

UCH-FC  
Biotecnología  
F363  
C.1



UNIVERSIDAD DE CHILE - FACULTAD DE CIENCIAS - ESCUELA DE PREGRADO

**“Descubrimiento de polimorfismos de un sólo nucleótido (SNPs)  
para *Salmo salar* en genes asociados a vías de señalización de  
sistema inmune”**

Seminario de Título entregado a la Universidad de Chile en cumplimiento parcial de los requisitos para optar al Título de Ingeniero en Biotecnología Molecular.

GONZALO FERNÁNDEZ SMITS

Dr. Víctor Martínez  
Director del Seminario de Título  
Patrocinante: Dr. Víctor Cifuentes

Julio de 2012  
Santiago - Chile

ESCUELA DE PREGRADO – FACULTAD DE CIENCIAS – UNIVERSIDAD DE CHILE



**INFORME DE APROBACIÓN SEMINARIO DE TÍTULO**

Se informa a la Escuela de Pregrado de la Facultad de Ciencias, de la Universidad de Chile que el Seminario de Título, presentado por el Sr. **Gonzalo Gabriel Fernández Smits**.

**“Descubrimiento de polimorfismos de un sólo nucleótido (SNPs) para *Salmo salar* en genes asociados a vías de señalización en sistema inmune”**

Ha sido aprobado por la Comisión de Evaluación, en cumplimiento parcial de los requisitos para optar al Título de Ingeniero en Biotecnología Molecular.

*Dr. Víctor Martínez Moncada*  
**Director Seminario de Título**

*Dr. Víctor Cifuentes*  
**Profesor patrocinante**

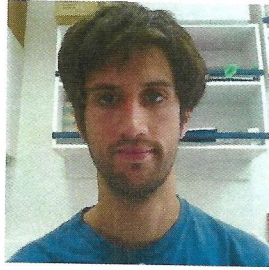
**Comisión de Evaluación:**

*Dr. Marcelo Baeza*  
**Presidente Comisión**

*Dra. Margarita Carú*  
**Evaluadora**



Santiago de Chile, julio de 2012



Gonzalo Gabriel Fernández Smits realizó todos sus estudios primarios y secundarios en el colegio The Grange School, desde el año 1991 hasta 2004. Luego ingresó al programa académico de Bachillerato en la Universidad de Chile, completándolo entre los años 2005 a 2006. Desde el programa de Bachillerato fue transferido a la carrera de Ingeniería en Biotecnología Molecular en la Facultad de Ciencias de la Universidad de Chile el año 2007. Realizó el año 2007 una pasantía en el laboratorio de Bioquímica con la Dra. Ana Preller. Por otro lado, el año 2009 completó una unidad de investigación con su tutor de seminario de título Dr. Víctor Martínez en la “Identificación y cuantificación del dinoflagelado *Alexandrium catenella* en muestras de *Mytilus edulis chilensis*” trabajo que se encuentra actualmente en vías de publicación. Durante el año 2011 completó un diplomado en la Facultad de Economía y Negocios de la Universidad de Chile en “Preparación y Evaluación de Proyectos” con el profesor Reinaldo Sapag.

## AGRADECIMIENTOS

Primero, quisiera agradecer el apoyo dado en el laboratorio INBIOGEN en la Facultad de Ciencias Veterinarias y Pecuarias de la Universidad. Especialmente a mi tutor, el Dr. Víctor Martínez, por su constante preocupación y apoyo para la realización del seminario de título. También a Ana Jedlicki, por sus consejos y guías, tanto en el ámbito personal como profesional.

A mi familia, por estar siempre presente y por el incondicional apoyo que brindaron.

A Paola, por la paciencia que ha tenido y la compañía y aliento brindado durante estos tres años.

Esta tesis fue realizada gracias al financiamiento del proyecto FONDECYT 1090632.

# INDICE DE CONTENIDOS

## Contenido

INTRODUCCION .....	1
1.1 Antecedentes .....	1
1.2 Salmónidos .....	2
1.2.1 Efectos de una duplicación genómica reciente .....	3
1.2.2 Antecedentes de genes ligados a respuesta inmunológica en salmónidos .....	4
1.3 Polimorfismos .....	6
1.3.1 Tecnologías de genotipado de SNPs .....	6
1.4 Búsqueda y análisis in silico de SNPs .....	7
1.4.1 Recolección de información: NCBI, UniGene .....	8
1.4.2 Algoritmos disponibles para el descubrimiento y corrección de SNPs <i>in silico</i> .....	8
1.4.3 Dificultad que presentan secuencias parálogas en el DNA de salmónidos .....	9
1.5 Validación y genotipado de SNPs mediante denaturación de curvas en alta resolución (HRM) .....	11
2 .....	13
HIPOTESIS Y OBJETIVOS .....	13
METODOLOGIA .....	14
3.1 Muestreo de individuos .....	14
3.2 Extracción de DNA .....	14
3.3 Descubrimiento de polimorfismos de un solo nucleótido <i>in silico</i> .....	14
3.4 Diseño y síntesis de partidores .....	15
3.5 Optimización de temperatura de hibridación de partidores para el PCR .....	15
3.6 Curvas de alta resolución .....	16
3.7 Secuenciación .....	17
3.8 Anotación de polimorfismos .....	17
RESULTADOS .....	18
3.9 Descubrimiento de SNPs .....	18
3.10 Optimización de temperatura de hibridación y corroboración del largo de amplicón .....	21
3.11 Validación de SNPs utilizando curvas de alta resolución .....	24
3.12 Visualización de genotipos mediante curvas de diferencias .....	27
3.13 Secuenciación de los polimorfismos validados .....	29
3.14 Anotación de SNPs <i>in silico</i> en genes encontrados .....	30
DISCUSION Y PROYECCIONES .....	31
CONCLUSION .....	37
BIBLIOGRAFIA .....	38
ANEXOS .....	41

# INDICE DE CONTENIDOS

## Contenido

INTRODUCCION .....	1
1.1 Antecedentes .....	1
1.2 Salmónidos .....	2
1.2.1 Efectos de una duplicación genómica reciente .....	3
1.2.2 Antecedentes de genes ligados a respuesta inmunológica en salmónidos. ..	4
1.3 Polimorfismos .....	6
1.3.1 Tecnologías de genotipado de SNPs .....	6
1.4 Búsqueda y análisis in silico de SNPs.....	7
1.4.1 Recolección de información: NCBI, UniGene. ....	8
1.4.2 Algoritmos disponibles para el descubrimiento y corrección de SNPs <i>in silico</i> . 8	
1.4.3 Dificultad que presentan secuencias parálogas en el DNA de salmónidos. .9	
1.5 Validación y genotipado de SNPs mediante denaturación de curvas en alta resolución (HRM). ....	11
2	13
HIPOTESIS Y OBJETIVOS.....	13
METODOLOGIA .....	14
3.1 Muestreo de individuos. ....	14
3.2 Extracción de DNA. ....	14
3.3 Descubrimiento de polimorfismos de un solo nucleótido in silico. ....	14
3.4 Diseño y síntesis de partidores. ....	15
3.5 Optimización de temperatura de hibridación de partidores para el PCR. ....	15
3.6 Curvas de alta resolución. ....	16
3.7 Secuenciación. ....	17
3.8 Anotación de polimorfismos. ....	17
RESULTADOS .....	18
3.9 Descubrimiento de SNPs.....	18
3.10 Optimización de temperatura de hibridación y corroboración del largo de amplicón.....	21
3.11 Validación de SNPs utilizando curvas de alta resolución.....	24
3.12 Visualización de genotipos mediante curvas de diferencias.....	27
3.13 Secuenciación de los polimorfismos validados. ....	29
3.14 Anotación de SNPs in silico en genes encontrados. ....	30
DISCUSION Y PROYECCIONES.....	31
CONCLUSION .....	37
BIBLIOGRAFIA .....	38
ANEXOS .....	41

## INDICE DE FIGURAS

<b>Figura 1.</b> Visualización de polimorfismos y electroferogramas en Consed.....	20
<b>Figura 2.</b> Gel de agarosa al 2% de PCR de agradiante de temperatura de partidor CTL 22	
<b>Figura 3.</b> Curvas de alta resolución (HRM) de 4 genes polimórficos y uno no polimórfico.....	25
<b>Figura 4.</b> Curvas de diferencias de los resultados contruidas a partir de curvas de altas resolución.....	28
<b>Figura 5.</b> Secuenciación del amplicón del polimorfismo $\beta$ 2- Microglobulina para los tres genotipos encontrado.....	30

## INDICE DE TABLAS

Tabla 1. Clasificación de SNPs por medio de la mutación nucleotídica generada. ....	11
Tabla 2. Resumen de partidores utilizados para el análisis de curvas en alta resolución	22
Tabla 3. Resumen del descubrimiento de SNPs <i>in silico</i> y posterior validación experimental .....	21
Tabla 4. Asignación de genotipos de los individuos muestreados por medio del análisis de curvas de alta resolución .....	29
Tabla 5. Anotación de SNPs encontrados usando información de la base de datos de NCBI .....	31



## LISTA DE ABREVIATURAS

**HRM:** High resolution melt o denaturación de curvas en alta resolución

**SNP:** Single nucleotide polymorphism o polimorfismo de nucleótido simple.

**PCR:** Polymerase chain reaction o reacción de la polimerasa en cadena.

**UTR:** Untranslated región o región no traducida.

**EST:** Expressed sequence tags o marcadores de secuencia expresada.

**IPN:** Infectious pancreatic disease o necrosis pancreática infecciosa.

**ISA:** Infectious salmon anemia o anemia infecciosa del salmón.

## ABSTRACT

In recent years, the production of *Salmo salar* or Atlantic Salmon in Chile has undergone a considerable expansion. Due to this, the implementation of research programs aimed at the innovation for developing competitive advantages has been regarded as an important role in Chilean governmental policies. Thus, the following work focuses on the discovery of novel single nucleotide polymorphisms in immune system signalization associated genes, for the generation of new genetic markers for the development of superior breeding lines which may show advantage in terms of disease resistance to local producers.

Numerous techniques have been established for the genotyping of single nucleotide polymorphisms or SNPs. High resolution melting curve analysis (HRMA) has proven to be a precise, low-cost and medium throughput technology, which fits all our requirements. The method was tested on a previously known polymorphism (SNP356), which was previously genotyped in our laboratory by microarray analysis. Results demonstrated full concordance, proving HRM reliability.

HRM was used to genotype 6 other polymorphisms from which we had no previous information. These polymorphisms were discovered by *in silico* analysis of EST sequences extracted from cGrasp database. Three out of six SNPs were polymorphic, giving a 50% success rate.

Sequences containing the polymorphisms were aligned using the blast algorithm against the NCBI database. The highest scoring genes were used to individualize the three SNPs discovered. Two of them, were in the 3'UTR region, and the other was in the

coding sequence (CDS) of the mRNA. As point mutations in the untranslated region do not tend to generate significant changes in protein structure or function, we suppose these as silent mutations. Nonetheless, as these genes have proven to be polymorphic, they could be very valuable mRNA associated molecular markers. On the other hand, TAP1 mutation was found to be in the CDS region, delivering a non-synonymous mutation from a glycine to a serine. One possibility is that these mutations could lead to the generation of susceptible, tolerant or resistant phenotypes.

Further studies are needed to assess the presence of these newly discovered polymorphisms in infected salmon populations, as they could generate novel genetic lines with resistant infectious disease phenotypes.

## RESUMEN

La salmonicultura durante la década del 90, ha tenido un auge en exportaciones y producción total. Es por ello que se ha fijado como área prioritaria en investigación científica, aquellos proyectos que fomenten aumentos productivos de *Salmo salar*. En este contexto los marcadores moleculares, microsatélites o polimorfismos de un solo nucleótido (SNP), han sido descritos como esenciales para la generación de nuevas líneas genéticas. Es por ello que mediante la identificación, validación, genotipificación y anotación de polimorfismos tipo SNPs, en genes asociados a vías de señalización en el sistema inmune en *Salmo salar*, se pretende dar un paso hacia la identificación de marcadores moleculares que ayuden a descubrir y reproducir líneas de selección que otorguen ventajas competitivas al producto nacional.

Numerosas técnicas de genotipado de polimorfismos simples han sido desarrolladas. Para este trabajo y basado en la extensión del análisis, se optó por una técnica económica, precisa y de rendimiento medio llamada “denaturación de curvas en alta resolución”. Esta se basa en la detección de los polimorfismos mediante la disminución de fluorescencia de un agente intercalante de ADN en el proceso de disociación de la doble hebra, siendo la señal fluorescente del fluoróforo más potente cuando se encuentra unida a la doble hebra de ADN.

Los ensayos de genotipado por esta técnica se realizaron en muestras de 4 poblaciones comerciales y 1 naturalizada. Estas poblaciones permitieron encontrar polimorfismos de más de un 7% de frecuencia. Se escogieron los genes a analizar examinando trabajos en los que identificaran cambios de expresión de mRNAs bajo

infecciones. Es así como se identificaron 8 genes putativos en los cuales buscar polimorfismos: Tiorredoxina, TAP1, TAP2,  $\beta$ 2-Microglobulina, PKR, IRF-1, C-type Lectin y Mx1. De un total de 6 SNPs descubiertos *in silico*, 3 fueron polimórficos. Esto no incluye un polimorfismo usado para estandarizar y optimizar la técnica (SNP356), cuyo genotipo para las muestras ya era conocido. Nuestros resultados finalmente demostraron un 50% de éxito entre SNPs descubiertos *in silico* y su validación.

Por otro lado, de los 4 polimorfismos analizados, solo el SNP356 no tuvo puntajes significativos al hacer un blast de las secuencias obtenidas *in silico* contra la base de datos de NCBI. En tanto, para los SNPs TAP1, Tiorredoxina y  $2\beta$ -Microglobulina hubo puntajes significativos, encontrándose secuencias responsables de su transcripción. Para los polimorfismos ligados a estos tres genes, se vio que dos están en regiones 3'UTR, mientras que para TAP1, su mutación se encuentra en la región funcional de la proteína, conllevando un cambio desde glicina a serina, cambio que trae modificaciones a la proteína generada.

Los polimorfismos obtenidos podrían ser analizados en poblaciones desafiadas a diferentes patógenos, con el objeto de ver si se corresponden con algún grado de resistencia, tolerancia o susceptibilidad a enfermedades infecciosas. De esta manera se podrían generar líneas de selección que aumenten la competitividad del salmón chileno.

# INTRODUCCION

## 1.1 Antecedentes

Durante la década del 90, la producción y exportación de salmonídeos prosperó notablemente. Es así como durante el año 2008, la industria del salmón generó ingresos por 2.392 millones de dólares (Salmon Chile: <http://estadisticas.intesal.cl/>; N° 3), representando 445 miles de toneladas de producto neto (Salmon Chile: <http://estadisticas.intesal.cl/>; N° 4). Desde el año 1998 hasta el 2008, la producción neta ha aumentado un 244%, convirtiendo a esta industria en una de las de más fuerte crecimiento del país. Además del progreso económico, esta industria ha contribuido a la generación de empleo y al progreso social en la zona centro sur y sur del país, especialmente en las regiones de Los Lagos y Aysén.

En los últimos años, la salmonicultura se ha visto seriamente afectada por enfermedades de tipo infeccioso. La más importante involucra una bacteria intracelular llamada *Piscirickettsia salmonis* que produce la *Piscirickettsiosis* o Síndrome Rickettsial del Salmón. Durante el año 2007, se observó que en promedio, un 57% de la mortalidad mensual en peces se debió a infecciones por *Piscirickettsiosis* (Leal y Woywood 2007). Por otro lado, el virus de la anemia infecciosa (ISA), durante un brote masivo, generó una reducción de un 48% de la producción de *Salmo salar* durante el año 2009 (Fitchratings: [www.fitchratings.cl](http://www.fitchratings.cl)). Es por ello, que buscando contener ambas infecciones, se han desarrollado vacunas y antibióticos, siendo estos suministrados individual (vacunas) o colectivamente (disueltos en estanques) (Olsen y col. 1997; Wilhelm y col. 2006). A pesar de generar una disminución en la mortalidad, estos

métodos provocan un número de problemas secundarios. En este contexto, las vacunas han demostrado tener una limitada efectividad, ya que el desarrollo e implementación de estas se encuentra restringida debido al escaso conocimiento existente respecto del sistema inmune de los peces (Míquel y col. 2003; Almendras y Fuentealba 1997) y al alto costo que representa la aplicación individual de cada inyección. Por otra parte, serias limitaciones medioambientales existen al suministro de antibióticos, ya que ellos provocan problemas a la microbiota endémica y al ecosistema. Además, las normas internacionales son exigentes respecto de la presencia de antibióticos en los alimentos.

Dado los pobres resultados obtenidos a partir del uso de vacunas y antibióticos (Burridge y col 2008; salmonchile), se han planteado otras estrategias para combatir la mortalidad en peces. Una muy atractiva es la selección fenotípica. Aplicando esta estrategia, empresas productoras de salmonídeos han implementado programas de mejoramiento que involucran la generación de líneas o familias de peces portadoras de determinadas características fenotípicas, elegidas por su capacidad de generar algún tipo de plusvalía comercial. Sin embargo, hoy en día el avance de la ciencia en el campo genético y de la biología molecular ha abierto un nuevo y amplio espectro en la producción animal. En este sentido, la disponibilidad de prácticamente la totalidad del genoma del Salmón del Atlántico, permite identificar polimorfismos en este genoma. Esto posibilita efectuar un análisis más acabado de resistencias a patógenos, estudiando las regiones del genoma involucradas con estos caracteres complejos.

## 1.2 Salmónidos

El Salmón del Atlántico es una especie de la familia *salmonidae*, del género *salmo*.

Dentro de la familia existen otras especies comercializadas, como es la trucha arcoíris (*Onchorhynchus mykiss*) y el Salmón coho (*Onchorhynchus kisutch*). Una característica peculiar de todos ellos es la presencia de un ancestro tetraploide en la familia *salmonidae*, producto de una duplicación genómica completa, lo que se puede deducir del tamaño del genoma y los brazos cromosómicos. En este contexto, la duplicación citada permite la evolución diferencial de genes originales y duplicados por las distintas presiones selectivas a las que puede estar sometido cada gen dependiendo de su posición y funcionalidad. En este sentido se ha propuesto que la rápida adaptación mostrada en algunos individuos de la familia *salmonidae* sea gracias a variaciones que pueden ocurrir en el genoma debido a este evento de duplicación (Taylor y col. 2001).

#### 1.2.1 Efectos de una duplicación genómica reciente

A raíz de eventos de duplicación genómico completa, Qumsiyeh (1994) planteó dos posibles escenarios. El primero es la mantención de la dotación cromosómica posibilitando eventos de recombinación más frecuentes, generando variación, permitiendo la colonización de nuevos hábitats con mayor eficiencia. El segundo es la disminución de la dotación cromosómica reduciendo los eventos de recombinación, pero seleccionando genes que otorguen alguna ventaja adaptativa al medio. Esto se traduce en una mejor adaptación al ambiente, sacrificando la posibilidad de colonizar nuevos nichos. La mayor parte de los miembros del género *Oncorhynchus* han mantenido su dotación cromosómica, conservando la capacidad de adaptarse a hábitats nuevos. El género *Salmo*, en tanto, ha reducido su dotación cromosómica (de 80 a 54-56), adaptándose a condiciones medio ambientales específicas, limitando la colonización de



nuevos nichos ambientales.

Por otro lado, se ha descrito que pueden suceder tres procesos génicos cuando ocurre una duplicación genética. El primero, y más común, es el silenciamiento de genes duplicados, esto implica la pérdida de funcionalidad génica (no-funcionalización) por mutaciones ocurridas en regiones relevantes como promotoras, de procesamiento y/o señalización. Segundo, que una de las copias adquiriera una mutación que produzca una ventaja por sobre el original, remplazándola con el tiempo (neo-funcionalización). Tercero, que ambos genes muten generando un producto que en su conjunto genera sinergia, aumentando la eficiencia que produce cada uno por separado (sub-funcionalización). A pesar de que el silenciamiento es lo más común, la segunda y tercera opción son las que generan nuevas líneas de individuos, logrando así la especiación o una ventaja selectiva a alguna especie (Lynch y Conery 2000).

#### 1.2.2 Antecedentes de genes ligados a respuesta inmunológica en salmónidos.

La aplicación de chips comerciales en salmónidos que permiten la cuantificación de expresión génica ha generado una gran cantidad de información disponible en las bases de datos. Rise y col. (2004), Jørgensen y col. (2008) y Young y col. (2007) hicieron uso de tres microarreglos para medir cambios de expresión en *Salmo salar* bajo condiciones de infección por *Piscirickettsia salmonis*, virus ISA y *Neoparamoeba perurans* respectivamente. Rise y col. (2004) investigaron genes que variaran su expresión en macrófagos y células de riñón anterior 14 días después de una inyección peritoneal con el patógeno *P. salmonis*. Por otro lado Jørgensen y col. (2008), a través de un microarreglo compuesto de 1800 cDNAs, analizaron cambios de expresión en peces

muertos entre los días 21-25, 28-31 y 37-48 pos infección con virus ISA. Young y col. (2007) describen una regulación de 99 genes inmunológicos de *Salmo salar* por parte de *Neoparamoeba perurans* en el área de infección, estos genes se encuentran ligados al procesamiento y presentación de antígenos. La información obtenida de estos trabajos permitió acotar los genes en los cuales se debe buscar polimorfismos para encontrar individuos resistentes o susceptibles a ambas enfermedades.

Larsen (2004), por otro lado, describe el perfil de expresión *in vitro* del gen Mx1 en cultivos celulares de Salmon Chinook (*Oncorhynchus tshawytscha*) infectados con el virus de la necrosis pancreática infecciosa (IPNV). En su trabajo demuestra que una sobreexpresión del gen Mx1 en células CHSE-214 aumenta la tasa de sobrevivencia celular desde un 10%, a aproximadamente un 99%, efecto similar al que produce una inducción de respuesta por interferones. Por último, describe que los clones AsMx1, que sobre expresan Mx1, son capaces de inhibir la síntesis proteica del IPNV, demostrando así una relación entre la traducción del gen Mx1 y la protección a IPNV.

La proteína quinasa activada por RNA (PKR) se describe como un modulador importante en la respuesta celular ante infecciones virales de la familia orthomyxovirus, a la cual pertenece el virus ISA. Por otra parte, está descrito el efecto positivo de la expresión de PKR en vías de señalización y respuesta celular inmunológica frente a agentes patógenos virales en organismos eucariontes (García y col. 2006). A raíz de estas investigaciones se espera que la sobre o sub expresión de esta proteína pueda otorgar resistencia o susceptibilidad a individuos expuestos a agentes virales de la familia orthomyxovirus, modulando la respuesta inmune.

### 1.3 Polimorfismos

Los polimorfismos genéticos, de cualquier tipo, permiten la diferenciación a nivel genómico. Esto ocurre ya que no todos los individuos, incluso de una misma familia, tienen la misma secuencia nucleotídica, sino que varía de individuo a individuo. Estas diferencias pueden ser usadas como marcadores moleculares; polimorfismos asociados a alguna característica en específico.

Existen dos tipos de polimorfismos ampliamente utilizados en los estudios genéticos: los microsatélites y los polimorfismos de un solo nucleótido (SNPs). Los microsatélites son secuencias que se encuentran repetidas en bloque en el genoma y los SNPs se definen como variaciones nucleotídicas en una sola base. Los primeros son sumamente variables, contando con varios alelos por polimorfismo. En tanto, los segundos, tienen normalmente dos alelos y la generación de nuevas mutaciones suele ser muy limitada (Smith y col, 2005). Sin embargo, los SNPs se encuentran con una frecuencia de 4.3 por cada 1000 bases permitiendo una cobertura genómica sumamente alta (Vignal y col, 2002). La elección de uno u otro marcador depende directamente del diseño experimental y los objetivos del trabajo.

#### 1.3.1 Tecnologías de genotipado de SNPs

La tecnología de punta para los estudios de genotipado en individuos es el uso de chips que contienen miles de diferentes SNPs, debido a su alto rendimiento (high-throughput) (Oliver y col. 2002; Kim y Misra 2007; Treff y col. 2011). Sin embargo, esta tecnología es de muy alto costo, siendo pocos los laboratorios que pueden enfrentar su

implementación. Es por ello, que se han desarrollado tecnologías más económicas, pero de menor envergadura para lograr el genotipado y validación de este tipo de polimorfismos, como por ejemplo la PCR alelo específico, detección por corte con enzimas de restricción, reacción en cadena por la ligasa o marcadores fluoroforados tipo TaqMan (Black y Vontas 2007).

Actualmente, se dispone de una novedosa tecnología para la detección de SNPs, denominada “denaturación de hebras con alta resolución” o por sus siglas HRM, método de menor costo, con capacidad de multiplex y rendimiento muestral medio (Montgomery y col. 2007; Landi y Tavtigian 2009). El interés que ha despertado esta técnica ha logrado que compañías fabricantes de equipamiento de laboratorio, como Illumina, Qiagen o Idaho, comenzaran a incluir protocolos y funciones que incorporan este tipo de análisis (Hermann y col. 2006). De la mano, se han desarrollado nuevas y mejores tinciones fluorescentes fuera de SYBR Green, entre las que están LCGreen+ y Evagreen, por nombrar algunos.

Se espera para los próximos años, un desarrollo exponencial en tecnologías para el genotipado para marcadores tipo SNPs, pues sus características son ideales para generar trazabilidad entre individuos. La alta cobertura genómica de estos marcadores, permiten estudios de asociación más precisos, teniendo una mayor cantidad de datos por kilo base. También la capacidad de efectuar multiplex y el desarrollo de nuevas plataformas fomenta el uso de estos polimorfismos, por sobre otros como los microsatélites.

#### 1.4 Búsqueda y análisis in silico de SNPs

Junto a los proyectos de expresión génica, y nuevas tecnologías de secuenciación,

se ha llevado a cabo el proyecto de secuenciación del salmón (<http://web.uvic.ca/grasp/>). Esto ha generado una base de datos importante de segmentos cortos de ESTs y DNA genómico (fragmentos de 700bp aproximadamente), con los cuales se pueden construir secuencias más largas uniendo los extremos sobrelapados de ellas (contigs). Al alinear varias secuencias en la misma región que provienen de diferentes individuos, existe la posibilidad de que se puedan encontrar polimorfismos, en especial SNPs. Para poder llevar a cabo el proceso anterior, se necesita de una base de datos que contenga varios cientos de secuencias asignadas al gen o región de mi interés, sino con la falta de información se puede caer en errores de asignación.

#### 1.4.1 Recolección de información: NCBI, UniGene.

La generación de cantidades masivas de información ha creado un requerimiento por métodos nuevos de análisis. La división de la información en segmentos especializados en la base de datos NCBI, ha facilitado la búsqueda y filtración de datos útiles. En este sentido, el segmento dedicado a UniGene, se encarga de establecer genes y asignar secuencias en forma no-redundante, permitiendo de esta manera unificar la información existente a cerca de una región específica, aunque esta provenga de distintos autores.

#### 1.4.2 Algoritmos disponibles para el descubrimiento y corrección de SNPs *in silico*.

En paralelo a la implementación de grandes bases de datos biológicas, el desarrollo de herramientas bioinformáticas que analizan esta información ha sido, durante los últimos años, esencial. Para el descubrimiento de nuevos SNPs haciendo uso

de secuencias redundantes de una misma región pero de distintos individuos se ha desarrollado polyBayes (Marth y col. 1999). Este programa utiliza la información generada por Phred y Phrap (Ewing y col. 1998; Ewing y Green 1998), algoritmos que asignan calidad de bases secuenciadas (secuenciación mediante metodología Sanger) y unión de ESTs en contigs. Phrap ocupa la asignación de calidad de bases hecha por Phred para generar el alineamiento de secuencias y la generación de contigs, para luego encontrar, a través de gigaBayes (algoritmo bayesiano recursivo), los polimorfismos. Consed (Gordon y col. 1998) se utiliza para la visualización de resultados en una forma gráfica, mostrando las posiciones de los polimorfismos junto con los electroferogramas correlativos, esto permite corregir manualmente los polimorfismos.

A pesar de que los algoritmos y programas aumentan la eficiencia y la eficacia del descubrimiento de SNPs, estos no son infalibles, sino que dependen de una serie de parámetros y aproximaciones que, mal aplicados, pueden llevar a equivocaciones. Varios autores han planteado distintas maneras de lograr polimorfismos *in silico* más precisos. Es así como la división de la base de datos de ESTs utilizando “cross-validation” (Hayes y col 2007) o la aplicación de una máquina que usando datos ya validados sea capaz de afinar el descubrimiento de futuros polimorfismos, son aproximaciones viables para aumentar la eficiencia y eficacia del descubrimiento de SNPs (Matukumalli y col. 2006).

#### 1.4.3 Dificultad que presentan secuencias parálogas en el DNA de salmónidos.

Uno de los principales desafíos que ha tenido el descubrimiento de SNPs ha sido la duplicación completa reciente del genoma del salmón (Allendorf y Thorgaard, 1984). La generación de diferencias genéticas con un ancestro común que luego generan

diversidad se le llama homología, en caso de que estos genes cumplan una función similar pero no exista un ancestro común, sino solo convergencia evolutiva, se le denomina analogía. Por otro lado, la homología puede dividirse en ortología y paralogía, la primera involucra un proceso de especiación y posterior diferenciación y modificación génica. La segunda, involucra la generación de duplicados génicos que, por tasas de mutación desiguales, conllevan modificaciones al organismo original. Es importante indicar que la ortología involucra un evento de especiación anterior que la paralogía no acarrea (Sonnhammer y Koonin, 2002; Fitch, 2000). En *Salmo salar* debido al evento de duplicación genómica reciente, existe una gran cantidad de secuencias duplicadas que se encuentran en lugares físicos separados de la original, permitiendo tasas de mutación diferentes debido a presiones selectivas desiguales (Conant y Wagner 2003). El problema es que al amplificar o secuenciar una región, se puede obtener un segmento no específico. En este contexto, si se hace un ensamblaje de secuencias amplificadas usando Phrap es muy probable que este las alinee todas, sin discriminar copias, esto hará que cuando se busquen polimorfismos, estos no sean debido a mutaciones en el mismo gen, sino a cambios en las copias del gen original, entregando SNPs falsos (Seeb y col. 2011). Este problema es general para todas las especies, pero en *Salmo salar* es preocupante por el conocimiento de un evento de duplicación genómica y porque no existe un mapa físico completo que permita comparar secuencias directamente. Aunque Phrap no tiene parámetros dirigidos directamente a corregir este problema, existen ciertas precauciones que se pueden tomar en el algoritmo que permiten disminuir la ocurrencia de alineamientos de secuencias paralogas. Por otro lado, el programa PolyBayes si está dirigido a corregir este problema, pero al usar de input los contigs

formados por Phrap, depende de este, de manera que si Phrap genera uniones imprecisas, la cadena de trabajo que sigue será errada. De hecho existe un algoritmo en la cadena de polyBayes, llamado gigaBayes, solamente dedicado evaluar la presencia de este tipo de secuencias, re-corrigiendo los errores en los que pudo haber incurrido Phrap.

### 1.5 Validación y genotipado de SNPs mediante denaturación de curvas en alta resolución (HRM).

El análisis de HRM, es una técnica que mide la disminución de fluorescencia de un agente intercalante en el proceso de disociación de la doble hebra de ADN. El método se basa en que la señal fluorescente del agente intercalante es más potente cuando se encuentra unido a la doble hebra de ADN, disminuyendo su intensidad más de mil veces en presencia de ADN de hebra simple. La medición se lleva a cabo inmediatamente después de un PCR convencional y la forma de la curva de denaturación depende de la secuencia y el polimorfismo contenido en el amplicón, lo que convierte al método en una forma poderosa, rápida y de bajo costo para la detección de mutaciones.

A raíz de esta metodología se generó una clasificación de SNPs que los divide en 4 clases como indica la Tabla 1.

**Tabla 1.** Clasificación de SNPs por medio de la mutación nucleotídica generada.

SNP	Mutación
Clase 1	C $\leftrightarrow$ T y G $\leftrightarrow$ A
Clase 2	C $\leftrightarrow$ A y G $\leftrightarrow$ T
Clase 3	C $\leftrightarrow$ G
Clase 4	A $\leftrightarrow$ T

Esta división se genera por el impacto que tiene cada mutación en la temperatura



de denaturación. Es así como la detección de SNPs clase 1 es más fácil que las de clase 4, ya que genera mayores cambios de temperatura de denaturación en un amplicón. La frecuencia de los polimorfismos de clase 1 y 2 es 0.838 mientras que los de clase 3 y 4 solo 0.162, esto quiere decir que aunque los últimos son los más difíciles de validar son también los que están presentes en menor cantidad (Liew y col. 2004).

Tanto el equipo como la tinción a utilizar son de vital importancia para la resolución del método, pues el análisis de los datos posterior al ensayo permite detectar diferencias que suelen ser ínfimas.

## HIPOTESIS Y OBJETIVOS

**Mutaciones puntuales de tipo SNP en transcritos de genes asociados al sistema inmune de *Salmo salar***, se encontrarían asociadas a cambios estructurales proteicos que conllevarían efectos importantes en el sistema inmunológico de esta especie.

**Este trabajo tiene como objetivo principal determinar mutaciones puntuales en regiones transcritas de genes asociados a vías de señalización al sistema inmune de *Salmo salar***. Para el cumplimiento de esto se debe:

- 1) Identificar SNPs *in silico* utilizando secuencias expresadas en tejidos (ESTs).
- 2) Validar y genotipificar los SNPs encontrados *in silico*, a través del análisis de curvas en alta resolución, en poblaciones comerciales y naturalizadas de *Salmo salar*.
- 3) Determinar la posición y la funcionalidad de los SNPs en genes validados *in vitro*.

## METODOLOGIA

### 3.1 Muestreo de individuos.

Las muestras utilizadas en este trabajo provienen de músculo de cinco poblaciones distintas de *Salmo salar* que tienen su origen en 4 centros de producción, y una naturalizada de Lago Ranco.

### 3.2 Extracción de DNA.

Las extracciones fueron hechas desde muestras de tejido muscular de *Salmo salar* usando un kit comercial (Qiagen) según las instrucciones del fabricante. La cantidad y calidad de la extracción fue analizada a través del kit Qubit de Invitrogen y geles de agarosa al 1% teñidos con bromuro de etidio respectivamente.

### 3.3 Descubrimiento de polimorfismos de un solo nucleótido in silico.

Los números de acceso correspondientes a los ESTs de los genes Mx1, Tap1, Tap2, C-type Lectin B, IRF-1, PKR, Tiorredoxina y 2- $\beta$  Microglobulina se extrajeron desde UniGene (Anexo 3). Se accedió a la base de datos de cGrasp en busca de los electroferogramas de cada una de las secuencias. Para determinar la región nucleotídica de cada EST que sería sometida al ensamblaje de regiones contiguas por Phrap, el algoritmo Phred fue aplicado sobre cada electroferograma para determinar valores de calidad de cada base. Se eliminaron las bases desde los extremos hasta llegar a una cuyo valor de calidad fuera sobre 30 (probabilidad de error en asignación de 1 en 1000). Por otro lado, se realizó un enmascaramiento de secuencias, con el objetivo de no alinear regiones contiguas a través de vectores de clonación, usando el algoritmo Crossmatch, de las secuencias pGEM-5Zf (+), pCMV-Sport 6, pBlueScrpt II SK(+), PBK-CMV,

pGEM-T Easy Vector, pDONR223, pDONR221 y una secuencia de reconocimiento de corte con número de acceso DD038280.1. Con las secuencias ya depuradas por Phred y Crossmatch, se generaron los “contigs”, o secuencias ensambladas de ESTs, utilizando Phrap (95% de identidad de alineamiento (stringency) y sobreposición de 50 bases (minmatch 50). Se aplicó gigaBayes, el cual informó solo aquellos polimorfismos putativos que alinearan por lo menos 4 secuencias, que el alelo menos presente estuviese al menos en 2 de ellas, cuya calidad de secuenciación fuere sobre 20 y probabilidad límite de polimorfismo sobre 90%. Consed permitió la visualización y corroboración manual de resultados. Un resumen de la metodología se puede encontrar en el Anexo 1.

#### 3.4 Diseño y síntesis de partidores.

El diseño de partidores para la validación de los SNPs se llevó a cabo usando el programa AmplifX (<http://ifjr.nord.univ-mrs.fr/AmplifX>). Para los ensayos en HRM los amplicones tuvieron una longitud máxima de 160bp y fueron diseñados para que el polimorfismo se encontrara en el centro del amplicón. La calidad de todos los partidores se optimizó *in silico* usando el programa NetPrimer analizando la formación de estructuras secundarias (horquillas y hetero y homo dímeros) y temperatura de hibridación ([www.premierbiosoft.com](http://www.premierbiosoft.com)).

#### 3.5 Optimización de temperatura de hibridación de partidores para el PCR.

Estos experimentos se llevaron a cabo en un termociclador Axygen (MaxyGene Gradiente), considerando un programa de PCR bajo una gradiente de temperatura. El protocolo utilizado para cada set de partidores involucró 10min de activación a 94°C,

seguido por 40 ciclos a 30s a 94°C, 30s a una gradiente de temperatura de entre 5°C por sobre el partidor de más alta  $T_m$  y 5°C por bajo el de menor y 30s de extensión a 72 ° C. Los ensayos se llevaron a cabo en 2mM de  $MgCl_2$ , 1X Buffer Taq, 0.02 U/ $\mu$ l Hotstart Fermentas Taq Pol, 0.5  $\mu$ g/ $\mu$ l BSA, 100uM dNTP, 0.5mM cada partidor y 20ng totales de DNA de las muestras Na 22 y Ca 40 en un volumen final de 10 $\mu$ l. Los resultados de cada experimento fueron evaluados mediante electroforesis en geles de agarosa al 2% teñidos con Gelred, comparando el producto de PCR con un estándar de tamaño molecular de 50pb. Se determinó la temperatura óptima de hibridación, por la presencia de un amplicón único del tamaño teórico esperado.

### 3.6 Curvas de alta resolución.

Todas las reacciones se llevaron a cabo en un equipo Eco Illumina por triplicado. El kit GoTaq qPCR MasterMix de Promega fue utilizado para hacer la PCR y visualizar las curvas de alta resolución. La amplificación fue llevada a cabo en 1X Mastermix, 1X ROX, 0.5 $\mu$ M de cada partidor y 20ng totales de DNA de cada individuo. La PCR se realizó con 10min de activación a 94°C, luego 50 ciclos a 94° por 30s, 30s de hibridación de partidores a temperaturas optimizadas para cada uno en la Sección 3.5 y 15s de extensión a 72°C. Una vez terminados los 50 ciclos se aumentó la temperatura a 94°C por 30s, disminuyéndose esta a 60°C rápidamente, comenzando el HRM subiendo la temperatura desde 60°C a 94°C de a 0.1°C/medición. Se corroboró la especificidad del amplicón generado a través de la temperatura de denaturación del producto. Se generaron curvas de diferencias de fluorescencia para corroborar los genotipos asignados. Los genotipos normal y mutante se definieron según su temperatura de

denaturación, siendo el mutante el de mayor temperatura y el normal el de menor. Los resultados fueron analizados con el software del equipo Eco Real Time PCR system v.3.1.

### 3.7 Secuenciación.

Se secuenció individuos que presentaban los tres genotipos alternativos (Normal, Mutante y Heterocigoto) para cada uno de los SNPs validados. Se amplificó el fragmento deseado, usando el protocolo de la Sección 3.5 de cada genotipo en un volumen total de 200µl. Se sometió a electroforesis en gel de agarosa el amplificado de PCR, cortando la banda y purificándola usando el MinElute gel extraction kit de Qiagen. Los productos extraídos del gel fueron secuenciados en un equipo ABI PRISM 3100 (Anexo 2).

### 3.8 Anotación de polimorfismos.

La anotación se llevó a cabo alineando los contigs formados en la sección 3.3 a la base de datos de NCBI mediante el algoritmo blast. Se alinearon los contigs de cada gen a la secuencia que arrojó el mejor puntaje en blast usando el algoritmo Clustal X (Larkin y col. 2007). Se utilizó BioEdit (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>) como interfaz gráfica para la visualización de los alineamientos y las mutaciones, aislando así el polimorfismo. La relevancia de la mutación se analizó comparando la información acerca de la traducción de la secuencia nucleotídica obtenida por el alineamiento de los ESTs y la proteína publicada en NCBI.

## RESULTADOS

### 4 Descubrimiento de SNPs.

A partir de la revisión bibliográfica de genes que varían su expresión frente a infecciones por los patógenos IPN, ISA, *Piscirickettsia salmonis* y *Neoparamoeba perurans* (Sección 1.2.2), se buscaron secuencias de ESTs asignados a estos genes en la base de datos de Unigene. Se comenzó el alineamiento y generación de contigs usando los parámetros estándar establecidos para el algoritmo de alineamiento, Phrap. Estos se ajustaron al input de secuencias que se introducían hasta llegar a los óptimos descritos en la Sección 3.3, que se definieron por medio de la relación entre el número generado de contigs y secuencias que no generan ningún solapamiento. De la misma manera se definieron los parámetros óptimos para el óptimo funcionamiento del algoritmo GigaBayes.

Se pudo apreciar que la probabilidad de encontrar alineamientos de secuencias paralogas aumentaba a medida que se acrecentaba el número de ESTs alineados para cada contig (resultados no mostrados). Estos resultados provocaron un aumento en la posibilidad de encontrar falsos positivos en la sección experimental. Por esto, la revisión manual de los polimorfismos encontrados fue un paso esencial, permitiendo eliminar varios polimorfismos por paralogía (Figura 1. B.)

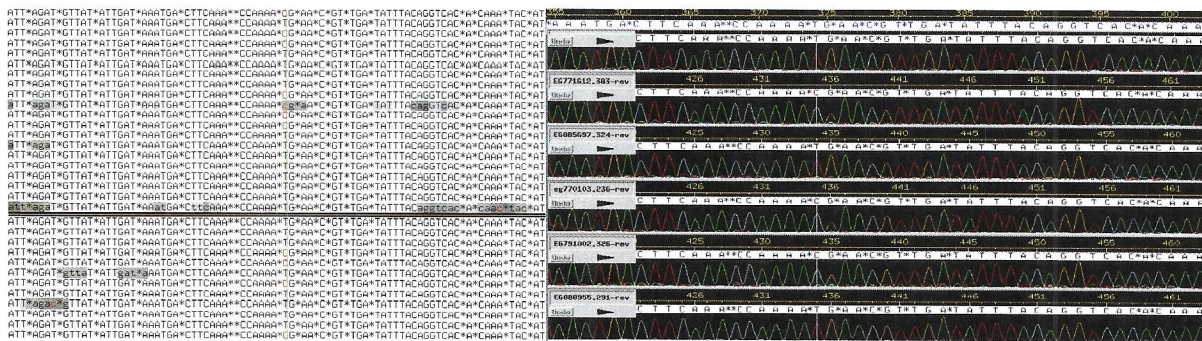
La alta heterogeneidad del número de secuencias de ESTs asignadas a cada gen (segunda columna Tabla 2) determinó que sólo algunos de los polimorfismos encontrados “*in silico*” pudiesen ser exitosamente validados *in vitro*. Es por ello que finalmente se optó por hacer partidores para 6 posibles SNPs: TAP1, Tiorredoxina, Mx1,

2 $\beta$ -Microglobulina, PKR e IRF-1 (Tabla 2). Estos partidores fueron elegidos debido a que permiten genotipar polimorfismos en regiones de alta cobertura, contar con una frecuencia alélica menos frecuente mayor al 20% y cuyas variantes tuvieran un valor de calidad sobre 30 (Figura 1.A.)



**Figura 1.** Visualización de polimorfismos en Consed. A) SNP validado, genotipado y secuenciado ( $\beta$ 2-Microglobulina). B) SNP debido a paralogía no validado ( $\beta$ 2-Microglobulina). C) Baja profundidad de análisis (Mx-1). D) Alta profundidad de análisis (TAP1).

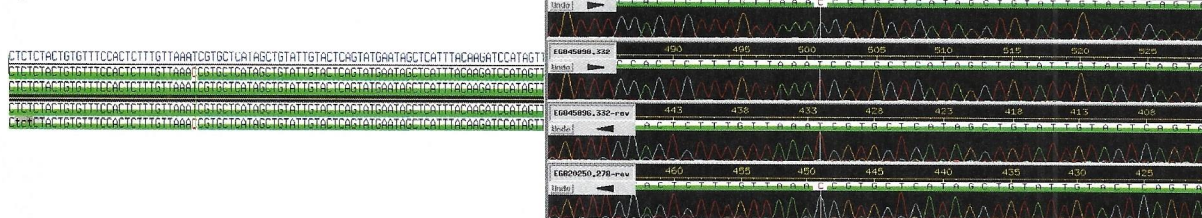
**A**



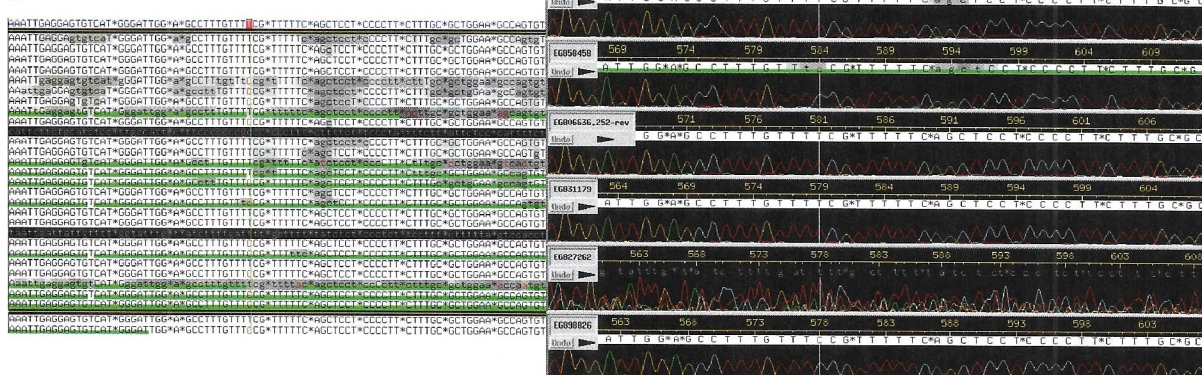
**B**



**C**



**D**



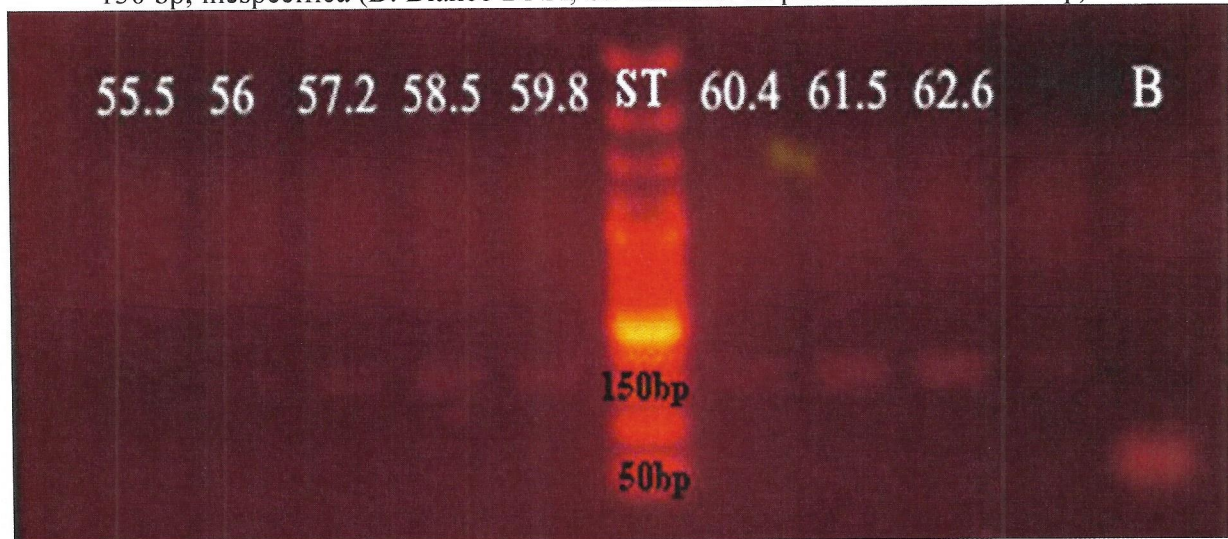
**Tabla 2.** Resumen resultados de descubrimiento de SNPs *in silico* y validación experimental.

	ESTs encontrados	SNPs encontrados <i>in silico</i>	SNPs validados por HRM
Transport associated protein 1	61	1	1
Tiorredoxina	111	1	1
2 $\beta$ -Microglobulina	175	1	1
Myxovirus protein 1	26	1	-
Protein Kinase RNA activated	71	1	-
Interferon regulatory factor-1	59	1	-
Total	521	6	3

#### 4.1 Optimización de temperatura de hibridación y corroboración del largo de amplicón.

A través de la PCR de gradiente se pudo establecer una temperatura de hibridación óptima para los partidores usados en los ensayos de HRM. También se pudo corroborar la especificidad del amplicón a través de su tamaño. Sin embargo, los sets de partidores HRM\_C\_type y HRM\_TAP-2 fueron descartados por mostrar una amplificación inespecífica (HRM\_C\_type se presenta como ejemplo en la Figura 2). Las temperaturas óptimas para todos los set de partidores usados se muestran en la Tabla 3.

**Figura 2.** PCR de gradiente (55,5-62,6 °C) con 8 temperaturas distintas usando partidor HRM\_C\_type-F (tamaño esperado 108bp). Se puede apreciar una banda tenue sobre 150 bp, inespecífica (B: Blanco DNA, St: Estándar de peso molecular de 50bp).



**Tabla 3.** Resumen partidores utilizados para HRM con temperaturas óptimas experimentales.

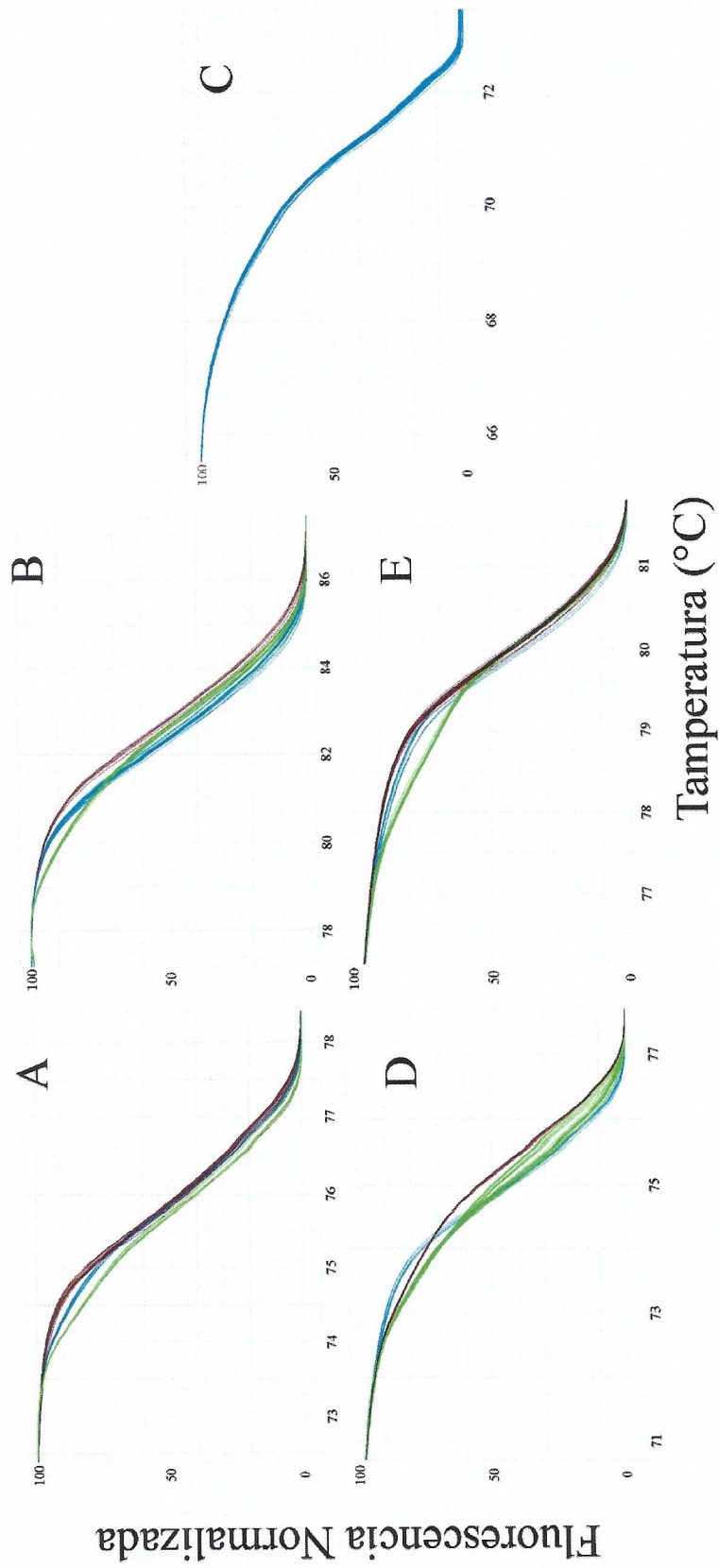
Fuente	Gen	Polimorfismo	Partidor	Secuencia Partidor	Tm (°C)	Temperatura óptima (°C)
INBIOGEN	SNP356	C/G	HRM_ss356-F	GAGAAATGGAAAAGTAGCCCGT	59.6	60.2
			HRM_ss356-R	TAATACATAGAGCACCAGCAAGG	57.8	
Rise et al. (2004)	TAP-2	A/G	HRM_TAP-2-F	GCCCTAAAATGTCAACGCTCCTAC	58.3	62.6
			HRM_TAP-2-R	CATGAGCAGAGTGAGCTTCCAG	60	
	TAP-1	C/T	HRM_Tap-1_2-F	GTTGAAAAATGTGAGATGGTGGCA	62.3	64.4
			HRM_Tap-1_2-R	GCAAAGAAAGGGGAGGAGCTGAA	64.7	
Tiorredoxina	A/G	HRM_Thio-F	GACCTGAATAAGAATGACTAATACACC	57.7	56.9	
		HRM_Thio-R	CGAAGGACCAGATTCCACCA	60.4		
C-type Lectin B	A/G	HRM_C_type-F	GTTACTCTCCATGACCACACAGTCC	61.7	62.7	
		HRM_C_type-R	GACAACACAAGGCTCACAAAAACC	63.5		
Jørgensen et al. (2008)	$\beta$ 2 microglobulina	A/G	HRM_Mb2-F	CCATGTGATATGTATTTGTGTGACC	59	61.4
			HRM_Mb2-R	ATGTTGAAGGATAGACAGAAGATTAG	56	
Larsen et al (2004)	Myxovirus protein 1	C/T	HRM_Mx1_1-F	GCTGTAGGTGATGTCCGTGTGTGA	60.6	62.6
			HRM_Mx1_1-R	GTCTACGGCTGAGTCCCATGTTGA	62.8	
García et al. (2006)	Protein Kinase RNA-ativated	A/T	HRM_PKR-F	TTCTGGCTGTGCAGGGTGG	62.3	62.7
			HRM_PKR-R	CATGCTGGTCATTTATGAACATTT	59	
Young et al. (2008)	Interferon Regulatory Factor 1	C/T	HRM_IRF1-F	CTCAGGACTACCCGATCATTGACC	61.3	64.4
			HRM_IRF1-R	CTGTTGGGGATAAAGCTGGATGTGT	60.5	

#### 4.2 Validación de SNPs utilizando curvas de alta resolución.

Estos experimentos se realizaron con DNA de 14 individuos escogidos desde las poblaciones comerciales y naturalizadas (13 y 2 individuos respectivamente). Se escogieron las muestras de manera que estas fueran lo suficientemente heterogéneas entre sí para asegurar variaciones alélicas entre ellas, pues poblaciones del mismo productor o del mismo desafío pueden no tener grandes divergencias. El análisis demostró claras variaciones en 4 de 7 polimorfismos probados.

La Figura 3 muestra los resultados obtenidos a través del análisis de curvas de alta resolución para SNPs en los genes PKR, SNP356,  $\beta$ 2-Microglobulina, Tiorredoxina y TAP1. Se puede ver para los cuatro últimos genes que se generaron tres grupos de curvas correspondientes a los genotipos normal (azul), mutante (rojo) y heterocigoto (verde). Por otro lado para el gen PKR existe solo una curva (azul), demostrando que no existe polimorfismo entre las 14 muestras probadas (Figura 3). Se puede ver también que en aquellos individuos verdaderamente polimórficos, la mutación que confiere mayor temperatura al amplicón (G o C) genera curvas que demoran más su denaturación; esto sustentado en la mantención de un nivel de fluorescencia durante incrementos constantes de temperatura en el tiempo (curva rojo). En tanto aquellas curvas que corresponden a individuos con mutaciones de menor temperatura producen curvas que sufren una denaturación sumamente rápida, apoyado por la rápida caída de denaturación a una cierta temperatura (curva azul).

**Figura 3.** Curvas de alta resolución (HRM) de los 4 genes polimórficos (A) Tiorredixina, B) TAP1, D)  $\beta$ -2 Microglobulina y E) ss356) y uno no polimórfico(C) PKR). Verde, azul y rojo corresponden a individuos heterocigotos, homocigoto normal y homocigoto mutante respectivamente.



En individuos heterocigotos, la presencia de ambas mutaciones en un mismo individuo genera una estructura denominada heterodúplex, estas son hibridaciones de hebras complementarias pero con diferente polimorfismo generando un cambio en aquella posición. Esto genera un perfil muy característico y de fácil discriminación que comienza con una denaturación rápida de las hebras híbridas con una mutación, seguida por la denaturación de los amplicones normales para finalizar con los mutantes de mayor temperatura. Finalmente, lo que se visualiza en los perfiles de HRM de individuos heterocigotos son dos segmentos: el primero correspondiente a una denaturación rápida (heterodúplex) que se estabiliza (cruzando por la curva del polimorfismo de menor temperatura) alcanzando la curva de más alta temperatura (curva rojo).

Al ser los individuos heterocigotos fácilmente distinguibles de sus pares homocigotos, la generación de “heterocigotos artificiales” para asegurar el genotipo asignado es una buena opción para la determinación de aquellos genotipos ambiguos. Este análisis es particularmente útil cuando las curvas de ambos homocigotos son muy similares (mutaciones A/T y G/C). La técnica involucra mezclar ADN del individuo de genotipo ambiguo con ADN de un individuo homocigoto contrario al que se cree es el genotipo ambiguo, la generación de un perfil de HRM heterocigoto u homocigoto comprueba el genotipo. De esta manera se pudo establecer el único genotipo normal de las muestras analizadas para el polimorfismo SNP356 (Tabla 4.).

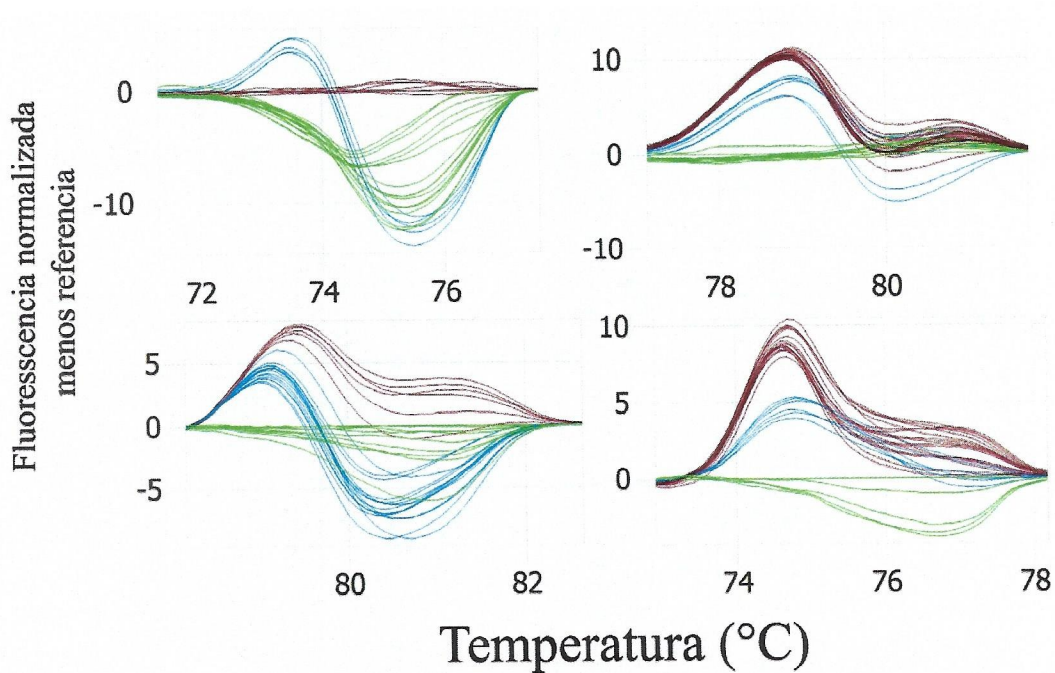
Los genes Mx1 e IRF-1 demostraron no ser polimórficos para los individuos muestreados (datos no mostrados) al igual que el gen PKR (Figura 1).

#### 4.3 Visualización de genotipos mediante curvas de diferencias

Para poder visualizar en forma más clara los genotipos encontrados se generaron curvas de diferencias. Estas curvas se generan restando el valor de fluorescencia en cada punto para cada individuo a uno de referencia definido por el usuario. Para el gen  $\beta 2$ -Microglobulina se tomó como referencia un individuo homocigoto mutante, en tanto para los otros tres SNPs polimórficos se tomó un heterocigoto. Se puede apreciar que las curvas de diferencias de fluorescencia permiten visualizar de forma más clara los grupos de genotipos (Figura 4).



**Figura 4.** Gráficos de diferencia de fluorescencia de los 4 genes polimórficos (verde, azul y rojo corresponden a individuos heterocigotos, homocigoto normal y homocigoto mutante respectivamente). A)  $\beta$ 2-Microglobulina, B) SNP356, C) TAP1 y D) Tiorredoxina.



De esta manera se pudo obtener los genotipos de los 14 individuos analizados para los 4 SNPs probados. Los resultados de todos ellos se encuentran en la Tabla 4.

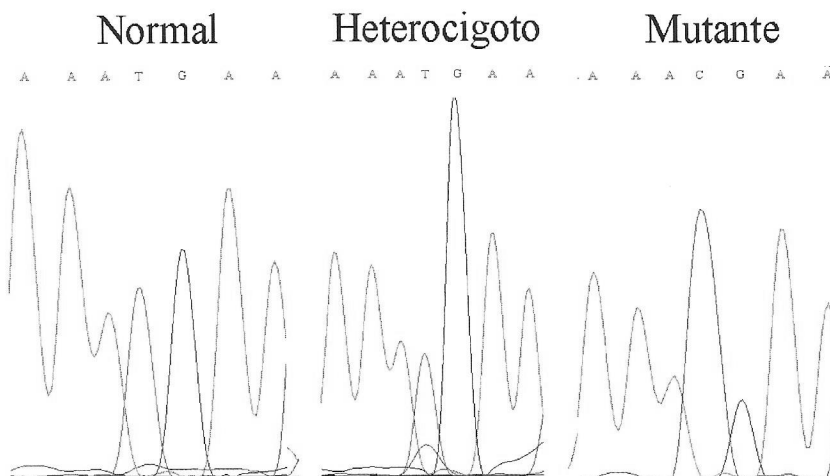
**Tabla 4.** Resumen del genotipado definidos en las muestras probadas por medio del análisis de curvas de alta resolución (HRM).

Población	Individuos	Genotipos			
		TAP1	SNP356	Tiorredoxina	2 $\beta$ -Microglobulina
Naturalizada	Na22	Normal	Mutante	Normal	Normal
	Na12	Heterocigoto	Mutante	Heterocigoto	Mutante
Comercial 1	Ca23	Normal	Mutante	Mutante	Normal
	Ca32	Normal	Normal	Mutante	Heterocigoto
	Ca40	Heterocigoto	Heterocigoto	Heterocigoto	Mutante
Comercial 2	Aq37	Normal	Heterocigoto	Mutante	Heterocigoto
	Aq15	Heterocigoto	Heterocigoto	Mutante	Heterocigoto
	Aq51	Mutante	Mutante	Mutante	Mutante
Comercial 3	LF9	Normal	Mutante	Normal	Heterocigoto
	LF43	Mutante	Mutante	Mutante	Heterocigoto
Comercial 4	PL1	Heterocigoto	Mutante	Mutante	Heterocigoto
	PL53	Normal	Heterocigoto	Heterocigoto	Heterocigoto
	PL80	Mutante	Mutante	Normal	Heterocigoto
	PL87	Heterocigoto	Mutante	Heterocigoto	Heterocigoto

#### 4.4 Secuenciación de los polimorfismos validados.

Para el genotipado de los SNPs descubiertos se procedió a secuenciar en forma dirigida cada polimorfismo. Se pudieron corroborar todos los SNPs validados bajo los ensayos de HRM. La Figura 5 muestra los genotipos secuenciados para el polimorfismo  $\beta$ 2-Microglobulina, se puede observar la mutación puntual (C/T) en la hebra secuenciada con el partidor HRM\_Mb2-R.

**Figura 5.** Electroferograma del polimorfismo validado  $\beta$ 2-Microglobulina (A/G) para cada uno de los genotipos secuenciados en la hebra complementaria (normal, heterocigoto y mutante).



#### 4.5 Anotación de SNPs in silico en genes encontrados.

Todos los contigs que contenían los polimorfismos validados, excepto el correspondiente al SNP356, tuvieron puntajes de hibridación significativos en un blast contra secuencias de la base de datos de NCBI. De esta manera, se pudo extraer información de los segmentos codificantes de los genes TAP1,  $\beta$ 2-Microglobulina y Tiorredoxina (números de acceso en Tabla 5). A partir de estas secuencias, se situó el polimorfismo validado en este trabajo en el mRNA procesado. Se pudo ver que los polimorfismos de los genes Tiorredoxina y  $2\beta$ -Microlobulina se encontraban en la región 3'UTR fuera del segmento traducible; por otro lado, el SNP TAP1 se encontraba dentro de la región codificante. Se pudo comprobar que la mutación proteica correspondía a un cambio en el aminoácido 233 desde glicina a serina (Tabla 5).

**Tabla 5.** Anotación de SNPs encontrados usando información de la base de datos de NCBI.

	Posición de la Mutación (bp)	Región codificante del gen (bp)	Número acceso gen referencia	Implicancia mutacion	Mutación a.a.
Transport associated protein 1	919	222-2441	NM_001123673.1	Codificante	G>S
Tiorredoxina	653	80-406	BT125521.1	3'UTR	NA
2 $\beta$ -Microglobulina	970	110-450	BT046451.2	3'UTR	NA

## DISCUSION Y PROYECCIONES

Para el descubrimiento y aislamiento de SNPs *in silico* se pudieron obtener un número de polimorfismos comparables a otros trabajos en la literatura dado el número de secuencias disponibles en la base de datos. Hayes *et al* (2007) describe una manera de mejorar el descubrimiento *in silico* de SNPs mediante la división de los electroferogramas encontrados en 2 sets, haciendo posteriormente una corroboración cruzada entre los SNPs encontrados para cada grupo. Sus resultados *in silico*, detectaron sobre 14.000 polimorfismos en los dos grupos de datos por separado, sin embargo, al entrecruzar los polimorfismos descubiertos entre ambos sets, sólo 856 polimorfismos se corresponden. A partir de estos datos, eligiendo 136 polimorfismos de los 856 para validar, un 72.5% de ellos mostró polimorfismo. Estos resultados son de gran eficiencia, debido a que existe una alta probabilidad de asignar en forma errónea los polimorfismos *in silico*. Por otro lado, en nuestro trabajo, se logró un 50% de éxito entre SNPs descubiertos *in silico* y validados experimentalmente. Para ambos trabajos, las tasas de éxito se explican por la exigencia aplicada en el análisis *in silico*, descartando gran parte de los polimorfismos que pueden presentar problemas a la hora de la validación.

Una limitante importante, ligada a la probabilidad de éxito de encontrar SNPs polimórficos, fue encontrar suficientes secuencias de ESTs para algunos genes en la base de datos de UniGene. Por un lado, Rise y col. (2004) describen secuencias que se sobre expresaban más de 30 veces bajo condiciones de infección que no pudieron ser utilizados por no tener ningún homólogo en la base de datos. Por otro, también describe secuencias asignadas a genes que, sin embargo, no cuentan con información suficiente

para llevar a cabo un análisis como el que hizo en este trabajo. En este sentido, la heterogeneidad de la información disponible en la base de datos provoca un riesgo diferente de descarte o admisión de polimorfismos verdaderos o falsos para cada gen. De esta manera, a pesar de que existen exigencias mínimas ligadas a la profundidad del polimorfismo y calidad de base analizada, algunos SNPs cumplen estas exigencias con creces, otros solamente apenas y muchos simplemente no las cumplen. Es por ello que antes de pasar a la fase de validación, los polimorfismos que cumplen con las condiciones mínimas impuestas tienen que ser analizados manualmente, evaluando el nivel de satisfacción en el cumplimiento de los requerimientos. Por otro lado, debido a la falta o escasez de información, es que a veces se forzó la búsqueda para que incluyera algunos genes que, aunque no hayan descripciones directas en peces, suponemos tendrán un comportamiento parecido a infecciones en humanos o ratones.

En los casos de la  $\beta$ 2-microglobulina y Tiorredoxina se pudieron visualizar en consed un número alto de secuencias alineadas para un contig (sobre 50). Contrario a lo que se puede pensar, el número de polimorfismos extraído fue muy bajo. Esto puede deberse a un número de factores, el primero y más importante es considerar que los ESTs, como su nombre indica, son secuencias procesadas expresadas en un tejido. Mutaciones sobre transcritos de DNA aumentan la probabilidad de modificaciones en la secuencia proteica, conllevando un posible cambio fenotípico, esto implica que mutaciones en estas regiones son más acotadas y restringidas que en ADN no codificante. Sin embargo, existen zonas del mRNA, como la 5'UTR y 3'UTR que no se traducen, pero pertenecen de igual forma a la secuencia procesada; mutaciones puntuales

en aquellas zonas pueden acarrear modificaciones nivel de expresión del mensajero. Es por ello que aunque se valide una mutación puntual en un mRNA la probabilidad de que esta se encuentre en un segmento codificante es relativamente baja, pues existe una probabilidad más alta encontrar un polimorfismo establecido en regiones no codificantes. Un segundo factor que influencia el bajo número de polimorfismos encontrados en los datos ocupados, se relaciona a secuencias parálogas encontradas dado el comprobado evento de duplicación genómica del salmón (Figura 1 A.). Esto dificulta en forma significativa la labor de encontrar polimorfismos reales pues es muy fácil confundirlos y al mismo tiempo fomenta el descarte de SNPs verdaderos por no cumplir con todas las exigencias propuestas. Finalmente, el afinamiento de la metodología hizo que la cantidad de falsos positivos fueran disminuyendo a medida que se encontraban y se probaban una mayor cantidad de SNPs putativos.

Otro contratiempo inesperado fue el de la amplificación de segmentos de ADN llevados a cabo en PCR convencional y vistas en geles de agarosa que no correspondían al tamaño esperado (Figura 2). Esto sucedió por la naturaleza de las secuencias en las que se diseñaron los partidores. Al ser secuencias de ESTs o secuencias expresadas (mRNA) lo que se utiliza como base para el diseño de oligos, puede que haya intrones interfiriendo que no pueden ser distinguidos a priori. Esto ocurre porque no se ha publicado el genoma del salmón completo, por lo tanto no se sabe con certeza si los oligos que se diseñan se encuentran en exones diferentes. A raíz de esto se tuvieron que descartar 3 set de partidores, uno de los cuales se muestra en la Figura 2.

En cuanto a la fase de validación de los polimorfismos, el SNP SNP356 se encontraba ya genotipado con anterioridad para algunos individuos a analizar. La ventaja

de ser un SNP para el cual existen individuos con genotipos conocidos, es que a través de la comparación de los genotipos conocidos con las curvas de muestras incógnitas se facilita la asignación de uno u otro alelo a través de HRM. Por otro lado, este SNP permitió también validar, optimizar y estandarizar la técnica. Se destaca este polimorfismo ya que la diferenciación de genotipos para este resultó representar una mayor dificultad dado por el tipo de mutación que presentaba (C/G). Se encontró una correspondencia completa entre los genotipos obtenidos por HRM y la información disponible para ellos, de esta manera se pudo estandarizar y rectificar la metodología para su uso en marcadores para los cuales no había ningún tipo de información.

Analizando los datos para el polimorfismo asignado al gen  $\beta$ 2-Microglobulina, esta forma curva peculiar ya que su homocigoto de mayor temperatura sufre una denaturación inicial de fragmentos bastante marcada que, por un rango de temperatura, pasa a estar por debajo de la curva del homocigoto de menor temperatura. Sin embargo, se estimó que el segmento final de la curva era más decisivo para la separación de curvas homocigotas, es por ello que se definieron las curvas normales y mutantes como se indica en la Figura 3, la asignación fue luego corroborada por secuenciación del fragmento (Figura 5).

La anotación de los SNPs validados se hizo buscando las secuencias de los genes disponibles para los contigs generados. Se pudo apreciar que de los 4 SNPs, solo el SNP356 no tiene concordancias en la base de datos como para asignar un gen. Para los genes  $\beta$ 2-microglobulina y Tiorredoxina la mutación se encuentra en la región 3'UTR. Esta zona contiene secuencias que no son traducidas, sin embargo sirven como



señalización celular para la maquinaria de corte y procesamiento del mRNA, esto influye en impedir la degradación, dirigir a algún organelo o región celular específica o unir algún ligando (Mazumder y col. 2003). Es por ello que mutaciones puntuales en aquella región, a pesar de no traer repercusiones a nivel funcional o macromolecular, si pueden interferir en la vida media o la actividad traduccional del mRNA procesado, modificando el fenotipo con respecto a un individuo normal.

En el caso de que haya un cambio aminoacídico la mutación puede conllevar un cambio estructural o funcional importante. Este caso es el que involucra la mutación en el gen TAP1, mutación que se encuentra en la región codificante de la proteína. El cambio corresponde de una glicina a una serina, siendo este cambio entre aminoácidos similares. En la base de datos de PDB existe un transportador ABC en *Homo sapiens* que tiene un 33% de identidad con TAP1. Para la posición del SNP encontrado se vio que el mensajero codificaba a una treonina, aminoácido muy parecido a la serina. Es por esto que se cree un cambio en este aminoácido no involucraría consecuencias graves en la estructura tridimensional de la proteína.

Se pretende en el futuro analizar las mutaciones encontradas, de manera de observar si existe alguna consecuencia a nivel celular y de organismo (especialmente la ligada al gen TAP1). Por otro lado, se piensa aplicar la metodología para analizar los 4 SNPs en una población de individuos desafiados a patógenos (IPN, *Piscirickettsia salmonis* o virus ISA), con el objetivo de ver si existen variaciones entre aquellos que demostraron ser susceptibles o resistentes a las infecciones. Esto último sería de suma importancia, pues generaría información suficiente para establecer si las mutaciones encontradas se encuentran relacionadas a algún fenotipo resistente.

## CONCLUSION

Se logró utilizar una cadena de trabajo para el descubrimiento, identificación, validación, genotipificación y anotación de SNPs por medio del análisis de curvas de alta resolución. Sin embargo, al analizar el impacto que tienen las mutaciones encontradas a nivel traduccional, se pudo comprobar que de tres polimorfismos solo el ligado al gen TAP1 se encontraba en región codificante. A partir de esto se cree que este polimorfismo podría ser el único que podría traer consecuencias a nivel fenotípico. Es por ello que en futuros experimentos se determinará los 4 polimorfismos validados en nuevas poblaciones desafiadas en número mayor de individuos.

## BIBLIOGRAFIA

- Allendorf, F.W. y Thorgaard, G.H. 1984. Tetraploidy, the evolution of salmonoid fishes. Turner, J.B. (Ed.) *Evolutionary Genetics of Fishes*:1–53.
- Almendras, F.E. y Fuentealba, I.C. 1997. Salmonid rickettsial septicemia caused by *Piscirickettsia salmonis* : a review. *Dis Aqual Org* 29:137-44.
- Black, W.C. y Vontas, J.G. 2007 Affordable assays for genotyping single nucleotide polymorphism in insects. *Insect Molecular Biology* 16:377-87.
- Burridge, L., Weis, J., Cabello, F. y Pizarro, J. 2008. Chemical use in salmon aquaculture: A review of current practices and posible enronmental effects. WWF
- Conant, G.C. y Wagner, A. 2003. Asymmetric sequence divergence of duplicate genes. *Genome Res.* 13:2052–58.
- Ewing, B. y Green P. 1998 Basecalling of automated sequencer traces using phred. II. Error probabilities *Genome Research* 8:186-194.
- Ewing, B., Hillier, L., Wendl, M. y Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.*8:175-185.
- Fitch, W.M. 2000. Homology a personal view of some of the problems. *Trends Genets.* 16:227-231.
- Garcia, M.A., Gil, J., Ventoso, I., Guerra, S., Domingo, E.,Rivas, C. y Esteban, M. 2006 Impact of protein kinase PKR in cell biology: from antiviral to antiproliferative action. *Microbiol Mol Biol Rev* 70:1032–1060.
- Gordon, D., Abajian, C. y Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Research.* 8:195-202.
- Hayes, B.J., Nilsen, K., Berg, P.R., Grindflek, E. y Lien, S. 2007. SNP detection exploiting multiple sources of redundancy in large EST collections improves validation rates. *Bioinformatics.* 23(13):1692-1693.
- Hermann, M.G., Durtschi, J.D., Bromley, L.K., Wittwer, C.T. y Voelkerding, K.V. 2006 Amplicon DNA melting analysis for mutation scanning and genotyping: Cross-platform comparison of instruments and dyes. *Clin Chem* 52:494-503.
- Jørgensen, S.V., Afanasyev, S. y Krasnov, A. 2008 Gene expression analyses in Atlantic salmon challenged with infectious salmon anemia virus reveal differences between individuals with early, intermediate and late mortality. *BMC Genomics.* 9:179-195.
- Kim, S. y Misra, A. 2007 SNP genotyping: Technologies and biomedical applications. *Annu. Rev. Biomed. Eng.* 9:289–320.
- Landi, S. y Tavtigian, S.V. 2009. Determining the effectiveness of High Resolution Melting analysis for SNP genotyping and mutation scanning at the TP53 locus. *BMC Genet.* 10:5-17.

- Larsen, R., Røkenes, T.P. y Robertsen, B. 2004. Inhibition of Infectious Pancreatic Necrosis Virus Replication by Atlantic Salmon Mx1 Protein. *J of Virology* 78:7938-44.
- Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. y Higgins D.G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948.
- Leal, J. y Woywood, D. 2007 Piscirickettsiosis en Chile: Avances y Perspectivas para su Control. *SalmoCiencia* vol. 2.
- Liew, M., Pryor, R., Palais, R., Meadows, C., Erali, E., Lyon, E. y Wittwer, C. 2004 Genotyping of single-nucleotide polymorphism by high-resolution melting of small amplicons. *Clin Chem* 50:1156-64.
- Lynch, M y Conery, J.S. 2000 The Evolutionary Fate and Consequences of Duplicate Genes. *Science*. 290:1151-5.
- Marth, G.T., Yandell M.D., Korf, I., Gu, Z., Yeh, R.T., Zakeri, H., Stitzel, N.O., Hillier, L., Kwok, P.Y. y Gish, W. 1999. A general approach to single-nucleotide polymorphism discovery. *Nature Genetics*. 23:452-456.
- Matukumalli, L.M., Grefenstette, J.J., Hyten, D.L., Choi, i.Y., Cregan, P.B. y Van Tassell, C.P. 2006 Application of machine learning in SNP discovery. *BMC Bioinformatics* 7:4.
- Mazumder, B., Seshadri, V. y Fox, P.L. 2003 Translational control by the 3'-UTR: the ends specify the means. *TRENDS in Biochemical Sciences* 28:91-98.
- Míquel, A., Müller, I., Ferrer, P., Valenzuela, P.D.T. y Burzio, L.O. 2003. Immunoresponse of Coho salmon immunized with a gene expression library from *Piscirickettsia salmonis*. *Biol Res*. 36:313-23.
- Montgomery, J., Wittwer, C.T., Palais, R. y Zhou, L. 2007 Simultaneous mutation scanning and genotyping by high-resolution DNA melting analysis. *Nature protocols* 2:59-66.
- Oliver, D.J., Nikolau, B. y Wurtele, E.S. 2002 Functional Genomics: High-Throughput mRNA, Protein, and Metabolite Analyses. *Metabolic Eng* 4:98-106
- Olsen, A.B., Melby, H.B., Speilberg, L., Evensen, O. y Hastein, T. 1997. *Piscirickettsia salmonis* infection in Atlantic salmon *Salmo Salar* in Norway-epidemiological, pathological and microbiological findings. *Dis Aquat Org*. 31:35-48.
- Qumsiyeh, M. B. 1994. Evolution of number and morphology of mammalian chromosomes. *Journal of Heredity* 85:455-465.
- Rise M.L., Jones S. R. M., Brown G.D., von Schalburg, K.R., Davidson, W.S. y Koop, B.F. 2004 Microarray analyses identify molecular biomarkers of Atlantic salmon macrophage and hematopoietic kidney response to *Piscirickettsia salmonis* infection. *Physiol Genomics* 20: 21-35

SalmonChile, INTESAL. Uso de antibióticos en la industria del salmón.

Seeb, J.E., Pascal, C.E., Grau, E.D., Seeb, L.W., Templin, W.D. y Roberts, S.B. 2011 Transcriptome sequencing and high-resolution melt analysis advance single nucleotide polymorphism discovery in duplicated salmonids. *Molecular Ecology Resources* 11:335-48.

Smith, C.T., Elfstrom, C.M., Seeb, L.W. y Seeb, J.E. 2005. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. *Mol Ecol* 14:4193:203

Sonnhammer, E.L. y Koonin, E.V. 2002. Orthology, paralogy and proposed classification for paralog subtypes. *Trends Genet.* 18:619-620.

Taylor, J.S., Van de Peer, Y. y Meyer, A. 2001 Genome duplication, divergent resolution and speciation. *Trends Genet.* 17:299-301.

Treff, N.R., Tao, X., M.S., Schillings, W.J., Bergh, P.A., Scott, R.T. y Levy, B. 2011 Use of single nucleotide polymorphism microarrays to distinguish between balanced and normal chromosomes in embryos from a traslocation carrier. *Fertility and Sterility*. In press.

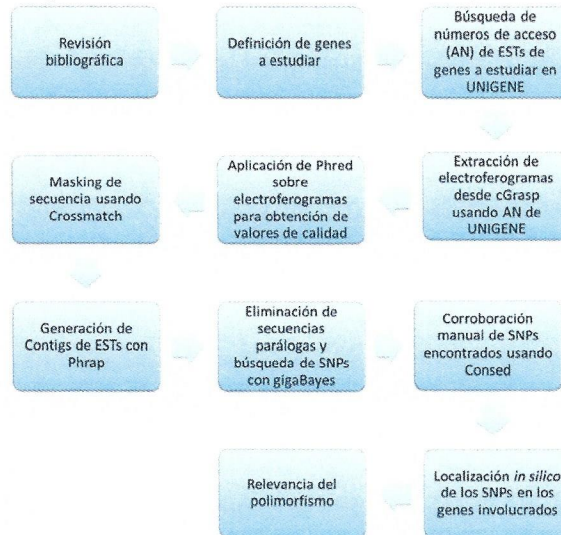
Vignal, A., Milan, D., Sancristobal, M. y Eggen, A. 2002 A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol.* 34:275-305

Wilhelm, V., Miquel, A., Burzio, L.O., Roseblatt, M., Engel, E., Valenzuela, S., Parada G. y Valenzuela, P.D.T. 2006. A vaccine against the salmonid pathogen *Piscirickettsia salmonis* based on recombinant proteins. *Vaccine* 24:5083-91.

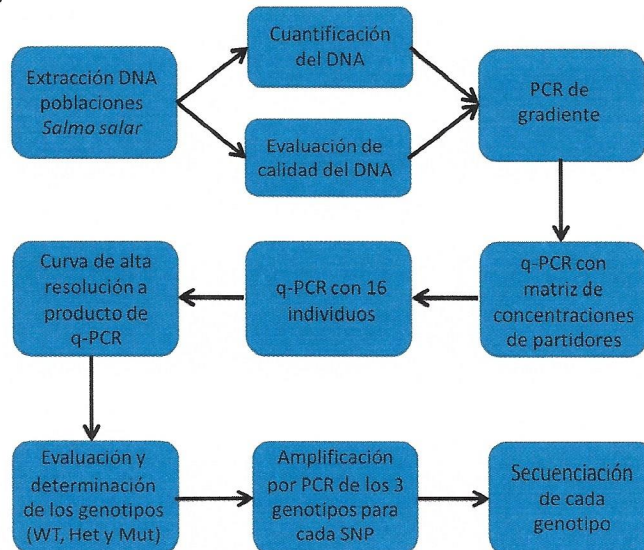
Young, N.D., Cooper, G.A., Nowak, B.F., Koop, B.F. y Morrison, R.N. 2007. Coordinated down-regulation of the antigen processing machinery in the gills of ameobic gill disease-affected Atlantic salmon (*Salmo salar L.*). *Molecular Immunology* 45:2581-97.

## ANEXOS

**Anexo 1.** Cadena de trabajo *in silico* para el descubrimiento y anotación de polimorfismos de un solo nucleótido (SNPs) en genes ligados al sistema inmune en *Salmo salar* (resumen sección 3.3).



**Anexo 2.** Cadena de trabajo experimental para la validación y genotipado de SNPs detectados *in silico* (resumen sección 3.6).



Anexo 3. Tabla con los Números de acceso de todos los ESTs ocupados en este trabajo para el descubrimientos *in silico* de polimorfismos tipo SNP

TAP2	TAP1	Mx1	PKR	IRF-1	Tiorredoxina						
CA040059.1	EG858744.1	CX355783.1	CA060728.1	CB507637.1	GO046132.1	CA048528.1	EG860118.1	EG767908.1	EG887092.1	EG926971.1	EG849687.1
CA052500.1	EG858745.1	DW540744.1	DW541730.1	CA038091.1	GO046133.1	CA063565.1	EG865660.1	EG767909.1	EG888258.1	EG928933.1	EG850342.1
CA060272.1	EG863828.1	DW551454.1	DY713889.1	CA039487.1	GO046983.1	CB517090.1	EG865661.1	EG784043.1	EG888259.1	EG928934.1	EG850343.1
DW545669.1	EG863829.1	DW563627.1	CB516446.1	CB510240.1	GO046984.1	DR696296.1	EG876677.1	EG784044.1	EG888860.2.1	EG930264.1	EG855130.1
DW553424.1	EG869093.1	DW563628.1	CX354492.1	CB510073.1	GO053534.1	DW542879.1	EG876688.1	EG785882.1	EG888860.3.1	EG930265.1	EG855131.1
DW559909.1	EG869094.1	DW571401.1	EG860664.1	CA039864.1	GO053935.1	DW565995.1	EG878881.1	EG785883.1	EG890339.1	EG935225.1	EG856172.1
DW576448.1	EG887912.1	DW580394.1	EG860675.1	CB499244.1	GO058385.1	DW565996.1	EG878882.1	EG797354.1	EG890350.1	EG935226.1	EG856173.1
DW576449.1	EG887913.1	DW580395.1	EG860675.1	CB499246.1	EG928660.1	DW568723.1	EG894160.1	EG797355.1	EG891376.1	EG936000.1	EG860619.1
DW584169.1	EG898826.1	DY716558.1	DW540178.1	CB498798.1	GE768160.1	DY708269.1	EG911343.1	EG811100.1	EG891377.1	EG936001.1	EG860621.1
DW584170.1	EG909445.1	DY730930.1	DW541731.1	CA044712.1	GE770933.1	DY712752.1	EG911344.1	EG811111.1	EG895484.1	EG937050.1	
DY706094.1	EG909824.1	DY730931.1	GE771962.1	CB503082.1	GE771770.1	DY712753.1	EG917337.1	EG825579.1	EG895486.1	EG938121.1	
DY712365.1	EG913827.1	EG831144.1	DW560429.1	CB505406.1	GE772487.1	DY729973.1	EG917338.1	EG825580.1	EG899565.1	EG938122.1	
DY724379.1	EG913828.1	GE770824.1	DW573128.1	CA059723.1	GE776776.1	DY729974.1	EG920070.1	EG825790.1	EG899566.1	EG938385.1	
DY739319.1	EG917114.1	GE770825.1	DY699946.1	CA060780.1	GE780922.1	EG761634.1	EG920071.1	EG825791.1	EG900404.1	EG938396.1	
DY739320.1	EG917115.1	GE778681.1	DY703480.1	CA059713.1	GE783792.1	EG761645.1	EG925729.1	EG826593.1	EG900405.1	EG941773.1	
EG801598.1	EG921139.1	GE778689.1	DY706638.1	CA061703.1	GE783850.1	EG772152.1	EG925732.1	EG826594.1	EG900937.1	EG942469.1	
EG806636.1	EG921140.1	GE784847.1	DY713888.1	CX354788.1	GE785622.1	EG772153.1	EG823381.1	EG838626.1	EG900938.1	EG942470.1	
EG806637.1	EG926839.1	GE784848.1	DY741139.1	DN048129.1	GE787601.1	EG787355.1	EG832127.1	EG838627.1	EG902579.1	EG875996.1	
EG826201.1	GE778108.1		DY741140.1	DW531555.1	GE789506.1	EG787357.1	EG832128.1	EG838826.1	EG902580.1	EG879972.1	
EG827261.1	GE778109.1		EG757847.1	DW534608.1	GE789900.1	EG790635.1	EG845896.1	EG838827.1	EG907068.1	EG879973.1	
EG827262.1	GE779558.1		EG757848.1	DW540764.1	GE791532.1	EG790636.1	EG845898.1	EG838927.1	EG907069.1	EG882057.1	
EG827512.1	GE779559.1		EG765746.1	DW547794.1	GE792301.1	EG791273.1	EG846728.1	EG839390.1	EG915807.1	EG882058.1	
EG827513.1	GE791524.1		EG765747.1	DW549147.1	GE795909.1	EG791274.1	EG846729.1	EG839391.1	EG915808.1	EG884177.1	
EG831178.1	GE791525.1		EG857335.1	DW551230.1	DY733401.1	EG791274.1	EG858089.1	EG839807.1	EG916376.1	EG884178.1	
EG831179.1	GE793500.1		EG857336.1	DW553554.1	DY737255.1	EG805960.1	EG858090.1	EG839810.1	EG916387.1	EG886269.1	
EG845383.1	GE793501.1		EG872871.1	DW554740.1	EG812117.1	EG811790.1	EG860117.1	EG841303.1	EG916659.1	EG886270.1	
EG845384.1			EG925240.1	DW565268.1	EG832165.1	EG811791.1		EG841304.1	EG916660.1	EG887091.1	
EG851023.1			GE771961.1	DW572470.1	EG832452.1	EG812052.1		EG841772.1	EG918230.1	EG861767.1	
EG853583.1			GO064289.1	DW582708.1	EG832454.1	EG812053.1		EG841773.1	EG918232.1	EG861892.1	
EG853584.1				DY706264.1	EG864751.1	EG812871.1		EG846890.1	EG918903.1	EG861893.1	
EG854841.1				DY717864.1	EG925840.1	EG812872.1		EG846891.1	EG918904.1	EG875995.1	
EG854852.1				DY718469.1	EG925843.1	EG820249.1		EG848774.1	EG920295.1	EG860686.1	
EG858457.1				DY721667.1	EG928659.1	EG820250.1		EG860698.1	EG920296.1	EG848776.1	
EG858458.1				DY725526.1		EG823380.1		EG861766.1	EG926970.1	EG849686.1	