

Tesis defendida por  
Rigoberto Delgado Vega  
y aprobada por el siguiente Comité

---

Dr. Miguel Ángel del Río Portilla  
Director del Comité

---

Dr. Axayácatl Rocha Olivares

Miembro del Comité

---

Dra. Carmen Guadalupe Paniagua  
Chávez

Miembro del Comité

---

Dr. Cristian Jorge Gallardo Escárte

Miembro del Comité

---

Dra. Fabiola Lafarga de la Cruz

Miembro del Comité

---

Dra. Beatriz Cordero Esquivel

Coordinador del Posgrado en Acuicultura

---

Dr. Jesús Favela Vara

Encargado del despacho de la Dirección  
de Estudios de Posgrado

Diciembre de 2013

**CENTRO DE INVESTIGACIÓN CIENTÍFICA Y DE EDUCACIÓN SUPERIOR  
DE ENSENADA, BAJA CALIFORNIA**



---

Programa de Posgrado en Ciencias  
en Acuicultura

---

Anotación de secuenciación masiva y variabilidad genética evaluada por microsatélites obtenidos mediante ESTs de la trucha de San Pedro Mártir *Oncorhynchus mykiss nelsoni*

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de

Maestro en Ciencias

Presenta:

Rigoberto Delgado Vega

Ensenada, Baja California, México

2013

Resumen de la tesis de Rigoberto Delgado Vega, presentada como requisito parcial para la obtención del grado de Maestro Ciencias en Acuicultura.

Anotación de secuenciación masiva y variabilidad genética evaluada por microsatélites obtenidos mediante ESTs de la trucha de San Pedro Mártir *Oncorhynchus mykiss nelsoni*

Resumen aprobado por:

---

Dr. Miguel Ángel del Río Portilla

## Resumen

La trucha de San Pedro Mártir *Oncorhynchus mykiss nelsoni* (Evermann, 1908) es una subespecie de trucha arcoíris, es endémica de Baja California y, debido al reducido número de organismos, distribución y ambiente altamente cambiante, se considera como especie vulnerable. Por lo anterior, actualmente se encuentra dentro de los planes de conservación y manejo impulsados por la Comisión Nacional de Acuicultura y Pesca (CONAPESCA), donde se busca fomentar su cultivo por encima de especies introducidas en la región. Para lograr este objetivo es necesario realizar investigación en el área de reproducción y el diseño de programas de seguimiento genético y selección a través de marcadores moleculares. En los últimos años se han desarrollado enormemente las técnicas secuenciación de ácidos nucleicos, dando lugar a una gran cantidad de secuencias o lecturas que facilitan el desarrollo de marcadores moleculares como por ejemplo los microsatélites o secuencias de repetición simple (SSR). También hay herramientas que facilitan la identificación y asignación de funciones en genes de interés comercial, lo cual ayuda en el desarrollo de programas de selección artificial asistida. El objetivo de este estudio incluyó la secuenciación masiva del genoma y anotación funcional genómica de *O. mykiss nelsoni*, así como el desarrollo de marcadores moleculares de tipo microsatélites a partir de marcas de secuencias expresables (EST-SSR) de la trucha arcoíris *Oncorhynchus mykiss mykiss*. La secuenciación de DNA se llevó a cabo con la plataforma MiSeq Sequencing Systems (Illumina) obteniendo 4 688 266 de lecturas con una longitud promedio de 240pb. Se ensamblaron las lecturas con el programa informático CLC Genomics Workbench V 6.0 para formar regiones contiguas (contigs), se anotaron genéticamente con el paquete informático Blast2GO y se identificaron microsatélites con el programa QDD de los cuales se determinó su posible asociación con la ontología genética de los loci que los contenían. Se obtuvieron 126 123 contigs con una longitud promedio de 293 pb. Sólo se pudieron anotar 9 800 contigs en términos de ontología genética y se diseñaron iniciadores para 6 842 posibles microsatélites, de los cuales, 6433 resultaron ser del tipo dinucleótidos, 318 trinucleótidos, 83 tetranucleótidos y 8 pentanucleótidos. Para los microsatélites a partir de EST (EST-SSR), se descargaron 287 967 EST de la trucha *O. mykiss mykiss* de la base de datos GenBank, se ensamblaron para obtener contigs, se anotaron genéticamente, se buscaron EST-SSR de manera similar a la secuenciación masiva y se amplificaron en *O. mykiss nelsoni*. Se ensamblaron 28 237 contigs, con una longitud promedio de 884 pb, de los cuales 12 500 se anotaron en términos de ontología genética. Se encontraron un total de 2 721 posibles microsatélites: 2452 dinucleótidos, 235 trinucleótidos y 34 tetranucleótidos. Se seleccionaron 20 EST-SSR (tri y tetranucleótidos) obtenidos de la base datos de ESTs de GenBank, de éstos 18 fueron monomórficos y dos polimórficos. El locus Omy17219 tuvo seis alelos, con una heterocigosidad esperada ( $H_e$ ) y observada ( $H_o$ ) de 0.692 y de 0.200 respectivamente. El coeficiente de endogamia ( $F_{IS}$ ) fue de 0.700, no se ajustó al Equilibrio de

Hardy-Weinberg (EHW) y mostró un déficit de heterocigotos. En el locus Omy17784 se detectaron siete alelos con una  $H_e$  de 0.805 y una  $H_o$  de 0.800 con valor de  $F_{IS}$  de -0.244 ajustándose al EHW. A partir de los datos de anotación genética se podrán desarrollar microsatélites relacionados con características de interés biológico, sin embargo los resultados de amplificación sugieren que la probabilidad de polimorfismo en regiones expresables puede ser muy baja en el caso de EST-SSR tri y tetranucleótidos. Esto da soporte a la hipótesis de que la trucha de SPM cuenta con una baja variabilidad genética en la región expresable de su genoma y por consiguiente debe de seguir en protección y esta información debe tomarse en cuenta para cualquier programa de cultivo y de mejoramiento genético.

Palabras clave: *Oncorhynchus mykiss nelsoni*, secuenciación masiva, anotación funcional, microsatélites.

Abstract of the thesis presented by Rigoberto Delgado Vega as a partial requirement to obtain the Master of Science degree with orientation in Aquaculture.

Annotation of massive sequencing and genetic variability evaluated by microsatellite obtained from ESTs of the San Pedro Martir trout *Oncorhynchus mykiss nelsoni*

Abstract approved by:

---

Dr. Miguel Angel del Río Portilla

## Abstract

The San Pedro Martir trout, *Oncorhynchus mykiss nelsoni* (Evermann, 1908), is a subspecies of the rainbow trout and is endemic of Baja California, Mexico; due to the low population size, limited distribution range, and the drastic changes in the environment where it lives, it is considered as a vulnerable species. However, nowadays it is currently under the conservation and management plans driven by the National Commission of Aquaculture and Fisheries (CONAPESCA), which seeks to promote its cultivation over the introduced species. To accomplish this goal it is necessary to do research to give support for genetic breeding, design genetic, selection and monitoring programs. In recent years, there has been a great development on nucleic-acid sequencing techniques, giving rise to a large number of sequences or reads, facilitating the development of molecular markers, such as microsatellites or simple sequence repeats (SSR). These tools also allow the identification and gene-sequence assignment of commercially important traits. In this sense the functional annotation is essential for the identification of quantitative trait loci useful on the development of assisted breeding programs. The objective of this study involved the massive sequencing, the genomic functional annotation of *O. mykiss nelsoni*, and the development of microsatellite molecular markers from expressible sequence tags (EST-SSR) of rainbow trout *Oncorhynchus mykiss mykiss*. Sequencing was carried out with the platform MiSeq Sequencing Systems (Illumina) obtaining 4 688 266 sequences with an average length of 240 pb. Reads were assembled with CLC Genomics Workbench software V 6.0 to form contiguous regions (contigs), and were genetically annotated with Blast2GO software. Microsatellites were identified with QDD software of which were determined the possible association to gene ontology. 126 123 contigs were obtained with an average length of 293 bp, only could annotate 9800 contigs in terms of Gene Ontology and designed primers for 6 842 potential microsatellites, of which 6 433 were dinucleotide, 318 trinucleotide, 83 tetranucleotide and 8 pentanucleotide. For microsatellites from ESTs (EST-SSR), 287 967 ESTs of rainbow trout *O. mykiss mykiss* were downloaded from GenBank database, reads were assembled to obtain contigs, were genetically annotated, were searched EST-SSR similarly to massive sequencing and were amplified in *O. mykiss nelsoni*, 28 237 contigs with an average length of 884 bp were assembled and 12 500 contigs were annotated in terms of Gene Ontology. We found a total of 2 721 potential microsatellites, of which 2 452 were dinucleotide, 235 trinucleotide and 34 tetranucleotide. 20 EST-SSR (tri and tetranucleotide) were selected, of these 18 were monomorphic and two polymorphic. The locus Omy17219 had six alleles with expected ( $H_e$ ) and observed ( $H_o$ ) heterozygosities of 0.692 and 0.200 respectively. The inbreeding coefficient ( $F_{IS}$ ) was 0.700, was not adjusted to Hardy-Weinberg equilibrium (HWE), showing a deficit of heterozygotes. In the Omy17 784 locus, seven alleles were detected with a  $H_e$  of 0.805 and  $H_o$  of 0.800, with  $F_{IS}$  value of -0.244 adjusting to HWE. From the annotation data, it is

possible to develop microsatellites related to characteristics of biological interest, however amplification results suggest that the probability of polymorphism in expressible regions is very low in the case of tri and tetranucleotide EST-SSRs in this species. This supports the hypothesis that the SPM trout has a low genetic variability in the expressed region of the genome and therefore the SPM trout must continue under protection and this information must be taken into account in any aquaculture and genetic improvement program.

**Keywords:** *Oncorhynchus mykiss nelsoni*, massive sequencing, functional annotation, microsatellite.

## **Dedicatorias**

A mis padres, Aurelia Vega Jacobi y Rigoberto Delgado Parra, por apoyarme en todo momento y en cada una de las etapas de mi formación humana y académica.

A mis hermanos, David, Fabricio y Mauricio, ya que soy un poco de cada uno de ellos.

A toda mi familia, en especial a mi abuela Socorro que sin su ayuda y soporte todo hubiera sido más difícil.

A Sara, por estar conmigo todo este tiempo y por su incondicional apoyo.

## **Agradecimientos**

Al Dr. Miguel Ángel del Río, por darme la oportunidad de ser su estudiante, por su paciencia, enseñanzas y por haber dirigido este trabajo con la dedicación y profesionalismo que lo caracteriza.

A los miembros del comité de tesis, Dr. Axayácatl Rocha Olivares, Dra. Carmen Paniagua Chávez, Dr. Cristian Gallardo Escárate y Dra. Fabiola Lafarga De La Cruz, por sus sugerencias, consejos, correcciones y por su valiosa ayuda que me brindaron durante el desarrollo de esta tesis.

Al Dr. Gorgonio Ruiz Campos por su ayuda en la colecta de los organismos utilizados en este trabajo, por sus consejos y por su siempre amable disponibilidad.

A mis compañeros de laboratorio, Edgar López Landavery y Luis Miguel Molina Carrillo, por las horas de ayuda prestada, sugerencias y compañía durante la realización de los trabajos de laboratorio.

A Carmen Vargas, técnico del laboratorio de genética, por su ayuda en momentos clave y por mantener siempre el orden en nuestra área de trabajo.

A todos mis compañeros generación, Araceli Cazares Salazar, David Eduardo Guzmán Lerma, Jorge Madrid Hernández, José Pablo Fuentes Quesada, Luis Miguel Molina Carrillo, Miriam Lecuanda Ontiveros, Omar Guillermo Montes González, Raquel Escuredo Vielba, Roberto Cruz Flores y especialmente a Paul Alberto Vásquez Gallegos por cuidar de las truchas.

Al Dr. Jorge Arturo Cid Becerra, a la Lic. Miriam Gicela Gastélum García y a Restauración de Ecosistemas A.C., por su invaluable e integral apoyo que me han ofrecido desde que nos conocimos.

Al Dr. Francisco Javier García De León por haberme permitido realizar una estancia en su laboratorio por su hospitalidad, atenciones, por sus enseñanzas y consejos.

Al Centro de Investigación Científica y de Educación Superior de Ensenada y al Departamento de Acuicultura, por brindarme las herramientas y recursos necesarios durante la realización de este trabajo y por el apoyo recibido de los investigadores, técnicos y personal administrativo.

Al consejo Nacional de Ciencia y Tecnología por el apoyo económico otorgado a través de la beca de estudios de posgrado.

## Tabla de Contenido

|  | <b>Página</b> |
|--|---------------|
| Resumen.....   | ii            |
| Abstract.....  | iv            |
| Dedicatorias.....  | vi            |
| Agradecimientos.....   | vii           |
| Lista de figuras.....  | xi            |
| Lista de tablas.....   | xiv           |
| 1.- Introducción.....  | 1             |
| 1.1.- Trucha de San Pedro Mártir.....  | 1             |
| 1.2.- Marcadores moleculares.....  | 2             |
| 1.3.- Microsatélites.....  | 3             |
| 1.4.- ESTs.....  | 4             |
| 1.5.- Secuenciación.....   | 5             |
| 1.6.- Anotación Genética.....  | 7             |
| 1.7.- Análisis Genético Poblacional.....   | 8             |
| 2.- Hipótesis.....   | 10            |
| 3.- Objetivos.....   | 10            |
| 3.1.- Objetivo general.....  | 10            |
| 3.2.- Objetivos específicos.....   | 10            |
| 4.- Materiales y métodos.....  | 10            |
| 4.1.- Muestreo.....  | 10            |
| 4.2.- Extracción de DNA.....   | 11            |
| 4.3.- Secuenciación Masiva de trucha de San Pedro Mártir.....                              | 11            |
| 4.4.- Anotación genética de los contigs de la trucha de San Pedro Mártir.....              | 11            |
| 4.5.- Búsqueda de microsatélites por secuenciación masiva.....                             | 12            |
| 4.6.- Selección de iniciadores para microsatélites obtenidos por secuenciación masiva..... | 13            |
| 4.7.- ESTs.....  | 13            |
| 4.8.- Anotación genética de ESTs de trucha arcoíris.....                                   | 13            |
| 4.9.- Búsqueda de Microsatélites y selección de iniciadores en ESTs.....                   | 13            |
| 4.10.- Amplificación de EST-SSR.....   | 14            |
| 4.11.- Análisis genético poblacional.....  | 14            |
| 5.- Resultados.....  | 16            |
| 5.1.- Muestreo.....  | 16            |

|   |    |
|---|----|
| 5.2.- Extracción de DNA .....   | 16 |
| 5.3.- Secuenciación masiva de trucha de San Pedro Mártir.....   | 18 |
| 5.4.- Anotación de contigs formados a partir de secuenciación masiva de la trucha de San Pedro Mártir.<br>..... | 18 |
| 5.5.-Búsqueda de microsatélites por secuenciación masiva .....  | 22 |
| 5.6.-Anotación genética de microsatélites obtenidos por secuenciación masiva. ....                              | 23 |
| 5.7.- Análisis de las secuencias EST de trucha arcoíris .....   | 29 |
| 5.8.- Búsqueda de microsatélites en secuencias EST.....   | 34 |
| 5.9.- Selección de iniciadores para EST-SSR.....  | 37 |
| 5.10.- Amplificación de EST-SSR.....  | 40 |
| 5.11.- Análisis genético poblacional .....  | 40 |
| 6.- Discusión.....  | 41 |
| 7.- Conclusiones .....  | 49 |
| 8.- Recomendaciones .....   | 51 |
| Referencias bibliográficas .....  | 52 |
| Anexo .....   | 59 |

## Lista de figuras

**Página**

|   |    |
|---|----|
| Fig. 1. Ubicación de la Sierra de San Pedro Mártir, en Baja California. ....  | 1  |
| Fig. 2. Costo en dólares por secuenciar un millón de bases (A, G, T, C). Tomada de NHGRI, 2011.....   | 6  |
| Fig. 3 Diagrama de flujo de secuenciación, ensamblaje, búsqueda de microsatélites, diseño de iniciadores y anotación genética. ....   | 12 |
| Fig. 4. Muestreo de la trucha de San Pedro Mártir <i>Oncorhynchus mykiss nelsoni</i> en el arroyo San Rafael. Captura de la trucha en el Arroyo San Rafael (A), jaulas donde se mantuvieron las truchas capturadas (B), hieleras donde se trasladaron los organismos con oxígeno saturado (C), bolsa con truchas de San Pedro Mártir en aclimatación para introducirlas a los estanques (D), estanques del sistema de recirculación (E). .... | 16 |
| Fig. 5. Truchas de San Pedro Mártir, en los círculos se muestra el corte realizado en aleta caudal.....   | 17 |
| Fig. 6. Integridad de DNA de trucha de San Pedro Mártir <i>Oncorhynchus mykiss nelsoni</i> . ....   | 17 |
| Fig. 7. Distribución de secuencias anotadas durante el análisis de los contigs obtenidos de la secuenciación masiva. ....   | 18 |
| Fig. 8. Distribución de especies por número de secuencias identificadas durante el BLAST. ....  | 19 |
| Fig. 9. Distribución de secuencias identificadas con respecto al valor E. ....  | 20 |
| Fig. 10. Distribución del número de anotaciones con respecto al nivel de ontología genética, donde P = Proceso biológicos, F = Función molecular y C = Componente celular. ....   | 20 |
| Fig. 11. Clasificación de Ontología genética (GO) de las secuencias de la trucha de San Pedro Mártir basada en los procesos biológicos.....   | 21 |
| Fig. 12. Clasificación de Ontología genética (GO) de las secuencias de trucha de San Pedro Mártir basada en la función molecular.....   | 22 |
| Fig. 13. Clasificación de Ontología genética (GO) de las secuencias de trucha de San Pedro Mártir basada en el componente celular.....  | 22 |
| Fig. 14. Distribución de microsatélites expresado en porcentaje según el tipo de repetición. ....   | 23 |

|   |    |
|---|----|
| Fig. 15. Clasificación de la categoría componente celular en términos de ontología genética para las secuencias que contienen microsatélites obtenidos por secuenciación masiva.....  | 23 |
| Fig. 16. Clasificación de la categoría función molecular en términos de ontología genética para las secuencias que contienen microsatélites obtenidos por secuenciación masiva.....   | 24 |
| Fig. 17. Clasificación de la categoría procesos biológicos en términos de ontología genética para las secuencias que contienen microsatélites obtenidos por secuenciación masiva.....   | 25 |
| Fig. 18. Distribución de la secuencia de repetición dentro de los microsatélites dinucleótidos, expresados en número y porcentaje. ....   | 25 |
| Fig. 19. Distribución de la secuencia de repetición dentro de los microsatélites trinucleótidos, expresados en número y porcentaje, no se muestra la categoría “otros”, que representa el 27.3% con 87 diferentes motivos de repetición.....  | 26 |
| Fig. 20. Distribución de la secuencia de repetición dentro de los microsatélites tetranucleótidos, expresados en número y porcentaje, no se muestra la categoría “otros”, que representa el 50.6% con 42 distintos motivos de repetición..... | 26 |
| Fig. 21. Clasificación de ontología genética para las secuencias que contienen microsatélites dinucleótidos.....  | 27 |
| Fig. 22. Clasificación de ontología genética para las secuencias que contienen microsatélites trinucleótidos. ....  | 28 |
| Fig. 23. Clasificación de ontología genética para las secuencias que contienen microsatélites trinucleótidos. ....  | 28 |
| Fig. 24. Distribución de secuencias anotadas durante el análisis de los contigs obtenidos de secuencias EST. ....   | 29 |
| Fig. 25. Distribución de especies por número de secuencias EST identificadas durante el BLAST. ....   | 30 |
| Fig. 26. Distribución de secuencias EST identificadas con respecto al valor E.....  | 30 |
| Fig. 27. Distribución del número de anotaciones de las secuencias EST con respecto al nivel de ontología genética, donde P = Proceso biológicos, F = Función molecular y C = Componente celular. ....   | 31 |

|   |    |
|---|----|
| Fig. 28. Clasificación de Ontología genética (GO) de los contigs obtenidos a partir de secuencias EST basada en los procesos biológicos. Total de secuencias 33 156. .... | 32 |
| Fig. 29. Clasificación de Ontología genética (GO) de los contigs obtenidos a partir de secuencias EST basada en la función molecular. Total de secuencias 17 497. ....    | 33 |
| Fig. 30. Clasificación de Ontología genética (GO) de los contigs obtenidos a partir de secuencias EST basada en el componente celular. Total de secuencias 21 862. ....   | 34 |
| Fig. 31. Distribución de microsatélites EST expresado en porcentaje según el tipo de repetición. ....   | 34 |
| Fig. 32. Distribución de la secuencia de repetición dentro de los microsatélites EST dinucleótidos, expresados en número y porcentaje. ....                               | 35 |
| Fig. 33. Distribución de la secuencia de repetición dentro de los microsatélites EST trinucleótidos, expresados en número y porcentaje. ....                              | 36 |
| Fig. 34. Distribución de la secuencia de repetición dentro de los microsatélites EST tetranucleótidos, expresados en número y porcentaje. ....                            | 36 |
| Fig. 35. Gel de agarosa con los productos de PCR al amplificar el locus Omy16187. ....  | 39 |

## Lista de tablas

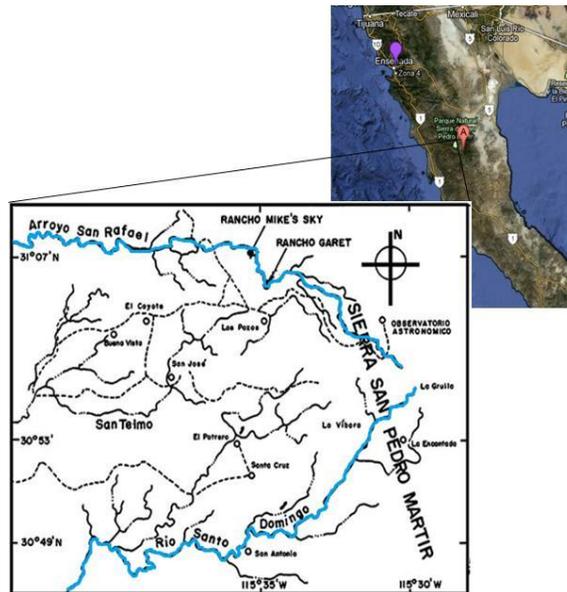
**Página**

|  |    |
|--|----|
| Tabla 1. Comparación de los instrumentos de secuenciación, ordenados según la relación del costo por Mb. (Glenn, 2011). .....                        | 7  |
| Tabla 2. Iniciadores (primers) para amplificar microsatélites obtenidos a partir de EST de trucha arcoíris <i>Oncorhynchus mykiss mykiss</i> .....   | 37 |
| Tabla 3. Anotación genética de secuencias EST con su anotación genética y porcentaje de similitud .....  | 39 |
| Tabla 4. Diversidad alélica de microsatélites polimórficos amplificados en la trucha de San Pedro Mártir <i>Oncorhynchus mykiss nelsoni</i> .....    | 40 |
| Tabla 5. Frecuencias alélicas de los microsatélites polimórficos. ....   | 40 |
| Tabla 6. Comparación de resultados de secuenciación y ensamblaje de ESTs en tres especies de salmonidos.....   | 43 |
| Tabla 7. Comparación de resultados de análisis de microsatélites en truchas.....   | 48 |
| Tabla 8. Iniciadores de microsatélites tetra y pentanucleótidos sin anotación genética.....  | 59 |
| Tabla 9. Iniciadores de microsatélites tetranucleótidos con anotación genética.....  | 61 |
| Tabla 10. Anotación genética de contigs con contenido de microsatélites tetra y pentanucleótidos de la secuenciación masiva de la trucha de SPM..... | 62 |

## 1.- Introducción

### 1.1.- Trucha de San Pedro Mártir

En México, existen por lo menos seis especies de truchas nativas distribuidas en Baja California, Sonora, Chihuahua, Durango y Sinaloa. Estas especies se encuentran amenazadas por la escasez de agua, pesca furtiva y deterioro del hábitat. Una de las especies es la trucha de San Pedro Mártir, *Oncorhynchus mykiss nelsoni* (Evermann, 1908), la cual es una subespecie de trucha arcoíris (Miller, 1950; Smith, 1984). La trucha de San Pedro Mártir (SPM) habita en la reserva de la biosfera de la Sierra de San Pedro Mártir en Baja California (Fig. 1), se encuentra en arroyos de poca profundidad con sustrato arenoso, donde se alimenta de larvas y pupas de insectos, es endémica de la región y es considerada como especie nativa de México (Ruiz-Campos, 1993). Originalmente, la trucha de SPM habitaba solamente en el arroyo de San Antonio y La Zanja dentro de la cuenca de Santo Domingo (Evermann, 1908; Nelson, 1921; Snyder, 1926; Smith, 1991). Sin embargo, entre los años de 1929 y 1941 el naturalista Charles Edward Utt realizó movimiento de organismos hacia los arroyos de La Grulla, Potreo y San Rafael (Ruiz-Campos y Pister, 1995). De esta manera, actualmente se reconocen cinco poblaciones establecidas entre el Río Santo Domingo y el arroyo San Rafael (Ruiz-Campos, 1993).



**Fig. 1. Ubicación de la Sierra de San Pedro Mártir, en Baja California. Recuperada <https://maps.google.com.mx/>, 2013.**

Sus poblaciones han sido reportadas como estables (Ruiz-Campos y Pister, 1995), aunque, se considera como especie vulnerable debido a las alteraciones ambientales (incendios, reducción del hábitat de origen antropocéntrica, pesca furtiva, contaminación de las aguas, utilización del agua por parte de la agricultura) que pudieran afectar su supervivencia (Jelks *et al.*, 2008). Este organismo presenta una baja tasa de crecimiento en vida silvestre, sin embargo, tiene características que le confieren potencial acuícola, por ejemplo: Puede sobrevivir en amplio intervalo de temperatura, su madurez la alcanza durante su primer año de vida, con una talla de alrededor de 100 mm en las hembras y alrededor de 80 mm en los machos (Ruiz-Campos, 1993), además no tiene comportamiento migratorio (Needham, 1938). En este sentido, recientemente se ha comprobado que puede ser mantenida en cautiverio en sistemas de recirculación (Aguilar-Juárez, 2010).

Actualmente la trucha de SPM se encuentra dentro de los planes de conservación y manejo impulsados por la Comisión Nacional de Acuicultura y Pesca (CONAPESCA), donde se busca fomentar su cultivo por encima de especies introducidas. Sin embargo, para lograr estos objetivos es necesario realizar investigación en el área de reproducción y el diseño de programas de seguimiento genético y selección, con el fin de lograr el desarrollo de su cultivo. Asimismo, se considera que para lograr su conservación y óptimo aprovechamiento es necesario que se realicen estudios de variabilidad genética y de identificación de genes con importancia comercial, utilizando técnicas de biología molecular y marcadores genéticos. No obstante, estas herramientas se consideraban hace poco más de 10 años caras y tediosas en su desarrollo, sin embargo con las nuevas tecnologías se han disminuido los costos y los procesos de secuenciación y análisis se han vuelto más rápidos (Métais *et al.*, 2002).

## **1.2.- Marcadores moleculares**

En los últimos años se han desarrollado tecnologías que nos ayudan en este tipo estudios, como es el caso de la secuenciación masiva y su uso en el desarrollo de marcadores moleculares como por ejemplo los microsatélites, estas herramientas también facilitan la identificación y asignación de funciones en genes de interés comercial (Pareek *et al.*, 2011). Al contar con esta tecnología se agilizan las labores de investigación y resultan muy útiles a la hora de sentar las bases para el diseño de planes de selección asistida (Martínez, 2005).

Para realizar los estudios de diversidad genética existen diferentes técnicas que abarcan distintos enfoques y también los marcadores moleculares utilizados son muy variados, donde su elección dependerá de los objetivos que hayan planteado y de las hipótesis que se intenten resolver. Dentro de los marcadores moleculares, los más utilizados son: las alozimas, RFLPs (Restriction Fragment Length Polymorphism), RAPDs (Randomly Amplified Polymorphic DNA), AFLPs (Amplified Fragment Length Polymorphism), SNPs (Single Nucleotide Polymorphisms) y microsatélites también conocidos como SSR (Single Sequence Repeats) (Liu y Cordes, 2004, Azofeifa, 2006).

### **1.3.- Microsatélites**

Los microsatélites son marcadores moleculares localizados en todo el genoma, tanto en regiones codificantes, como no codificantes (Liu y Cordes, 2004). Están formados por secuencias cortas de nucleótidos repetidas en tándem (una detrás de la otra). También se les conoce como secuencias simples de repetición (SSR, por sus siglas en inglés, simple sequence repeats). Los motivos de repetición que forman a los microsatélites pueden ser de dos a seis pares de bases, clasificándolos en di, tri, tetra, penta y hexanucleótidos, siendo más comunes los de tipo dinucleótido. Por otro lado, también se clasifican en puros y compuestos según estén formados por uno o más motivos de repetición, o se encuentren interrumpidos ya sea por el cambio de uno o más nucleótidos o por la falta de algún nucleótido (Zane *et al.*, 2002; Aranguren *et al.*, 2005). También se dividen en tipo I, que están vinculados a los genes con función conocida, y tipo II, asociados a regiones genómicas no anotadas (O'Brien, 1991). Los SSRs de tipo I son generalmente más conservados, siendo más valioso su desarrollo en vez de marcadores de tipo II (Siemon *et al.*, 2005).

El análisis de microsatélites es una de las herramientas moleculares más usadas en la actualidad por sus diversas aplicaciones y cualidades, debido a que son marcadores genéticos altamente variables (Liu y Cordes, 2004). Estas variaciones están dadas por la longitud del fragmento entre individuos que dan lugar a diferentes alelos, exhibiendo una amplia variación alélica que los vuelve altamente polimórficos (Tautz, 1989; Litt y Lutty, 1989; Wright, 1993). Otras características que los hacen altamente informativos y prácticos son: herencia mendeliana simple, codominancia, selección neutra, tasa de mutación alta, los cuales son amplificables por PCR, fáciles de medir, analizar, y pueden ser altamente reproducibles y confiables (Aranguren *et al.*, 2005).

Debido a las anteriores características estos loci son usados comúnmente para la identificación de individuos y análisis poblacionales para detectar estructura genética, patrones de subdivisión geográfica y estimar niveles de flujo genético (Tautz 1989; Avise, 2004; Astaneï, *et al.*, 2005; Geist y Kuehn, 2005; Lind *et al.*, 2007; Zhan *et al.*, 2009).

A través de la comparación de genotipos se pueden realizar estudios en ecología y evolución, como la adaptación de las poblaciones a cambios ambientales, el efecto de la fluctuación del tamaño poblacional sobre caracteres adaptativos, la hibridación entre especies y otros.

El aislamiento tradicional de microsatélites con el fin de desarrollar iniciadores específicos del locus, requería mucho tiempo y mano de obra intensiva (Metzker, 2005). Sin embargo, actualmente las bases de datos públicas como GenBank, contienen un gran número de Expressed Sequence Tags (ESTs), las cuales además de su utilidad original, también permiten la búsqueda de microsatélites y el diseño de iniciadores para su amplificación. Esta estrategia, ha demostrado ser un método eficaz de bajo costo y con ahorro de tiempo para desarrollar microsatélites (Serapion *et al.*, 2004, Yue *et al.*, 2004, Wang *et al.*, 2005, Chen *et al.*, 2005, Pérez *et al.*, 2005; Maneeruttanarungroj *et al.*, 2006).

#### **1.4.- ESTs**

Las EST se obtienen al extraer el mRNA de distintos tejidos para después sintetiza DNA complementario (cDNA) por retrotranscripción, estos fragmentos de cDNA son clonados insertándolos en un vector para transformar células bacterianas y obtener bibliotecas que posteriormente serán secuenciadas (Bouck y Vision, 2007). Las secuencias obtenidas son depositadas en las bases de datos y quedan disponibles para realizar análisis y comparaciones por medio de programas de cómputo especializados.

Las ESTs proporcionan una gran cantidad de información acerca de las regiones expresadas en el genoma de los organismos que seguramente corresponden a genes. Además, son una fuente valiosa para el desarrollo de marcadores genéticos, como microsatélites y SNP (Liu y Cordes, 2004; Boussaha, 2012). En teleósteos, el organismo con más ESTs es el pez cebra, con más de 1300 millones de secuencias. Los salmónidos también están bien representados, con más de 500 mil ESTs, de los cuales casi 300 mil pertenecen a trucha arcoíris (Salem, *et al.*, 2010).

Con este nuevo método se pueden desarrollar microsatélites con un tercio del tiempo y con una cuarta parte del costo requerido al generarlos con el método tradicional. Además, al analizar

grandes cantidades de secuencias se pueden abarcar un mayor número de loci identificando miles de posibles microsatélites, con lo cual se pueden ajustar los criterios de selección y diseño de iniciadores a unos más estrictos con el fin de aumentar la probabilidad de éxito en la amplificación o seleccionar productos de PCR que en teoría tendrán tamaños diversos entre sí, buscando la optimización de PCR múltiple para analizar más loci con menos reacciones de PCR (Lepais y Bacles, 2001).

### **1.5.- Secuenciación**

Otra de las tendencias en el desarrollo de microsatélites es el diseño de iniciadores a partir de secuenciación masiva. La secuenciación del DNA fue descrita por primera vez por Maxam y Gilbert y por Sanger en el año 1977 y con este avance cambió la forma de estudiar y comprender a la genética (Pareek *et al.*, 2011). Desde su invención, la secuenciación ha mejorado continuamente, sin embargo la técnica de secuenciación tradicional es lenta y costosa.

Los avances en la tecnología de secuenciación de nueva generación ha revolucionado la manera de hacer ciencia biológica, ya que se obtiene una gran cantidad de fragmentos secuenciados (lecturas) en una sola corrida (Nowrousian, 2010; McCormack *et al.*, 2013). El acceso a grandes cantidades de datos de secuenciación, así como las mejoras posteriores de longitud de lectura y el rendimiento de las diferentes plataformas de secuenciación, está dando lugar a un mejor aprovechamiento a un costo razonable (Metzker, 2010).

Las nuevas tecnologías se están desarrollando rápidamente y tienen el potencial de acelerar dramáticamente la investigación en ecología, genética de poblaciones, así como también acelerar los trabajos en materia de selección (Shokralla *et al.* 2013). A consecuencia de esta revolución, han surgido distintos métodos de secuenciación, como la de secuenciación por síntesis después de la amplificación, en donde se incluyen técnicas como la pirosecuenciación, tecnología 454, Illumina y Solid (Nowrousian, 2010). Otro método es la secuenciación a partir de moléculas de DNA de cadena sencilla. La principal ventaja de estas nuevas tecnologías es que ofrecen un rendimiento mucho mayor con respecto a la secuenciación clásica, además se sustituye la clonación previa, por amplificación de fragmentos de DNA por PCR y en las tecnologías más recientes se elimina este paso, ya que las moléculas de DNA de cadena sencilla son secuenciadas de manera directa, reduciendo aún más los costos (NHGRI, 2013).



**Fig. 2. Costo en dólares por secuenciar un millón de bases (A, G, T, C). Recuperada de <http://www.genome.gov/sequencingcosts/>. NHGRI, 2013.**

Actualmente existen diferentes plataformas de secuenciación masiva en paralelo, como Roche/454, SOLiD Life Technologies, y las plataformas de Illumina. (Xiong *et al.* 2010). Estas han aumentado drásticamente el número de bases obtenidas por secuenciación, mientras que al mismo tiempo se han disminuido los costos por base (Nowrousian, 2010; Davey y Blaxter, 2010). En comparación con la secuenciación de Sanger, la secuenciación masiva o de nueva generación (NGS por sus siglas en inglés) produce lecturas de corta longitud, pero la diferencia más clara entre la secuenciación Sanger y NGS es la cantidad de datos que generan. Normalmente la secuenciación Sanger genera alrededor de 500 lecturas, mientras que un equipo Illumina genera de 1 millón a 2 millones (Stein, 2010). Por otro lado, estas cifras podrían ser de alguna manera engañosas, porque otra diferencia importante entre NGS y la secuenciación Sanger es el tamaño de las lecturas que van de 100 a 1000 dependiendo de la plataforma (Tabla 1). En este punto, un buen ensamblaje es primordial para tener una alta cobertura en las secuencias consenso.

Así, estas tecnologías han simplificado la secuenciación, reduciendo enormemente los costos y la velocidad en la que se puede secuenciar un genoma completo, permitiendo además secuenciar muchas muestras a la vez. Sin embargo, estas tecnologías también han requerido cambios profundos en el análisis de datos. El principal obstáculo en la investigación genómica ya no es la

producción de datos. Después del ensamblaje de las lecturas, el siguiente reto es analizar e interpretar las variantes genéticas, para ello se utilizan herramientas bioinformáticas capaces de realizar la anotación funcional de los datos, donde uno de los aspectos más importantes es asociar las secuencias con su función biológica (Cantacessi *et al.*, 2010).

Tabla 1. Comparación de los instrumentos de secuenciación, ordenados según la relación del costo por Mb. (Glenn, 2011), página 6.

| Plataforma            | Tiempo de corrida | Millones de lecturas/corrida | Longitud de lecturas (pb) | Costo de reactivos por corrida (dólares) | Costo de reactivos por Mb | Mb producidas por corrida |
|-----------------------|-------------------|------------------------------|---------------------------|--|---------------------------|---------------------------|
| 3730xl (capillary)    | 2h                | 0.000096                     | 650                       | \$96                                     | \$1,500.00                | 0.06                      |
| Ion Torrent chip 314  | 2h                | 0.1                          | 100                       | \$500                                    | <50                       | >10                       |
| 454 GS Jr. Titanium   | 10h               | 0.1                          | 400                       | \$1,100                                  | \$22.00                   | 50                        |
| Starlight             | nd                | <0.01                        | >1000                     | nd                                       | nd                        | nd                        |
| PacBio RS             | 0.5 - 2h          | 0.01                         | 860-1100                  | \$110-900                                | 11-180                    | 5-10                      |
| 454 FLX Titanium      | 10h               | 1                            | 400                       | \$6,200                                  | \$12.40                   | 500                       |
| 454 FLX+              | 18 - 20h          | 1                            | 700                       | \$6,200                                  | \$7.00                    | 900                       |
| Ion Torrent chip 316  | 2h                | 1                            | >100                      | \$750                                    | <7.5                      | >100                      |
| Helicos               | nd                | 800                          | 35                        | nd                                       | nd                        | 28000                     |
| Ion Torrent chip 318  | 2h                | 4-8                          | >100                      | \$925                                    | <0.93                     | >1000                     |
| Illumina MiSeq        | 26h               | 3.4                          | 150 + 150                 | \$750                                    | \$0.74                    | 1020                      |
| Illumina iScanSQ      | 8 días            | 250                          | 100 + 100                 | \$10,220                                 | \$0.20                    | 50000                     |
| Illumina GAIIx        | 14 días           | 320                          | 150 + 150                 | \$11,524                                 | \$0.12                    | 96000                     |
| SOLiD 4               | 12 días           | >840                         | 50 + 35                   | \$8,128                                  | <0.11                     | 71400                     |
| Illumina HiSeq 1000   | 8 días            | 500                          | 100 + 100                 | \$10,220                                 | \$0.10                    | 100000                    |
| Illumina HiSeq 2000   | 8 días            | 1000                         | 100 + 100                 | \$20,120                                 | \$0.10                    | 200000                    |
| SOLiD 5500            | 8 días            | >700                         | 75 + 35                   | \$6,101                                  | \$<0.08                   | 77000                     |
| SOLiD 5500xl          | 8 días            | >1410                        | 75 + 35                   | \$10,503                                 | \$<0.07                   | 155100                    |
| Illumina HiSeq 2000v3 | 10 días           | ≤3000                        | 100 + 100                 | \$23,470                                 | \$≥0.04                   | ≤600000                   |

## 1.6.- Anotación Genética

La anotación funcional automática es un método que permite la categorización de los genes en las clases funcionales y puede ser muy útil para comprender el significado fisiológico de grandes cantidades de genes y para evaluar las diferencias funcionales entre los subgrupos de secuencias (Conesa *et al.*, 2005). En este sentido, la anotación genética es fundamental para la identificación de loci asociados a caracteres cuantitativos (QTLs) y ayuda en el desarrollo de programas de selección artificial asistida por marcadores (Féral 2002, Liu y Cordes 2004, Schlotterer 2004, Hayes *et al.* 2006, Arif y Khan 2009).

## 1.7.- Análisis Genético Poblacional

Las actividades antropogénicas como la contaminación, cacería (pesca furtiva) o destrucción del hábitat usualmente resultan en reducciones sustanciales del tamaño de la población provocando cuellos de botella y una consecuente reducción de la heterocigosidad, lo cual lleva a las poblaciones a la pérdida de la variabilidad genética (Hartl y Clark, 1997; Hedrick, 2000). La variabilidad genética resulta esencial para la correcta adaptación y supervivencia de una población (Caro y Laurenson, 1994; Hedrick, 1996; O'Brien 1994; O'Brien *et al.*, 1987).

Una forma de medir la variabilidad genética en una población es a través de la proporción de individuos que poseen copias distintas de sus alelos dentro del genoma, los cuales son llamados heterocigotos. Al medir la proporción de organismos heterocigotos con respecto al total de organismos analizados, obtenemos el índice de Heterocigosidad (H). La riqueza alélica (número de alelos observados en una población), la heterocigosidad observada ( $H_o$ ) y la heterocigosidad esperada ( $H_e$ ), son los parámetros más utilizados en la evaluación de la diversidad genética. La  $H_o$  es calculada directamente de la población analizada, es comparada estadísticamente con la  $H_e$ , la cual es obtenida a partir del modelo de equilibrio de Hardy y Weinberg (EHW). Este modelo considera una población infinita de individuos con reproducción aleatoria, en donde no hay selección natural, ni mutación, ni migración. En caso de que la  $H_o$  no sea estadísticamente diferente de  $H_e$  se sugiere que la población está en EHW. Por otro lado, si la  $H_e$  difiere de  $H_o$ , se dice que la población está en desequilibrio, y se considera que alguno de los supuestos del modelo del EHW no se cumplen y por lo tanto se deben estudiar cuales son los factores (selección, migración, etc.) que están influyendo en la distribución de la estructura y la diversidad genética observada en esta población. Otros parámetros utilizados en la evaluación de la diversidad genética son el coeficiente de consanguinidad o endogamia ( $F_{IS}$ ) y el índice de fijación o diferenciación genética entre poblaciones ( $F_{ST}$ ), conocidas como las estadísticas F de Wright (Weir y Basten, 1990), que miden el grado de diferenciación genética intra e interpoblacional, calculando las varianzas estandarizadas de frecuencias alélicas entre poblaciones. Se puede calcular la significancia estadística para los valores de  $F_{ST}$  entre pares de poblaciones (Weir y Cockerham, 1984) para comprobar la hipótesis nula de una falta de diferenciación genética entre poblaciones, y por tanto, la división de la diversidad genética (Mburu *et al.*, 2003), también puede realizarse un análisis de varianza molecular (AMOVA) (Excoffier *et al.*, 1992) para evaluar la distribución de diversidad dentro y entre las poblaciones.

Hasta el momento, es poco lo que se ha hecho en el ámbito genético en esta especie, por ejemplo: Nielsen *et al.* (1996) encontraron que la trucha de San Pedro Mártir posee un alelo exclusivo determinado por microsatélites que permite diferenciarla de las truchas cabeza de acero y de la trucha arcoíris, por otro lado Berg y Gall (1988) y Villarreal-Zazueta (2012), determinaron mediante alozimas, que la especie se encuentra estable y no hay diferencia genética entre sus poblaciones, también se ha encontrado que al comparar los salmónidos del noroeste de México por medio de PCR-RFLP, la trucha de San Pedro Mártir tiene los valores de diversidad genética más bajos (Camarena-Rosales *et al.*, 2008).

Como se observa, aún es necesario realizar estudios más profundos, con el fin de desarrollar marcadores moleculares específicos, con los cuales se podrá evaluar el estado de la variabilidad genética de la trucha de San Pedro Mártir e identificar a las poblaciones mejor conservadas, además se necesita contar con información aprovechable en el diseño de planes de conservación adecuados, que permitan la fundación de grupos reproductores y sistemas de apareamiento que conserven la variabilidad genética bajo condiciones de cultivo. Por otro lado, se deben identificar secuencias genómicas específicas con posible valor ecológico y comercial.

## 2.- Hipótesis

---

La secuenciación masiva y anotación genética permitirá identificar genes de interés ecológico o comercial en la trucha *Oncorhynchus mykiss nelsoni* y se podrán diseñar iniciadores de microsatélites dentro de estas secuencias.

Debido al reducido tamaño de la población efectiva, la trucha *Oncorhynchus mykiss nelsoni* presentará una baja variabilidad genética en la población estudiada, al evaluarse mediante microsatélites desarrollados a partir de ESTs.

## 3.- Objetivos

---

### 3.1.- Objetivo general

Realizar la secuenciación masiva, anotación funcional, desarrollo de microsatélites y evaluación de la variabilidad genética de la trucha *O. mykiss nelsoni*.

### 3.2.- Objetivos específicos

Obtener la secuencia parcial del genoma de la trucha *O. mykiss nelsoni*, mediante secuenciación masiva

Desarrollar marcadores genéticos tipo microsatélites para la trucha *O. mykiss nelsoni* a partir de secuenciación masiva y bibliotecas de ESTs.

Evaluar la diversidad alélica, grado de polimorfismo y la heterocigosidad en la trucha *O. mykiss nelsoni* por medio de microsatélites derivados de ESTs.

## 4.- Materiales y métodos

---

### 4.1.- Muestreo

La recolecta de truchas de SPM, *O. mykiss nelsoni*, se realizó en la cuenca del arroyo San Rafael en la localidad del Rancho Mike's Sky de la Sierra de San Pedro Mártir (31°05'51.7" N, 115°47'16.3" O). Para ello se utilizó un equipo de electropesca Smith Root 15-B POW (110V CA). Una vez que se recolectaron las truchas, se colocaron en bolsas de plástico con agua tomada del mismo arroyo. Se inyectó oxígeno a las bolsas y se trasladaron en hieleras las cuales

contenían hielo para mantener la temperatura a 16°C hasta llegar al Departamento de Acuicultura del CICESE.

#### **4.2.- Extracción de DNA**

La extracción de DNA se realizó utilizando el kit de extracción por columnas “EZNA Tissue DNA KIT” (Omega Bio-Tek, USA) según las instrucciones del fabricante. Para ello se utilizaron 40 organismos, tomado 100 mg de aleta caudal de 12 organismos vivos y músculo de 28 congelados. El tejido se maceró en un mortero con la ayuda de nitrógeno líquido. Después de realizar la extracción de DNA se procedió a estimar su integridad por electroforesis en gel de agarosa al 1.5% y tinción con bromuro de etidio. A su vez, se midió la concentración y pureza del DNA por espectrofotometría con el equipo Nanodrop 2000 (Thermo Scientific) en las longitudes de onda 260/280, la cual indica la presencia de contaminación por proteínas y 260/230, que indica presencia de contaminaciones por reactivos como fenoles. Al terminar la extracción de DNA, se diluyeron las muestras a la concentración de trabajo, que fue de 50 ng  $\mu\text{l}^{-1}$  y se almacenó a -20°C, por último el DNA concentrado restante se almacenó a -80°C.

#### **4.3.- Secuenciación Masiva de trucha de San Pedro Mártir**

Para realizar la secuenciación se utilizaron ocho muestras de DNA a 50 ng  $\mu\text{l}^{-1}$  en un volumen de 20  $\mu\text{l}$  y se enviaron a secuenciar con la plataforma MiSeq Benchtop Sequencer (Illumina) a la Universidad de Georgia, E.U.A. Las lecturas recibidas a partir de la secuenciación se ensamblaron con el programa CLC Genomics Workbench V. 6.0 para formar los contigs y singletons, utilizando el algoritmo de Bruijin, la longitud mínima de contigs se fijó en 200 pb, el tamaño de las palabras se fijaron en 20 b, el tamaño de burbuja en 50 b, el costo de desfase fue de dos, el costo de inserción y deleción se fijó en tres, la longitud de fracción se ajustó en 0.5 y la fracción de similitud en 0.8. Una vez concluido el ensamblaje, se generó un archivo en formato FASTA para los análisis de anotación de los contigs obtenidos.

#### **4.4.- Anotación genética de los contigs de la trucha de San Pedro Mártir**

El archivo FASTA que contenía los contigs se analizó con el paquete informático Blast2GO (Conesa *et al.*, 2005), con el fin de encontrar genes de interés ecológico y/o comercial y de ser posible asociarlos a algún tipo de marcador molecular. El análisis se llevó a cabo buscando similitud de las secuencias con las depositadas GenBank y otras bases de datos públicas con Annex e Interpro. Las secuencias anotadas se clasificaron bajo los términos de ontología genética

(GO), asociando los resultados a las categorías de componentes celulares (CC), función molecular (FM), y procesos biológicos (PB).



**Fig. 3 Diagrama de flujo de secuenciación, ensamblaje, búsqueda de microsatélites, diseño de iniciadores y anotación genética.**

Después de realizar anotación para el total de contigs formados a partir de la secuenciación masiva, se llevó a cabo por separado la anotación únicamente para las secuencias que contienen microsatélites y se formaron dos grupos, anotadas y no anotadas. Una vez terminado lo anterior, se realizó la anotación genética de las secuencias con microsatélites, separándolas según su motivo de repetición (di, tri, tetra y pentanucleótidos). Lo anterior se llevó a cabo con el fin de tener distintos grupos de microsatélites, los que pertenecen a secuencias con función conocida y reportada y los que no cuentan con alguna función conocida hasta el momento, además de identificar de manera práctica las funciones a las que estén relacionados. De esta manera se podrá optar la utilización de marcadores moleculares provenientes de regiones expresables o bien provenientes de regiones en teoría neutras.

#### **4.5.- Búsqueda de microsatélites por secuenciación masiva**

Como siguiente paso, con el archivo FASTA de los contigs se procedió a buscar motivos de repetición de tipo microsatélite con ayuda del paquete informático QDD (Meglecz *et al.*, 2010), para posteriormente y utilizando el programa Primer3 (Rozen y Skaletsky, 2000) diseñar los iniciadores para la amplificación de los microsatélites seleccionados.

La búsqueda de microsatélites se realizó en secuencias con una longitud mínima de 80 pb (pares de bases), la longitud mínima y máxima del producto de PCR a amplificar se fijó en 90 y 300 pb respectivamente, la longitud mínima de los iniciadores se fijó en 18 pb, longitud máxima de 27 pb, con una longitud óptima de 20 pb. El contenido de GC dentro de los iniciadores se fijó en un óptimo del 50%, con 20% como mínimo y un máximo de 80%. La temperatura de alineamiento de los iniciadores se fijó entre 57 y 63°C, siendo la óptima 60°C

#### **4.6.- Selección de iniciadores para microsatélites obtenidos por secuenciación masiva**

Del total de microsatélites encontrados, se realizó una selección posterior, con los siguientes criterios: Microsatélites de tipo perfectos (no compuestos), tri y tetranucleótidos, con una profundidad de al menos de 4X en las lecturas que forman la región donde se encontraron los microsatélites e iniciadores. Otras condiciones de diseño fueron, no tener más de 2°C de diferencia en las temperaturas de alineamiento entre los iniciadores pares, contenido de GC mayor al 40% y longitud de iniciadores entre 18 y 27 pb.

#### **4.7.- ESTs**

Con el fin de enriquecer la base de datos de marcadores SSR en la trucha de SPM, *Oncorhynchus mykiss nelsoni*, se descargaron 287 967 ESTs de trucha arcoíris *Oncorhynchus mykiss mykiss* de la base de datos GenBank. Estas secuencias se ensamblaron con el programa informático CLC Genomics Workbench V 6.0, de la misma manera en que fueron ensambladas lecturas provenientes de la secuenciación masiva.

#### **4.8.- Anotación genética de ESTs de trucha arcoíris**

El archivo FASTA que contenía los contigs se analizaron con el paquete informático Blast2GO (Conesa *et al.*, 2005), con el fin de encontrar genes interés ecológico y/o comercial, y de ser posible asociarlos a algún tipo de marcador molecular. El análisis se llevó a cabo de igual forma que las lecturas obtenidas por secuenciación masiva. Además, se generaron gráficas combinadas de ontología genética independientes para cada tipo de microsatélites según su motivo de repetición, además de realizarse esta anotación para el total de contigs generados por ESTs.

#### **4.9.- Búsqueda de Microsatélites y selección de iniciadores en ESTs**

La búsqueda de motivos de repetición de tipo microsatélite dentro de los contigs formados por ESTs y la selección de iniciadores para su amplificación se llevaron a cabo con los mismos

programas informáticos y criterios utilizados en los contigs generados a partir de las lecturas de secuenciación masiva.

#### **4.10.- Amplificación de EST-SSR**

Se buscó la amplificación de los EST-SSR por medio de la amplificación en cadena de la polimerasa (PCR). Las amplificaciones se llevaron a cabo en el termociclador Veriti (Applied Biosystem), Para determinar las condiciones óptimas de amplificación se probaron cuatro temperaturas de alineamiento (58, 60, 62 y 64°C) y dos concentraciones de Cloruro de Magnesio (MgCl<sub>2</sub>), 1.5 y 2.5 mM, la concentración de la polimerasa estuvo a 0.01 U/ $\mu$ l<sup>-1</sup>, el amortiguador de reacción a 1X, 0.2mM de dNTPs, 0.5  $\mu$ M de cada iniciador y 4 ng  $\mu$ l<sup>-1</sup> de DNA genómico en un volumen final de 12.5  $\mu$ l.

Con el fin de comprobar la amplificación de los EST-SSR, los productos obtenidos se visualizaron por electroforesis en gel de agarosa al 1.5%, teñido con bromuro de etidio durante 20 min, las condiciones de corrida se ajustaron en 100 V por 120 min, el tamaño de los productos se compararon con un marcador de peso molecular de 50 pb y otro de 100 pb (PROMEGA). El gel se observó sobre un transiluminador de luz ultra violeta en el equipo GelDoc (BIORAD) y se visualizaron los EST-SSR con el programa Quantity One (Bio Rad).

#### **4.11.- Análisis genético poblacional**

Una vez que se comprobó la amplificación de los EST-SSR, se amplificaron los microsatélites con resultados positivos en 40 muestras de DNA previamente extraído de los 12 organismos vivos y los 28 organismos congelados de trucha de SPM y realizó la asignación de genotipos por locus e individuo de forma visual. Para ello se visualizaron los productos de PCR en el equipo de electroforesis capilar Qiaxcel (Qiagen), se utilizó un cartucho de alta resolución, el método OM500, utilizando marcadores de alineamiento de 15 y 1000 pb con la escalera de referencia de 50 a 800 pb. Después de cada corrida se comprobaron los tamaños de los fragmentos obtenidos en cada capilar, asignando en primer lugar de manera automática y después de forma manual los fragmentos correspondientes a los marcadores de alineamiento, según el tiempo de detección.

Al final, los alelos de cada locus se asignaron de acuerdo a su tamaño y se obtuvo el genotipo de cada organismo para cada locus, con los datos obtenidos se generó una matriz de datos con los genotipos de la población y se revisaron posibles errores de asignación por medio del paquete informático MICROCHECKER v2.2.3 (Oosterhout *et al.*, 2004). Las variables de diversidad

genética como número de alelos por locus, riqueza alélica, heterocigosidad observada y esperada y ajuste al EHW se calcularon por medio del paquete informático ARLEQUIN 3.5 (Excoffier *et al.*, 2005) y por último el coeficiente de endogamia poblacional ( $F_{IS}$ ) se evaluó mediante el paquete informático FSTAT 2.9.3 (Goudet, 1995).

## 5.- Resultados

### 5.1.- Muestreo

Se recolectaron en total 69 organismos, los cuales se encuentran dentro de un sistema de recirculación (Fig. 4.). Se tomó la muestra de la aleta caudal de 12 organismos y con el fin de obtener tejido para el análisis genético y no sacrificar a los animales con lo cual no hubo mortalidad como resultado de este muestreo (Fig. 5).



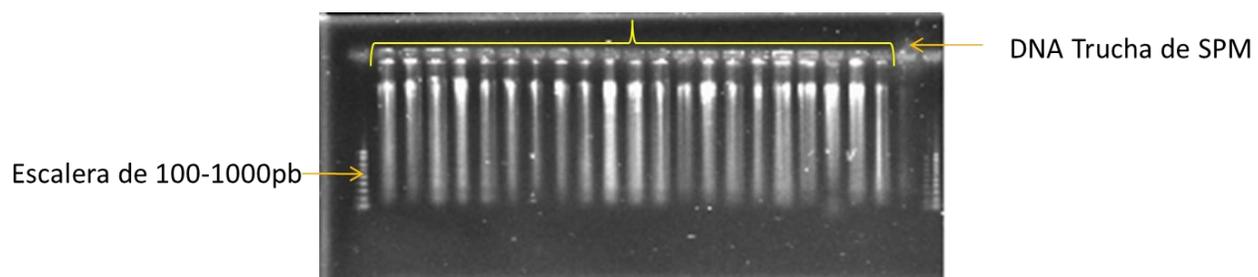
**Fig. 4.** Muestreo de la trucha de San Pedro Mártir *Oncorhynchus mykiss nelsoni* en el arroyo San Rafael. Captura de la trucha en el Arroyo San Rafael (A), jaulas donde se mantuvieron las truchas capturadas (B), hieleras donde se trasladaron los organismos con oxígeno saturado (C), bolsa con truchas de San Pedro Mártir en aclimatación para introducirlas a los estanques (D), estanques del sistema de recirculación (E).

### 5.2.- Extracción de DNA

Se obtuvo una concentración promedio de  $376.2 \text{ ng } \mu\text{l}^{-1}$ , los valores de absorbancia 260/280 y 260/230 fueron de 1.99 y 2.14 respectivamente. La integridad del DNA evaluada por electroforesis indica que se obtuvo material genético de alto peso molecular (Fig. 6).



**Fig. 5. Truchas de San Pedro Mártir, en los círculos se muestra el corte realizado en aleta caudal.**



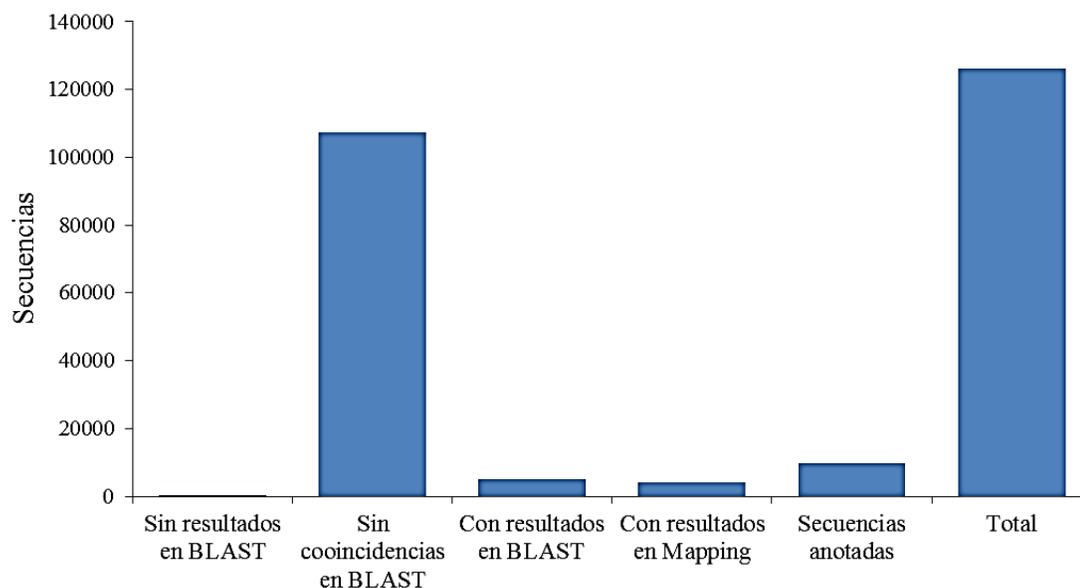
**Fig. 6. Integridad de DNA de trucha de San Pedro Mártir *Oncorhynchus mykiss nelsoni*.**

### 5.3.- Secuenciación masiva de trucha de San Pedro Mártir

Se recibieron 4 688 266 de lecturas en crudo con una longitud promedio de 240 pb, con 1 127 millones de bases, de las cuales, se lograron ensamblar el 54.48% de las secuencias recibidas. Se obtuvieron 126 123 contigs con una longitud promedio de 293 pb, el 45.52% de las secuencias restantes, conformaron un total de 2 133 862 singletons.

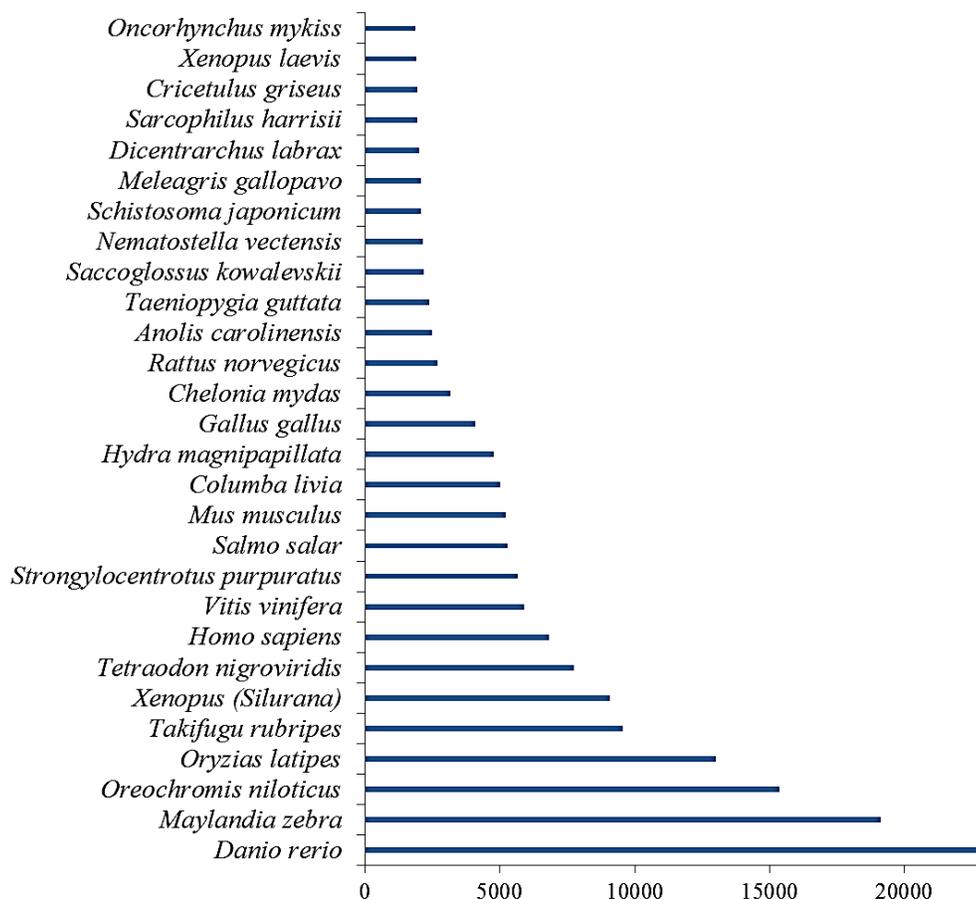
### 5.4.- Anotación de contigs formados a partir de secuenciación masiva de la trucha de San Pedro Mártir.

Se analizaron los 126 288 contigs, de los cuales siete no tuvieron resultados en BLAST, mientras que 107 193 no tuvieron similitud con secuencias reportadas en el banco de genes, 5 170 tuvieron resultados en BLAST, el número de secuencias mapeadas fue de 4 052 y el número de secuencias anotadas correspondió a 9 866 (Fig. 7).



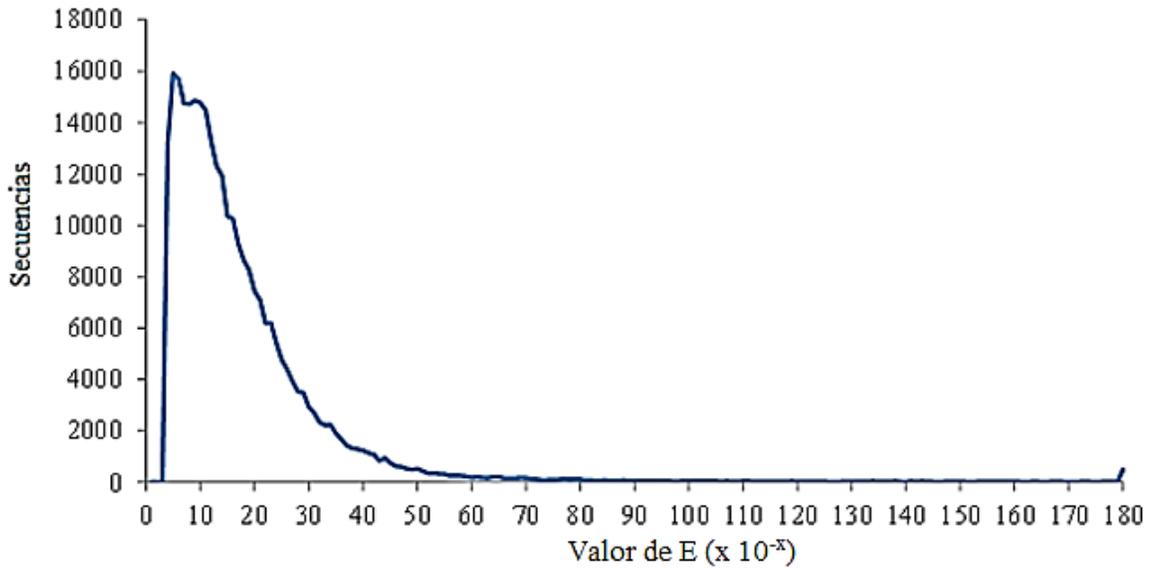
**Fig. 7. Distribución de secuencias anotadas durante el análisis de los contigs obtenidos de la secuenciación masiva.**

Al finalizar la anotación genética, las especies con las que se encontró mayor similitud entre las secuencias fueron el pez cebra *Danio rerio* con 22 973 coincidencias, el pez cebra del Lago Malawi, *Maylandia zebra*, con 19 124, la tilapia, *Oreochromis niloticus*, con 15 372 y el medaka, *Oryzias latipes*, con 13 008. Dentro de los salmónidos, el salmón, *Salmo salar*, tuvo 5 270 coincidencias y la trucha arcoíris *Oncorhynchus mykiss* 1 883 (Fig. 8).



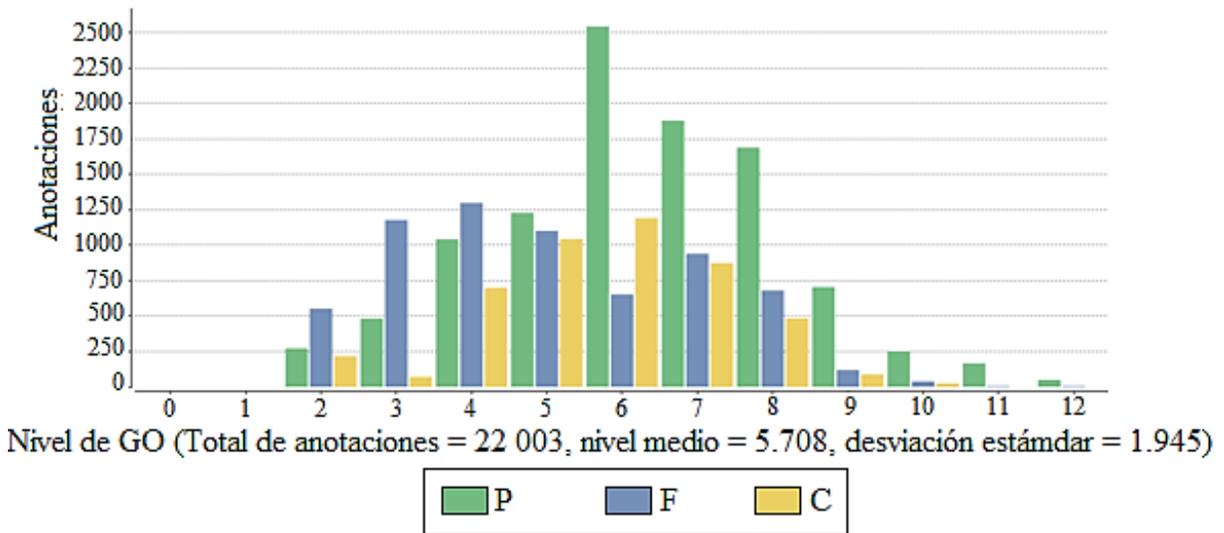
**Fig. 8. Distribución de especies por número de secuencias identificadas durante el BLAST.**

La distribución del valor de E (el cual representa la probabilidad de que la similitud de las secuencias se haya encontrado por azar) mínimo obtenido fue de  $1 \times 10^{-4}$  con casi 13 201 secuencias anotadas con este valor, mientras que el máximo fue de  $1 \times 10^{-180}$ , con 475 secuencias anotadas con este valor de E. El valor de E con más secuencias anotadas fue el de  $1 \times 10^{-5}$  con 15 904 secuencias (Fig. 9).



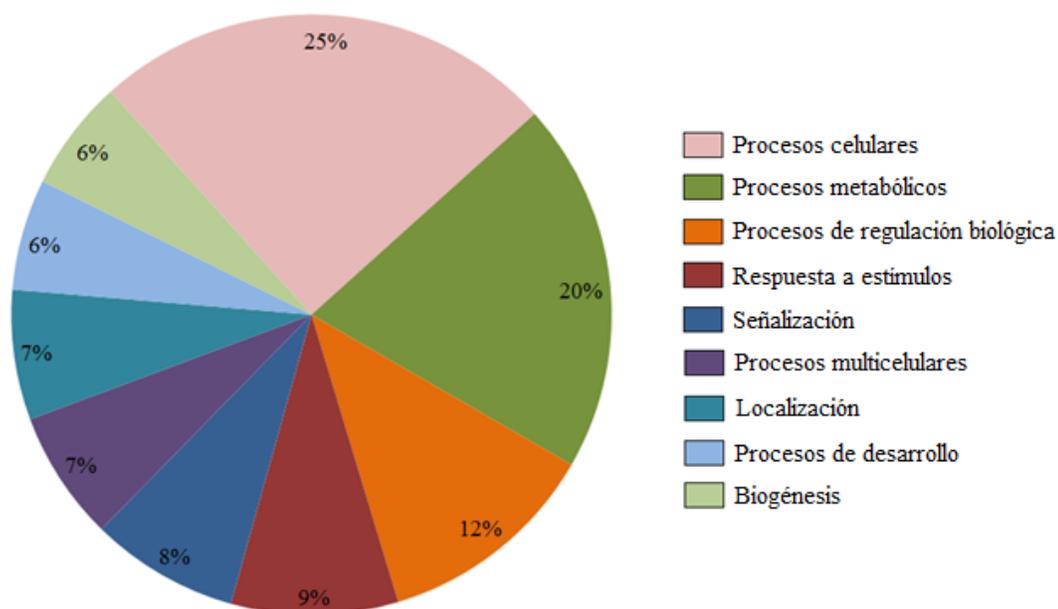
**Fig. 9. Distribución de secuencias identificadas con respecto al valor E.**

La anotación de secuencias con respecto al nivel de ontología genética muestra que el máximo de secuencias anotadas en las categorías de proceso biológico y componente celular se alcanzó en el nivel seis, mientras que para la categoría función molecular el máximo de secuencias anotadas ocurrió en el nivel cuatro (Fig. 10).



**Fig. 10. Distribución del número de anotaciones con respecto al nivel de ontología genética, donde P = Proceso biológicos, F = Función molecular y C = Componente celular.**

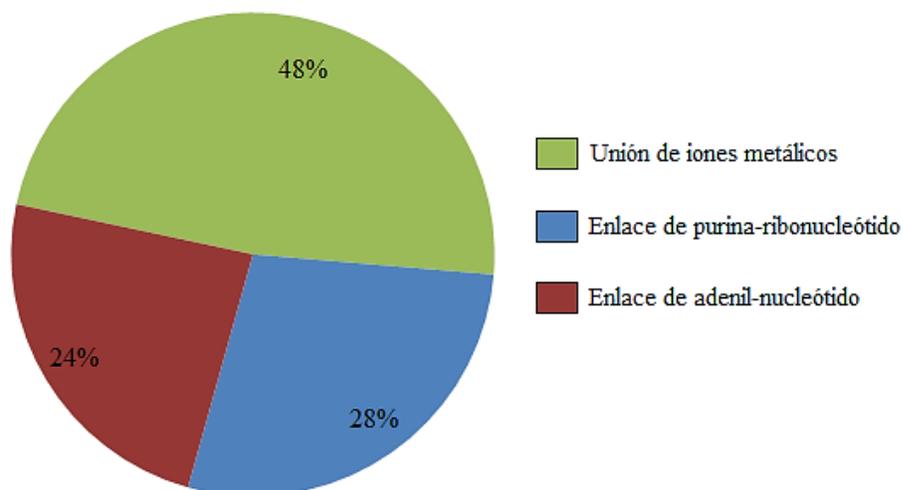
En términos de GO de la categoría procesos biológicos (Fig. 11), el mayor de número de secuencias anotadas estuvieron en los procesos celulares, con el 25% (7 176), seguido por procesos metabólicos con un 20% (5 710) y procesos de regulación biológica con un total de 12% (3 542), la respuesta a estímulos tuvo un 9% (2 736) de la anotación, con un 8% (2 205) estuvo la categoría señalización, con 7 % (2 126) estuvo la categoría procesos multicelulares y localización también obtuvo el 7% de la anotación (1 930), por último representado el 6% estuvieron procesos de desarrollo y biogénesis con 1 881 y 1 687 secuencias anotadas respectivamente.



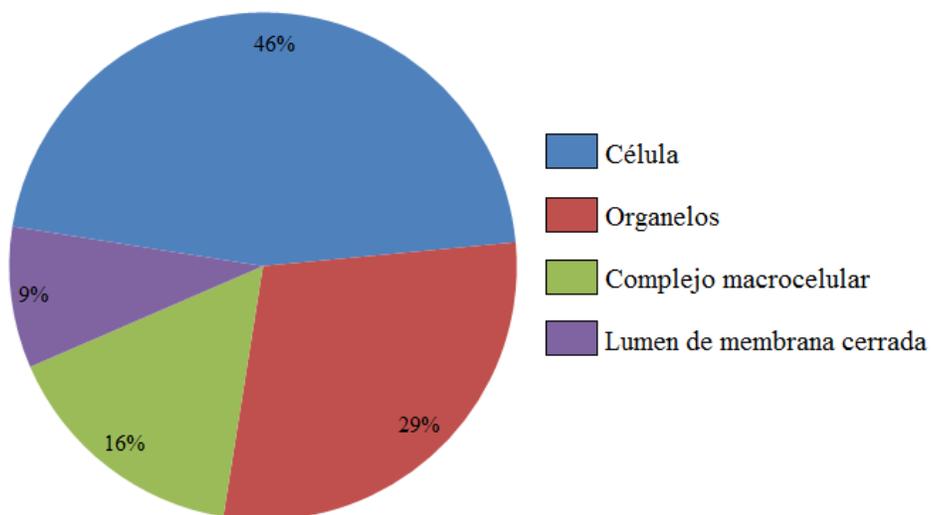
**Fig. 11. Clasificación de Ontología genética (GO) de las secuencias de la trucha de San Pedro Mártir basada en los procesos biológicos.**

En la categoría función molecular, el mayor número de secuencias anotadas correspondió a las funciones de unión de iones metálicos con el 48% (2 065), seguido por la categoría enlace de purina-ribonucleótidos que tuvo un total de 28% (1 174), en tercer lugar representando un 24% (1012) estuvo la categoría enlace de adenil-nucleótido (Fig. 12).

En la categoría de componente celular la función de célula tuvo el mayor número de anotaciones con un 46% (5 922), seguida por la categoría organelos con un 29% (3 668) en tercer lugar estuvo la categoría de complejo macromolecular con un 16% (2 072) y por último la categoría lumen de membrana cerrada tuvo el 9% (1 119) (Fig. 13).



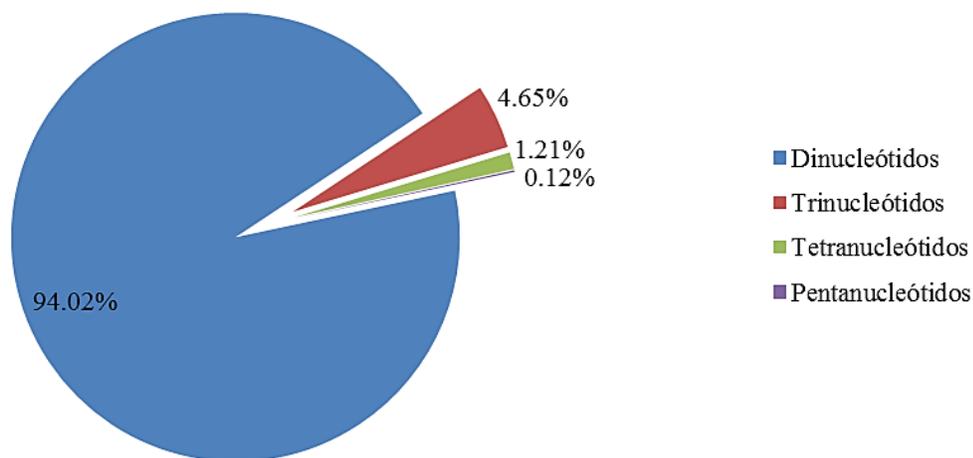
**Fig. 12.** Clasificación de Ontología genética (GO) de las secuencias de trucha de San Pedro Mártir basada en la función molecular.



**Fig. 13.** Clasificación de Ontología genética (GO) de las secuencias de trucha de San Pedro Mártir basada en el componente celular.

### 5.5.-Búsqueda de microsatélites por secuenciación masiva

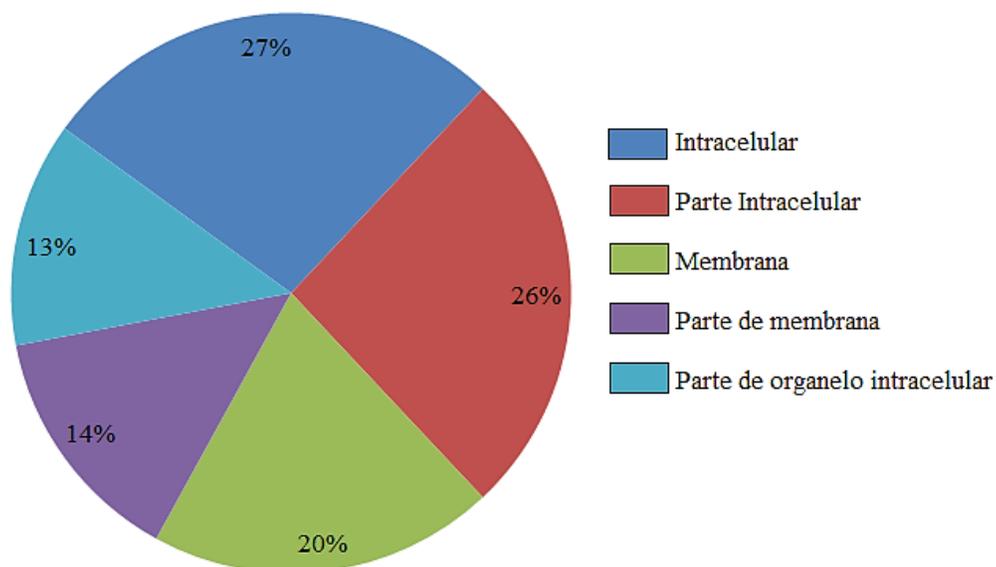
A partir de los 126 288 contigs se encontraron 10 100 secuencias que contenían microsatélites, de los cuales se pudieron diseñar iniciadores para 6 842 posibles microsatélites dentro de los 6 138 contigs. Del total de microsatélites 6 433 resultaron del tipo dinucleótidos representando el 94.02% del total, los de tipo trinucleótidos representaron el 4.65% con 318 microsatélites, se encontraron 83 microsatélites tetranucleótidos y 8 pentanucleótidos, siendo el 1.21% y 0.12% respectivamente (Fig. 14).



**Fig. 14. Distribución de microsatélites expresado en porcentaje según el tipo de repetición.**

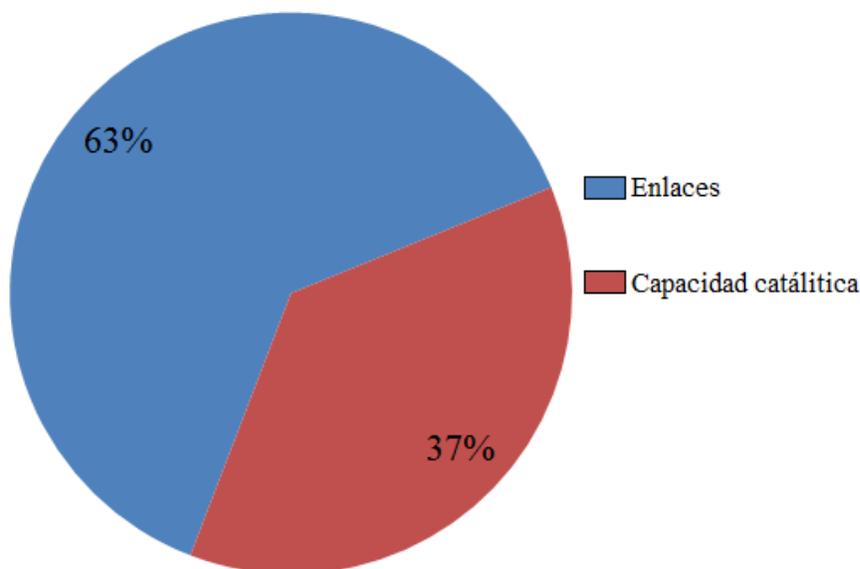
### 5.6.-Anotación genética de microsatélites obtenidos por secuenciación masiva.

En la categoría de componente celular la función intracelular tuvo el mayor número de secuencias anotadas con un 27% (226), seguida por la categoría parte intracelular con un 26% (212), en tercer lugar estuvo la categoría de membrana con un 20% (162), la categoría parte de organelo intracelular tuvo el 14% (119) de las anotaciones y por último representado el 13% (106) de la anotación estuvo la categoría parte de membrana (Fig. 15).



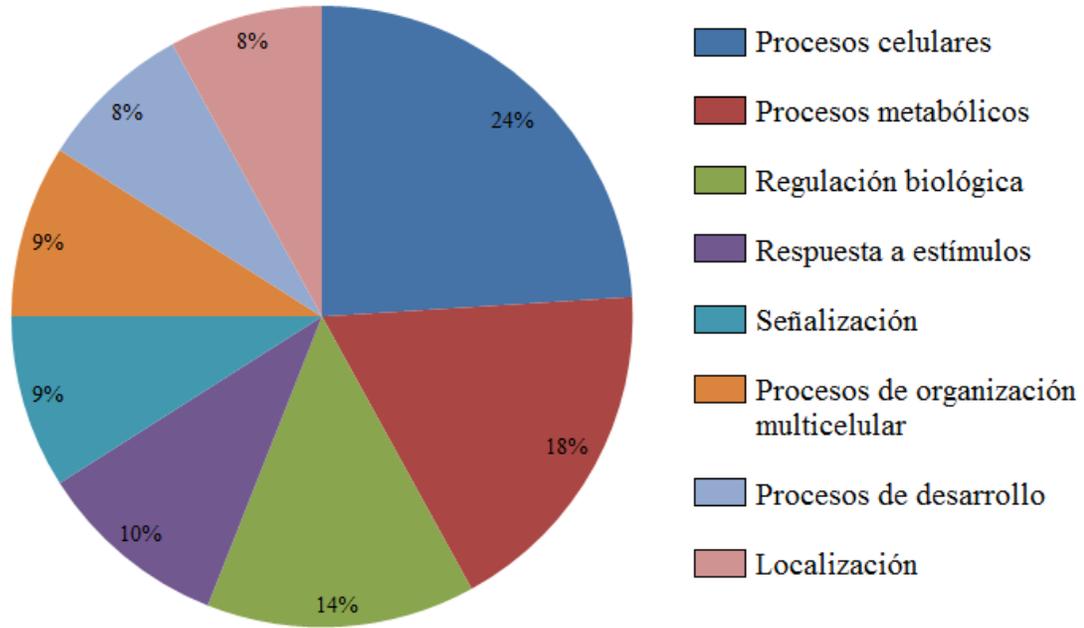
**Fig. 15. Clasificación de la categoría componente celular en términos de ontología genética para las secuencias que contienen microsatélites obtenidos por secuenciación masiva.**

En la anotación genética para las secuencias que contienen microsatélites, la categoría de GO función molecular sólo arrojó dos resultados, enlaces con 294 secuencias anotadas y capacidad catalítica con 173 secuencias anotadas, que representan un 63 y 37% de la anotación respectivamente (Fig. 16).



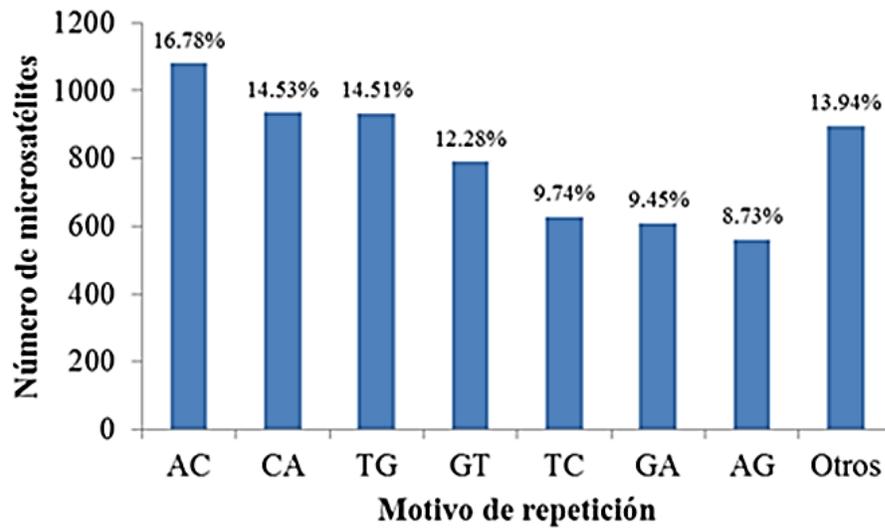
**Fig. 16. Clasificación de la categoría función molecular en términos de ontología genética para las secuencias que contienen microsatélites obtenidos por secuenciación masiva.**

En términos de GO de la categoría procesos biológicos (Fig. 17), el mayor número de secuencias anotadas estuvieron en los procesos celulares con el 24% (303), seguido por procesos metabólicos con un 18% (231) y procesos de regulación biológica con un total de 14% (179). La respuesta a estímulos tuvo un 10% (130) de la anotación, con un 9% (113), estuvo la categoría procesos de organización multicelular. Con el 9% (109) estuvo la categoría de señalización. Por último las categorías procesos de desarrollo y localización tuvieron un 8% de anotación cada una con 104 y 101 secuencias anotadas respectivamente.



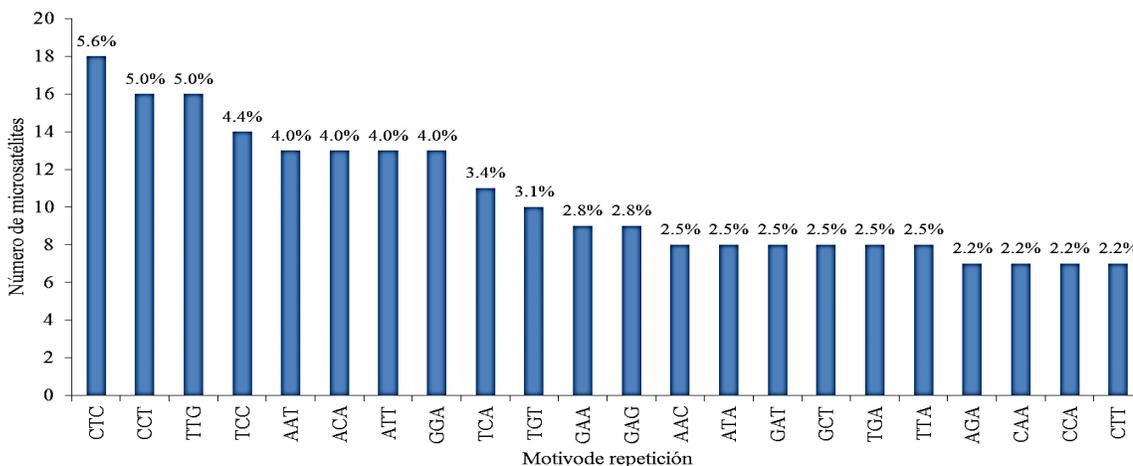
**Fig. 17.** Clasificación de la categoría procesos biológicos en términos de ontología genética para las secuencias que contienen microsatélites obtenidos por secuenciación masiva.

Dentro de los microsatélites dinucleótidos, las secuencias de repetición más abundantes fueron las de tipo AC con 1 080 microsatélites, que representa el 16.78%. Los de tipo CA representaron el 14.53% con 935 microsatélites y los de motivo de repetición TG representaron el 14.51% con 934 microsatélites (Fig. 18).



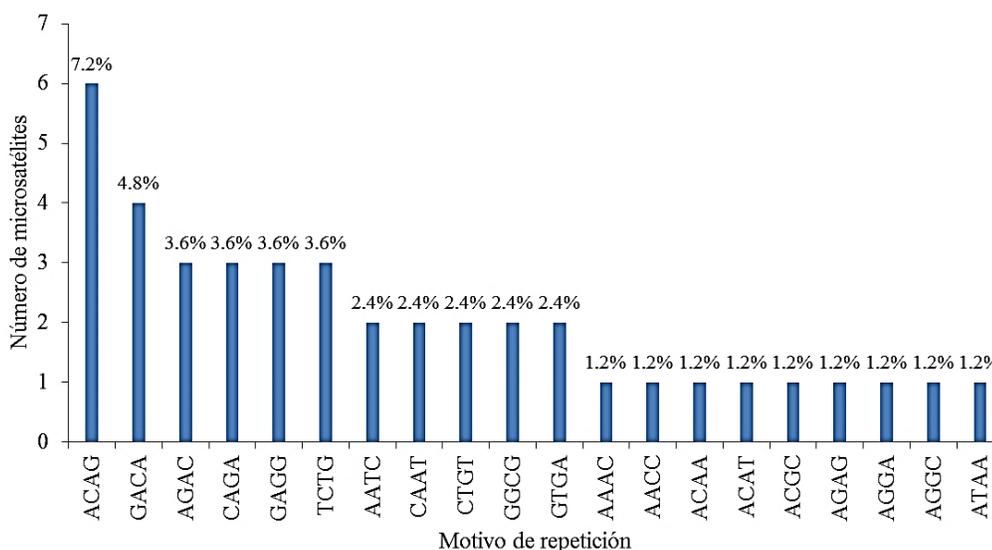
**Fig. 18.** Distribución de la secuencia de repetición dentro de los microsatélites dinucleótidos, expresados en número y porcentaje.

Al realizar un análisis similar con microsatélites trinucleótidos, el motivo CTC fue el más abundante con 18 microsatélites que representaron el 5.66%. Los motivos de repetición CCT y TTG representaron el 5.03% con 16 microsatélites cada uno, el motivo de repetición TCC representó el 4.40% del total de trinucleótidos con 14 microsatélites (Fig. 19).



**Fig. 19. Distribución de la secuencia de repetición dentro de los microsatélites trinucleótidos, expresados en número y porcentaje, no se muestra la categoría “otros”, que representa el 27.3% con 87 diferentes motivos de repetición.**

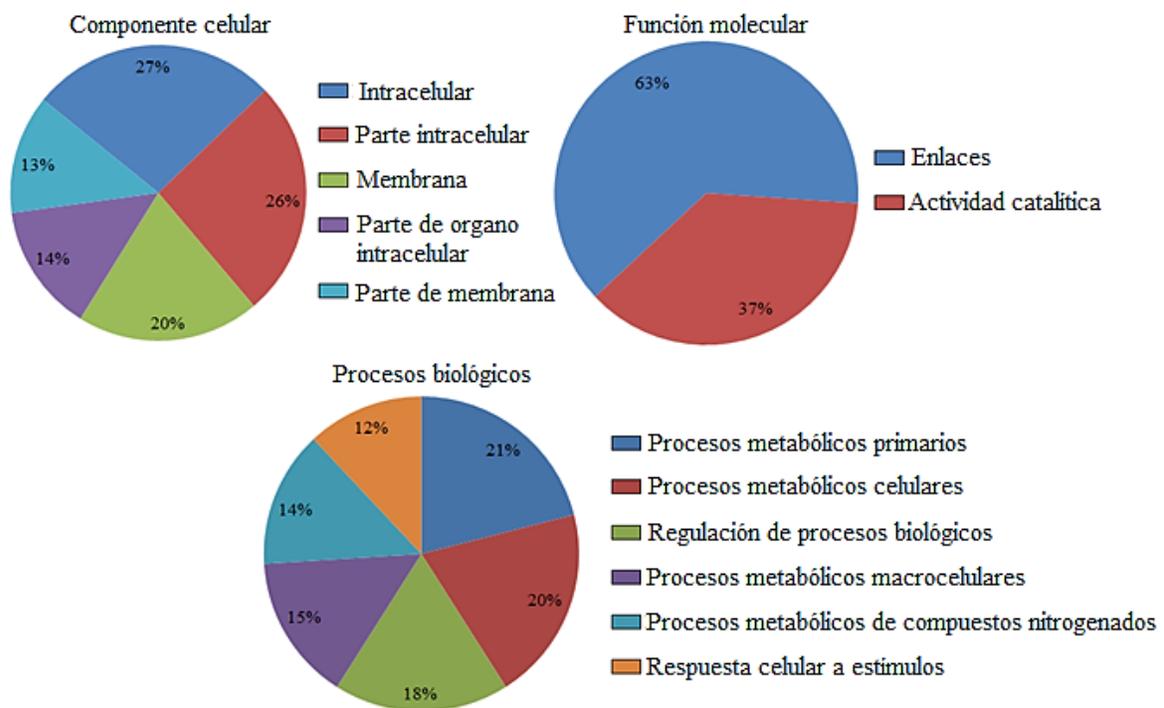
En los microsatélites de tipo tetranucleótido el motivo de repetición ACAG representó el 7.22% con seis microsatélites, seguido del motivo GACA representando el 4.81% con cuatro microsatélites, los motivos de repetición AGAC, CAGA, GAGG y TCTG representaron el 3.61% con tres microsatélites cada uno (Fig. 20).



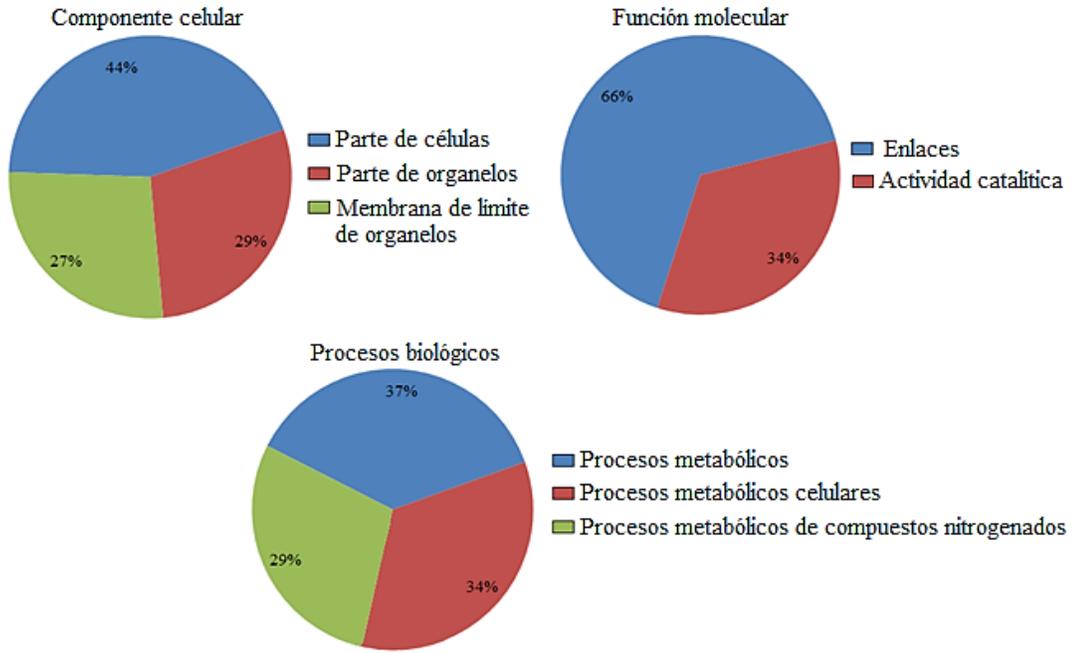
**Fig. 20. Distribución de la secuencia de repetición dentro de los microsatélites tetranucleótidos, expresados en número y porcentaje, no se muestra la categoría “otros”, que representa el 50.6% con 42 distintos motivos de repetición.**

Los microsatélites de tipo pentanucleótidos fueron ocho diferentes (ACAGT, CACTA, CCCTA, CTCTC, GTGTA, TACAT, TCCTC, TCGTG) con el 12.5% cada uno.

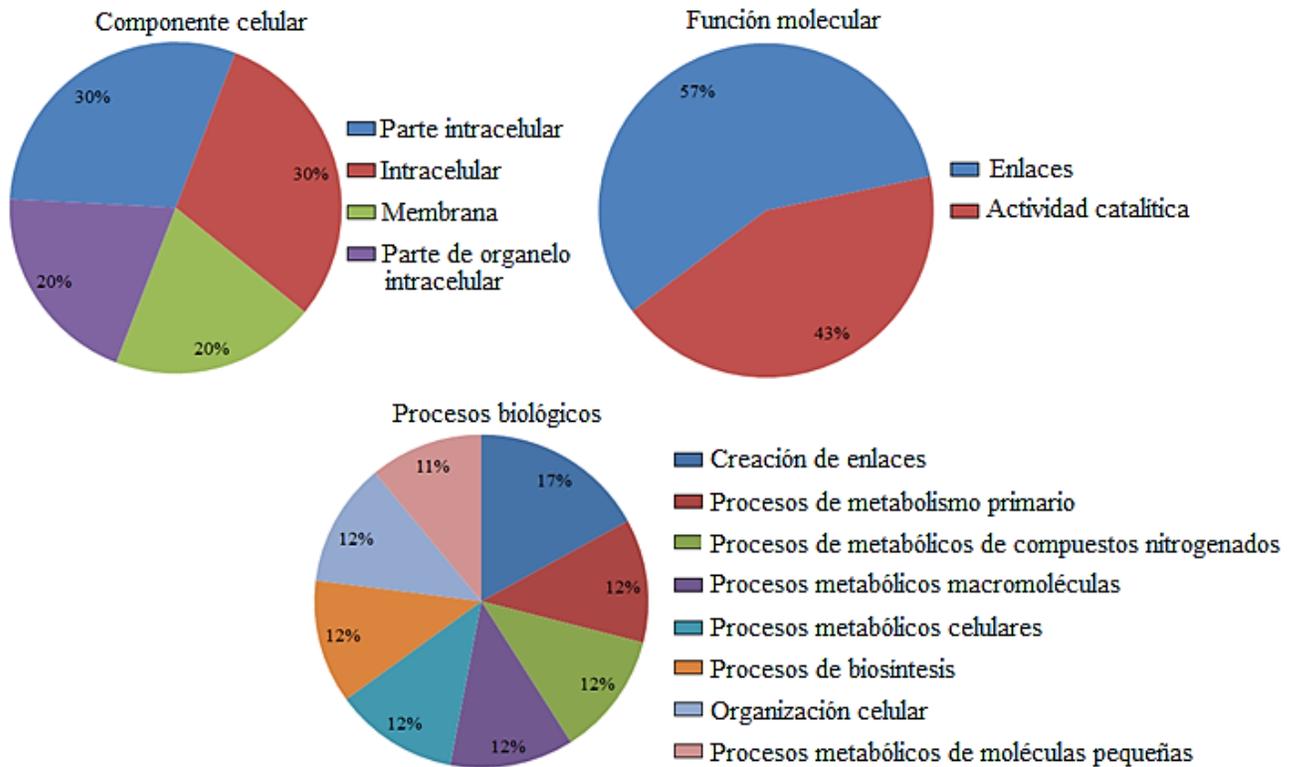
Por último se obtuvieron los resultados de anotación de los contigs que contienen microsatélites según tipo de repetición. En los dinucleótidos se encontró en la categoría de componente celular 755 secuencias, en función molecular 430 y en procesos biológicos 923 (Fig. 21) En los trinucleótidos se anotaron en la categoría componente celular 510 secuencias, en función molecular 430 y en procesos biológicos 38 (Fig. 22). En los de tipo tetranucleótido se anotaron en la categoría de componente celular 10 secuencias, en función molecular siete y en procesos biológicos 17 (Fig. 23). En las secuencias que contenían los microsatélites pentanucleótidos no se anotó ninguna secuencia.



**Fig. 21. Clasificación de ontología genética para las secuencias que contienen microsatélites dinucleótidos.**



**Fig. 22. Clasificación de ontología genética para las secuencias que contienen microsatélites trinucleótidos.**



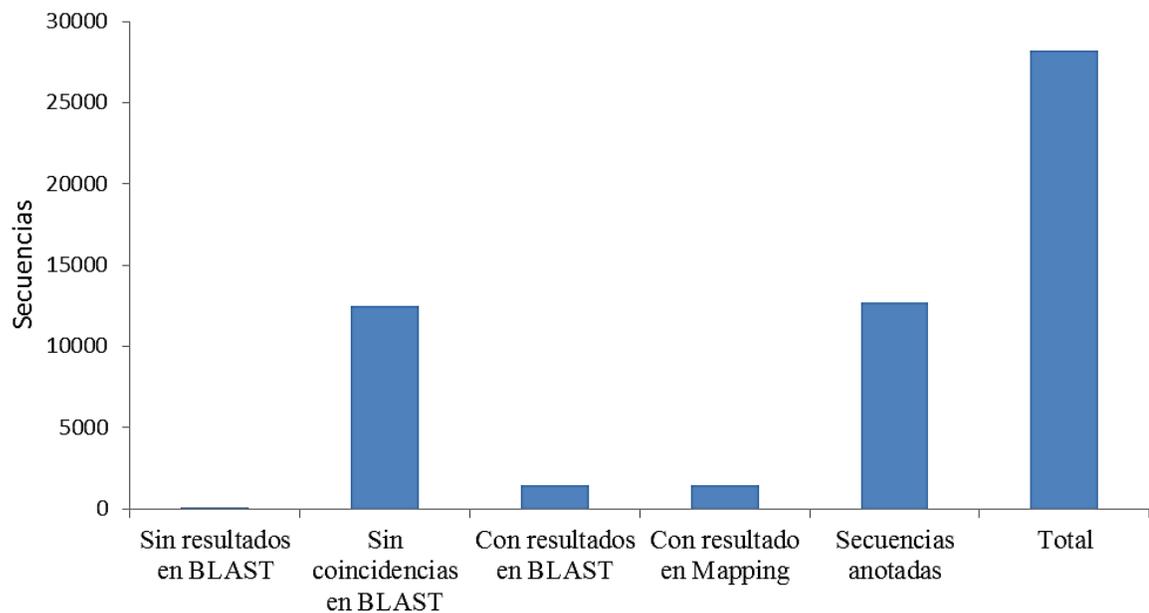
**Fig. 23. Clasificación de ontología genética para las secuencias que contienen microsatélites tetranucleótidos.**

### 5.7.- Análisis de las secuencias EST de trucha arcoíris

A partir de 287 967 secuencias EST descargadas se ensamblaron 28 237 contigs, con una longitud promedio de 884 pb.

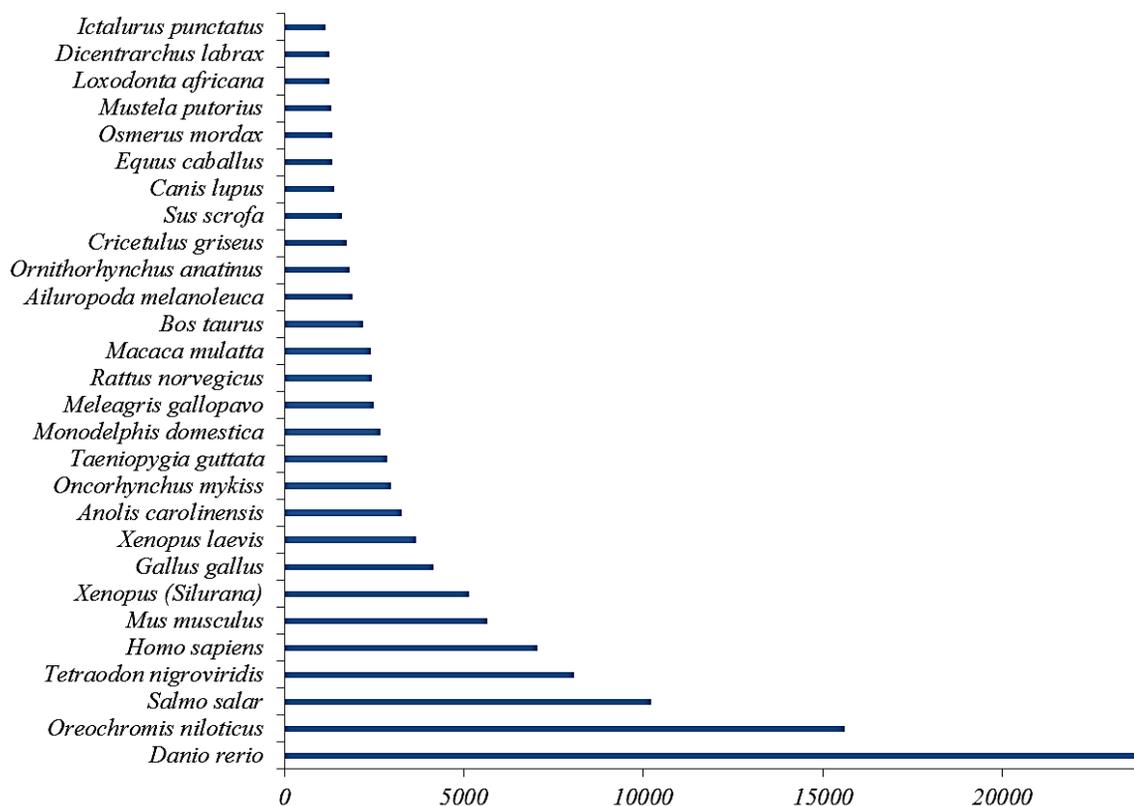
De estos fragmentos se ensamblaron un total de 185 000 secuencias que formaron 28 237 contigs correspondientes a casi 25 millones de bases una longitud promedio de 884 pb. El número total de fragmentos no ensamblados fue de 103 mil con una longitud promedio de 666 b. La proporción de nucleótidos fue predominada por la Adenina (27%) seguida por la timina (26%), citosina (22.5%) y guanina (22.3%).

Durante la anotación genética, se encontró que 12 528 de secuencias no tuvieron una similitud en el banco de genes. A 1 250 secuencias se les encontró similitud en el banco de genes. Un número similar se pudo mapear y alrededor de 12 500 contigs se pudieron anotar encontrando su función con respecto a los datos de bases de datos existentes (Fig. 24).



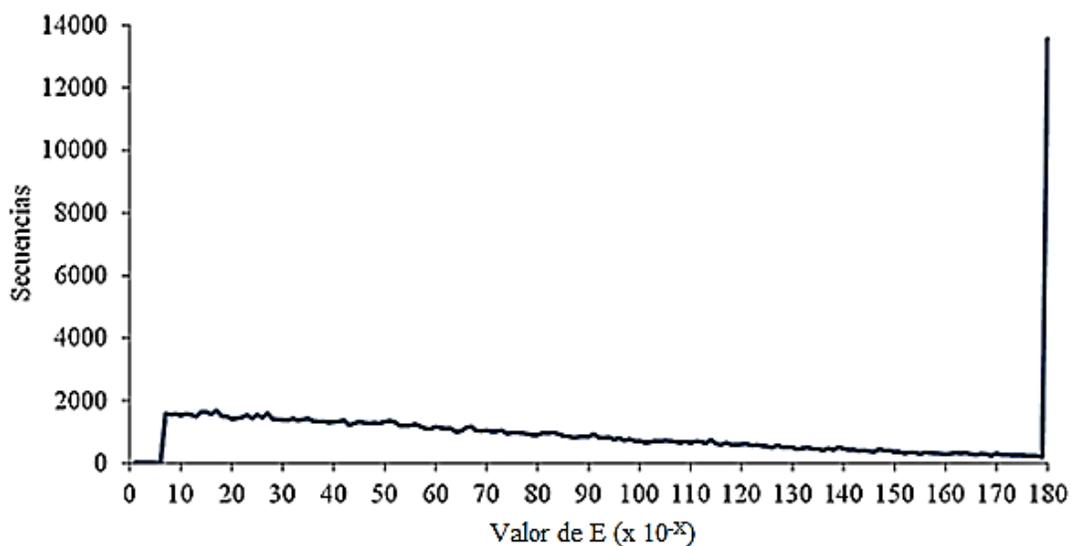
**Fig. 24. Distribución de secuencias anotadas durante el análisis de los contigs obtenidos de secuencias EST.**

Al término del BLAST, la mayoría de los EST mostraron similitud con el pez cebra (*Danio rerio*) seguido de la tilapia (*Oreochromis niloticus*) y el salmón del Atlántico (*Salmo salar*) (Fig. 25).



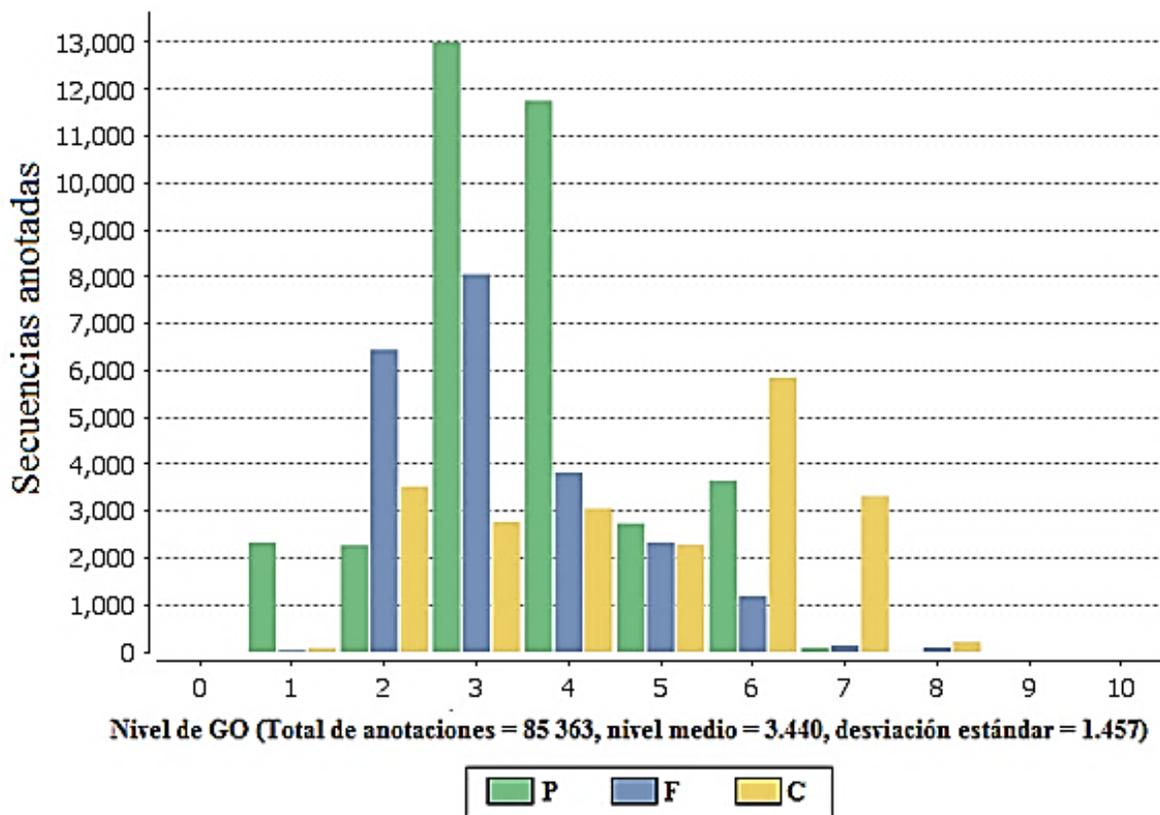
**Fig. 25. Distribución de especies por número de secuencias EST identificadas durante el BLAST.**

La distribución del valor de E mínimo obtenido fue de  $1 \times 10^{-7}$  con casi 1 578 secuencias anotadas con este valor. El máximo fue de  $1 \times 10^{-180}$ , con 13 547 que a su vez fue el valor de E con más secuencias anotadas (Fig. 26).



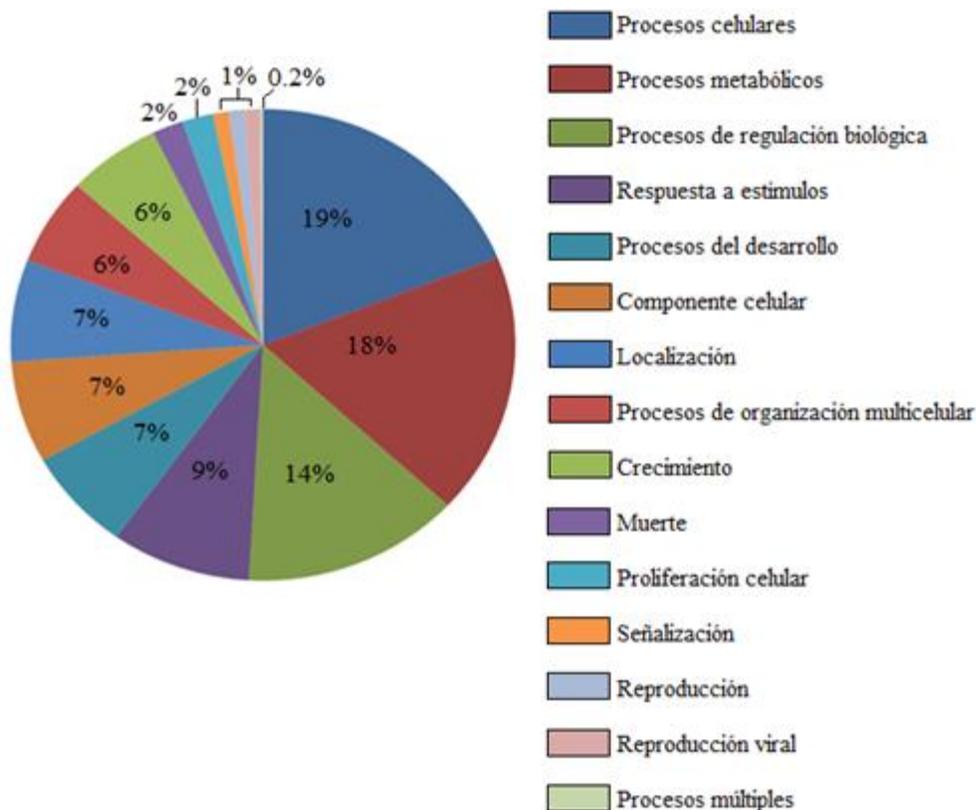
**Fig. 26. Distribución de secuencias EST identificadas con respecto al valor E.**

La anotación de secuencias EST con respecto al nivel de ontología genética muestra que el máximo de secuencias anotadas en las categorías de proceso biológico y función molecular se alcanzó en el nivel tres, mientras que para la categoría componente celular el máximo de secuencias anotadas ocurrió en el nivel seis (Fig. 27).



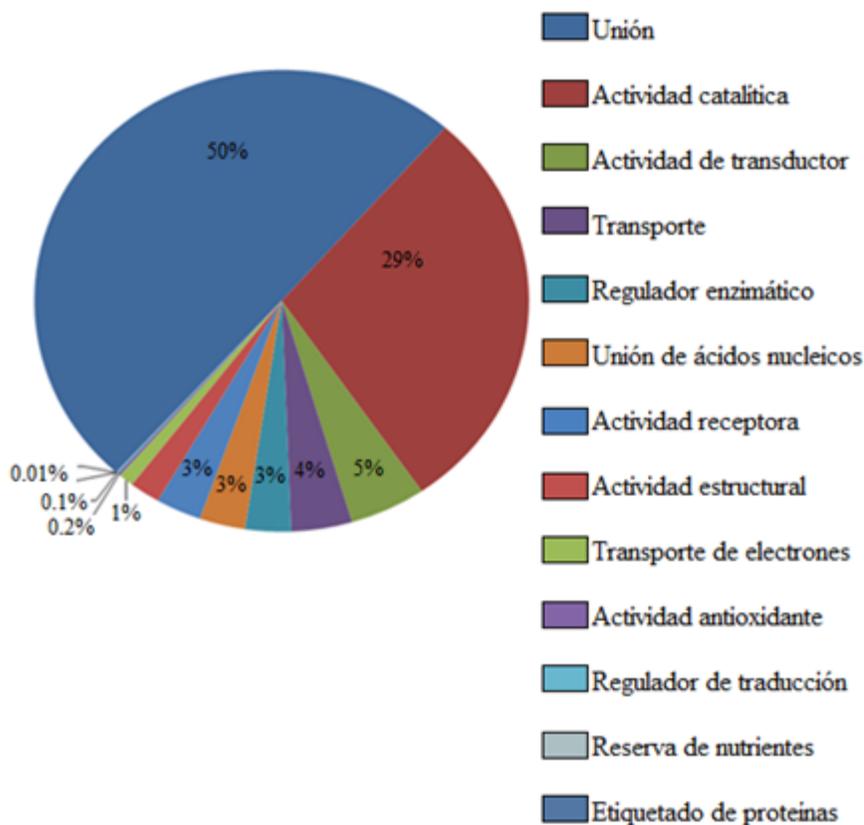
**Fig. 27. Distribución del número de anotaciones de las secuencias EST con respecto al nivel de ontología genética, donde P = Proceso biológicos, F = Función molecular y C = Componente celular.**

En términos de GO de la categoría procesos biológicos (Fig. 28), el mayor de número de secuencias anotadas estuvieron en los procesos celulares, con el 19% de secuencias representadas por un total de 6 403, seguido por procesos metabólicos con 6 067 secuencias anotadas que representan un 18% y procesos de regulación biológica con 4 504 secuencias anotadas dando un total de 14%. La respuesta a estímulos tuvo un 9% de la anotación con 3 012 secuencias anotadas. Con poco más de 2 mil secuencias anotadas que representan un 7%, estuvieron las categorías procesos del desarrollo, componente celular y localización.



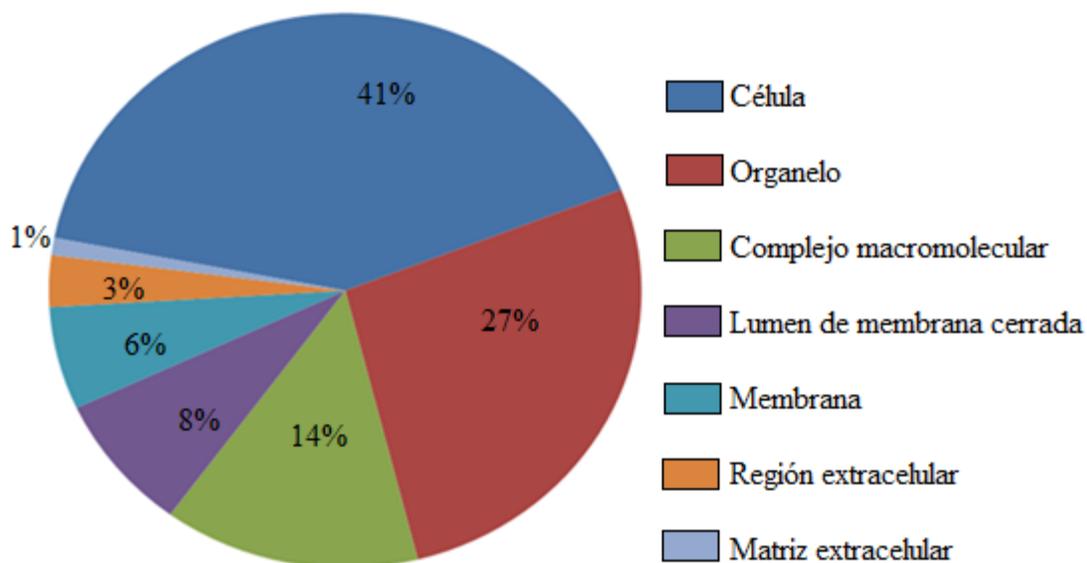
**Fig. 28. Clasificación de Ontología genética (GO) de los contigs obtenidos a partir de secuencias EST basada en los procesos biológicos. Total de secuencias 33 156.**

En la categoría función molecular, el mayor número de secuencias anotadas correspondió a las funciones de unión con 8 714 secuencias que representan el 50%, seguido por la actividad catalítica que tuvo un total de 5 082 secuencias anotadas para obtener el 29% de la anotación dentro de esta categoría, en tercer lugar con 729 secuencias anotadas representando un 5% está la actividad de transductor, todas las demás categorías tuvieron menos de 700 secuencias anotadas (Fig. 29).



**Fig. 29. Clasificación de Ontología genética (GO) de los contigs obtenidos a partir de secuencias EST basada en la función molecular. Total de secuencias 17 497.**

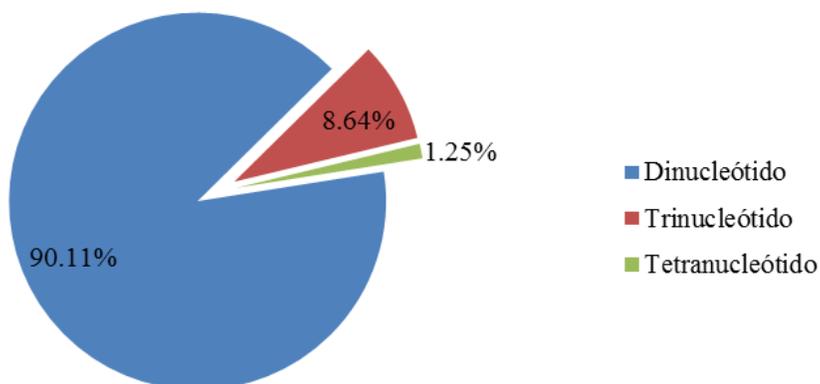
En la categoría de componente celular la función de célula tuvo el mayor número de anotaciones con 8 979 secuencias anotadas que representan un 41%, seguida por la categoría organelo con 5 805 secuencias anotadas para un 27%. En tercer lugar estuvo la categoría de complejo macromolecular con 3 054 secuencias representando un 14%, las demás funciones dentro de esta categoría tuvieron menos de 2 mil secuencias anotadas, donde matriz extracelular fue la que menos secuencias tuvo con apenas el 1% siendo 160 secuencias anotadas (Fig. 30).



**Fig. 30.** Clasificación de Ontología genética (GO) de los contigs obtenidos a partir secuencias EST basada en el componente celular. Total de secuencias 21 862.

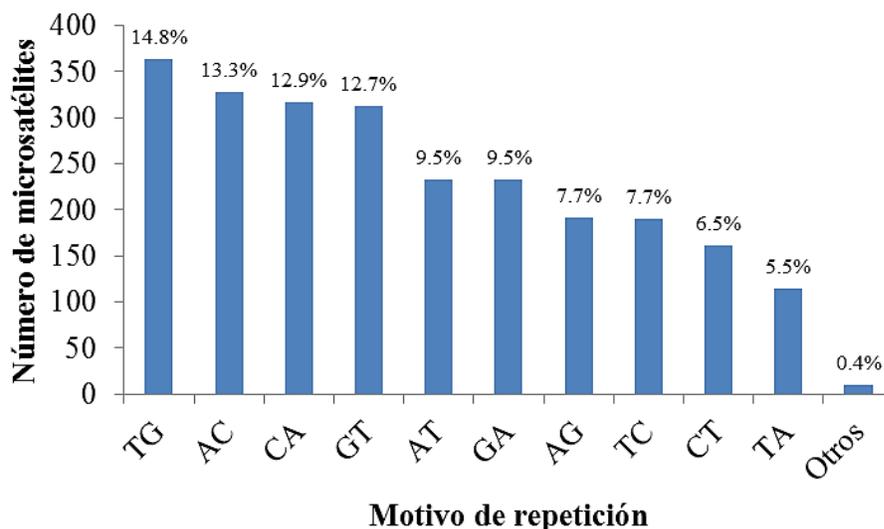
### 5.8.- Búsqueda de microsatélites en secuencias EST

Dentro de los 28 237 contigs formados por las secuencias EST, se encontraron un total de 2721 posibles microsatélites, de los cuales 2452 son de tipo dinucleótidos representando el 90.11%, el 8.64% resultaron trinucleótidos con 235 y solamente el 1.25% son del tipo tetranucleótidos con 34. No se detectaron microsatélites de tipo penta ni hexanucleótidos (Fig. 31).



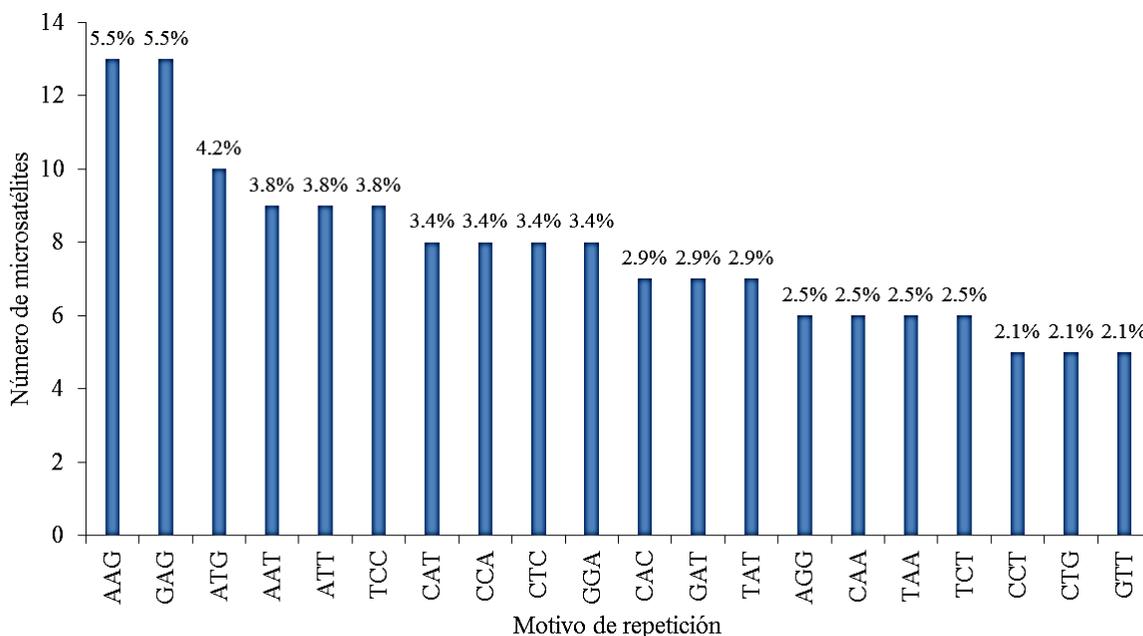
**Fig. 31.** Distribución de microsatélites EST expresado en porcentaje según el tipo de repetición.

Dentro de los microsatélites dinucleótidos se encontraron 363 microsatélites con el motivo de repetición TG, que representa el 14.80% siendo el más abundante, seguido del motivo de repetición AC con el 13.38% representados por 328 microsatélites, los microsatélites con motivos de repetición CA y GT tuvieron el 12.93 y 12.72% con 317 y 312 microsatélites respectivamente (Fig. 32).



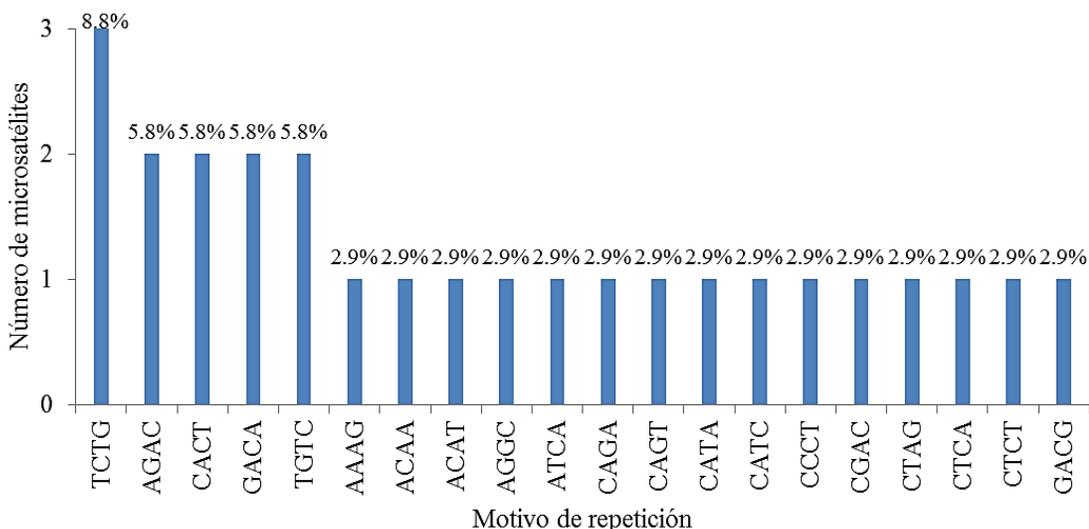
**Fig. 32. Distribución de la secuencia de repetición dentro de los microsatélites EST dinucleótidos, expresados en número y porcentaje.**

En los microsatélites trinucleótidos, los motivos de repetición más abundantes fueron los de tipo AAG y GAG, con 13 microsatélites cada uno representando el 5.53%, seguidos de los motivos de repetición AAT, ATT y TCC con nueve microsatélites que representan el 3.83% para cada uno de ellos (Fig. 33).



**Fig. 33. Distribución de la secuencia de repetición dentro de los microsatélites EST trinucleótidos, expresados en número y porcentaje, no se muestra la categoría “otros”, que representa el 34.0% con 80 diferentes motivos de repetición.**

Por último en los microsatélites de tipo tetranucleótidos, el motivo de repetición más abundante fue el TCTG, con tres microsatélites que representan el 8.82%, seguido de los motivos de repetición AGAC, CACT, GACA y TGTC con dos microsatélites que representan el 5.88% para cada uno de ellos (Fig. 34).



**Fig. 34. Distribución de la secuencia de repetición dentro de los microsatélites EST tetranucleótidos, expresados en número y porcentaje, no se muestra la categoría “otros”, que representa el 23.5.0% con 8 diferentes motivos de repetición.**

### 5.9.- Selección de iniciadores para EST-SSR

De los 2 721 posibles microsatélites se sintetizaron 24 pares de iniciadores (Tabla 2) que cumplieran con las características deseadas para amplificar microsatélites, de los cuales 3 son de tipo tetranucleótidos y 21 trinucleótidos.

Tabla 2. Iniciadores (primers) para amplificar microsatélites obtenidos a partir de EST de trucha arcoíris *Oncorhynchus mykiss mykiss*.

|    | Contig   | Motivo de repetición | Secuencia de iniciadores                              | TM °C        | Tamaño esperado (pb) | GC %         |
|----|----------|----------------------|---|--------------|----------------------|--------------|
| 1  | Omy10016 | ACC (5)              | L:TCAGGACCAGTCAGAAGGAAA<br>R:CCATCCTCCCAATGGTCATA     | 59.8<br>60.5 | 145                  | 47.6<br>50.0 |
| 2  | Omy1017  | AAG (5)              | L:AACCGGCACATCACACTTCT<br>R:CACTGAGCCTGGAGGAAAGA      | 60.5<br>60.5 | 115                  | 50.0<br>55.0 |
| 3  | Omy10717 | AGC (5)              | L:ATTCCACCAACTAGCCGACA<br>R:CTCTATCCTCATCTTTCTGTCCC   | 60.5<br>58.3 | 131                  | 50.0<br>47.8 |
| 4  | Omy10773 | AAG (6)              | L:CCTGACTGTGAACCTCCTGC<br>R:TGTTGTCCTTCAAAGGCGAT      | 60.8<br>60.6 | 127                  | 60.0<br>45.0 |
| 5  | Omy11300 | AAT (5)              | L:TATGGGCAGTGCTGAGACAG<br>R:TCTCCCATCATTCCACAAC       | 60.0<br>60.7 | 118                  | 55.0<br>50.0 |
| 6  | Omy11979 | AAC (5)              | L:TCAGAAGCAATTCAGAATCAACC<br>R:TCAATGACCGGTGTATTTGTTC | 60.6<br>59.7 | 147                  | 39.1<br>53.2 |
| 7  | Omy12323 | ACC (7)              | L:TCCTCAGTGCCCTCTCCTGC<br>R:CTGTGTTGTGCGGCTATCTG      | 61.7<br>60.4 | 126                  | 63.2<br>55.0 |
| 8  | Omy12395 | ACT (7)              | L:AGAGAAGAATGAATCACCCCTGG<br>R:AACACTTGGGACACCTGCC    | 59.5<br>60.9 | 126                  | 45.5<br>57.9 |
| 9  | Omy12448 | ACG (5)              | L:GCTGAGTGAGGAAGAGCTGC<br>R:GCTACTCCTGGGTCTCCCAT      | 60.4<br>60.4 | 204                  | 60.0<br>60.0 |
| 10 | Omy13166 | AGG (5)              | L:AAAGAGCTGCCAGTGGTGAG<br>R:CACAGTCCCACTTCTCTGCC      | 60.5<br>60.8 | 97                   | 55.0<br>60.0 |
| 11 | Omy14477 | AGG (7)              | L:GGTGTCCAAATGCTTTCAGC<br>R:CGTCAGCTAAGACGGCTGTT      | 60.6<br>60.5 | 238                  | 50.0<br>55.0 |
| 12 | Omy1474  | ACG (5)              | L:GCTCTTCTCTGAGCTGCTGG<br>R:GTGGAGGTGACGAAGGTGTC      | 60.5<br>60.5 | 123                  | 57.8<br>57.7 |
| 13 | Omy14968 | AAT (6)              | L:TCAAACCTCCTGGAACAACACAA<br>R:CTCTGAACGAGATGTCGCTGT  | 60.5<br>60.6 | 173                  | 40.9<br>52.4 |
| 14 | Omy15027 | AGT (6)              | L:GCTCCTCCACTTCTACCGCT<br>R:CTCATGTTCGTAGTCCCAGGC     | 60.9<br>60.6 | 129                  | 60.0<br>60.0 |
| 15 | Omy15106 | AAC (7)              | L:CGGTGATGGTAATTCTTAACGG<br>R:CCATTCCAACATTTGCATCC    | 60.6<br>61.1 | 119                  | 45.5<br>45.0 |
| 16 | Omy15365 | AAC (5)              | L:CCTCTCAGGAGACAAGCACAC<br>R:AAAGTGATTGAGGGAGACAGCTT  | 60.0<br>60.6 | 108                  | 57.1<br>43.5 |
| 17 | Omy17098 | AGC (5)              | L:TGTCCTCCTGGTTCTGCTTC<br>R:TCCCGTCTGTGTGTTTGATG      | 60.3<br>60.5 | 133                  | 55.0<br>50.0 |

Tabla 2. Cont.

|    | Contig   | Motivo de repetición | Secuencia de iniciadores                           | TM °C        | Tamaño esperado (pb) | GC %         |
|----|----------|----------------------|--|--------------|----------------------|--------------|
| 18 | Omy17219 | ACT (6)              | L:TACCAAATGGGCTTTGGACT<br>R:CCAAACCTCTCCTCAACTGC   | 59.4<br>59.8 | 165                  | 45.0<br>55.0 |
| 19 | Omy1727  | AAG (5)              | L:CGCTGCTCTTCTTGGTCTTC<br>R:AAGGAGAGGTGGCCAAGGT    | 60.2<br>61.0 | 123                  | 55.0<br>57.9 |
| 20 | Omy17784 | ACT (5)              | R:AATAGGGTGTGGGTGGGAGT<br>L:CTAAACCGACTAGCGCTCCA   | 60.5<br>60.4 | 223                  | 55.0<br>55.0 |
| 21 | Omy18329 | ACC (5)              | L:GAGCCATGAGATTCCTGTGG<br>R:GGGACAGAGTTCTTGGGCTT   | 60.6<br>60.6 | 148                  | 55.0<br>55.0 |
| 22 | Omy3249  | ACAG (5)             | L:GCKCTTTAGTCATCCATAGCA<br>R:TCCAGGACAGAGAGGATTGTG | 60.3<br>60.1 | 94                   | 45.2<br>52.4 |
| 23 | Omy3851  | AATC (5)             | F:ATGAGAAGCAAGCCAACCAA<br>R:GCTTCTTCACCTGCATTGCT   | 60.7<br>60.5 | 183                  | 45.0<br>50.0 |
| 24 | Omy16187 | ACAG (6)             | L:CCATTACCAACGGAGAACCA<br>R:AGGTGTCTGTCTGCGAAAG    | 60.7<br>60.4 | 284                  | 50.0<br>55.0 |

De las 24 secuencias seleccionadas con contenido de microsatélites, 16 estuvieron relacionadas con alguna función molecular, procesos biológicos, componente celular o se le encontró similitud con alguna proteína reportada en las bases de datos públicas, los porcentajes de similitud durante la anotación oscilaron entre el 53% en el caso de la proteína hipotética: NEMVEDRAFT\_v1g156722 y el 100% como la proteína MuRF1 y el gen Precursor de hemoproteína tipo 2, estos dos últimos por su porcentaje de similitud son considerados como secuencias homólogas, lo anterior según el índice de anotación del The Institute for Genomic Research (TIGR por sus siglas en inglés) que incluye tres niveles de significancia basada en el porcentaje de identidad, donde de 90 a 100% se clasifican como "homólogos en el rango de 70-90% se clasifican como "similar", y las secuencias que coinciden en menos del 70% se clasifican como "débilmente similares " (Tabla 3).

Tabla 3. Anotación genética de secuencias EST con microsatélites, su descripción ontológica y porcentaje de similitud.

|    | Contig | Resultado de anotación  | % de similitud |
|----|--------|---|----------------|
| 1  | 10016  | Organización mitocondrial. Translocasa de importación de la membrana interna, subunidad Tim17.  | 95             |
| 2  | 1017   | Proceso metabólico celular que implica bases nitrogenadas, nucleósidos, nucleótidos y ácidos nucleicos. Proteína asociada a espermatogénesis  | 93             |
| 3  | 10717  | Sin resultado   | -              |
| 4  | 10773  | Sin resultado   | -              |
| 5  | 11300  | Unión de proteínas. Proteína asociada a BAL1 (Inhibidor de angiogénesis cerebral).  | 86             |
| 6  | 11979  | Proteína hipotética: NEMVEDRAFT_v1g156722   | 53             |
| 7  | 12323  | Proteína predicha: SCAF11   | 70             |
| 8  | 12395  | Sin resultado   | -              |
| 9  | 12448  | Factor de licencia de replicación del DNA: MCM7   | 97             |
| 10 | 13166  | Anticuerpo LTV1   | 96             |
| 11 | 14477  | Proteína hipotética LOC100704557  | 58             |
| 12 | 1474   | Interacción selectiva y de forma no covalente con cualquier proteína o complejo de proteínas (un complejo de dos o más proteínas que pueden incluir otras moléculas no proteicas). Subunidad 12 del complejo integrador.  | 77             |
| 13 | 14968  | Sin resultado   | -              |
| 14 | 15027  | Ocludina, isoforma CRA_a  | 98             |
| 15 | 15106  | Sin resultado   | -              |
| 16 | 15365  | Sin resultado   | -              |
| 17 | 17098  | Nucleósido fosforilasa  | 96             |
| 18 | 17219  | Proteína MuRF1  | 100            |
| 19 | 1727   | La interacción selectiva y de forma no covalente con cualquier proteína o complejo de proteínas (un complejo de dos o más proteínas que pueden incluir otras moléculas no proteicas). Proteína relacionada al tipo MARCKS | 78             |
| 20 | 17784  | Sin resultado   | -              |
| 21 | 18329  | Proteína precursora del factor de crecimiento de fibroblastos   | 60             |
| 22 | 3249   | Sin resultados  | -              |
| 23 | 3851   | Precursor de hemoproteína tipo 2  | 100            |
| 24 | 16187  | Gen EXT1 (Exostosis múltiple)   | 93             |

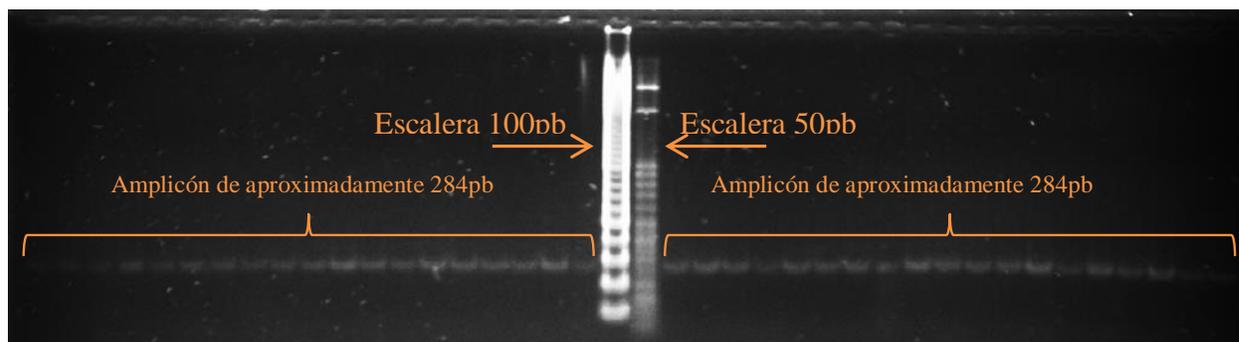


Fig. 35. Gel de agarosa con los productos de PCR al amplificar el locus Omy16187.

### 5.10.- Amplificación de EST-SSR

Se lograron amplificar 20 de 24 de los microsatélites utilizados, obteniendo un 83.3% de éxito, donde los microsatélites que no se lograron amplificar fueron Omy12448, Omy13166, Omy3249 y Omy3851.

### 5.11.- Análisis genético poblacional

De los 20 microsatélites amplificados 18 resultaron monomórficos y solamente dos polimórficos, los cuales son los loci Omy17219 con seis alelos, una  $H_e$  de 0.692 y  $H_o$  de 0.200 con valor de  $F_{IS}$  de 0.700, mostrando un déficit de heterocigotos. En el locus Omy17784 se detectaron siete alelos diferentes con una  $H_e$  de 0.805 y  $H_o$  de 0.800 con valor de  $F_{IS}$  de -0.244 ajustándose al EHW (Tabla 4).

Tabla 4. Diversidad alélica de microsatélites polimórficos amplificados en la trucha de San Pedro Mártir *Oncorhynchus mykiss nelsoni*.

| Locus    | intervalo de tamaño | # alelos | $H_e$ | $H_o$ |
|----------|---------------------|----------|-------|-------|
| Omy17219 | 160-178             | 6        | 0.692 | 0.200 |
| Omy17784 | 218-236             | 7        | 0.805 | 0.800 |

Las frecuencias alélicas oscilaron entre 0.0125 y 0.4000 en el locus Omy17219 y 0.0125 y 0.3000 en locus Omy17784 (Tabla 5).

Tabla 5. Frecuencias alélicas de los microsatélites polimórficos.

| Locus    | Alelos (pb) | Frecuencia | Locus    | Alelos (pb) | Frecuencia |
|----------|-------------|------------|----------|-------------|------------|
| Omy17219 | 160         | 0.0125     | Omy17784 | 218         | 0.0125     |
|          | 163         | 0.4000     |          | 221         | 0.3000     |
|          | 169         | 0.3625     |          | 224         | 0.0875     |
|          | 172         | 0.1500     |          | 227         | 0.0375     |
|          | 175         | 0.0250     |          | 230         | 0.3000     |
|          | 178         | 0.0250     |          | 233         | 0.0250     |
|          |             |            | 236      | 0.0750      |            |

## 6.- Discusión

---

La recolecta de individuos se llevó a cabo durante aproximadamente cinco horas, en ese tiempo se colectaron 69 organismos, lo cual es un número considerablemente bueno, tomando en cuenta el difícil acceso a su hábitat. Por otro lado, la técnica de captura empleada durante el muestreo, resulto adecuada y eficiente, ya que todos los animales recolectados se recuperaron del choque eléctrico a los pocos minutos de ser colocados en las trampas.

Nielsen y Sage (2001) al estudiar el complejo de truchas mexicanas utilizaron en promedio 11.41 organismos por población estudiada y mencionan que los tamaños de muestra en la mayoría de las poblaciones son pequeños debido al difícil acceso del área y al limitado número de peces que pueden ser recogidos en lugares poco accesibles. En otro estudio, Camarena-Rosales y colaboradores (2007) utilizaron en promedio 14 truchas por población en un análisis de haplotipos en truchas silvestres del noroeste mexicano. Por otro lado, Ruiz-Campos y colaboradores (1997) reportan que realizaron 17 muestreos en la sierra de San Pedro Mártir utilizando también la técnica de electropesca, donde en promedio capturaron 29.94 organismos por muestreo y la mayoría de las localidades muestreadas presentaron bajas densidades.

Además de los animales capturados, se contaba con un lote de truchas congeladas, por lo tanto se decidió extraer DNA a partir de 28 de estos y 12 de los organismos vivos, con el fin de estresar al menor número posible de organismos al momento de tomar la muestra. Por otro lado, el tejido utilizado para la extracción de DNA fue aleta, ya que con ello, no es necesario matar a los animales. Sin embargo, una de las variables más importantes durante la preparación de muestras para secuenciación masiva es la cantidad y calidad requerida para realizar el análisis, por lo tanto fue necesario utilizar a los organismos más grandes para este fin.

Para llevar a cabo los objetivos planteados en este trabajo, era suficiente secuenciar a un solo individuo, un ejemplo de ello es el trabajo de García de León y colaboradores (2013), en el cual secuenciaron un único organismo de trucha dorada *Oncorhynchus crysogaster*, logrando el desarrollo de 17 microsatélites tetranucleótidos a partir de 1 425 encontrados. Además, Gardner y colaboradores (2011), mencionan que en el desarrollo de microsatélites a partir de secuenciación masiva un solo organismo es suficiente, donde lo importante es obtener la cantidad necesaria de DNA para la secuenciación. Sin embargo, en algunas ocasiones es necesario utilizar más de un organismo para obtener una suficiente cantidad de DNA. No obstante, al momento de realizar la

secuenciación de la trucha de SPM, se decidió utilizar DNA de ocho muestras para una secuenciación múltiple, con el fin de poder alinear sus secuencias y buscar SNPs en el futuro. Además de permitir la búsqueda de SNPs, la secuenciación facilita el desarrollo de otros marcadores moleculares como los microsatélites y ayuda en la identificación de loci asociados a caracteres cuantitativos útiles en el desarrollo de programas de selección asistida.

Al realizar una búsqueda de los diferentes trabajos de secuenciación masiva en trucha, se detectó que se han utilizado distintas plataformas (Gahr y Rexroad, 2007, Sánchez et al., 2009, Salem et al., 2010, Le Cam et al., 2012, Tabla 6). De estos trabajos se resaltan el de Le Cam y colaboradores (2012), quienes secuenciaron a la trucha arcoíris con la plataforma 454-Titanium (Roche), produjeron 3 millones de secuencias con un promedio de 328 pb y lograron obtener un total de 219 814 contigs. Por otro lado, Gahr y Rexroad (2007) obtuvieron 3 840 EST al secuenciar una biblioteca de cDNA de trucha arcoíris, a partir de estas secuencias, ensamblaron un total de 524 contigs con una longitud promedio de 852 pb, también muy cercana a la alcanzada por este estudio. Como se puede observar la longitud de los fragmentos generados por secuenciación 454 y los obtenidos por EST es mayor que los generados por Illumina, pero en menor número, no obstante permiten producir un mayor número de contigs y también de mayor tamaño, debido a que resulta más fácil su ensamblaje que al utilizar fragmentos pequeños, sin embargo conforme se tiene un mayor número de lecturas es posible aumentar la profundidad o cobertura de los contigs.

Una de las dificultades que se presentan durante el ensamblaje de secuencias es debido al tamaño y a la complejidad del genoma de los salmónidos (Costa *et al.*, 2010), lo que resulta en bajos porcentajes de bases ensambladas y por ende en bajas coberturas finales. Por ejemplo Sánchez y colaboradores (2009) secuenciaron la trucha arcoíris *O. mykiss mykiss* con el equipo GS FLX System (Roche) obteniendo 459.825 Mb que, aunque representan casi un 20% del total de bases del genoma de la trucha, al ensamblarlas obtuvieron una cobertura total de 1.6 %, muy parecida a la alcanzada en este trabajo que fue de 1.5% en la secuenciación masiva, mientras que al ensamblar las ESTs recogidas de GenBank se obtuvo 1.0%. Estos porcentajes tan bajos son debidos también a la falta actual de secuencia genoma completo en los salmónidos que sirva de referencia (Le Cam *et al.*, 2012).

Tabla 6. Comparación de resultados de secuenciación y ensamblaje de ESTs en tres especies de salmónidos.

| Autores y año                    | Adzhubei et al., 2007 | García de León et al., 2013 | Sánchez et al., 2009 | Gahr y Rexroad 2007       | Salem et al., 2010 | Le Cam et al., 2012 | Presente estudio |                          |
|----------------------------------|-----------------------|-----------------------------|----------------------|---------------------------|--------------------|---------------------|------------------|--------------------------|
|                                  |                       |                             |                      |                           |                    |                     | EST              | Sec. masiva              |
| Especie                          | <i>S. salar</i>       | <i>O. chrysogaster</i>      | <i>O. mykiss</i>     |                           |                    |                     |                  |                          |
|                                  |                       |                             | <i>mykiss</i>        | <i>mykiss</i>             | <i>mykiss</i>      | <i>mykiss</i>       | <i>mykiss</i>    | <i>nelsoni</i>           |
| <b>Plataforma utilizada</b>      | ABI Prism 377         | GS FLX System               | GS FLX System        | Applied Biosystems 3730xl | GS FLX System      | Roche 454-Titanium  | GenBank          | MiSeq Sequencing Systems |
| <b>Lecturas</b>                  | 57212                 | 184458                      | 2044864              | 3840                      | 1380311            | 3067356             | 287967           | 4688266                  |
| <b>Longitud Promedio</b>         | nd                    | 328                         | 217                  | nd                        | 344                | 328                 | 666              | 240                      |
| <b>Número de bases (Mb)</b>      | nd                    | 60.502                      | 459.825              | nd                        | 447.826            | 1007.449            | 191.786          | 1125.183                 |
| <b>Contigs</b>                   | 6424                  | nd                          | 149096               | 524                       | 161818             | 242187              | 28237            | 126123                   |
| <b>Longitud Promedio</b>         | nd                    | nd                          | 259                  | 852                       | 758                | nd                  | 884              | 293                      |
| <b>Bases ensambladas (Mb)</b>    | nd                    | nd                          | 38.615               | 0.446                     | 122.658            | nd                  | 24.961           | 36.954                   |
| <b>% ensamblado</b>              | nd                    | nd                          | 8.397                | nd                        | 27.389             | nd                  | 13.015           | 3.284                    |
| <b>Singletons</b>                | 13595                 | nd                          | nd                   | 1256                      | 261071             | 22373               | 102967           | 2133862                  |
| <b>Longitud Promedio</b>         | nd                    | nd                          | nd                   | nd                        | 344                | 328                 | 666              | 240                      |
| <b>Bases no ensambladas (Mb)</b> | nd                    | nd                          | 421.21               | nd                        | 89.808             | 7.338               | 68.576           | 512.126                  |
| <b>% no ensamblado</b>           | nd                    | nd                          | 91.603               | nd                        | 72.611             | nd                  | 86.985           | 96.716                   |
| <b>% de cobertura total</b>      | nd                    | nd                          | 1.645                | 0.019                     | 5.225              | nd                  | 1.063            | 1.574                    |

Tamaño genómico de la trucha arcoíris *Oncorhynchus mykiss mykiss* = 2347.2 Mb. Calculado a partir de un valor C de 2.40 (Hardie y Hebert, 2004) y multiplicando 1pg por 978 Mb. nd= no hay datos.

Durante la anotación de los contigs obtenidos por secuenciación masiva, se pudo observar que las especies con las que se encontró mayor similitud entre las secuencias fueron el pez cebra *Danio rerio*, el pez cebra del Lago Malawi *Maylandia zebra*, la tilapia *Oreochromis niloticus* y el medaka *Oryzias latipes*, curiosamente dentro de los salmónidos aparece el hasta el onceavo lugar el salmón del Atlántico con solamente 5 270 relacionadas y hasta el último lugar la trucha arcoíris *Oncorhynchus mykiss* con 1882 secuencias. Esto se debió a que no se cuenta con una secuencia genómica de salmónidos que sirva de referencia y las especies que están mejor representadas en las bases de datos son las en las que se encuentra mayor similitud. En la anotación de las secuencias EST la mayoría de las secuencias coinciden con el pez cebra *Danio rerio* seguido de la tilapia *Oreochromis niloticus* y el salmón del Atlántico *Salmo salar*, a diferencia de la anotación de los contigs de secuenciación masiva, en las secuencias EST sí aparece en los primeros planos un salmónido, esto porque en las bases de datos existe un gran número de secuencias provenientes del salmón del Atlántico.

De la misma manera, al comparar los valores de E de la anotación de secuenciación y ESTs, observamos que la mayoría de las secuencias oscilan en  $1 \times 10^{-5}$  en la anotación de los contigs obtenidos por secuenciación masiva y la mayoría de los contigs provenientes de ESTs tienen valores de  $1 \times 10^{-180}$ , lo cual también se explica por la ausencia de una secuenciación de genoma completo que sirva de referencia para los salmónidos, caso contrario las secuencias EST si son abundantes en este grupo de peces.

Adzhubei y colaboradores (2007) anotaron genéticamente secuencias EST del salmón Atlántico y encontraron en la categoría de procesos biológicos: un 15.03% (4 659) de secuencias asociadas a procesos fisiológicos, seguidos de 10.94% (3 392) de secuencias para el metabolismo y 10.35% (3 207) para metabolismo celular. En nuestro estudio al anotar los contigs obtenidos por secuenciación se encontró que en la categoría procesos biológicos el mayor de número de secuencias anotadas estuvieron en los procesos celulares, con el 25% (7 176), seguido por procesos metabólicos con un 20% (5 710) y procesos de regulación biológica con un total de 12%. En la anotación de secuencias EST procesos biológicos tuvo el mayor de número de secuencias anotadas estuvieron en los procesos celulares con el 19% (6 403), seguido por procesos metabólicos con 18% (6 067) y procesos de regulación biológica con un total de 14% (4 504). En este caso la tendencia en la anotación de los contigs por secuenciación y los obtenidos por EST es la misma. Sin embargo, difieren de los resultados de Adzhubei y colaboradores

(2007) en cuanto a la distribución de las categorías de anotación encontrados en salmón. Salem y colaboradores (2010) en la anotación de trucha arcoíris encontraron que en la categoría de procesos biológicos, 25% de los genes están asociados con los procesos celulares, siendo el mismo porcentaje de anotación que en nuestro trabajo de secuenciación. Dentro de la categoría de función molecular, el 46% está relacionado con la unión de proteínas seguida de la actividad catalítica con 32%, muy parecido a nuestro trabajo con EST, donde el mayor número de secuencias anotadas correspondió a las funciones de unión con el 50% (8 714), seguido por la actividad catalítica que tuvo un total de 29% (5 082) y en componentes celulares, 59% de los genes que están relacionados con el componente célula y el 24% estuvieron relacionados con organelo, misma tendencia que en la notación tanto de contigs por secuenciación como en los de EST donde el mayor número de secuencias asociadas fue para célula seguida por organelo con un 46% (5 922) y 29% (3668) en secuenciación y 41% (8 979) y 27% (5 805) en EST.

Al realizar una anotación genética de los microsatélites amplificados, 13 tuvieron similitud con secuencias genéticas reportadas, desafortunadamente sólo dos de ellos resultó polimórfico. Por lo tanto los demás microsatélites aunque están relacionados con funciones conocidas no podrán ser aprovechados marcadores de selección asistida, pero es posible que contengan polimorfismo de un solo nucleótido (SNPs). Sin embargo, entre los loci polimórficos el locus Omy17219 resultó estar asociado al gen que codifica la proteína MuRF1, este dato es muy interesante, ya que MuRF1 se ha identificado como un conector de metabolismo de la energía muscular y la síntesis de proteínas (Koyama *et al.*, 2008) y existen reportes en porcinos donde han desarrollado microsatélites asociados a esta proteína (Shen *et al.*, 2011). Por lo anterior Omy17219 puede ser considerado como un marcador molecular candidato a ser utilizado en selección asistida donde se busque aumentar las tasas de crecimiento muscular.

En otro estudio, García de León y colaboradores (2013) secuenciaron la trucha dorada, *Oncorhynchus chrysogaster*, donde obtuvieron 184 458 lecturas, con longitud promedio de 328 pb, misma longitud alcanzada por el estudio de Le Cam y colaboradores (2012). Dentro de estas lecturas, encontraron 18 570 loci microsatelitales, lo que resulta en un 10.07% de lecturas con contenido de secuencias microsatelitales. En nuestro trabajo a partir de 28 237 contigs formados por secuencias EST de trucha arcoíris *Oncorhynchus mykiss mykiss*, 3 845 contig tuvieron al menos un microsatélite, lo que resulta en un 13.62% de contenido de secuencias microsatelitales. En las secuencias obtenidas por secuenciación masiva, se encontraron 10 100 secuencias con

motivos de repetición a partir de los 126 288 contigs. Lo que significa que casi un 8% de los contigs contiene microsátélites, un bajo porcentaje en comparación a los casos anteriores, debido posiblemente a una menor longitud de los contigs. Siemon y colaboradores (2005) utilizaron 143 753 EST de salmón del Atlántico para formar 73 389 contigs, en donde identificaron 1 975 contigs que contenían motivos de repetición, resultando en una frecuencia de microsátélites del 2.7%. Por otro lado, Wang y colaboradores (2007) encontraron microsátélites en el 5.5% de las secuencias EST de la carpa común que utilizaron. Esto nos lleva pensar que la presencia de los microsátélites puede variar entre especies y también depende de la calidad de las secuencias originales, de los programas bioinformáticos que hayan sido utilizados y qué tan estrictos sean los parámetros fijados para la búsqueda de microsátélites.

Los microsátélites son lo suficientemente abundantes en el genoma de los organismos y normalmente dominan los de tipo dinucleótido que aunque los más comunes son, a la vez, los más difíciles de genotipificar (Abdelkrim *et al.*, 2009). Por esta razón es que nuestros criterios de selección estuvieron encaminados al diseño de iniciadores para microsátélites tri y tetra nucleótidos. Como se ha mencionado, se diseñaron iniciadores a partir de dos fuentes; una a partir de ESTs y otra de secuenciación masiva, donde se utilizaron los microsátélites encontrados en las ESTs para su validación y análisis, ya que éstas al estar asociadas al transcriptoma tienen la ventaja de estar correlacionados al fenotipo ofreciendo la oportunidad de utilizarlos como herramienta para el análisis de variabilidad genética y como marcadores útiles en selección asistida (Serapion *et al.*, 2004, Aguilar - Espinoza, 2010).

En la amplificación de EST-SSR, Siemon y colaboradores (2005) diseñaron iniciadores para amplificar 158 SSR, de los cuales lograron amplificar 76 en el salmón del Atlántico. El 96% de estos marcadores también amplificaron en otros salmónidos y 61 resultaron polimórficos. En nuestro caso, con EST-SSR en la trucha de San Pedro Mártir sólo se encontró polimorfismo en el 20% de los loci, mientras que García de León y colaboradores (2013), con 17 loci, tuvieron el 47% polimórficos en la trucha de SPM, por lo que se observa que el polimorfismo en EST-SSR que obtuvimos en este estudio es muy bajo al compararlos con otros salmónidos e incluso en la misma especie con microsátélites desarrollados en regiones genómicas, lo que podría sugerir un alto grado de endogamia en esta trucha. Sin embargo, hay que tomar en cuenta que en nuestro estudio se utilizaron microsátélites tri y tetranucleótidos y que se ha reportado en varias especies que los microsátélites tetranucleótidos tienen tasas de mutación más bajas que los de tipo

dinucleótidos (Kruglyak et al. 1998; Schug et al. 1998). Por consiguiente es necesario realizar el análisis con SSR dinucleótidos y además con SSR compuestos.

En cuanto a otros parámetros de diversidad genética y al analizar diversos trabajos (Tabla 7) se encontró que Arslan y Fevzi (2010) amplificaron cinco loci de microsatélites en la trucha café *Salmo trutta* obteniendo en promedio 7.4 alelos por locus una  $H_o$  de 0.254 y una  $H_e$  0.565 y se ajustaron al EHW, García de León y colaboradores (2013) en 19 ejemplares de trucha dorada y 30 ejemplares de truchas de seis diferentes cuencas de la Sierra Madre Occidental (cinco muestras de cada uno), en cinco de Trucha de SPM y cinco *O. mykiss* de cultivo, obtuvo una media de 16 alelos por locus. Encontraron una  $H_o$  promedio de 0.650, y una  $H_e$  promedio de 0.749, en cuanto a la trucha de SPM, se obtuvo un promedio de 6.5 alelos por locus, una  $H_e$  media de 0.748 y una  $H_o$  de 0.500, donde al igual que en los otros trabajos se ajustan al EHW. La diversidad alélica de los 6.5 alelos por locus en promedio para los dos loci polimórficos obtenida en la trucha de SPM es muy parecida a la que encontraron Nielsen y Sage (2001) en 11 loci de microsatélites en 12 poblaciones de trucha distribuidas en el Río Yaqui, Río Casas Grandes y Río Mayo, con una media de 6.6 alelos. Pero nuevamente quedan por debajo en comparación con otras truchas nativas como la trucha arcoíris del río McCloud en California, Estados Unidos, donde se encontró una media de 17 alelos por locus al utilizar 11 loci de microsatélites (Nielsen et al., 1999). En otro trabajo, Villarreal-Zazueta (2012), al estudiar seis localidades de trucha de San Pedro Mártir por medio de alozimas, encontró que la localidad de Mike's Sky tiene una heterocigosidad observada más alta que la esperada con una media de 0.586, muy parecida a la encontrada en el presente estudio, que fue de 0.500, esto se explica ya que ambas muestras de muestra provienen de la misma población.

Tabla 7. Comparación de resultados de análisis de microsatélites en truchas y salmón.

| Fuente                              | n    | Especie                  | #Alelos | H <sub>O</sub> | H <sub>E</sub> | EHW |
|-------------------------------------|------|--------------------------|---------|----------------|----------------|-----|
| Arslan y Fevzi, 2010                | 358  | <i>Salmo trutta</i>      | 7.4     | 0.254          | 0.565          | +   |
| García de León <i>et al.</i> , 2013 | 19   | <i>O. chrysogaster</i>   |         |                |                |     |
|                                     | 5    | <i>O. mykiss</i>         | 16*     | 0.650*         | 0.749*         | +   |
|                                     | 5    | <i>O. mykiss nelsoni</i> |         |                |                |     |
| Nielsen <i>et al.</i> , 1999        | 1023 | <i>O. mykiss stonei</i>  | 17      | 0.680          | nd             | +   |
| Nielsen y Sage, 2001                | 124  | <i>O. mykiss</i>         | 6.6     | 0.346          | 0.347          | +   |
| Presente estudio                    | 40   | <i>O. mykiss nelsoni</i> | 6.5     | 0.5            | 0.748          | +   |

+ Poblaciones que se encuentran en Equilibrio de Hardy Weinberg.

\* En promedio de las tres especies utilizadas por García de León *et al.* (2013)

## 7.- Conclusiones

---

La técnica de captura por electropesca empleada durante el muestreo, resulta adecuada y eficiente, ya que todos los animales recolectados se recuperaron del choque eléctrico a los pocos minutos de ser atrapados.

Se corrobora que la técnica de secuenciación masiva es un método eficiente para el análisis genético, anotación y desarrollo de marcadores moleculares en especies no modelo.

Tanto en las lecturas de secuenciación masiva como en las ESTs, el tipo de microsatélites dominantes fueron los dinucleótidos, seguido de los tri, tetra y pentanucleótidos, estos últimos sólo se encontraron en las lecturas de secuenciación masiva.

El tamaño y la complejidad del genoma de los salmónidos, combinado con la falta de un genoma de referencia o de un organismo estrechamente relacionado con la trucha de SPM, volvieron difícil el ensamblaje y anotación de secuencias de la trucha.

Se identificaron las funciones de algunos de los microsatélites a través de la búsqueda de similitud de secuencias donde destaca el gen MurF1, el cual puede ser considerado como un candidato potencial como marcador molecular de selección asistida.

La utilización de microsatélites derivados de EST permitirá amplificar microsatélites asociados a funciones de interés comercial.

Al contar con dos conjuntos de iniciadores para la amplificación de microsatélites, recomendamos la utilización de los provenientes de regiones neutras para la realización de análisis genéticos poblacionales con el fin de aumentar la probabilidad de amplificación polimórfica.

Se cuenta con una base de datos con 6842 secuencias que contienen microsatélites que al desarrollarse en trabajos futuros pueden resultar útiles en la conservación de la trucha de San Pedro Mártir, ayudando entender la estructuración de las poblaciones silvestres y su parentesco.

Por otro lado, estos marcadores también podrían ser de gran ayuda en los programas de cría selectiva más tradicionales, por su posible utilidad en la asignación de parentesco entre crías y reproductores, en la estimación de endogamia y variabilidad genética.

De los 20 EST-SSR 18 resultaron fueron monomórficos. Estos resultados dan soporte a la hipótesis, de que la trucha de SPM estudiada cuenta con una baja variabilidad genética, en la región expresable de su genoma.

Lo anterior sugiere que la presión de selección a la que está sometida la trucha de SPM, la ha llevado a una deriva genética donde se conservan características muy específicas y necesarias para su supervivencia, aunado a la fragmentación del ambiente provocando el aumento en el número de homocigotos y la expresión de genes deletéreos.

## 8.- Recomendaciones

Validar la amplificación y grado de polimorfismo de los microsatélites desarrollados a partir de secuenciación masiva.

Realizar un análisis genético poblacional, utilizando marcadores moleculares desarrollados de manera específica para la trucha de SPM y donde se abarque el total de localidades que conforman a la población. Planificar programas de manejo y conservación de la trucha de SPM con base a los resultados que se obtengan en cuanto a su estructura y variabilidad genética.

Comparar mediante un análisis genético poblacional a la trucha de SPM con la trucha arcoíris (*Oncorhynchus mykiss mykiss*) para determinar si los microsatélites derivados de las EST tiene el mismo grado de polimorfismo.

Evitar la introducción y el asentamiento de piscifactorías de otras especies de truchas en el hábitat de la trucha de SPM.

Promover el cultivo de la trucha de SPM como especie regional.

Utilizar los marcadores moleculares que cuentan con anotación genética y determinar si hay polimorfismo con el fin de utilizarlos en el desarrollo de programas de cría selectiva para el cultivo de la trucha de SPM.

Incrementar el número de marcadores moleculares para la caracterización de los gametos criopreservados provenientes de reproductores de esta especie.

Buscar, desarrollar y validar marcadores de tipo SNPs dentro de la base de datos generada en este estudio que sean útiles para el establecimiento de un programa de selección asistida de la trucha de SPM.

## Referencias bibliográficas

---

- Adzhubei, A., Vlasova, A., Hagen, H., Ruden, T., Laerdahl, J., K. y Høyheim, B. 2007. Annotated expressed sequence tags (ESTs) from pre-smolt Atlantic salmon (*Salmo salar*) in a searchable data resource. *BMC genomics*, 8(1): 209 p.
- Abdelkrim, J., Robertson, B., Stanton, J., Gemmell, N. 2009. Fast, cost- effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques*, 46(3): 185–192p.
- Aguilar-Espinoza, Andrea. 2010. Caracterización de marcadores EST-SSR en *H. rufescens* desde secuencias heterólogas. Tesis de pregrado en Ingeniería en Biotecnología Marina y Acuicultura. Universidad de Concepción. Concepción, Chile.
- Aguilar-Juárez, Marisela. 2010. Inducción a la maduración gonádica y conservación del espermatozoide de la trucha de San Pedro Mártir *Oncorhynchus mykiss nelsoni* (Evermann). Tesis de Doctorado en Ciencias en Ecología Molecular y Biotecnología. Universidad Autónoma de Baja California. Ensenada, México.
- Andreassen, R., Lunner, S., y Høyheim, B. 2009. Characterization of full length sequenced cDNA inserts (FLICs) from Atlantic salmon (*Salmo salar*). *BMC genomics*, 10(1): 502 p.
- Aranguren, J. A., Román, R., Isea, W., Villasmil, Y. y Jordana, J. 2005. Los microsatélites (STR's), marcadores moleculares de ADN por excelencia para programas de conservación: Una revisión. *Archivos Latinoamericanos de Producción Animal*, 13(1): 1-6 p.
- Arif, I., A. y Khan, H., A. 2009. Molecular markers for biodiversity analysis of wildlife animals: a brief review. *Animal Biodiversity and Conservation*, 32(1): 9-17 p.
- Arslan, S. y Bardakci, F. 2010. Genetic structure of brown trout (*Salmo trutta*) populations from Turkey based on microsatellite data. *Biochemical genetics*, 48(11-12): 995–1014 p.
- Astaneï, I., Gosling, E., Wilson, J. y Powell, E. 2005. Genetic variability and phylogeography of the invasive zebra mussel, *Dreissena polymorpha* (Pallas). *Molecular Ecology*, 14(16): 1655-1666 p.
- Awise, J., C. 2004. Molecular markers, natural history, and evolution. Sinauer. Segunda edición. Sunderland, Massachusetts, 541p.
- Azofeifa, A., 2006. Uso de marcadores moleculares en plantas: aplicaciones en frutales del trópico. *Agronomía Mesoamericana*, 17(2): 221-242 p.
- Berg, W., J. y Gall, G., A., E. 1988. Gene flow and genetic differentiation among California coastal rainbow trout populations. *Canadian Journal of Fisheries and Aquatic Sciences*, 45(1): 122-131 p.

- Bouck, A., Vision, T. 2007. The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology*, 16(5): 907-924 p.
- Boussaha, M., Guyomard, R., Cabau, C., Esquerre, D. y Quillet, E. 2012. Development and characterization of an expressed sequence tags (EST) derived single nucleotide polymorphisms (SNPs) resource in rainbow trout. *BMC Genomics*, 13(1): 238-248 p.
- Camarena-Rosales, F., Ruiz-Campos, G., De la Rosa, J., Mayden R., L., Hendrickson D., A., Varela, A. y García-de-León, F. J. 2008. Mitochondrial haplotype variation in wild trout populations (Teleostei: Salmonidae) from Nort Western Mexico. *Reviews in Fish Biology and Fisheries*, 18(1): 33-45 p.
- Cantacessi, C., Jex, A., R., Hall, R., S., Young, N., D., Campbell, B., E., Joachim, A., Nolan, M., J., Abubucker, S., Sternberg, P., W., Ranganathan, S., Mitreva, M. y Gasser, R., B., 2010. A practical, bioinformatic workflow system for large data sets generated by next-generation sequencing. *Nucleic Acids Research*, 38(17): 1-12 p.
- Chen, S., L., Liu, Y., G., Xu, M., Y., Li, J., 2005. Isolation and characterization of polymorphic microsatellite loci from an EST library of red sea bream (*Chrysophrys major*) and cross species amplification. *Molecular Ecology Notes*, 5(2): 215-217 p.
- Conesa, A., Gotz, S., García, J., M., Terol, J., Talon, M. y Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18): 3674-3676 p.
- Costa, V., Angelini, C., De Feis, I., Ciccodicola, A. 2010. Uncovering the complexity of transcriptomes with RNA-Seq. *Journal of Biomedicine and Biotechnology*, 2010(2010), 1-20p.
- Davey, J., L., Blaxter, M., W. 2010. RADSeq: next generation population genetics. *Briefings in Functional Genomics*, 95(5-6): 416-423 p.
- Evermann, B., W. 1908. Descriptions of a new species of trout (*Salmo nelsoni*) and a new cyprinodont (*Fundulus meeki*) with notes on other fishes from Lower California. *Proceedings of the Biological Society of Washington*, 21(1): 19-30 p.
- Excoffier, L., Laval, G. y Schneider, S. 2005. Arlequin version 3.0: An integrated software package for population genetic data analysis. *Evolutionary Bioinformatics Online*, 1: 47-50 p.
- Excoffier, L., Smouse, P. y Quattro, J., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, 131(2): 479-491 p.
- Féral J., P. 2002. How useful are the genetic markers in attempts to understand and manage marine biodiversity? *Journal of Experimental Marine Biology and Ecology*, 268(2): 121-145 p.

- Gahr, S., Rexroad, C., E., Rise, M., L., Hunt, P., y Koop, B. 2007. A survey of expressed sequence tags from the rainbow trout (*Oncorhynchus mykiss*) pituitary. *Animal biotechnology*, 18(3): 213-30 p.
- García-de León, F., J., Lozano-Garza, O., A., Mayden, R., L., Delgado-Vega, R., Camarena-Rosales, F., Ruiz-Campos, G., Escalante-Sánchez, M., Ruiz-Luna, A., Del Río-Portilla, M., Barriga-Sosa, I., Arcos-Ortega, G. y Paniagua-Chávez, C. 2013. Isolation and characterization of 17 tetranucleotide microsatellite loci in the Mexican golden trout (*Oncorhynchus chrysogaster*, Needham y Gard 1964) derived from 454 pyrosequencing. *Conservation Genetics Resources*, 5(4): 1-4p.
- Gardner, G., Fitch, J., Bertozzi, T. y Lowe, A. J. 2011. Rise of the machines recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources*, 11(6); 1093-1101p.
- Geist, J. y Kuehn, R. 2005. Genetic diversity and differentiation of central European freshwater pearl mussel (*Margaritifera margaritifera* L.) populations: implications for conservation and management. *Molecular Ecology*, 14(2): 425- 39 p.
- Glenn, T., C. 2011. Field guide to next generation DNA sequencers. *Molecular Ecology Resources*, 11(5): 759-69 p.
- Goudet J. 1995. Fstat: A computer program to calculate F-statistics. *Journal of Heredity*, 86(6): 485-486 p.
- Hartl, D., L. y Clark, A., G. 1997. Principles of population genetics. Sinauer. Tercera edición. Sunderland., Massachusetts, 542 p.
- Hayes, B., He, J., Moen, T. y Bennewitz, J. 2006. Use of molecular markers to maximise diversity of founder populations for aquaculture breeding programs. *Aquaculture*, 255(1): 573-578.
- Hedrick, P., W. 1996. Bottlenecks or metapopulation in cheetahs. *Conservation Biology*, 10(3): 897-899 p.
- Hedrick, P., W. 2000. Genetics of populations. Jones y Bartlett. Boston, Segunda edición, 545 p.
- Jelks, H., L., Walsh, S., J., Burkhead N., M., Contreras, S., Díaz, E., Hendrickson, D. A., Lyons, J., Mandrak N. E., McCormick, F., Nelson, J. S., Platania, S. P., Poter, B. A., Renaud, C. B., Schmitter, J. J., Taylor, E. B. y Warren, M., L. 2008. Conservation Status of Imperiled North American Freshwater and Diadromous Fishes. *Fisheries*, 38(8): 372-407 p.
- Koyama, S., Hata, S., Witt, CC, Ono, Y, Lerche, S., Ojima, K., Chiba, T., Doi, N., Kitamura, F, Tanaka, K., Abe. K., Witt, S.H., Rybin, V, Gasch, A., Franz, T., Labeit, S. y Sorimachi, H., 2008. Muscle RING-finger protein-1 (MuRF1) as a connector of muscle energy metabolism and protein synthesis. *Journal of Molecular Biotechnology*, 376(5): 1224-1236p.

- Kruglyak, S., Durrett, R., Schug, M. y Aquadro, C. 1998 Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences*, 95(10): 774– 778 p.
- Le Cam, A., Bobe, J., Bouchez, O., Cabau, C., Kah, O., Klopp, C. y Guiguen, Y. 2012. Characterization of rainbow trout gonad, brain and gill deep cDNA repertoires using a Roche 454 Titanium sequencing approach. *Gene*, 500 (1), 32-9 p.
- Lepais, O. y Bacles, C. 2011. Comparison of random and SSR enriched shotgun pyrosequencing for microsatellite discovery and single multiplex PCR optimization in *Acacia harpophylla*. *Molecular Ecology Resources*, 11(4):711-724 p.
- Lind, C., E., Evans, B., S., Taylor, J., J. y Jerry, D., R. 2007. Population genetics of a marine bivalve, *Pinctada maxima*, throughout the Indo-Australian Archipelago shows differentiation and decreased diversity at range limits. *Molecular Ecology*, 16(24): 5193-203 p.
- Litt, M. y Luty, J., A. 1989. A hypervariable microsatellite revealed by in vitro amplification of dinucleotide repeats within the cardiac muscle actine gene. *American Journal of Human Genetic*, 44(3): 397-401 p.
- Liu, Z., J., Cordes, J., F. 2004. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*, 238(1): 1-37 p.
- Maneeruttanarungroj, C., Pongsomboon, S., Wuthisuthimethavee, S., Klinbunga, S., Wilson, K.J., Swan, J., Li, Y., Whan, V., Chu, K., H., Li, C., P., Tong, J., Glenn, K., Rothschild, M., Jerry, D., Tassanakajon, A. 2006. Development of polymorphic expressed sequence tag derived microsatellites for the extension of the genetic linkage map of the black tiger shrimp (*Penaeus monodon*). *Animal Genetics*, 37(4): 363-368 p.
- Martínez, P. 2005. Aplicaciones de la genética para la mejora de la acuicultura. *Boletín del Instituto Español de Oceanografía*, 21 (1-4): 225-238 p.
- Maxam, A., M. y Gilbert, W. 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2): 560-564 p.
- Mburu, D., N., Ochieng, J., W., Kuria, S., G., Jianlin, H. y Kaufmann, B. 2003. Genetic diversity and relationships of indigenous Kenyan camel (*Camelus dromedarius*) populations: implications for their classification. *Animal Genetics*, 34(1): 26-32 p.
- McCormack, J., E., Hird S., M., Zellmer, A., J., Carstens, B., C., Brumfield, R., T. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics, *Molecular Phylogenetics and Evolution*, 66 (2): 526-538 p.
- Meglecz, E., Costedoat, C., Dubut, V., Gilles, A., Malausa, T., Pech, N., Martin, J., F. 2010. QDD: a user friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics*. 26(3): 403-404 p.

- Métais, I., Hamon, B., Jalouzot, R., y Peltier, D. 2002. Structure and level of genetic diversity in various bean types evidenced with microsatellite markers isolated from a genomic enriched library. *Theoretical and Applied Genetics*, 104(8); 1346-1352p.
- Metzker, M., L. 2005. Emerging technologies in DNA sequencing. *Genome Research*, 15(12): 1767-1776 p.
- Metzker, M., L., 2010. Sequencing technologies the next generation. *Nature Reviews Genetics*, 11(1): 31-46 p.
- Miller, R., R. 1950. Notes on the cutthroat and rainbow trouts with description of a new species from the Gila River, New Mexico. *Occasional Papers of the Museum of Zoology, University of Michigan*, 529, 1-42 p.
- Needham, P., R. 1938. Notes on the introduction of *Salmo nelsoni* Evermann into California from Mexico. *Transactions of the American Fisheries Society*, 67(1) 139-146 p.
- Nelson, E., W. 1921. Lower California and its natural resources. *Memoirs of the National Academy of Sciences*, 16(1):1-194 p.
- NGHRI. 2013. Sequencing cost. <http://www.genome.gov/sequencingcosts/>.
- Nielsen, J., L. 1996. Molecular genetics and the conservation of salmonid biodiversity: *Oncorhynchus* at the edge of their range. En: T. Smith y B. Wayne (ed.) *Molecular Genetics Approaches in Conservation*, Oxford University Press, London, 383–398 p.
- Nielsen, J., L., y Sage, G., K. 2001. Microsatellite analyses of the trout of northwest Mexico. *Genetica*, 111(1-3): 269-278 p.
- Nowrousian, M. 2010. Next generation sequencing techniques for eukaryotic microorganisms: sequencing based solutions to biological problems. *Eukaryotic cell*, 9(9): 1300-10 p.
- O'Brien, S., J. 1991 Molecular genome mapping lessons and prospects. *Current Opinion in Genetics and Development*, 1(1): 105-111 p.
- O'Brien, S., J. 1994. The cheetha's conservation controversy. *Conservation Biology*, 8(4): 1153-1155 p.
- O'Brien, S., J., Wildt, D., E., Bush, M., Caro, T., M., Fitzgibbon, C., Aggundey, I. y Leakey. R., E. 1987. East African cheetahs: evidence for two population bottlenecks *Proceedings of the National Academy of Sciences*, 84(2): 508-511 p.
- Oosterhout, C., Hutchinson, W., Wills, D. y Shipley P. 2004. Microchecker: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4(3):535-538 p.
- Pareek, C., S., Smoczynski, R., y Tretyn, A. 2011. Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, 52(4), 413-35 p.

- Pérez, F., Ortiz, J., Zhinaula, M., Gonzabay, C., Calderón, J., Volckaert, J. 2005. Development of EST-SSR markers by data mining in three species of shrimp: *Litopenaeus vannamei*, *Litopenaeus stylirostris*, and *Trachypenaeus biridy*. *Marine Biotechnology*, 7(5): 554-569 p.
- Rozen, S., Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Molecular Biology*. 132(1): 365-386 p.
- Ruiz-Campos, G. 1993. Bionomía y ecología poblacional de la trucha arcoíris *Oncorhynchus mykiss nelsoni* (Evermann), de la Sierra San Pedro Mártir, Baja California, México. Tesis de Doctorado en Ciencias con Especialidad en Ecología Acuática y Pesca. Universidad Autónoma de Nuevo León. Monterrey, México.
- Ruiz-Campos, G. y Pister E., P. 1995. Distribution, habitat, and current status of the San Pedro Martir rainbow trout, *Oncorhynchus mykiss nelsoni* (Evermann). *Bulletin of the Southern California Academy of Sciences*, 94(2): 131-148 p.
- Salem, M., Rexroad, C., E., Wang, J., Thorgaard, G., H., Yao, J. 2010. Characterization of the rainbowtrout transcriptome using Sanger and 454 pyrosequencing approaches. *BMC Genomics*, 11(1): 564 p.
- Sánchez, C., Smith, T., Wiedmann, R., Vallejo, R., Salem, M., Yao, J., y Rexroad, C. 2009. Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *BMC Genomics*, 10(1); 559p.
- Sanger, F., Nicklen, S., Coulson, A., R. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci.*, 74(12): 5463-5467 p.
- Schlötterer, C. 2004. The evolution of molecular markers just a matter of fashion? *Nature reviews. Genetics*, 5(1):63-9 p.
- Serapion, J., Kucuktas, H., Feng, J.N., Liu, Z., J. 2004. Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). *Mar. Biotechnol*, 6(4): 364-377 p.
- Shen, H., Zhao, S., Cao, J., Li, X. y Fan, B. 2011. Porcine MuRF2 and MuRF3: molecular cloning, expression and association analysis with muscle production traits. *Molecular biology reports*, 38(8), 5115–23p.
- Shokralla, S., Spall J., Gibson, J., Hajibabaei, M. 2013. Next generation sequencing technologies for environmental DNA research. *Molecular Ecology*, 21(8): 1794-1805 p.
- Schug, M., Hutter C., Wetterstrand, K., Gaudette, M., Mackay, T. y Aquadro, C. 1998 The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Molecular Biology and Evolution*, 15(12), 1751–1760 p.
- Siemon, N., G., Chang, A., Brown, G., D., Koop, B., F. and Davidson, W., S. 2005. Type I microsatellite markers from Atlantic salmon (*Salmo salar*) expressed sequence tags. *Molecular Ecology Notes*, 5(4): 762-766 p.

- Smith, C., 1984. Estimated costs of genetic conservation in farm animals. En: FAO/UNEP Technical Consultation (ed). Animal Genetic Resources Conservation and Management. Proceedings of the Animal Production and Health, 44:1, 44p
- Smith, R., H. 1991. Rainbow trout, *Oncorhynchus mykiss*. En: Stolz, J. and Schnell, J. (eds.), Trout. The Wildlife Series, Stackpole Books, Harrisburg, 304-322 p.
- Snyder, J., O., 1926. The trout of the Sierra San Pedro Martir, Lower California. University of California Publications in Zoology, 21(1): 419-426 p.
- Stein, L., D. 2010. The case for cloud computing in genome informatics. Genome biology, 11(5), 207 p.
- Tautz, D., 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research, 17(16): 6463-6471 p.
- Villarreal-Zazueta, N. A. 2012. Análisis de la estructura genético poblacional de la Trucha Arcoíris de la Sierra San Pedro Mártir *Oncorhynchus mykiss Nelsoni* (Evermann, 1908). Tesis de Maestría en Ciencias en Ecología Molecular y Biotecnología. Facultad de Ciencias Marinas, Universidad Autónoma de Baja California, Baja California, México. 116p.
- Wang, D., Liao, X., Cheng, L., Yu, X., y Tong, J. 2007. Development of novel EST-SSR markers in common carp by data mining from public EST sequences. Aquaculture, 271(1-4), 558-574 p.
- Wang, H.X., Li, F.H., Xiang, J.H., 2005. Polymorphic EST-SSR markers and their mode of inheritance in *Fenneropenaeus chinensis*. Aquaculture 249(4): 107-114 p.
- Weir, B. S. y C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution, 38(6): 1358-1370 p.
- Weir, B., S. y Basten, C., J. 1990. Sampling strategies for distances between DNA sequences. Biometrics, 46(1): 551-582 p.
- Wright, J., M. 1993. DNA fingerprinting of fishes. En: P.W. Hochachka y T. Mommsen (eds.). Biochemistry and Molecular Biology of Fishes. Elsevier, Amsterdam, 57-91 p.
- Xiong, M., Zhao, Z., Arnold, J., y Yu, F. 2010. Next-generation sequencing. Journal of Biomedicine and Biotechnology, 2010. Editorial, doi:10.1155/2010/370710.
- Yue, G.H., Ho, M.Y., Orban, L., Komen, J., 2004. Microsatellites within genes and ESTs of common carp and their applicability in silver crucian carp. Aquaculture 234(4): 85-98 p.
- Zane, L., L. Bargelloni y T. Patarnello. 2002. Strategies for microsatellite isolation: A review. Molecular Ecology, 11(1): 1-16 p.
- Zhan, A., J. Hu, X. Hu, Z. Zhou, M. Hui, S. Wang, W. Peng, M. Wang y Z. Bao. 2009. Fine-scale population genetic structure of Zhikong scallop (*Chlamys farreri*): do local marine currents drive geographical differentiation? Marine Biotechnology, 11(2): 223-35 p.

## Anexo

En la tabla 8 se muestra una serie de iniciadores tetra y pentanucleótidos pertenecientes a la base de datos que se encuentra en el laboratorio de genética acuícola, los cuales serán sintetizados para con el fin de validar su uso en análisis genético poblaciones de la trucha de SPM.

Tabla 8. Iniciadores de microsatélites tetra y pentanucleótidos sin anotación genética.

|    | Contig | Motivo de repetición | Puro/compuesto | Secuencia de iniciadores                                     | TM °C        | Tamaño esperado (pb) | GC %           |
|----|--------|----------------------|----------------|--|--------------|----------------------|----------------|
| 1  | 17592  | TCCTC                | C              | F:CCCGCTCCTCTCTCCTCT<br>R:GGGAGGAAAGGAATGAGGTAAT             | 61.0<br>59.6 | 150                  | 63.2%<br>45.5% |
| 2  | 17799  | TCGTG                | C              | F:CGTTCCATTCGAGTCTTTCG<br>R:ATCCTCTCGGCCAATTCAGT             | 60.7<br>60.9 | 228                  | 50.0%<br>50.0% |
| 3  | 54266  | GTGTA                | P              | F:ACCTGGGTAGTGCAGTGTG<br>R:ACAGGAACTACATCGCACC               | 58.6<br>57.6 | 151                  | 55.0%<br>50.0% |
| 4  | 55207  | CACTA                | C              | F:CACTACACTCCACTCCACTACT<br>R:GATAAGTCCTCGTCTCGTGGG          | 57.4<br>61.0 | 288                  | 50.0%<br>57.1% |
| 5  | 64556  | CCCTA                | P              | F:CCGTGGGACATTCTTACCCT<br>R:ATCCTTGGGATGTCCCTACC             | 61.1<br>60.0 | 141                  | 55.0%<br>55.0% |
| 6  | 66501  | CTCTC                | P              | F:ACCACATGTAGACGGCTCCC<br>R:AGGCAACCATATCTCTGGCA             | 62.2<br>60.6 | 166                  | 60.0%<br>50.0% |
| 7  | 81036  | ACAGT                | P              | F:CCCTGGCTACCAATAAACCTC<br>R:AAGCATGGAGACTGCAAACC            | 59.8<br>60.2 | 171                  | 52.4%<br>50.0% |
| 8  | 89890  | TACAT                | C              | F:AACGAACCGAAATGCTTCAA<br>R:GAGCGGAGTAGGCTACCCAT             | 60.6<br>60.6 | 146                  | 40.0%<br>60.0% |
| 9  | 822    | GATG                 | P              | F:GAAGAGCAAGAGGGAGGAGG<br>R:CCTTCCTCTGGTAGGCCTGT             | 60.4<br>60.6 | 226                  | 60.0%<br>60.0% |
| 10 | 31109  | CCTC                 | P              | F:TGTATGTATGTATGTATGCACTGTCTC<br>R:GAGAGACACAGAGAGACAAATTGAA | 57.3<br>59.0 | 107                  | 37%<br>40%     |
| 11 | 1673   | GTTT                 | P              | F:CCATGGGAACAATTTGTATGG<br>R:ATTCGGAGCAGATGTTGCAT            | 59.9<br>60.6 | 154                  | 42.9%<br>45.0% |
| 12 | 26143  | GCAC                 | P              | F:TAGCACATGCAGGCAGGAAG<br>R:TTCTCCTCTCGTTTCTCCTCC            | 62.4<br>59.9 | 184                  | 55.0%<br>52.4% |
| 13 | 3625   | AAAC                 | P              | F:GCCAGTTGGAGCTTCTCCT<br>R:TGTGGCTCTGGGTGGTTGT               | 60.9<br>62.6 | 90                   | 55.0%<br>57.9% |
| 14 | 115284 | AATC                 | P              | F:TCGCACCGTTCATATCACAG<br>R:CCTGTCACCACAGTCTCGGT             | 60.6<br>61.2 | 127                  | 50%<br>60%     |
| 15 | 11191  | AGAC                 | P              | F:GACACGGAGACATGTAGAGGC<br>R:TCAGATATCCGTCTGGTCCG            | 59.7<br>61.0 | 222                  | 57.1%<br>55.0% |
| 16 | 11627  | GAGT                 | P              | F:TTGGACAAAGTGTGTGAGTGAGA<br>R:CCATGTTCCACTGTAGCG            | 60.7<br>60.1 | 131                  | 43.5%<br>55.0% |

Tabla 8. Cont.

|    | Contig | Motivo de repetición | Puro/compuesto | Secuencia de iniciadores                                | TM °C        | Tamaño esperado (pb) | GC %           |
|----|--------|----------------------|----------------|---|--------------|----------------------|----------------|
| 17 | 25220  | GAGG                 | P              | F:GAGAGACAGAGAGAGACAAATTGAA<br>R:CTATGGTGGACAGCCGAAGA   | 58.3<br>61.1 | 134                  | 40.0%<br>55.0% |
| 18 | 12892  | CAGA                 | P              | F:CAGAGACAGACACAGAGATGCAG<br>R:ATCCACACTGCCTTGTGTCC     | 60.2<br>60.9 | 128                  | 52.2%<br>55.0% |
| 19 | 121386 | ACAG                 | P              | F:AGTCTGCTGACGGAGACAGG<br>R:CAGACAGACGGTCAATGGT         | 60.6<br>60.9 | 96                   | 60%<br>55%     |
| 20 | 13193  | AATC                 | P              | F:GGCAGGCATTCCCAGTAGAT<br>R:GCGCAGGTTGAAGAGGTATG        | 61.3<br>60.7 | 133                  | 55.0%<br>55.0% |
| 21 | 14743  | CAGG                 | P              | F:GCGAATGATCCAGTCAGTCC<br>R:CTGCAGATGGCACACTTGAG        | 60.6<br>60.6 | 144                  | 55.0%<br>55.0% |
| 22 | 15100  | TGGA                 | P              | F:GTTACGTGAGCTGCGTTTC<br>R:GCCACAAAGGCCATCAAATA         | 60.9<br>60.8 | 288                  | 55.0%<br>45.0% |
| 23 | 15884  | CAAA                 | P              | F:CATGCAGCTGCCATAATGAA<br>R:CCCTTGCCTTATACTCCAATTACA    | 60.7<br>60.5 | 218                  | 45.0%<br>41.7% |
| 24 | 16238  | TAGC                 | P              | F:GGAGGTGCCAATGTCAAAGA<br>R:GCTAGCTGACTACATAATGCCA      | 61.0<br>57.2 | 162                  | 50.0%<br>45.5% |
| 25 | 16925  | AGAC                 | P              | F:ATGGAGTCACTGAGGGTGGGA<br>R:TTCTAACATTACCTGAATTGAAGGAT | 60.5<br>58.6 | 95                   | 55.0%<br>30.8% |
| 26 | 17363  | GACA                 | P              | F:ACATGCAGCTTCCCCTGTT<br>R:TGGGACGGCAGTTTACAAAT         | 60.7<br>60.3 | 166                  | 50.0%<br>45.0% |
| 27 | 22332  | CAGC                 | P              | F:AGTGAAGGCTCTGGAACGAA<br>R:ATAGAGGTGTGGAGTGGGCA        | 59.9<br>60.5 | 105                  | 50.0%<br>55.0% |
| 28 | 22405  | ACAG                 | P              | F:TGTCATCTGTTGAGTGTGCTTT<br>R:TCTAGCAGCAGTACATTGGTTG    | 59.8<br>58.1 | 90                   | 39.1%<br>45.5% |
| 29 | 22602  | CTCT                 | P              | F:TTGGGTGATGGCTTAAGAGC<br>R:CGAGGGAGGATGAATGAGTG        | 60.2<br>60.6 | 285                  | 50.0%<br>55.0% |
| 30 | 25099  | CATC                 | P              | F:CATCTGTGGGCTATTCCTGC<br>R:TGGATATCTCAGGAAGAGGGA       | 60.6<br>58.6 | 95                   | 55.0%<br>47.6% |

Se cuenta también con un conjunto de ocho iniciadores de microsatélites tetranucleótidos los cuales tuvieron anotación genética y que en futuras investigaciones podrán validarse para su uso como marcadores de selección asistida (Tabla 9).

Tabla 9. Iniciadores de microsatélites tetranucleótidos con anotación genética.

|   | Contig | Motivo de repetición | Puro/compuesto | Secuencia de iniciadores                             | TM °C        | Tamaño esperado (pb) | GC %           |
|---|--------|----------------------|----------------|--|--------------|----------------------|----------------|
| 1 | 3643   | GGCG                 | P              | F:GCCTTTAGTGTGTTGCGGTC<br>R:ATGGACCAACTGCCAGGTAA     | 60.9<br>60.3 | 110                  | 55.0%<br>50.0% |
| 2 | 13051  | ACAA                 | P              | F:TTAGATCACTGCGCCACTCA<br>R:TCAACAAAGCATGAGAAGTCTACC | 60.5<br>59.8 | 94                   | 50.0%<br>41.7% |
| 3 | 28825  | TCAT                 | P              | F:GTGAAATTATCCACCAGCCG<br>R:TGGAGCGCCTTCAAATAGTG     | 60.3<br>61.2 | 255                  | 50.0%<br>50.0% |
| 4 | 32506  | ATAA                 | P              | F:TTGTCAGCTGCCACAGAGAG<br>R:GCCCATACCATCAGCAACAT     | 60.3<br>60.7 | 213                  | 55.0%<br>50.0% |
| 5 | 46367  | TAAT                 | C              | F:AGGCATTCTTTCCAACGTGA<br>R:AACGCACTCCCTATCACCAA     | 60.6<br>60.5 | 127                  | 45.0%<br>50.0% |
| 6 | 57399  | TCAA                 | P              | F:CAGCTTTCCTCTCCAGGTT<br>R:TTCGCATTAACGGCAAACCT      | 60.7<br>60.6 | 213                  | 55.0%<br>40.0% |
| 7 | 59308  | GGAT                 | P              | F:TGGGCTTGTTCTCTTGTTT<br>R:TAATCCATCCACCCAACCAC      | 60.2<br>60.4 | 157                  | 50.0%<br>50.0% |
| 8 | 59523  | CTGT                 | P              | F:TCATGGCTTTCCTGGGACTA<br>R:TTTGGATGCAGAGAAGGAGTG    | 60.5<br>60.3 | 131                  | 50.0%<br>47.6% |

Los contigs de donde provienen los microsatélites cuentan con su descripción de acuerdo a la ontología genética y su porcentaje de similitud (Tabla 10).

Tabla 10. Anotación genética de contigs con contenido de microsatélites tetra y pentanucleótidos de la secuenciación masiva de la trucha de SPM.

|   | <b>Contig</b> | <b>Resultado de anotación</b>   | <b>% de similitud</b> |
|---|---------------|---|-----------------------|
| 1 | 3643          | F: zinc ion binding, ligase activity, protein binding. C: cytoplasm   | 88.5                  |
| 2 | 13051         | P: chromatin assembly or disassembly, RNA-dependent DNA replication. C: chromatin, nucleus, F: metal ion binding, helicase activity, chromatin binding, Nucleic acid binding, zinc ion binding, chromatin assembly or disassembly, RNA-directed DNA polymerase activity, ATP binding. | 54.95                 |
| 3 | 28825         | PREDICTED: ATP-dependent RNA helicase DDX54-like isoform 2 [ <i>Oryzias latipes</i> ]   |                       |
| 4 | 32506         | putative nuclease HARBI1  |                       |
| 5 | 46367         | C: integral to plasma membrane. P: neurotransmitter transport, serotonin transport, cation transport. F: serotonin transmembrane transporter activity, neurotransmitter:sodium symporter activity.  | 74.8                  |
| 6 | 57399         | thrombospondin type-1   | 79.67                 |
| 7 | 59308         | P: determination of dorsal identity, otic placode formation, oocyte differentiation, midbrain development F: binding, C: extracellular region.  | 92.75                 |
| 8 | 59523         | P: lysosomal transport, protein oligomerization, acyl-carrier-protein biosynthetic process. C: lysosomal membrane   | 70.63                 |