza del cromatograma de fluorescencia generados del secuenciador de ADN, como a los archivos de calidad generados por Phred (Gordon et al. 1998). Los últimos, son representados con esquemas de colores diferentes; las representaciones de alta calidad son vistas con un fondo blanco y letras mayúsculas. De manera tal que la calidad disminuye conforme el fondo se oscurezca y las letras se hagan minúsculas. Adicionalmente, Consed presenta los datos de las secuencias repetidas, representadas por diferentes tonos de verde, los cuales pueden ser ligados directamente a repetidas presentes en diferentes contigs (Gordon et al. 1998). Las discrepancias de alta calidad pueden ser representadas con un color rojo, el cual puede indicar clonas quiméricas, las cuales pueden ser ligadas a diferentes contigs.

3.3 Rellenamiento de huecos “Gap Closure”

3.3.1 Caminata de iniciadores “Primer- walking” – Clonas de insertos grandes

El principio detrás de la caminata de iniciadores “primer walking” es usar secuencias de clonas con insertos grandes generadas usando iniciadores universales anversos y reversos con el fin de delimitar clonas con un potencial de ser traslapadas y unidas de manera tal que sea posible cubrir cualquier hueco entre los contigs que se encuentren en la secuencia ensamblada. Las secuencias repetidas están comúnmente presentes en las regiones flaqueadas por los huecos. El programa de ensamblaje Phred ensambla de manera incorrecta estas regiones ya que traslapa las secuencias repetidas dentro de las secuencias contiguas acortando la secuencia original o ensamblando incorrectamente las regiones que flanquean diferentes copias de secuencias repetidas. Este programa traslapa clonas,

lo cual permite crear puentes de secuencia para que finalmente los huecos sean rellenados. Esto se realiza gracias a una síntesis de iniciadores selectiva la cual tiene el propósito de hacer que los iniciadores caminen al momento de rellenar los huecos. Este método asegura que no ocurran ensamblajes incorrectos, pues los huecos son cerrados haciendo uso de la secuencia de cada clona y sus sucesivas amplificaciones. Los iniciadores de caminata “Walking primers” fueron elegidos utilizando el programa PrimOU, el cual es una adaptación mejorada del programa Primo desarrollado en la Universidad de Texas-Southwest Medical Center. PrimOU identifica secuencias iniciadoras únicas las cuales han sido filtradas a partir de las secuencias contiguas de las cuales se originadas o de todas las secuencias contiguas recolectadas. Dicho programa es automáticamente activado cuando Phrap termina de ensamblar los datos. El output de las secuencias iniciadoras recolectadas es almacenado como un archivo de texto, el cual contiene el número del contig, la dirección y la secuencia del iniciador. Adicionalmente, PrimOU utiliza el control de calidad que permite asegurar con exactitud la secuencia de ADN. Una vez realizado esto, los archivos output generados por PrimOU fueron mandados electrónicamente al sintetizador de oligonucleotidos MerMade (Avantec, Inc). Todos los iniciadores de caminata “walking primers” fueron sintetizados en la versión comercial de un instrumento originalmente diseñado en la Universidad de Texas Southwestern Medical Centre. El Mermade tiene una capacidad de sintetizar 192 iniciadores en 2 placas con filtros de 96 pozos en alrededor de 16 horas. Este instrumento es un catalizador que permite el uso extensivo de un sistema de síntesis de iniciadores a gran escala aplicado al rellenamiento de secuencias de ADN, lo cual permite un bajo costo (~\$1/primer), velocidad, producción y confiabilidad. Típicamente, la síntesis de cada iniciador es producida en una concentración de 2 nmole por cada pozo de una placa de 96 pozos.

3.3.2 Reacción múltiple en cadena de la polimerasa (MPCR)

Aunque la secuenciación, el traslapamiento de clonas y el rellenamiento de huecos usando la estrategia de caminata de iniciadores “walking primers” fue bastante exitosa, algunos huecos quedaron sin traslapamiento con inserto alguno de subclona o secuencia. Esto indicó que el origen de dicho huecos fue debido a una gran cantidad de secuencias

repetidas en estas clonas, las cuales quizás dieron origen a reacomodamientos que cambiaron la secuencia original. En este caso, los huecos físicos fueron rellenados haciendo uso de una estrategia basada en el PCR aunque la técnica de PCR uniplex empleada con un solo par de iniciadores es muy útil, debido a que el orden de los contigs es conocido. Cuando el orden de los contigs es incierto es de mayor utilidad el usar la estrategia múltiple de PCR (MPCR). Este método fue desarrollado específicamente cuando el orden de los contigs era incierto. Los iniciadores fueron diseñados a partir de los extremos finales de todo los contigs, los cuales fueron sintetizados y combinados para el MPCR, usando ADN genómico como molde. Después de secuenciar los productos de PCR generados, los iniciadores que fueron producidos a partir de la secuencia, fueron sustraídos sucesivamente para otra ronda de MPCR.

3.3.3 Secuenciación de regiones difíciles

Cuando la estrategia de caminata de iniciadores “primer walking” o MPCR no funcionó como resultado de secuencias altamente repetitivas y difíciles de secuenciar o de segmentos genómicos ricos en GC, se consideró el uso de Seven Deaza. Seven Deaza es una mezcla de nucleótidos alternativos los cuales contiene menos puentes de hidrógeno que los nucleótidos comunes. Esta característica disminuye gran parte de la formación de estructuras secundarias e incrementa la eficiencia de secuenciación.

La siguiente mezcla de reacción fue utilizada cuando este tipo de regiones difíciles fueron un problema.

- 1) ADN objetivo (SAP-ExoI cleaned PCR 7-deaza-dG containing product) 4ul
Primer (from mermaid) (200 pmoles of primer) 1ul
- 2) Polyoxyethylenesorbitan monolaurate (1% v/v with sterile double distilled water)* 1ul
Igepal CA-630 (1% v/v with sterile double distilled water)* 1ul
- 3) BigDye V3.0 (un-diluted) 1ul, Final volume 9ul.

Usando el termociclador GeneAmp PCR System 9700, las condiciones de las reacciones

de secuenciación fueron realizadas de la siguiente manera:

95 grados C por 5 minutos,

99 o 60 ciclos utilizando el siguiente programa de secuenciación:

- 1) 95 grados C por 30 segundos,
- 2) 50 grados C por 20 segundos,
- 3) 60 grados C por 4 minutos.

3.4 Métodos computacionales y análisis de datos

3.4.1 Análisis de secuencia

Cuando la cobertura de la secuencia cada fósido llegó a 6 X, el proceso de anotación fue comenzado. Existen tres razones principales para iniciar la anotación antes de finalizar al 100% el proceso de secuenciación. Primero, al llegar a una cobertura de 6X Alrededor del 95% de los datos útiles de la secuencia han sido obtenidos, permitiendo una aproximación bastante precisa de la observación del contenido génico. La segunda razón se sustenta en que los datos obtenidos de la anotación pueden ser útiles para cerrar los huecos pues ayudan a indicar en que contigs se encuentran porciones de cierto gene. Finalmente, la cuarta razón es atribuida a que es posible realizar un análisis comparativo. Alineando los contigs con una secuencia conocida de un genoma relacionado, lo que permite ayudar a cerrar o rellenar los huecos gracias a que el orden genico puede ser conservado.

El protocolo básico para anotación preliminar de los once fosmidos que traslapan el Fu/HC estan ilustrados como diagrama de flujo (**Figura. 16**). El análisis de los datos inició al crear un archivo que contiene todos los contigs de una extensión mayor a 2kb. Esto fue realizado al unir los extremos finales de cada contig de manera aleatoria mediante 20 N s representando cualquier tipo de base de acuerdo a los códigos de la IUB. Los archivos generados fueron sometidos a dos programas buscadores de genes; FGENESH (Solovyev, 2001,2002 38,39) and GENSCAN (Burge 1998). Ambos programas fueron usados en vista del hecho de que la predicción de genes suele ser variable y complementaria usando ambos. Así, GENSCAN mostró una mayor sensibilidad al ser capaz de predecir genes no predichos por FGENESH. Sin embargo, muchas de las predicciones hechas por FGENESH fueron precisas en contraste a las

imprecisiones que GENSCAN solía hacer, pues en ocasiones sobre predecía dando origen a una estimaciones erróneas de genes pequeños o incompletos.

Los archivos generados por GENSCAN y FGENESH fueron visualizados en Artemis (Rutherford et al., 2000), o una interfase gráfica de secuencias de ADN y una herramienta de anotación, la cual permitió exponer las secuencias e identificar los marcos abiertos de lectura (ORFs) basados en la distancia entre los codones de paro consecutivos sobre toda la extensión de una secuencia. Además, es posible modificar los parámetros de visualización generando un zoom positivo o negativo de 100 bases. Los ORFs entonces fueron analizados por una serie de operaciones controladas a través de instrucciones o “Scripts” efectuadas por Perl y escritas por Fares Najjar y James White. El primer script se encargó de realizar una visualización de los datos obtenidos por FGENESH y GENSCAN al mismo tiempo. **(Figura. 18).**

El segundo script realizó búsquedas haciendo uso de BLASTX y BLASTP, es decir, de secuencias potencialmente protéicas y/o expresadas en relación a cada ORF (Altschul et al., 1997). Gracias a este script, se filtraron sólo los más altos valores E, las más altas homologías de información de cada archivo resultado (output file) obtenido por Blast. Una vez teniendo estos resultados, los archivos se convirtieron del formato Fasta al formato de EMBL para hacerlos compatibles con el tipo de archivos que utiliza Artemis. **(Figura. 17).**

To further analyze the BLASTP homology results, a separate search was performed to look for conserved domains at NCBI by using Rps-Blast (Figura. 19).

Uno de los objetivos finales de este proyecto de investigación fue el buscar una molécula de tipo MHC o un posible receptor membranal envuelto en las primeras etapas del reconocimiento propio/no propio. Con este objetivo en mente, se realizó una búsqueda de proteínas transmembranales, haciendo uso del programa bioinformática SOSUI (Mitaku et al. 1998). SOSUI es una herramienta predictiva de proteínas transmembranales la cual se basa en la secuencia de amino ácidos para predecir estructuras secundarias y proteínas membranales (Figura. 20). La idea básica de predicción en este sistema esta basada en las propiedades físico químicas de las secuencias de amino ácidos tales como hidrofobicidad y cargas. El sistema usa tres tipos de predicción: a) Discriminación de proteínas membranales de solubles b) Predicción de hélices tranmembranales y c) Determinación de regiones tranmembranales helicoidales.

La precisión de este sistema, la discriminación de proteínas membranales, la existencia de hélices transmembranales y de regiones helicoidales transmembranales es de alrededor de 99%, 96% y 85% respectivamente.

CAPÍTULO 4 – RESULTADOS Y DISCUSIÓN

4.1 Secuenciación y anotación de 11 fósmidos que traslapan el locus Fu/HC

Durante los últimos años, A. W. De Tomaso e Irving Weissman caracterizaron y mapearon genéticamente el locus Fu/HC. Este locus altamente polimórfico fue mapeado al identificar inicialmente marcadores genéticos ligados a Fu/HC. Estos marcadores son llamados “Polimorfismos de longitud de fragmentos amplificados” (AFLPs) (Vos et al., 1995), los cuales están directamente correlacionados con los fenotipos de fusión y rechazo mediados por los ensayos de transplatación *versus* una cepa probando de Fu/HC (**Figura. 7**) (De Tomaso et al., 1998). Gracias a los AFLPs polimórficos y otros marcadores en las 725-Mega bases que conforman el genoma de *B. schlosseri*, el Fu/HC ha sido delineado en una región de aproximadamente 1-centi Morgan (cM). (De Tomaso and Weissman, 2003) (**Figura. 6**).

Aproximadamente, 500 kb han sido secuenciadas y anotadas dentro de la región de Fu/HC delineada en 1cM. Esto incluye los 100kb~ de secuencia completamente ensamblada y anotada, incluida en los tres fósmidos completamente secuenciados; subsf-217g15, subsf-371a15 and subsf-531d19 ([AC136512](#);32Kb, [AC140854](#)30Kb, [AC140854](#) 30Kb). Las 400 kb restantes han sido incluidas en ocho fósmidos con un porcentaje completado de entre 70-90%. Esta secuencia ha sido organizada en 30 contigs (secuencias contiguas) de varios tamaños, mismos que fueron reportados en el NCBI (**Figura. 21**).

Actualmente, el segmento contiguo más grande es de 4.4kb en subsf-555i12. Los contigs remanentes oscilan entre 2 kb y 3kb.

Una vez ensamblada la secuencia borrador “draft sequence”, fue anotada. Lo que permitió descubrir genes putativos (**Tabla 3**) que posteriormente fueron identificados y localizados físicamente (**Figura. 21**).

Cuarenta y cinco genes putativos han sido identificados en el borrador de la secuencia ensamblada y anotada. Todos ellos fueron identificados por FGENESH a y/o por GENSCAN. Sin embargo, solo dieciocho genes fueron encontrados altamente similares ($E < 25$) a proteínas conocidas. Las cuales tienen el potencial de estar involucradas en sistemas de histocompatibilidad, procesos celulares, rutas metabólicas y transposición de ADN. Además, fueron identificadas 6 proteínas con dominios transmembranales los cuales reportaron valores E “E values” intrigantes, pues dadas las características transmembranales del TCR, Ig, etc. Existe la posibilidad de que estas proteínas tengan alguna función en la primer etapa de reconocimiento propio/no propio.

Es posible que seis genes se encuentren directa o indirectamente relacionados a un sistema de histocompatibilidad.

Una Fosfolipasa A2 la cual tiene un dominio conservado de fosfolipasa, Tres proteínas desconocidas pero con varios dominios WD40, Una serin threonin kinasa de especificidad dual con un dominio conservado de Serin Treonin kinasa.

Adicionalmente, se encontraron siete genes que pueden participar en una variedad de procesos celulares y rutas metabólicas.

Una GMP syntetasa con un dominio conservado de Glutamino Amidotransferasa y otro de ATP-asa, una Beta-lactamasa con un dominio conservado de lactamasa, una Adenyl succinato liasa con un dominio conservado de Adenyl succinato lyasa, una N-acetyl Neuramato Pyruvato lyasa un dos dominios conservados; Dihydropicolinato syntetasa y N-acetylneuraminato, Una ubiquinona, con un dominio conservado de NADH₂ dehydrogenase y un sitio de acarreamiento de acilos y de union a fosfopantenina y finalmente un par de proteínas hipotéticas and 2 Hypotetical con una breve homología a proteínas de tipo clatrina.

Cinco elementos genéticos móviles con valores $E > 10^{-6}$ fueron identificados en Fu/HC; dos elementos Ty1-Copia clasificados como retro elementos, los cuales poseen dominios conservados de integrasa y de Poliproteina, un elemento retrotransponible del tipo Non LTR similar a CRI el cual contiene un dominio conservado de reverso transcriptasa, un elemento transponible del tipo Tip 100 con un dominio dedo de zinc

ZF “zinc finger” y un dominio de transposasa y finalmente una transposasa putativa del tipo DDE endonucleasa aunada a un dominio conservado PHD del tipo dedo de zinc. Seis proteínas fueron identificadas con hits limitados pero intrigantes ($E > 10^{-1}$), con dominios transmembranales pero una con función intrigante aunque desconocida en terminos reales del valor E. La baja homología probablemente se deba a la calidad de la secuencia en borrador o bien a una predicción incompleta o a la falta de puntos de comparación, es decir, a la falta genomas de organismos marinos completamente secuenciados.

En este grupo de proteínas con dominios transmembranales se encuentran; Una pectinerasa con una helice primaria transmembranal, una proteína del tipo Hole con dos dominios transmembranales, uno con estructura de hélice primaria y el otro de hélice secundaria, un Acetyl coenzima A carboxylasa con una hélice transmembranal primaria, una proteína hipotética sin similitud a proteína existente pero con una hélice transmembranal primaria.

Por otro lado, se encontró un elemento interesante, un gene putativo similar al factor del complemento BF/C2 en el pez *Medaka*, en el cual se localizaron dos hélices primarias transmembranales (Fig20). Con respecto a estos genes, la homología real era muy baja para realizar análisis comparativos fidedignos. Sin embargo, este tipo de análisis se llevarán acabo, una vez que los “gaps” remanentes sean rellenados completamente. Cuando esto suceda, habrá una re anotación obligatoria de los genes mencionados en este párrafo y de esta manera se afirmará o descartará la homología encontrada en el presente análisis. Además, será de gran importancia hacer más análisis predictivos para saber si las proteínas transmembranales aquí identificadas son secretadas o tienen algún papel en los eventos de reconocimiento como el de unión a péptidos antigénicos o de presentación de moléculas en superficie.

| FÓSMIDO | GENSCAN | Score Blast p | FGENESH HMM | Score Blastp | DOMINIOS CONSERVADOS Blast phi/rpsBlast | Score Blast phi/rps | DOMINIOS TRANSMEMBRANALES SOSUI |
|---------|-------------------------------------|---------------|--------------|--------------|---|---------------------|---------------------------------|
| 217g15 | 1. GMP synthase <i>Mus musculus</i> | 555 / e-156 | GMP synthase | 566 / e-160 | GuaA PP-ATPase & Grutamine Amidotransferase | 198/1 e-51g | SOLUBLE |

| | | | | | | | |
|--------|---|-------------|--|--------------|---------------------------|------------|--------------------------------|
| 555i12 | 1. *Putative glucokinase <i>Salmonella Enterica</i> | 34/1.4 | ORF PUTATIVO | 0 | 0 | 0 | Transmembranal Hélice primaria |
| | 2. zinc finger protein <i>Mus musculus</i> | 36/0.11 | ORF PUTATIVO | 0 | 0 | 0 | SOLUBLE |
| | 3. galactose ABC transporter periplasmic-binding protein <i>Haemophilus Influenzae</i> Rd | 33/1.1 | ORF PUTATIVO | 0 | 0 | 0 | SOLUBLE |
| 533i14 | 1. Ty1-Copia Type Retroelement Polyprotein <i>Arabidopsis Taliana</i> | 251 - 3e-65 | Ty1-Copia Type Retroelement Polyprotein Arabidopsis Taliana | 251-3e-65 | Rve Integrase Core domain | 112/8e-26 | SOLUBLE |
| | 2. Putative Transposase <i>Oryza Sativa Japonica cultivar</i> | 40/0.98 | ORF PUTATIVO | ORF PUTATIVO | 0 | 0 | SOLUBLE |
| | 3. Reverse Transcriptase like CR1like non LTR retrotransposon. <i>Paralichthys Olivaceus</i> | 167/4e-40 | Reverse Transcriptase like CR1like non LTR retrotransposon. Paralichthys Olivaceus | 152/1e-35 | Rvt Reverse transcriptase | 81.2/3e-16 | SOLUBLE |
| | 4. Chromosome assembly protein homolog <i>Aquifex Aeolicus</i> | 41/0.022 | ORF PUTATIVO | ORF PUTATIVO | 0 | 0 | SOLUBLE |
| | 5. Ty1 Copia type retroelement Polyprotein <i>Arabidopsis Thaliana</i> | 227/2e-58 | Gag Pol Multigenic A1-Sh2 Zea mays | 223/4e57 | 0 | 0 | SOLUBLE |
| | 6. Hypotetical Protein <i>Oryza Sativa Japonica cultivar</i> | 39/0.052 | ORF PUTATIVO | 0 | 0 | 0 | SOLUBLE |

| | | | | | | | |
|--------|--|-------------------|--|-------------------|---|--------------------|-------------------------------------|
| | 7. Hypotetical Protein <i>Ratus Novergicus</i> | 33/ 4.1 | ORF PUTATIVO | 0 | 0 | 0 | SOLUBLE |
| 564g21 | 1. Predicted Reverse transcriptase 5A739 <i>C, elegans</i> | 62/ 1e- 09 | Reverse Transcriptase Anopheles | 55 /2e- 09 | Rvt,Reverse transcriptase. | 41.2/ 2e | SOLUBLE |
| | 2. PUTATIVE PROTEIN | 0 | ORF PUTATIVO | 0 | 0 | 0 | SOLUBLE |
| 568h21 | 1. Polyprotein (reverse transcriptase) <i>Tomato</i> | 61/ 1e- 07 | Polyprotein Tomato and Oryza Sativa Japonica cultivar | 60/2e -07 | 0 | | SOLUBLE |
| | 2. PUTATIVE TRANSPO SABLE ELEMENT Tip100 <i>Oryza Sativa Japonica cultivar</i> | 183 /1e- 44 | PUTATIVE TRANSPO SABLE ELEMENT Tip100 <i>Oryza Sativa Japonica cultivar</i> | 178/ 4e- 43 | ZnF_TTF Transposases and Transcriptases | 48.5/ 3e-06 | SOLUBLE |
| | 3. Ac like transposase <i>Arabidopsis</i> | 34/ 0.7 5 | ORF PUTATIVO | 0 | 0 | 0 | SOLUBLE |
| | 4. Reverse Transcriptas e synthetic <i>Anopheles Gambiae</i> | 76/ 8e- 13 | . Reverse Transcriptase | 0 | 0 | | SOLUBLE |
| 532h24 | 1. Beta- lactamase <i>Staphylococ us Aureus</i> and several Vectors | 541 /1e- 53 | Beta-lactamase Staphylococcus Aureus and several Vectors | 541/ 1e- 53 | *PenP, Beta Lactamase | 134/1 e-32 0 | SOLUBLE |
| | 2. Hypotetical Protein | 32/ 2.7 | ORF PUTATIVO | 0 | 0 | 0 | SOLUBLE |
| | 3. *Acetyl Coenzyme A Carboxylase <i>Xantomas Campestris</i> | 34/ 2.2 | ORF PUTATIVO | 0 | 0 | 0 | Transmembranal 1 Hélice primaria |

| | | | | | | | |
|--------|---|------------|--|-----------|--|-------------------------|--|
| 371a15 | 4. Similar to Jacalin Lectin <i>Arabidopsis Thaliana</i> | 34/1.9 | ORF PUTATIVO | 0 | 0 | 0 | SOLUBLE |
| | 5. Hypotetical protein | 33/5.4 | ORF PUTATIVO | 0 | 0 | 0 | SOLUBLE |
| | 1. ORF PUTATIVO | 0 | ORF PUTATIVO | 0 | 0 | 0 | SOLUBLE |
| | 2. Similar to ubiquitin-ligase E3 and NEDD4 <i>Homo sapiens</i> | 252/1e-655 | Similar to ubiquitin ligase | 246/9e-64 | HECTc C-terminal Ubiquitin ligase E3 | 197/2e-51F no g | SOLUBLE |
| | 3. Dyrk1a, Dual Specificity Serin/threonin-kinase. Similar to FAS/FADD interacting Kinase Inhibits fas mediated jun <i>Mus musculus</i> | 405/e-112 | Dual –specificiy tyrosine phosphorylation regulated kinase | 323/8e-87 | S_TKc, Serine/Threonine protein kinases, | 96/1e-20f | SOLUBLE |
| 557e15 | 1. Adenylosuccinatelyase (ASL) <i>Gallus gallus</i> | 348/e-135 | Adenylosuccinate lyase | 407e-112 | PurB, Adenylosuccinate lyase | 168/8e-43 f | SOLUBLE |
| | 2. Putativetransposase Fungus <i>Phaeosphaeria nodorum.</i> | 121/6e-26 | Transposase | 177/1e-42 | DDE endonuclease PHD, zinc finger | 132/1e-31f 38.3/0.005 f | SOLUBLE |
| | 3. Hole protein 2Transmembranal domains <i>Homo sapiens</i> | 77/3e-13 | Hole gallus | 82/5e-14 | 0 | | 2 dominios Transmembranales una Hélice Primaria y una Secundaria |
| | 4. Putative polysacaride biosynthesis protein <i>Bacteroides Heliothecae</i> | 32/2.9 | ORF PUTATIVO | 0 | 0 | 0 | SOLUBLE |

| | <i>n</i> | | | | | | |
|-------|---|---------------------|--|---------------------|---|-------------------------|------------------------------------|
| | 5. N-acetyl Neuramate Pyruvate lyase Cell envelope biogenesis <i>Mus musculus</i> | 130/1e-29 | N-acetylneuraminat e pyruvate lyase | 124/9e-28 | DapA, Dihydrodipicolinate synthase/N-acetylneuraminat e lyase | 124/9e-28f | SOLUBLE |
| | 6. N-acetyl Neuramate Pyruvate lyase <i>Mus musculus</i> | 30/0.017 | ORF PUTATIVO | 0 | 0 | 0 | SOLUBLE |
| | 7. Low affinity glucose transporter <i>Saccharomyces cerevisiae</i> | 30/7 | ORF PUTATIVO | 0 | ORF PUTATIVO | 0 | SOLUBLE |
| | 8. Hypotetical protein | 32/7.4 | 0 | 0 | ORF PUTATIVO | 0 | SOLUBLE |
| 560b6 | 1. Medaka Complement BF/C2 Complement Factor Sushi repeat SCR domain <i>Medaka Fish</i> | 36/0.51 | Polydom, Control Complement Protein CCP. | 33/8.1 | ORF PUTATIVO | 0 | Transmembranal 2 Hélices primarias |
| 53719 | 1. Similar to hypothetical protein | No similarity found | Similar to hypothetical protein | No similarity found | ORF PUTATIVO | 0 | Transmembranal 1 Hélice primaria |
| | 2. Unknown protein 353aa <i>Xenopus laevis</i> | 109/9e-23 | Similar to hypothetical protein | 182/1e-44 | WD40 & WD40 G beta | 88.1/3e-18f 40.6/6e-04f | SOLUBLE |
| | 3. Hypothetical protein FLJ31438 <i>Homo sapiens</i> | 150/5e-35 | Hypothetical protein FLJ31438 | 133/1e-29 | ORF PUTATIVO | 0 | SOLUBLE |

| | | | | | | | |
|--------|--|-------------|---|-------------|---|--------------------------------|----------------------------------|
| | 4. Unknown protein 697 aa <i>Xenopus laevis</i> | 197 / 6e-49 | Unknown protein | 107 9e-23 | WD40 WD40.G-beta | 68.9.4/3e-13f 33.83 / 6e-04 | SOLUBLE |
| | 5. Hypotetical Protein Flj31438 Clathrin like <i>Homo sapiens</i> | 184 / 4e-45 | Hypotetical Protein | 184 / 4e-45 | ORF PUTATIVO | 0 | SOLUBLE |
| | 6. NADH2dehydrogenase (ubiquinone) <i>Mus musculus</i> | 112 / 2e-24 | Mitochondrial NADH2dehydrogenase (ubiquinone) | 109 / 1e-23 | Acyl carrier & Phosphopantethin attachment site | 57.2 / 7e-10f 44.4 / 5e-06f | SOLUBLE |
| | 7. *G-box, Dead Box <i>Soy bean</i> | 40 / 0.028 | ORF PUTATIVO | 0 | ORF PUTATIVO | 0 | SOLUBLE |
| 531d19 | 1. Unknown protein <i>Xenopus laevis</i> | 150 / 1e-35 | Unknown protein | 107 9e-23 | WD40 WD40.G-beta | 68.9.4/3e-13f 33.83 / 6e-04 | SOLUBLE |
| | 2. Similar to hypothetical protein <i>Homo sapiens</i> | 154 / 4e-35 | Similar to hypothetical protein | 154 / 4e-35 | ORF PUTATIVO | 0 | SOLUBLE |
| | 3. Fosfolipasa A2 pkP2 precursor <i>Snake</i> | 58 / 3e-07 | Fosfolipasa A2 jerodoxine precursor | 46 / 7e-04 | Fosfolipasa A2 | 62.9 / 2e-08f | SOLUBLE |
| | 4. Pectinerase family <i>Arabidopsis thaliana</i> | 38 / 0.13 | 0 | 0 | ORF PUTATIVO | ORF PUTATIVO | 1 Transmembranal Hélice primaria |

Tabla 3. Anotación de once fósidos que sobrelapane 1 locus Fu/HC. *Posible contaminación de Vector.

4.3 No existe evidencia de elementos de inmunidad adaptativa en el Fu/HC

Los elementos más representativos del sistema inmuno-adaptativo y por tanto de su emergencia son: los receptores antígeno-específicos de vertebrados, dentro estos se

encuentra la super familia de las Inmunoglobulinas (Ig) de la cual forman parte los dominios constantes y variables. Los genes que codifican para los dominios variables son heredados como segmentos que sufren arreglos somáticos dentro del linaje celular de los linfocitos, con el objetivo de crear genes variables (V), alternativos y funcionales (Twonegawa S. 1976). Por otro lado, los dominios constantes que pertenecen a un número restringido de moléculas tales como el MHC de clase I y el de clase II, la fracción constante de las Inmunoglobulinas, y los receptores de Células T (TCR). Todas estas moléculas esenciales en la especificidad y memoria de la respuesta inmunoadaptativa, solo han sido identificadas en vertebrados ancestrales (vertebrados con mandíbulas) Sin embargo, su origen no ha sido elucidado aún (**Figura.1**) (Du Pasquier et al., 2000). El borrador de la secuencia y la anotación de Fu/HC en paralelo con el análisis genómico de el sistema inmune en *C. intestinalis* (Azumi et al., 2003) dio a conocer el primer análisis relativo a los orígenes de la respuesta inmunoadaptativa en los urocordados.

Hasta ahora, en Fu/HC no ha sido identificada molécula alguna que esta directamente relacionada con el sistema inmuno-adaptativo. Con respecto al hallazgo de elementos de un sistema inmune ancestral, siete elementos potenciales fueron encontrados; tres genes modelo codificantes de proteínas con varios dominios WD40, una E3 ubiquitin ligasa, una Fosfolipasa A2, una Serin Treonin Kinasa y una proteína con dos dominios transmembranales pero con una baja homología a la proteína C2-Bf del sistema de complemento. Aunados a los elementos antes mencionados, se encontraron seis genes en los que se incluyeron dominios transmembranales. Dentro de estas proteínas, se encontraron dos que tienen una breve pero intrigante homología con genes que codifican para proteínas del sistema de complemento en vertebrados e invertebrados; Bf2/C2 y un receptor del complemento con dominios de secuencias consenso cortas y repetidas (SCR) (De Tomaso and Nyholm., 2003).

Hasta ahora no hay estudios funcionales que validen a estos genes como candidatos a ser receptores y probables elementos durante la etapa de reconocimiento en un sistema inmune ancestral. Sin embargo, existe la posibilidad de que participen en la emergencia de la respuesta inmuno adaptativa o de que formen parte de un sistema de histocompatibilidad ancestral.

Aunque no existe relación directa de los genes aquí reportados con el sistema inmuno-

adaptativo, pero cabe la posibilidad de que en un futuro sean asociados con este sistema que hasta ahora sido exclusivo de los vertebrados. Los resultados reportados hasta el momento, sugieren que el ancestro del MHC o elementos relacionados con la emergencia del sistema inmuno-adaptativo faltan por ser confirmados. Sin embargo, es probable que más de estos elementos sean encontrados en el futuro dentro del segmento hasta ahora no caracterizado de Fu/HC, o en otro segmento del genoma de *B.schlosseri* o bien, en otro invertebrado cordado y colonial(**Figura 1**).

4.2 Regiones sinténicas del Fu/HC con el MHC

A través de búsquedas metódicas utilizando Blastn y Locus Link y mediante el uso de los modelos génicos descritos en la tabla 3. Fue posible analizar la posible existencia de una región sinténica del Fu/HC en el genoma de cualquier organismo vertebrado o invertebrado. A la fecha, no se ha encontrado ningunos de esos genes ligados o relacionados. Además, tampoco se ha encontrado región sinténica de Fu/HC a los genes del MHC.

4.3 ¿Existe algún gene en Fu/HC envuelto en reacciones inflamatorias durante el aloreconocimiento?

En los mamíferos, la inflamación es iniciada por la hidrólisis de fosfolípidos araquidónicos efectuada por la fosfolipasa A2, ocasionando la subsecuente separación del ácido araquidónico. Las prostaglandinas y los leucotrienos los cuales son responsables de varios efectos inflamatorios, son subsecuentemente sintetizados a partir de ácido araquidónico por cicloxigenasas y lipoxigenasas. La fosfolipasa A2 modula positivamente la respuesta inflamatoria (Gabay, 2001). Esta enzima es un proteína positiva de la fase aguda “positive acute phase protein” y parte de la respuesta sistémica de la fase aguda “systemic acute phase response”. Tal y como sucede con el genoma de

C. intestinallis el cual contiene varios modelos de genes que potencialmente codifican homólogos, mamíferos de la fosfolipasa A2 (Azumi et al., 2003). El locus Fu/HC contiene una fosfolipasa A2 misma que posee un dominio conservado de fosfolipasa. Es importante mencionar que cuando dos colonias de *B. schlosseri* incompatibles en Fu/HC locus se funcionan, se acciona una respuesta de aloreconocimiento. Este evento de reconocimiento propio/no propio es caracterizado por rechazo vascular e inflamación (Scofield et al., 1982). Entonces la fosfolipasa A2 codificada por Fu/HC probablemente sea la responsable de iniciar la respuesta inflamatoria como resultado del aloreconocimiento.

4.5 Elementos genéticos móviles en plantas en Fu/HC

Con respecto a la situación geográfica de *B. schlosseri* (en el mar), se encontró de manera inesperada que al menos dos elementos transponibles homólogos en plantas (en la tierra) fueron encontrados en Fu/HC.

Cabe señalar como hecho particular que los elementos genéticos móviles encontrados en *B. schlosseri* están directamente relacionados con elementos transposones de *Arabidopsis thaliana*, *Zea mays* y *Oriza sativa*. Resulta interesante que estos elementos móviles no posean relación filogenética o geográfica ya que la relación a pesar de la distancia filogenética es algo no común. Una posible explicación a este hecho es la transferencia horizontal, esta se da gracias a la promiscuidad común en la gran mayoría de los elementos genéticos móviles. Por el contrario, es interesante hacer notar el hecho de que un análisis previo al genoma de *Ciona intestinallis* (Simmen, 2000), considerada como otra ascidea marina con una relación evolutiva estrecha a *B. schlosseri*, fueron detectados elementos genéticos móviles i.e. Transposones y retrotransposones. Como era de esperarse, estos elementos fueron directamente relacionados con vertebrados e invertebrados marinos y terrestres tales como; *C. elegans*, *Anopheles gambiae*, Zebra fish, Puffer fish, *Fugu rubripes* and *Xenopus laevis* (Goodwin 2002, Simmen, 2000), pero no plantas. Tiene sentido la relación de

estos genéticos móviles con los transposones y retrotransposones de vertebrados e invertebrados de origen marino y terrestre. Debería de esperarse el mismo escenario para los elementos genéticos móviles de *B.schlosseri*. Sin embargo, no fue así.

Otra posible explicación a este hecho es la falta de información genómica referente a organismos marinos. Quizás existan otros casos similares. Entre más genomas marinos sean secuenciados, más información robustecerá los análisis evolutivos que eventualmente ayudaran a explicar el origen inesperado de estos elementos genéticos móviles homólogos en plantas.

4.6 La paradoja de un locus altamente polimórfico.

Es importante mencionar que a pesar de que existen algunas explicaciones teóricas relacionadas a la elevada tasa polimórfica de Fu/HC. No se ha encontrado explicación convincente alguna (Grosberg 2000). Sin embargo, cabe la posibilidad de que el origen de esta variación dinámica y polimórfica sea explicada por la elevada concentración de elementos genéticos móviles encontrados en este análisis

[12 elementos móviles/0.5Mb].

4.7 De ascideas, hongos y plantas a la utilidad de la teoría evolutiva. cuál es la relación de tres proteínas con varios dominios WD40 y la E3 ubiquitin ligasa encontrada en el fu/HC?

Los tres genes modelo codificantes de proteínas con varios dominios WD40 y la E3 ubiquitin ligasa encontrada en el locus Fu/HC son candidatos potenciales para formar parte del sistema de histocompatibilidad en *B.schlosseri*. La razón de esta aseveración es que por una parte; aunque el hongo del limo *Dictyostelium Discoideum*, el hongo filamentoso *Podospora anserina*, la planta *Brasica Pollen* y la ascidea marina *B.schlosseri* e no tienen una directa relación filogenética (Dec. N. Dao 2000 and Buss

1999). por otra parte, todos ellos tienen en común tanto la proteína E3 ubiquitin ligasa como el dominio WD40, los cuales forman parte tanto de sistemas de histocompatibilidad como de sistemas de reconocimiento propio y no propio en todos ellos. Además, todos ellos confrontan un problema de rechazo y de parasitismo por células de su propia especie. Este problema se ubica probablemente en la raíz de la evolución de una variedad de mecanismos de incompatibilidad (Buss 1999). Todos los organismos antes mencionados tienen una Fbox o un complejo SRC que se forma por proteínas con varios dominios WD40 domains las cuales inician el reclutamiento de las proteínas incluidas en el complejo de rechazo donde las ubiquitin ligasas E2 y E3 son esenciales también (Buss 1999, Burnet 1971, Sövan 2002, Michiko 1999, Sven J. Saupé 2000, Sophia L. Stone 2003, Dee N. Dao 2000, Espagne Eric 2002, Shigetsugu 1999). Basados en estos hechos, existe una posibilidad de que Fu/HC conlleve algunos elementos importantes de un sistema de histocompatibilidad o una estrategia de reconocimiento conservada en la evolución.

CAPÍTULO 5 –CONCLUSIONES Y PERSPECTIVAS

CONCLUSIONES

El presente análisis da a conocer la primera descripción en el reino animal de un locus de histocompatibilidad no vertebrado. A pesar del hecho de que aún no está completamente secuenciado, se ha analizado un segmento de ~0.5 Mb de Fu/HC.

El presente análisis es la base tanto de estudios contemporáneos como futuros en Fu/HC (De Tomaso et al., 1998; 2003 and 2004). Adicionalmente, la base de una más detallada estructura génica del Fu/HC está parcialmente establecida por lo que resultados son trascendentales.

En resumen, la secuencia borrador y la anotación de los once fósidos que traslapan parte del locus Fu/HC ha sido reportado. En el curso del proceso de anotación, cuarenta y cinco genes fueron elucidados. Por el contrario, solo el 40% de los genes putativos aquí descritos, tuvieron homología con valores E sostenibles ($E < 10^{-25}$) a genes codificantes de proteínas con función conocida encontrados en el banco de genes de NCBI. Todos ellos tienen funciones variadas en procesos celulares diferentes que van desde codificar para enzimas encontradas en rutas metabólicas, como proteínas que activan respuestas inflamatorias, de histocompatibilidad y de recombinación de ADN. El restante 60% de los genes anotados por FGENESH y GENSCAN no se les encontró relación fehaciente de función alguna. Sin embargo 6, de estos genes desconocidos, es posible que tengan algún papel en los eventos de reconocimiento dentro de la respuesta de aloreconocimiento. Estos genes tienen dominios transmembranales que los convierte en receptores potenciales. Es importante mencionar que a pesar de no existir prueba alguna de que los elementos genéticos móviles encontrados en Fu/HC son funcionales, quizás exista alguna relación entre el funcionamiento de los mismos y las características altamente dinámicas y polimórficas de este locus (De Tomaso et al. 4, 2003). No obstante, el 40% de los genes putativos con función desconocida están en proceso de ser definidos y confirmados hasta que los huecos “gaps” faltantes en la

secuencia sean rellenados o bien, más genomas sean anotados.

La información genómica revelada a través de la secuenciación y anotación de Fu/HC será de gran importancia para análisis de genómica comparativa. Como el primer locus de histocompatibilidad metazoano analizado fuera de los vertebrados, este estudio será de gran valor para futuros análisis evolutivos que pretendan entender lo que sucedió en el “big bang” inmunológico y por tanto en los orígenes de el “sistema inmuno adaptativo”.

PERSPECTIVAS

La relación filogenética de los protocordados en el inicio de la evolución de los vertebrados y las similitudes genéticas entre los sistemas de reconocimiento propio y no propio en el Fu/HC y el MHC han implicado una incisiva relación entre los dos sistemas (Scofield et al., 1982; De Tomaso et al., 2003). Sin embargo, los resultados aquí analizados, no expresan lo mismo partiendo de una perspectiva de genómica evolutiva. No existe evidencia de homología o sintenia entre el Fu/HC y el MHC. Posiblemente las similitudes son solo funcionales y no a nivel de la secuencia de ADN. Posiblemente, *B.schlosseri* utiliza otras tácticas para encender las estrategias de reconocimiento propio/no propio. Sin embargo, aún si no existiese homología entre a nivel de ADN. Es de gran importancia elucidar la estrategia que Fu/HC utiliza para activar los fenotipos de fusión/histocompatibilidad.

Una estrategia para descubrir el mecanismo de funcionamiento de los genes de Fu/HC es la de realizar estudios funcionales en cada uno de los genes aquí reportados, con el fin de sustentar su papel en el reconocimiento propio/no propio. Esta estrategia estaría basada en el uso de la información obtenida a lo largo de este estudio, con el fin de hacer sondas de cADN para eventualmente realizar hibridaciones in situ o micro hilera de ADN con el propósito de analizar si los genes Fu/HC se expresan y son funcionales, y si son prendidos o apagados alternativamente cuando se expresan los fenotipos de fusión o de rechazo. Además, haciendo uso de herramientas genómicas como los RNAsi, podrá ser descubierta la función del 60% de los genes putativos con función desconocida, elucidados por GENSCAN y FGENESH.

La incógnita de la etiología de la elevada tasa polimórfica en FU/HC esta aún vigente. Sería atractivo saber si los elementos móviles encontrados en el presente análisis son funcionales, y si lo son, en que medida contribuyen a la alta tasa polimorfita en este locus.

Debido al descubrimiento de una E2 ubiquitin ligasa y de tres proteínas con varios dominios WD40 y su relación directa a sistemas de aloreconocimiento e histocompatibilidad. Una incógnita se origina: ¿Son evolutivamente requeridas tanto la E2 ubiquitin ligasa como las 3 proteínas con varios dominios WD40 para accionar sistemas de Incompatibilidad ancestrales y contemporáneos? Esta pregunta puede ser respondida haciendo uso tanto de las herramientas funcionales previamente mencionadas como usando árboles filogenéticos. Así, será imprescindible hacer un análisis basado en genómica comparativa para definir la incisiva correlación de sistemas de histocompatibilidad y estrategias de aloreconocimiento en plantas, hongos, vertebrados e invertebrados. (Buss 1999, Brunet 1971).

REFERENCIAS

- 1 Altschul., Lipman,DJ. Et all 1997. Gapped BLAST and PSI'BLAST: A new generation of protein database search programs. *Nucleic Acid Res* 25:3389-3402.
- 2 Anthony W. De Tomaso, Yasunori Saito, Katherine J. Ishzuka, Karla J. Palmeri and Irving L Weissman., Mapping the Genome of a Model Protochordate. I. A low Resolution Genetic Map Encompassing the Fusión/Histocompatibility (Fu/Hc) Locus of *B. schlosseri* *Genetics* (1998)149:277-287
- 3 Anthony W. De Tomaso, S V Nyholm, IL Weissman. Cloning and Expression of a Novel, Highly Polymorphic Transmembranal Protein from a Protochordate Histocompatibility Locus. http://cyo.edu.au/meeting_2002/Abstracts/deTomaso.html
- 4 Anthony W. De Tomaso and Irving L. Weissman .Construction and characterization of large-insert genomic libraries (BAC and fosmid) from the Ascidian *B. schlosseri* and initial physical mapping of a histocompatibility locus. *Mar Biotechnol* (NY). (2003) Mar-Apr;5(2):103-15
- 5 Anthony W. De Tomaso and Irving L. Weissman. Initial characterization of a protochordates histocompatibility locus. *Immunogenetics*. (2003) Oct;55(7):480-90
- 6 Anthony W. De Tomaso and Irving L. Weissman. Evolution of a Protochordate Allorecognition Locus. *SCIENCE* (2004)303:977.
- 7 Baruch Rinkevich. The colonial urochordate *B. schlosseri* : from stem cells and natural tissue transplantation to issues in evolutionary ecology *BioEssays* (2002) 24:730-740
- 8 Brent Ewing and Phil Green. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research* (1998) 8:186-194.
- 9 Brent Ewing, LaDeana Hillier, Michael C. Wendl, and Phil Green. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*(1998) 8:175-185.
- 10 Bodenteich, S.Chisoe, YF Wang and B.A. Roe. Shotgun Cloning as the Strategy of Choice to Generate Templates for High-throughput Dideoxynucleotide Sequencing in Automatic ADN sequencing and Analysis Techniques (Academic Press. London). (1994):42-50

- 11** Burge, C. B. and Karlin, S. Finding the genes in genomic ADN. *Curr. Opin. Struct. Biol.* (1998) 8, 346-354.
- 12** Burnet, F.M. "Self-recognition" in colonial marine forms and flowering plants in relation to the evolution of immunity (1971). *Nature*, 232:230-235.
- 13** Buss Leo W. Slime molds, ascidians, and the utility of evolutionary theory. (1999) 96,8801-8803.
- 14** Chadwick-Furman, N.E., & Weissman, I.L. Life histories and senescence of *Botryllus schlosseri* (Chordata: Ascidiacea) in Monterey Bay. *Biological Bulletin, Marine Biological Laboratory, Woods Hole*, (1995)189, 36-41.
- 15** Christopher M. Hess, Joe Gasper, Hopi E. Hoekstra, Christopher E. Hill, and Scott V. Edwards, MHC Class II Pseudogene and Genomic Signature of a 32-kb Cosmid in the House Finch (*Carpodacus mexicanus*) *Genome Research* (2000)10:613-623,
- 16** Crick, F. The double helix: a personal view. *Nature* (1974) 248, 766–769.
- 17** David A Micklos & Greg A. Freyer, *ADN Science: A First Course in Recombinant ADN Technology*. January 1990, pp 400.
- 18** N. Dee N.Dao, Richard H. Kessin and Herbert L. Ennis. Developmental cheating and the evolutionary biology of *Dictyostelium* and *Myxococcus*. *Microbiology* (2000), 146, p.1505–1512
- 19** Du Pasquier L (2000) The phylogenetic origin of antigen-specific receptors. In : Du Pasquier L, Litman GW (eds) *Origin and evolution of the vertebrate immune system*. Springer, Berlin Heidelberg New York, pp. 160-185.
- 20** DuPasquier L, Flanjnik M (1999) Origin and evolution of the vertebrate immune system. In: Paul WE (ed) *Fundamental immunology*. Lippincott-Raven, Philadelphia.
- 21** Edwin L Cooper. Immune response evolution. *Encyclopedia of life sciences* Nature publishing group 1-8 (2001). www.els.net
- 22** Espagne ERic ,Pascale Balhade`re, Marie-Louise Penin, Christian Barreau and Be´atrice Turcq. HET-E and HET-D Belong to a New Subfamily of WD40 Proteins Involved in Vegetative Incompatibility Specificity in the Fungus *Podospora anserina*. *Genetics* (2002) **161**: 71–81
- 23** Flajnik MF, Kasahara M Comparative genomics of the MHC: glimpses in to the evolution of the adaptive immune system. *Immunity* (2001) 15:351-62
- 24** Gabay Cem and Kushner I Acute-Phase Proteinsand other systemic responses to

inflammation N England J Med (2001) 340:448-454 1999.

- 25** Godinez Ricardo. Biotecnología, Jugando a ser Dios? Gaceta del Colegio de Ciencias y Humanidades, Noviembre 2001,15-16.
- 26** Goodwin TJD, Poulter R.T.M. A group of deuterostome Ty3/gipsy-like retrotransposons with Ty1/*copia*-like *pol*-domain orders (2002) 267:481-491
- 27** Gordon, D., C. Abajian, and P. Green.. Consed: A graphical tool for sequence finishing. Genome Research. (1998) 8:195-202
- 28** Janeway, C.A., Travers, P., Wallport, M., Shlomchik, M. (2002) Immunobiology, 5 ed., Churchill Livingstone, 752 p
- 29** Kaoru Azumi, et al Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot". Immunogenetics .(2003) Nov 55: (8):570-81.
- 30** Khalturin K. Molecular self non-self recognition in the Urochordate *B. schlosseri* . (November 2002) Thesis
- 31** K Khalturin, M Becker, B Rinkevich and TCG Bosch Urochordates and the origin of natural killer cells: identification of a CD94 / NKR-P1 related receptor in blood cells in *Botryllus*. (2003)Proc. Natl. Acad. Sci. USA, 100, 622-627.
- 32** Laird DJ, De Tomaso AW, Cooper MD, Weissman IL. 50 million years of chordate evolution: seeking the origins of adaptive immunity.Proc Natl Acad Sci U S A. 2000 Jun 20;97(13):6924-6.s
- 33** Leo W. Buss Slime molds, ascidians, and the utility of evolutionary theory. Proc. Natl. Acad. Sci. USA 1999 August 3; 96 (6) 8801-8803.
- 34** Michiko Shirane
- 35** Smith, Lloyd M.; Sanders, Jane Z.; Kaiser, Robert J.; Hughes, Peter; Dodd, Chris; Connell, Charles R.; Heiner, Cheryl; Kent, Stephen B. H.; Hood, Leroy E.; Fluorescence detection in automated ADN sequence analysis. Nature. v 321: p 674-679, 1986.
- 36** Magor BG, De Tomasso A, Rinkevich B, Weissman IL. Allorecognition in colonial tunicados: Protection against predatory cell lineages? Immunol Rev (1999) 167:6979
- 37** Muller WEG, Pancer Z Rinkevich B. Molecular cloning and localization of a novel serine protease from the colonial tunicate *B. schlosseri* . Mol Mar Biol Biotechnol , (1994)3:70-77

- 38** Michiko Shirane, Shigetsugu Hatakeyama, Kimihiko Hattori, Keiko Nakayama, and Kei-ichi Nakayama. Common Pathway for the Ubiquitination of I κ B α , I κ B β , and I κ B ϵ Mediated by the F-Box Protein FWD1. *The Journal of Biological Chemistry* (1999) Vol. 274, No. 40, Issue of October 1, pp. 28169–28174,
- 39** Watson, J. D. & Crick, F. H. C. Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171, 964–967 (1953)
- 40** Oka H, Watanabe H. Colony-specificity in compound ascidians as tested by fusion experiments. *Proc Jpn Acad Sci* 1957; 33:657-664
- 41** Paramvir Dehal et al. The Draft Genome of *Ciona intestinalis*: Insights into Chordate and Vertebrate Origins. *Science* (2002)298:2157-2167
- 42** Richard K. Grosberg and Michael W. Hart. Mate Selection and the Evolution of Highly Polymorphic Self/Non-self Recognition Genes. *Science* (2000) 289:2111-2114.
- 43** Rinkevich B, Porat R, Goren M, Allrecognition elements on a urochordate histocompatibility locus indicate unprecedented extensive polymorphism. *Proc. Roy Soc Lond B* (1995)259:319-342
- 44** Rutherford, J, Parkhill, J, Crook, T, Horsnell, P, Rice, M-A, Rajandream and B. Barrell Artemis: sequence visualization and annotation." *Bioinformatics* (2000) 16 (10) 944-945.
- 45** S. Blair Hedges. The origin and evolution of model organisms. *Nature Reviews Genetics* (2002) 3:838-849
- 46** Sanger F, Nicklen S & Coulson A. R. Sanger, F., Nicklen, S., Coulson, A.R. ADN sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* (1977) 74, 5463-5467.
- 47** Satoh N, Jeffrey WR, Chasing tails in ascidians: developmental insights into the origin and evolution of chordates. *Trends Genet* (1995)11:354-359
- 48** Scofield, VL., Schlumpberger, J., West, LA & Weissman IL Protochordate allrecognition is controlled by a MHC-like gene system *Nature* (1982) 295, 499-502
- 49** Shigetsugu Hatakeyama et al. Ubiquitin-dependent degradation of I κ B α is mediated by a ubiquitin ligase Skp1 γ Cul 1 γ F-box protein FWD1. *Proc. Natl. Acad. Sci. USA* (1999) Vol. 96, pp. 3859–3863,
- 50** Simmen W, Martin and Bird Adrian. Sequence Analysis of Transposable Elements in the Sea Squirt *Ciona intestinalis*. (2000) *Mol.Biol.Evol* 17(11):1685-1694.

- 51** Simmen W Martin, Adrian Bird et al. Nonmethylated Transposable Elements and Methylated Genes in a Chordate Genome (1999) *Science* 283:1164-1167
- 52** Solovyev V.V. Statistical approaches in Eukaryotic gene prediction. In: *Handbook of Statistical genetics* (2001) John Wiley & Sons, Ltd., pp. 83-127.
- 53** Solovyev V.V. Finding genes by computer: Probabilistic and discriminative approaches. In: *Current Topics in Computational Biology* . (2002) MIT Press, p. 365-401.
- 54** Solovyev V.V. Structure, Properties and Computer Identification of Eukaryotic genes. In: *Bioinformatics from Genomes to Drugs. V.1. Basic Technologies.* (2002) pp.59 - 111.
- 55** Sovan Sarkar, Gopal Iyer, Jennifer Wu and N.Louise Glass. Nonspecific recognition is mediated by HET-C heterocomplex formation during vegetative incompatibility. *The EMBO Journal* (2002) Vol. 21 No. 18 pp. 4841-4850.
- 56** Stern, C., and E.R Sherwood. *The Origin of Genetics: A Mendel Source Book.* New York: W.H Freeman. A Short collection of important early papers, including Mendel's papers and correspondence. (1966) pp.300
- 57** Sophia L. Stone, Erin M. Anderson, Robert T. Mullen, and Daphne R. Goring. ARC1 Is an E3 Ubiquitin Ligase and Promotes the Ubiquitination of Proteins during the Rejection of Self-Incompatible Brassica Pollen . *The Plant Cell*, April 2003, Vol. 15, 885–898,
- 58** Sven J. Saupe. Molecular Genetics of Heterokaryon Incompatibility in Filamentous Ascomycetes. *Microbiology and Molecular Biology Reviews*, Sept. 2000, Vol 64 , No3. 489-502
- 59** Swerdlow H, Wu S, Harke H, Dovichi NJ. Capillary gel electrophoresis for ADN sequencing: laser-induced fluorescence detection with the sheath flow cuvette. *J Chromatogr* 1990; 516:61.
- 60** Mitaku Takatsugu Hirokawa, Seah Boon-Chieng and Shigeki, SOSUI: Classification and Secondary Structure Prediction System for Membrane Proteins, *Bioinformatics (formerly CABIOS)* (1998) May;14(4):378-379.
- 61** Tonegawa S Hozumi N. Evidence for somatic rearrangement of immunoglobulin genes for variable and constant regions. *Proc Nat Acad Sci USA* (1976) Oct 73(10)

3268-32.

62 Vos P et al. AFLP: a new technique for ADN fingerprinting. *Nucleic Acid Research* (1995) 23:4407-4414.

63 Watson, J.D. / Crick, F.H.C. A structure for deoxyribose nucleic acid *Nature* 171, 737-738 (1953).

64 Weissman IL Saito Y, Rinkevich B. Allorecognition histocompatibility in a protochordate species: is the relationship to MHC semantic or structural? *Immunol Rev* (1990) 113:227-241

65 Yund PO, Feldgarden M. Rapid proliferation of historecognition alleles in populations of a colonial ascidian. *J. Exp Zool* (1992) 264 : 442-452

FIGURAS

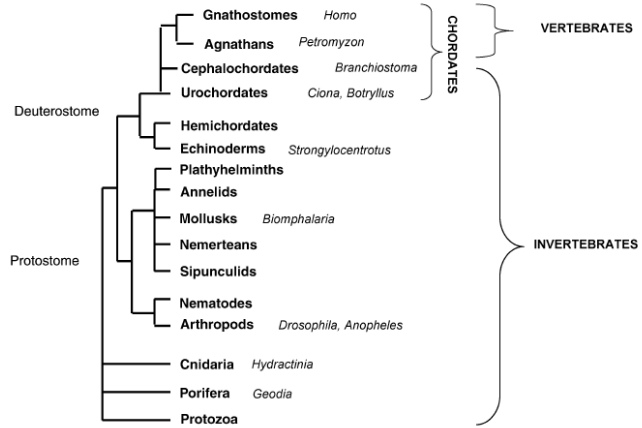


Figura 1. Clasificación filogenética de los metazoanos.

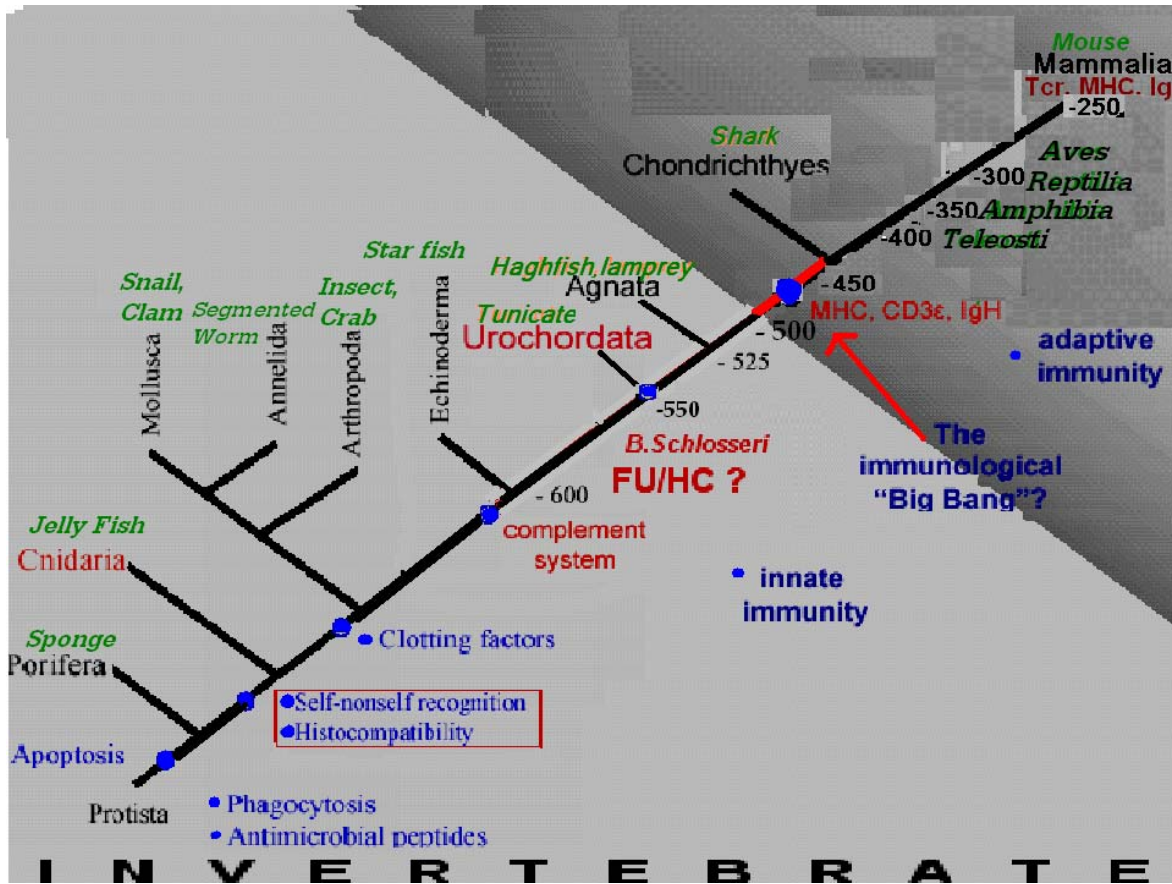


Figura 2. Evolución del sistema inmune. Los protocordados o tunicados como *B.Schlosseri* y *C.intestinalis* son parte del *subphylum* de los Urocordados, los cuales están directamente relacionados con el origen de los cordados hace 550 millones de años. Durante la evolución de los invertebrados a los vertebrados, varios mecanismos de defensa fueron acumulándose (puntos azules). La frontera de 50 millones de años entre los organismos que poseen el sistema de inmunidad innata y los que tienen inmunidad adaptativa se le llama "Big Bang" inmunológico. Aquí, Fu/HC puede jugar un papel importante como un locus de histocompatibilidad cercano a la emergencia del sistema inmuno-adaptativo.

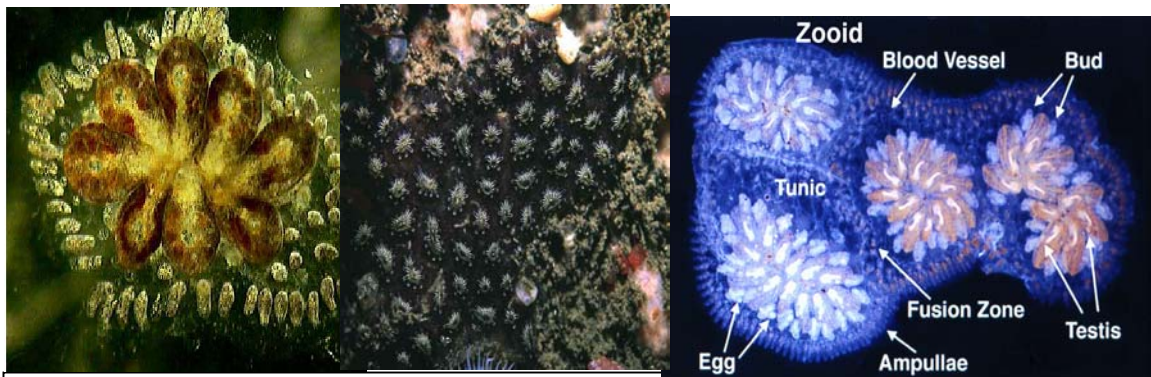


Figura 3. Anatomía de *Botryllus schlosseri*. Especimen individual (izquierda). Colonia de *B.schlosseri* conespecíficos (centro). Esquema anatómico de un zooid de *B.schlosseri*. (derecha).

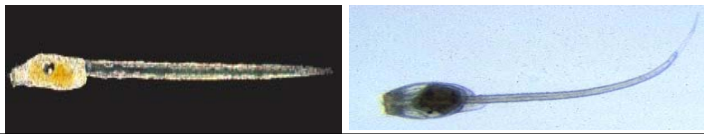


Figura 4. Características anatómicas de las larvas de los protocordados. Los urocordados *C.intestinalis* (izquierda) y *B.schlosseri* (derecha). Ambas larvas se caracterizan por tener el evento rasgo particular en la evolución de los cordados

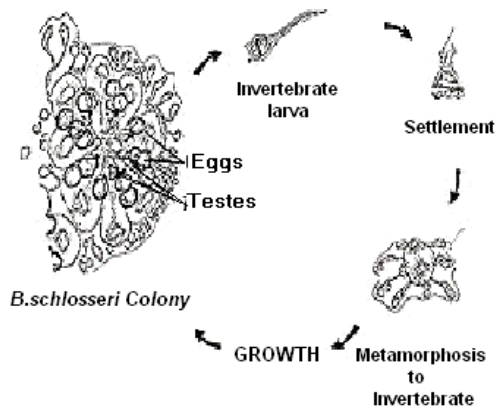


Figura 5. El ciclo de vida de *B.schlosseri*.

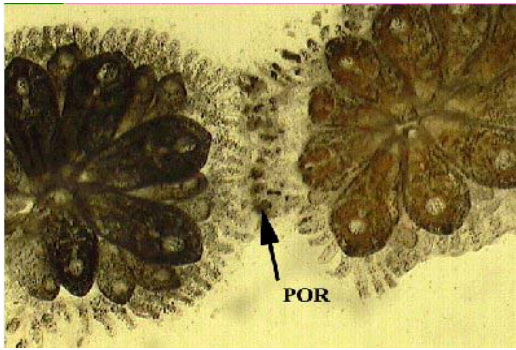


Figura 6. El aloreconocimiento i.e. Reconocimiento propio/no propio. El reconocimiento no propio originado por una incompatibilidad en los hapotipos Fu/Hc se caracteriza por un rechazo vascular entre colonias de *B.schlosseri*

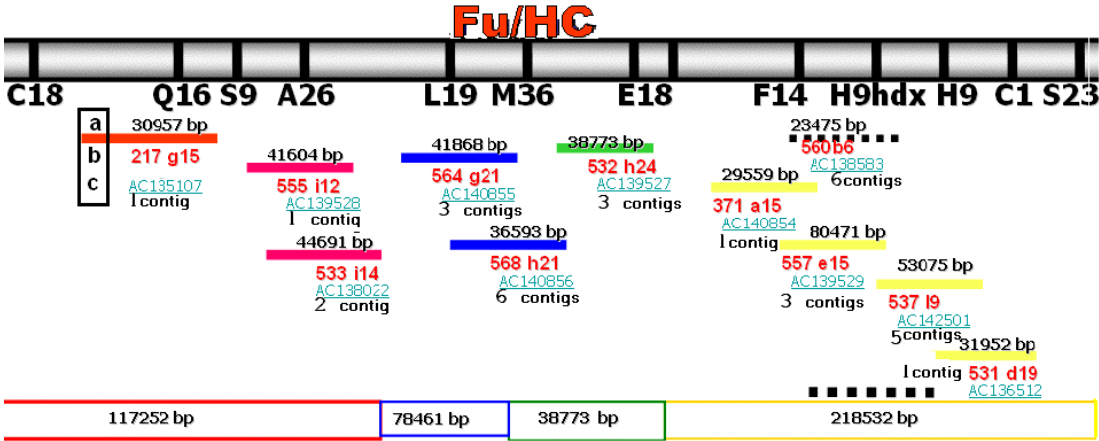


Figura 7. El Fu/Hc, locus de Fusión Histocompatibilidad delineado a aproximadamente 1 -cm. Once fósidos secuenciados. a) NÚMEROS DE ACCESO EN NCBI, b) Nombre del de fósido, c) Longitud del fósido en bp (pares de bases), MAYÚSCULAS, MARCADORES GENÉTICOS Y Minúsculas, número de contigs

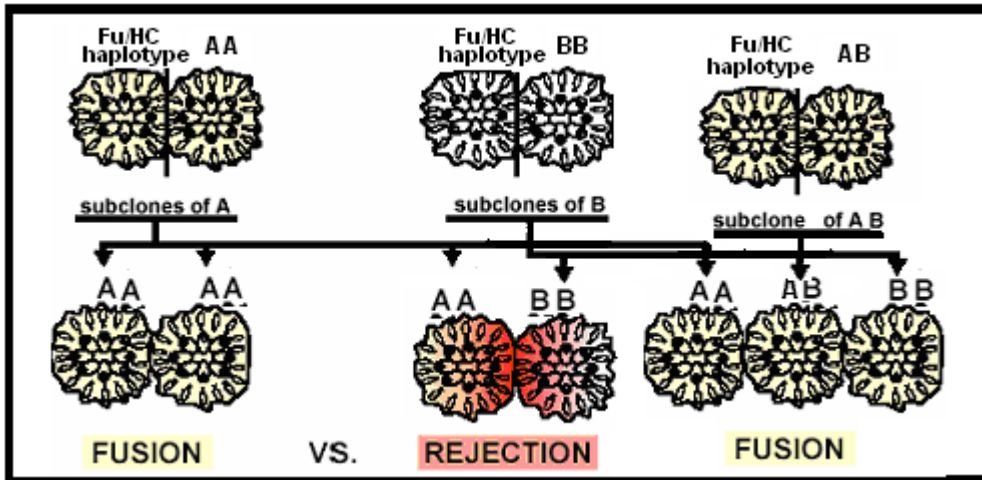


Figura 8. Los fenotipos de fusión y rechazo activados por Fu/HC. La *Fusion* es activada cuando las ámpulas que interactúan entre las colonias de zoooides de *B. schlosseri* comparten al menos un alelo. El Rechazo es activado cuando las colonias no comparten alelo alguno.

Shotgun Sequencing Method

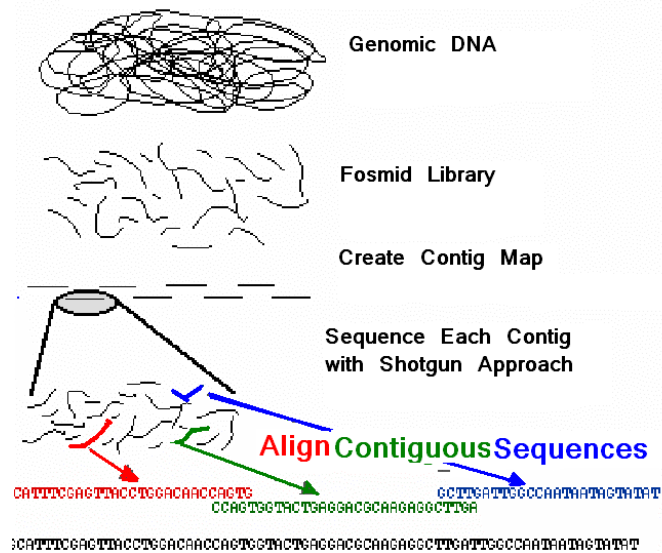


Figura.9. Resumen del Método de Secuenciación por Shot Gun.

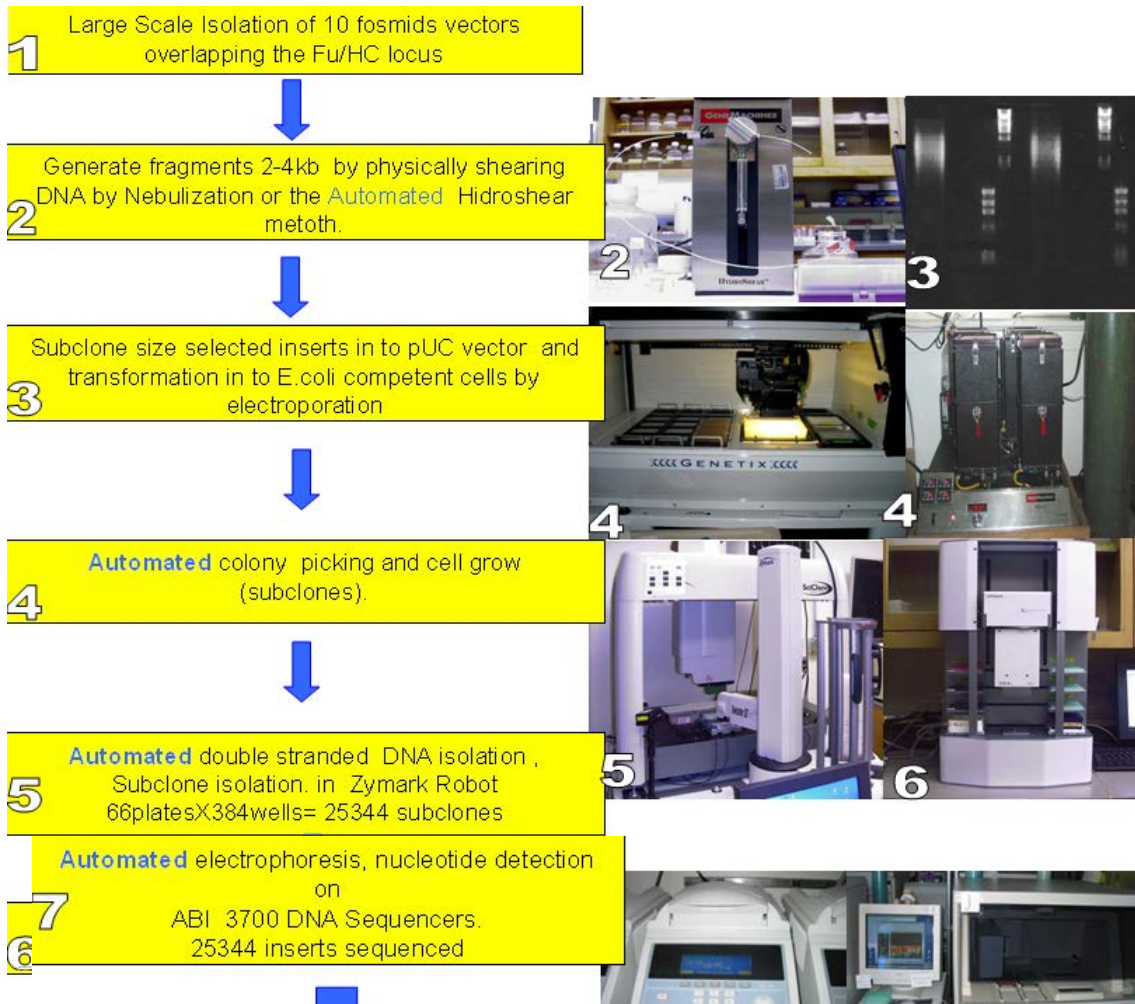


Figura 10. Resumen detallado del método de secuenciación. Descripción de los pasos 1-6 de la estrategia de secuenciación usada para secuenciar los 11 fósmidos que traslapan parte del Fu/HC

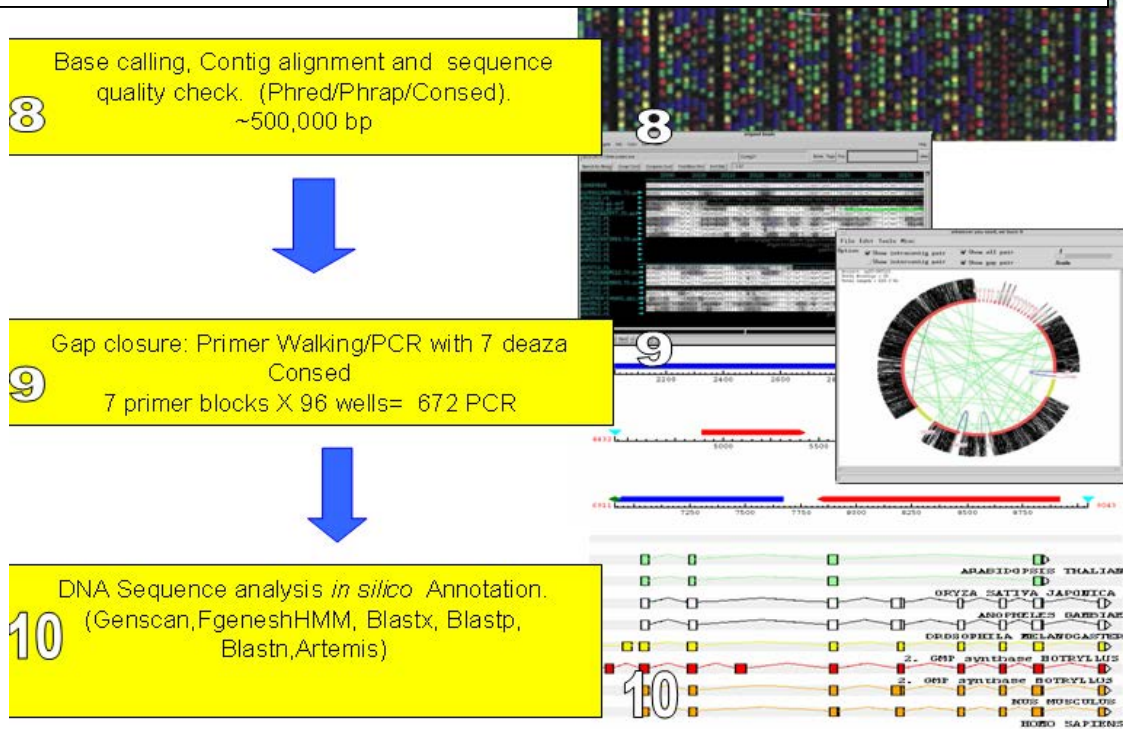


Figura 11. Resumen detallado del método de secuenciación. Descripción de los pasos 7-10 de la estrategia de secuenciación usada para secuenciar los 11 fósidos que traslapan el Fu/HC.

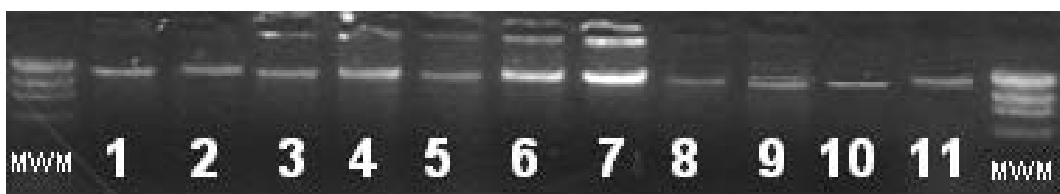


Figura 12. Muestras de ADN extraído de cada uno de los once fósidos. **1.** subsf-217g15, **2.** subsf-533i14, **3.** subsf-555i12, **4.** subsf-564g21, **5.** subsf-568h21, **6.** subsf-532h24, **7.** subsf-560b6, **8.** subsf-371a15, **9.** subsf-557e15, **10.** subsf- 537l9 y **11.** subsf-531d19. MWM(Marcadores de peso molecular lambda hind III y HaeI).

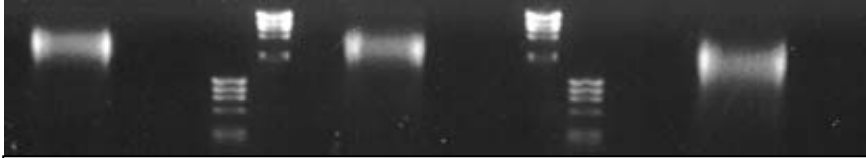


Figura 13. Shot gun de ADN hacienda uso de dos métodos de fragmentación aleatoria; a) HIDROSHEAR en las dos primeras bandas (de derecha a izquierda) b) NEBULIZACION en la última banda. MWM, (Marcador de peso molecular lambda hind III y HaeI).

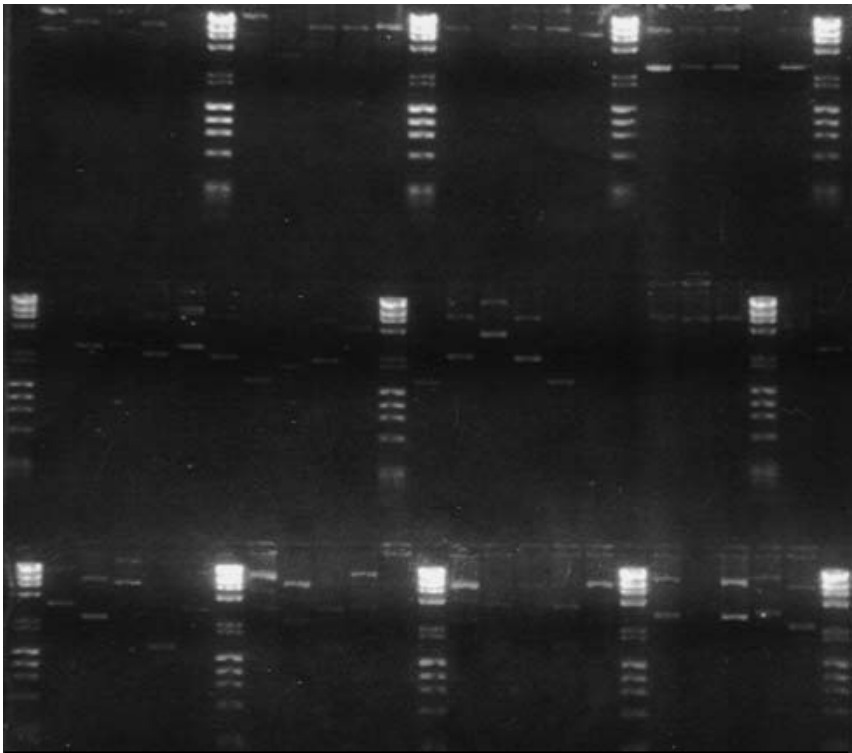


Figura.14. Muestras de extracciones de ADN a partir de Subclonas.

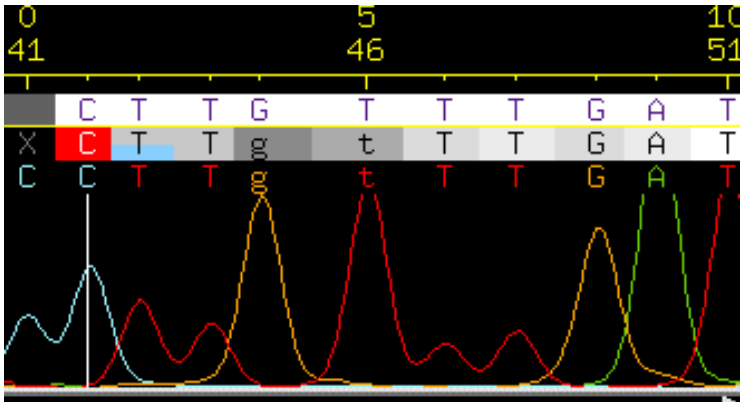


Figura 15. Resultados de Phred visualizados en Consed.

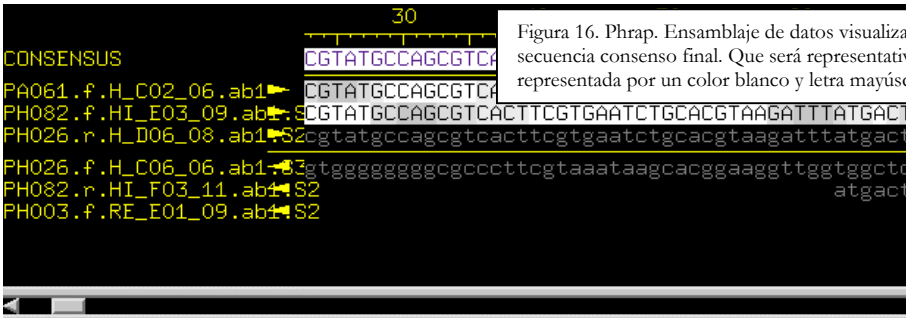
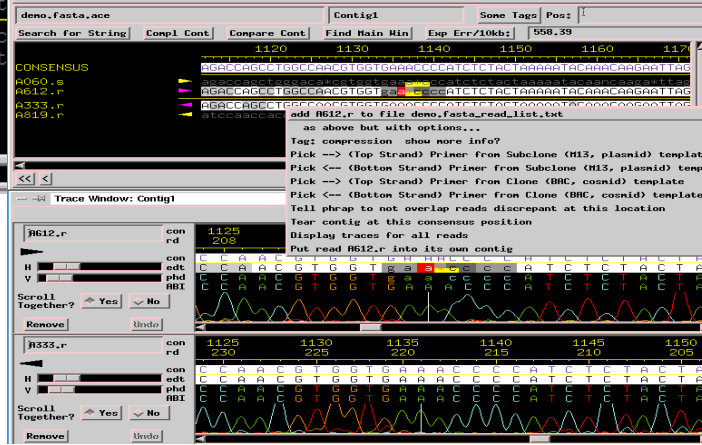


Figura 16. Phrap. Ensamblaje de datos visualizados en Consed. Cada lectura es traslapada y alineada para dar como resultado una secuencia consenso final. Que será representativa de la secuencia real. Una secuencia consenso confiable, de buena calidad, estará representada por un color blanco y letra mayúsculas.



Phred/Phrap.

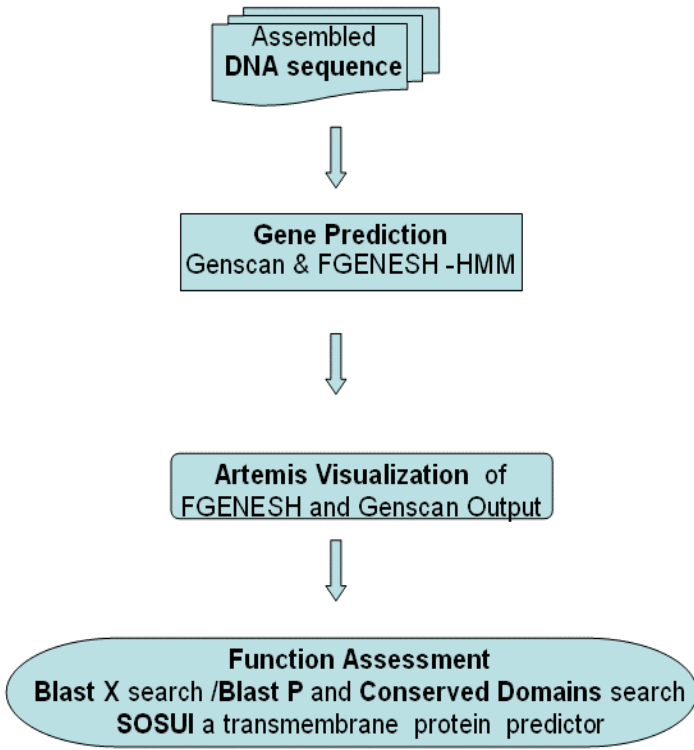


Figura 18. Resumen para el proceso de análisis y la usado para analizar los doce fósidos que traslapan parte del locus Fu/HC.

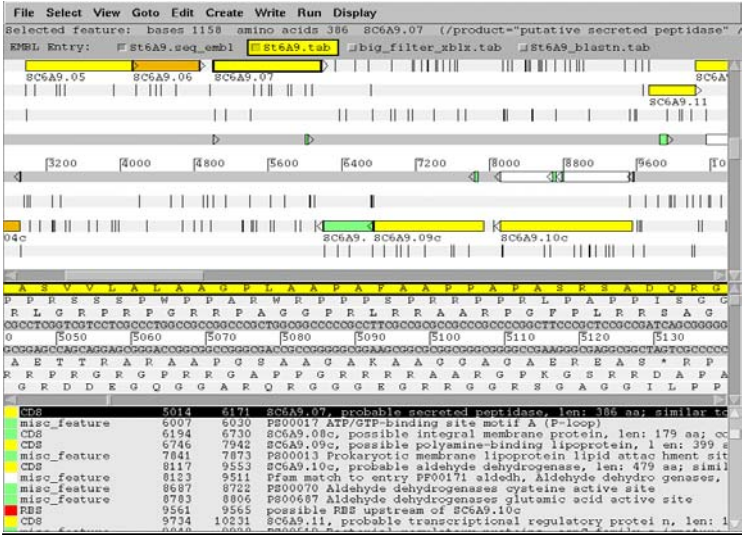


Fig 19. Resultados de Blast P visualizados en Artemis.

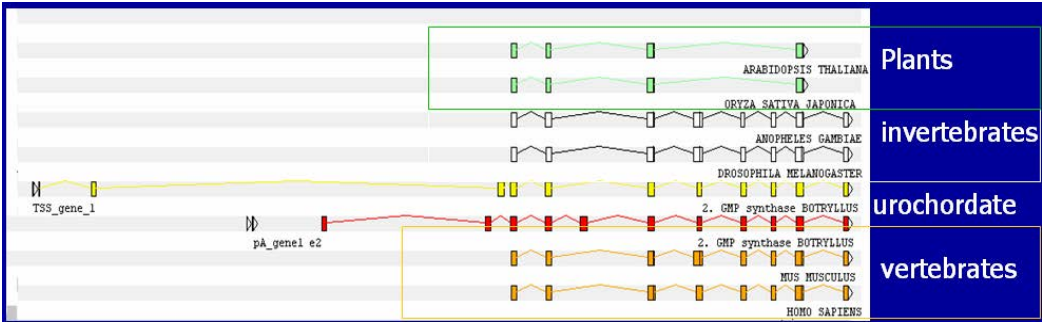


Figura 20. Predicción de Genes y estructuras de exones visualizadas en Artemis. La predicción del gene GMP syntetasa en el urocordado *B.schlosseri*, fue analizada por GENESCAN (rojo) y FGENESH/HMM (amarillo). La predicción y estructura exónica es consistente con los reultados de Blast P , representados en artemis. La estructura del gene en diferentes phylum; plantas , invertebrados, vertebrados y el gene GMP syntetasa del urocordado, tienen una relación estructural parecida.

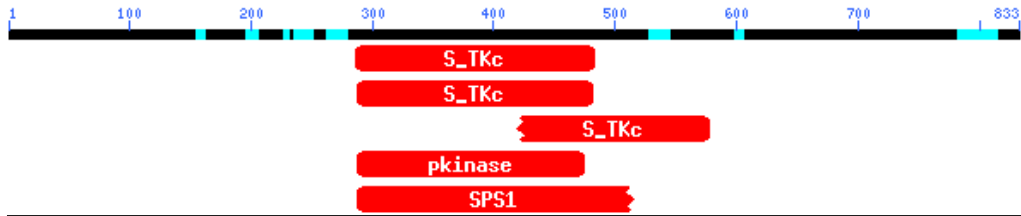


Figura 21. Un dominio conservado para el gene de la Serin Treonin Kinasa. Resultados de Rps-Blast. Los dominios conservados fueron buscados en la base de datos de NCBI.

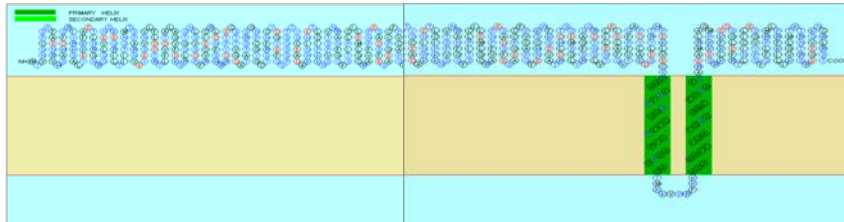


Figura 22. El programa SOSUI es una herramienta capaz de predecir estructuras secundarias y proteínas transmembranales. Aquí se representa una proteína con una breve homología a la proteína del sistema del complemento Bf2/c2.

Figura. 23. Resultados de la Secuenciación y anotación de los once fósidos que traslapan el locus Fu/HC.

