



UNIVERSIDAD PERUANA
CAYETANO HEREDIA
ESCUELA DE POSTGRADO

**IDENTIFICACION DE MARCADORES
MICROSATELITES PARA EL ESTUDIO DE LA
DIVERSIDAD GENETICA DE *Taenia solium***

**TESIS PARA OPTAR EL GRADO DE
MAGISTER EN BIOQUIMICA Y BIOLOGIA
MOLECULAR**

MARIA LISSETH EGUILUZ MOYA

LIMA – PERÚ

2014

JURADO EXAMINADOR:

Presidente: Dr. Hugo Garcia Lescano

Vocal: Dr. Holger Mayta Malpartida

Secretario: Mg. Jorge Rodriguez Bailon

Asesor: Dr. Mirko Zimic Peralta

DEDICATORIA

A mis padres, por su apoyo incondicional en los momentos más difíciles del desarrollo de esta tesis y por su amor brindado todos los días de mi vida.

A mi hermana, por su aliento constante para terminar la redacción de esta tesis y por ser mi motor en varias etapas de mi vida.

AGRADECIMIENTOS

A mi asesor de tesis, Dr. Mirko Zimic, por todos los conocimientos y la comprensión brindados a lo largo de la realización de mi tesis.

A mi co-asesora, Dra. Mónica Pajuelo por brindarme su confianza desde el primer momento que nos conocimos y ser todo el tiempo tan amable y tolerante conmigo.

A la Dra. Patricia Sheen, por sus alcances brindados para el desarrollo de los experimentos de mi tesis.

A los miembros de mi comité asesor, Dra. Vanessa Adauí y Dra. Manuela Verastegui, por apoyarme y ayudarme desde el inicio hasta la redacción de la tesis.

A los miembros de mi jurado por todos sus alcances durante el proceso de elaboración del proyecto y por su comprensión y diligencia para la revisión del manuscrito.

A todos los chicos del laboratorio de la Unidad de Bioinformática y Biología molecular, por brindarme su amistad y apoyo en la realización de los experimentos.

I. RESUMEN

La diversidad genética en parásitos está orientada hacia el esclarecimiento de la epidemiología y transmisión de las enfermedades. Muchos aspectos de la variación genética de *Taenia solium* se mantienen aún desconocidos. El estudio de la variación genética de este parásito permitiría comprender las diferencias observadas en la infectividad, patogenicidad y respuesta al tratamiento contra la neurocisticercosis. El polimorfismo de los loci microsatélites es un método utilizado ampliamente para el análisis de la estructura genética poblacional. Los marcadores microsatélites se encuentran distribuidos en las regiones codificantes y no codificantes de muchas especies, con una variación suficiente para poder diferenciar distintas cepas. El presente trabajo presenta por primera vez una investigación usando marcadores microsatélites para el análisis de la diversidad genética y estructura poblacional presente en parásitos *T.solium* provenientes de distintas regiones geográficas del Perú.

Las muestras de *T.solium* fueron obtenidas de siete provincias endémicas distribuidas por todo el Perú: Cajamarca, Cusco, Ayacucho, Ancash, Piura, Tumbes y Puno. El ADN de 156 *Taenias* fue evaluado con 30 marcadores microsatélites recientemente identificados y caracterizados, y un conjunto de 7 loci fue seleccionado basado en su polimorfismo para su análisis en toda la población. Nuestros resultados muestran que la mayor fuente de variación se encuentra dentro de las subpoblaciones, siendo la provincia de Puno la que presenta mayor diversidad genética. Además, se observó un alto grado de consanguinidad con una baja heterocigosidad observada para cada subpoblación. Interesantemente, la agrupación de los genotipos mostró una fuerte correlación con la distribución geográfica de las muestras del parásito. Todas las muestras del norte mostraron una baja diferenciación genética entre ellas pero un alto grado de diferenciación genética con respecto a las muestras del sur. Estos resultados refuerzan la idoneidad de los microsatélites para los estudios de genética de poblaciones de *T.solium* y sugieren que podrían haber varios genotipos distintos de *T.solium* circulando en nuestro país.

Palabras clave: variación genética, estructura poblacional, ADN microsatelite, *T.solium*, Perú

ABSTRACT

Genetic diversity in parasites is oriented to the clarification of the epidemiology and transmission of diseases. Many aspects of the genetic variation of *Taenia solium* remain unclear. Understanding the genetic variation of the parasite will let us better know the differences in infectivity, pathogenicity and response to treatment to neurocysticercosis. Analysis of the polymorphism of microsatellite loci is a widely used method for the analysis of population genetic structure. Microsatellite markers have been found broadly distributed in coding and noncoding regions in several species, with enough variation to differentiate strains. This study presents the first microsatellite investigation into the analysis of genetic diversity and population structure among *T.solium* from different geographical regions in Peru.

T. solium isolates were obtained from seven endemic provinces of Peru: Cajamarca, Cusco, Ayacucho, Ancash, Piura, Tumbes and Puno, distributed across the southern and northern coast and highland region. DNA from 156 tapeworms was screened against 30 *T. solium* microsatellites recently identified and characterized, and a set of seven loci were selected to be applied to the whole population based on their polymorphism. Our results show that the most source of variation is found within subpopulations, being Puno the province with the highest genetic diversity. Moreover, a high inbreeding coefficient was observed in most of the loci with low observed heterozygosity values for each population. Interestingly, genotype clustering showed strong correlation with the geographic origin of the parasite samples. All the strains from the North showed low genetic differentiation among them but high levels of differentiation were encountered between the North and South strains. These results reinforce the suitability of microsatellites for population genetic studies of *T. solium* and suggest that there may be different strains of *T.solium* circulating in our country.

Key words: genetic variation, population structure, microsatellite DNA, *Taenia solium*, Perú

INDICE	Página
I. INTRODUCCION	01
II. PLANTEAMIENTO DE LA INVESTIGACION	
2.1. Planteamiento del problema	04
2.2. Marco teórico	
2.2.1. <i>Taenia solium</i>	06
2.2.2. Reproducción en <i>T.solium</i>	08
2.2.3. Marcadores moleculares empleados en el estudio de la diversidad	10
2.2.4. Diversidad genética en <i>T.solium</i>	12
2.2.5. Microsatélites	
2.2.5.1. Base genética	16
2.2.5.2. Ventajas	16
2.2.5.3. Contenido de información polimórfica	16
2.2.5.4. Probabilidad de exclusión	17
2.2.5.5. Probabilidad de identidad	17
2.2.6. Estimación de la diversidad genética	
2.2.6.1. Diversidad alélica	
a) Heterocigosidad	19
b) Índice de Garza-Williamson(G-W)	20
c) Equilibrio de Hardy-Weinberg(HWE)	20
2.2.6.2. Estructura y diferenciación genética	
a) Estadística F	21
b) Análisis de varianza (AMOVA)	23
2.2.6.3. Distancia genética	
a) Método UPGMA	26
b) Neighbor joining	26
2.2.7. Factores que influyen la diversidad genética	
a) Flujo génico y migración	27
b) Mutación	29
c) Deriva génica	30
d) Endogamia	31
e) Selección natural	31
III.JUSTIFICACION	33
IV.OBJETIVOS	

4.1.	OBJETIVO GENERAL	35
4.2.	OBJETIVOS ESPECIFICOS	35
V. MATERIALES Y METODOS		
MATERIAL BIOLÓGICO		35
PROCEDIMIENTO EXPERIMENTAL		
5.1.	Identificación de marcadores a partir del genoma de <i>T.solium</i>	40
5.2.	Diseño de cebadores	42
5.3.	Metodología de los microsatélites	
5.3.1.	Extracción de ADN genómico a partir de proglótidios de <i>T.solium</i>	42
5.3.2.	Cuantificación y dilución de ADN genómico	42
5.3.3.	Amplificación de los microsatélites por PCR	43
5.3.3.1.	Estandarización de las condiciones por PCR	43
5.3.3.2.	Evaluación del polimorfismo de los marcadores microsatélites	45
5.3.3.3.	Electroforesis	45
5.3.3.4.	Genotipificación	46
5.3.4.	Ensayo de reproducibilidad	47
5.3.5.	Elaboración del archivo Input para el análisis de la diversidad	48
5.4.	Análisis de diversidad genética dentro de las subpoblaciones de <i>T.solium</i>	48
5.5.	Análisis de diversidad genética entre las subpoblaciones de <i>T.solium</i>	49
5.6.	Análisis de estructura poblacional	50
5.7.	Análisis de varianza molecular (AMOVA)	50
5.8.	Análisis de asociación-Test de Mantel	50
VI. RESULTADOS		
6.1.	Determinación de la abundancia, distribución y composición de microsatélite en el genoma de <i>Taenia solium</i>	51
6.2.	Diseño de cebadores presentes en el genoma <i>T.solium</i>	53
6.3.	Amplificación de los marcadores microsatélites de <i>T.solium</i>	55
6.4.	Ensayo de reproducibilidad	57
6.5.	Análisis de la diversidad genética dentro de las subpoblaciones	58
6.6.	Análisis de la diversidad entre las subpoblaciones de <i>T.solium</i>	61
6.7.	Análisis de estructura poblacional	71
6.8.	Análisis de varianza molecular (AMOVA)	72
VII. DISCUSION		74
VIII. CONCLUSIONES		87
IX. RECOMENDACIONES		88
X. BIBLIOGRAFIA		89

LISTA DE ABREVIATURAS

RAPDS: DNA Polimórfico Amplificado Aleatoriamente

ITS: Regiones espaciadoras internas de transcripción

rADN: *ADN ribosomal*

SCCP: *Polimorfismo de la conformación de una sola hebra*

RFLP: Polimorfismo en los tamaños de los fragmentos de restricción

AFLP: Polimorfismos de Longitud de Fragmentos Amplificados

ESTs: Secuencias cortas de expresión

LISTA DE TABLAS

Tabla 01. Origen geográfico de las muestras de *T.solium*, número de parásitos genotipificados por cada comunidad y región, así como las subpoblaciones asignadas

Tabla 02. Componentes del master mix del PCR

Tabla 03. Distribución de los microsátélites con respecto al motivo de repetición y tamaño del genoma en *T.solium*

Tabla 04. Frecuencia relativa de los diferentes tipos de motivos en cada clase de microsátélites con repeticiones tipo mono-, di- y tri- nucleótidos analizados en los genomas de *T.solium*.

Tabla 05. Lista de cebadores diseñados a partir de las secuencias microsátélites usando el programa Primer 3.

Tabla 06. Ensayo de reproducibilidad

Tabla 07. Número de alelos, heterocigosidad observada (HO), heterocigosidad esperada (HE), contenido de información polimórfica (PIC), en la población total de Taenias peruanas (N=153)

Tabla 08. Valores FIS, FST, FIT (Weir y Cockerham, 1984) y RhoIS, RhoST , RhoIT (Michalakis y Excoffier, 1996) para cada locus microsátélite analizado.

Tabla 09. Número promedio de alelos, heterocigosidad observada (HO), heterocigosidad esperada (HE) e índice de Garza-Williamson (G-W) en las 13 subpoblaciones de Taenias.

Tabla 10. Valores de FIS y FST (Weir y Cockerham, 1984) calculada para cada una de las 12 subpoblaciones de *T. solium*

Tabla 11. Matriz pareada con los valores de F_{ST} usando el programa Arlequin v3.1

Tabla 12. Matriz pareada con los valores de R_{ST} usando el programa Arlequin v3.

Tabla 13. Matriz pareada del número de migrantes calculado entre las catorce subpoblaciones de *T. solium*

Tabla 14. Matriz pareada de distancias genéticas D_A encima de la diagonal y D_S debajo de la diagonal entre las catorce subpoblaciones de *T. solium*

Tabla 15. Inferencia del valor de K (número de poblaciones) a partir de los 7 loci microsatélites en la población total de Taenias.

Tabla 16. Proporción estimada de contribución de cada grupo genético inferido (K=3) en las 14 subpoblaciones de Taenias.

Tabla 17. Promedio de los resultados del AMOVA para los 7 loci analizados en la población total de *T. solium*

LISTA DE FIGURAS

FIGURA 01. Diez parásitos de *T. solium* recobrados de un sólo paciente después del tratamiento con niclosamida y la purga con electrolitos.

FIGURA 02. Ciclo de vida de *Taenia solium*

FIGURA 03. Sistema reproductor de *Taenia solium*

FIGURA 04. Esquema representativo de los marcadores microsatélites

FIGURA 05. Localización de las comunidades de Tumbes empleadas en este estudio.

FIGURA 06. Localización de las comunidades de Piura empleadas en este estudio.

FIGURA 07. Localización de las comunidades de Puno empleadas en este estudio.

FIGURA 08. Recomendaciones especificadas en el manual del fabricante (Qiagen)

FIGURA 09. Gel de electroforesis y electroferograma con los productos de PCR de los microsatélites TS_SSR09 evaluados en las *Taenias* adultas.

FIGURA 10. Gel de electroforesis con los productos de PCR de los microsatélites TS_SSR03 y TS_SSR05 evaluados en las 12 *Taenias* adultas.

FIGURA 11. Gel de electroforesis y electroferogramas de los productos de PCR de los individuos amplificados con el marcador TS_SSR07.

FIGURA 12. Matriz pareada en base a los valores de Fst calculado entre cada par de subpoblaciones de *T. solium*.

FIGURA 13. Dendograma generado a partir del método de UPGMA calculado en base a la distancia genética estándar (Ds) por el programa Dispan.

FIGURA 14. Dendograma generado a partir del método de UPGMA calculado en base a la distancia genética de Nei (Da) por el programa Dispan.

FIGURA 15. Dendograma generado a partir del método de Neighbord joining (NJ) calculado en base a la distancia genética de Nei (Da) por el programa Dispan.

FIGURA 16. Dendograma generado a partir del método de Neighbord joining (NJ) calculado en base a la distancia genética estandar (Ds) por el programa Dispan.

I. INTRODUCCION

La Cisticercosis porcina causada por el céstode *Taenia solium* es una zoonosis de amplia distribución en países en vías de desarrollo y en otros de Asia, África y Oceanía. Esta enfermedad compromete la economía de pequeños criadores de cerdos en zonas endémicas y genera uno de los problemas más serios de salud pública por ser causante de la neurocisticercosis (Garcia HH, 2003)

Taenia solium es un parásito cosmopolita que incluye en su ciclo de vida al cerdo como el hospedero intermediario obligatorio del estadio larval denominado cisticerco y al hombre como único hospedero definitivo del gusano o estadio adulto. Sin embargo, bajo condiciones de insalubridad, el hombre puede convertirse en hospedero intermediario accidental y de esta manera desarrollar la neurocisticercosis.

Hasta hace poco la variación genética entre poblaciones de *T. solium* era un fenómeno virtualmente desconocido, a pesar que Thompson and Lymbery (1995) sugirieron que el conocimiento de la estructura genética podría ser aplicado a la epidemiología y control de este parásito. Los primeros estudios de diversidad genética fueron realizados empleando marcadores mitocondriales, como COI y Citocromo b, y reportaron la existencia de sólo dos genotipos distintos (Nakao *et al.*, 2002; Hancock *et al.*, 2001).

Los marcadores de ADN son las técnicas moleculares empleadas por excelencia para estudios de estructura genética. Se basan en el análisis de las diferencias, en pequeñas secuencias, del ADN entre individuos. En *T. solium*, hasta el momento, los datos de

genética poblacional, empleando este tipo de marcadores, se han realizado con marcadores RAPDs y regiones ITS del rDNA los cuales han demostrado la existencia de una variación genética en *T. solium* (Campbell *et al.*, 2006; Vega *et al.*, 2003; Bobes *et al.*, 2010) y ello podría influenciar en la presentación de la enfermedad debido a una fertilización cruzada o autofertilización del parásito (Maravilla *et al.*, 2003). Otras investigaciones usando marcadores tipo RAPDs encontraron una relación entre los polimorfismos genéticos de los cisticercos de *T. solium* de diferentes áreas en Brazil y las diferencias en la detección de anticuerpos en pacientes con neurocisticercosis (Da Costa *et al.*, 2012).

En la actualidad, con la disponibilidad del genoma secuenciado de *T. solium* se pueden desarrollar marcadores loci específicos como los microsatélites, los cuales al presentar mayores ventajas como su codominancia y alto grado de polimorfismo, permitirían evidenciar de manera más real la estructura genética de poblaciones de *Taenia* y aplicarlos en estudios de epidemiología molecular orientados hacia el esclarecimiento de la distribución geográfica y del proceso de transmisión de este parásito.

La metodología de este estudio estuvo basada en el diseño de cebadores que amplifiquen regiones microsatélites identificadas a partir del genoma secuenciado de *T. solium* con herramientas bioinformáticas. Por medio de una prueba de PCR convencional se amplificaron los microsatélites con el ADN previamente extraído; y por una electroforesis capilar se probaron las combinaciones de cebadores para escoger a las más polimórficas y luego ser amplificadas en la población total de

parásitos. Se evaluaron la variabilidad y estructura genética de las poblaciones a través de indicadores estadísticos disponibles en los programas informáticos de genética poblacional.

En el presente trabajo se ha logrado desarrollar por primera vez marcadores microsatélites a partir del genoma secuenciado de *T.solium*, evaluarlos y emplearlos en el análisis de la diversidad genética y estructura poblacional presente en poblaciones peruanas de *Taenias*.

II. PLANTEAMIENTO DE LA INVESTIGACION

2.1. PLANTEAMIENTO DEL PROBLEMA

La cisticercosis es una enfermedad causada por *Taenia solium*, parásito que infecta cerdos y humanos durante su ciclo de vida. La manifestación más severa es la infección del sistema nervioso central o neurocisticercosis (NCC), siendo la principal causa de epilepsia adquirida en adultos a nivel mundial (García H 1999; Schenone H 1982; Garcia & Del Brutto 2005; Nicoletti 2002). La Organización Mundial de la Salud (OMS), estima que más de 2,5 millones de personas presentan teniasis y alrededor de 50 millones presentan cisticercosis, con una tasa de mortalidad de 50,000 muertes al año a nivel mundial (Gilman, R.H. 1999).

La cisticercosis es una enfermedad endémica en países no desarrollados (García H, 2005). En nuestro país, la propagación de *T.solium* persiste en comunidades rurales (5% - 20%) donde se crían cerdos al pastoreo y en condiciones sanitarias deficientes (García H, 2003). En las regiones altoandinas, se ha reportado hasta un 7% de prevalencia de teniasis (Garcia HH, 2003). A pesar de la relativamente baja prevalencia de teniasis, la seroprevalencia de cisticercosis en humanos y cerdos es mucho más alta.

Actualmente, el conocimiento que se tiene de la variabilidad genética de *T.solium* es mucho menor en comparación con otros céstodes de la misma familia. Las investigaciones iniciales de diversidad genética han estado basadas en el uso de marcadores mitocondriales, los cuales tienen una utilidad limitada debido al

fenómeno de ligamiento, en el cual los genes mitocondriales son heredados como un sólo grupo y por tanto la inferencia evolutiva es parcial. Además que su análisis implica la subestimación de la diversidad genética (Zhang y Hewitt, 2003). Posteriormente, se emplearon marcadores moleculares nucleares como los RAPDs para el estudio de la variabilidad pero en estadio “cisticerco”. Dichos marcadores son altamente polimórficos y demostraron ser útiles en determinar la presencia de cierto nivel de polimorfismo. Sin embargo, brindan una información limitada de la diversidad genética real debido a su carácter dominante que no permite la detección de heterocigotos y a la falta de reproducibilidad en los ensayos. Las frecuencias alélicas obtenidas con estos marcadores no permiten analizar las diferencias genéticas dentro de una población y entre los individuos. En la actualidad no existe una herramienta de genotipificación de alta resolución con capacidad de diferenciar aislados de *T.solium* provenientes de zonas geográficas cercanas.

Los estudios de diversidad genética en parásitos están orientados hacia el esclarecimiento de la epidemiología y el proceso de transmisión de los mismos. En el caso de *T.solium*, existe poca información disponible acerca de la epidemiología de las infecciones causadas por el parásito adulto debido a que no se cuenta con herramientas moleculares adecuadas que ayuden en el mejor estudio de los patrones de transmisión del parásito en comunidades endémicas y con ello ayudar en el planteamiento de estrategias de control más eficaces para la erradicación de la enfermedad.

2.2.MARCO TEORICO

2.2.1. *Taenia solium*

Taenia solium es un céstode que, conjuntamente con otras 20 especies, pertenecen al género Taenidae (Pawlowski ZS, 2002). *Taenia solium* es también conocida como “solitaria” debido a su supuesta característica de que sólo un parásito se aloja en un hospedero. Sin embargo, el grupo de estudio de neurocisticercosis en el Perú encontró en un estudio previo que el 20% de los portadores de *Taenias* eliminan más de un parásito (Figura 01) (Jeri et al., 2004).



Figura 01. Diez parásitos de *T. solium* recobrados de un solo paciente después del tratamiento con niclosamida y la purga con electrolitos. Fotografía brindada por Javier Bustos.

En el ciclo usual de transmisión (Figura 02), el estadio adulto de *T. solium* habita por periodos prolongados (hasta 25 años) en el intestino delgado del hombre infectado. El parásito se adhiere a la mucosa intestinal por medio del escólex (cabeza) el cual está

seguido de un cuello delgado y un cuerpo denominado estróbilo, formado por varios centenares de segmentos, llamados proglótidos. El hospedero intermediario normalmente es el cerdo quien es infectado por la ingestión de huevos de *T.solium* a través de las heces humanas (Correa *et al.*, 1991). Las enzimas proteolíticas del estómago e intestino destruyen al embrióforo (una de las envolturas que protegen al embrión para sobrevivir en el medio ambiente) liberando y activando al embrión denominado oncósfera. El embrión activado alcanza los capilares sanguíneos y linfáticos que lo llevan al músculo, tejido subcutáneo, sistema nervioso central (SNC) o al ojo (White C, 1997). El embrión ya establecido aumenta de tamaño y en aproximadamente 10 semanas se desarrolla a cisticerco de *T. solium*. Cuando el ser humano ingiere carne de cerdo infectada cruda o mal cocida que contiene los cisticercos viables, el cisticerco es activado por las enzimas y sales biliares del tubo digestivo, y luego de 12 semanas desarrolla el gusano adulto. Cuando el portador es infectado por el gusano adulto, puede liberar proglótidos grávidos a través de las heces completando así el ciclo (Pawlowski ZS, 2002). Cabe resaltar que los proglótidos son aquellos que contienen gran cantidad de huevos en su interior.

El hombre puede actuar como un hospedero intermediario “accidental” cuando ingiere material contaminado con los huevos del parásito, las oncósferas son liberadas, penetran la mucosa intestinal, migran a través del cuerpo y se localizan principalmente en el sistema nervioso central causando la neurocisticercosis.

Figura 02. Ciclo de vida de *Taenia solium*

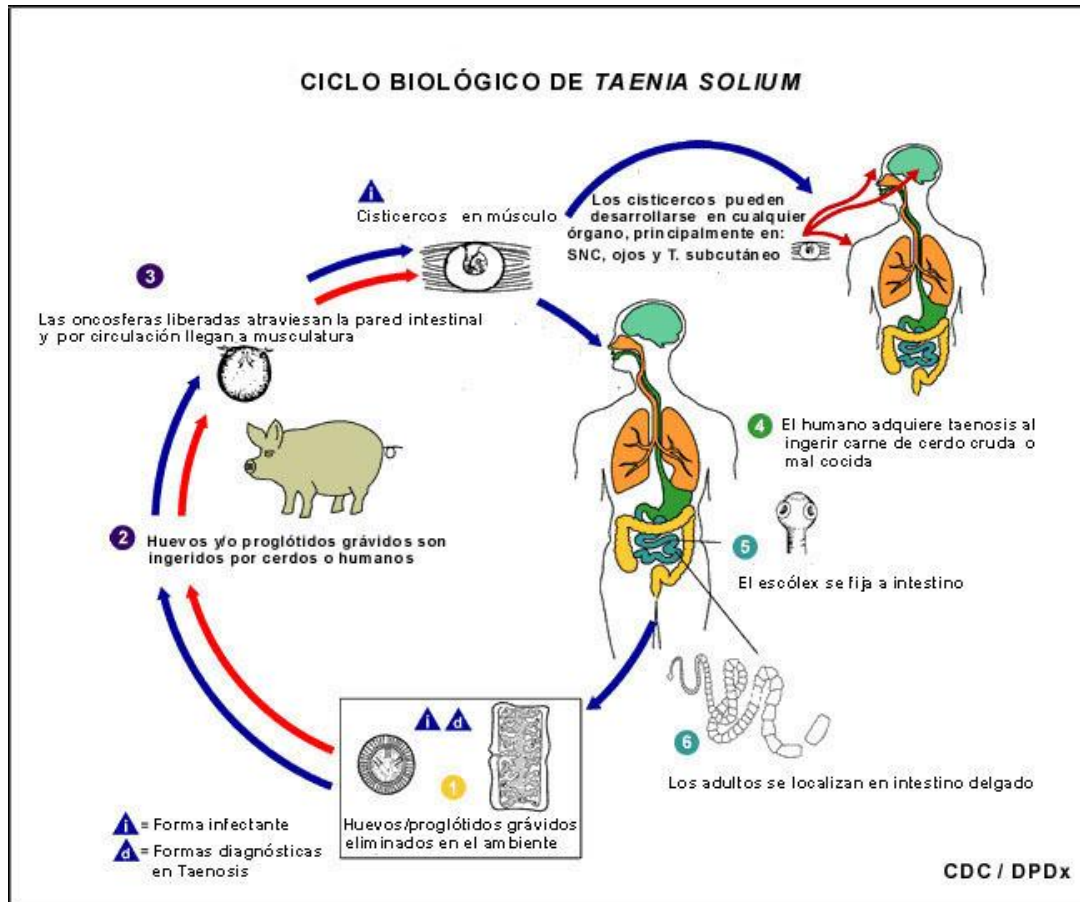


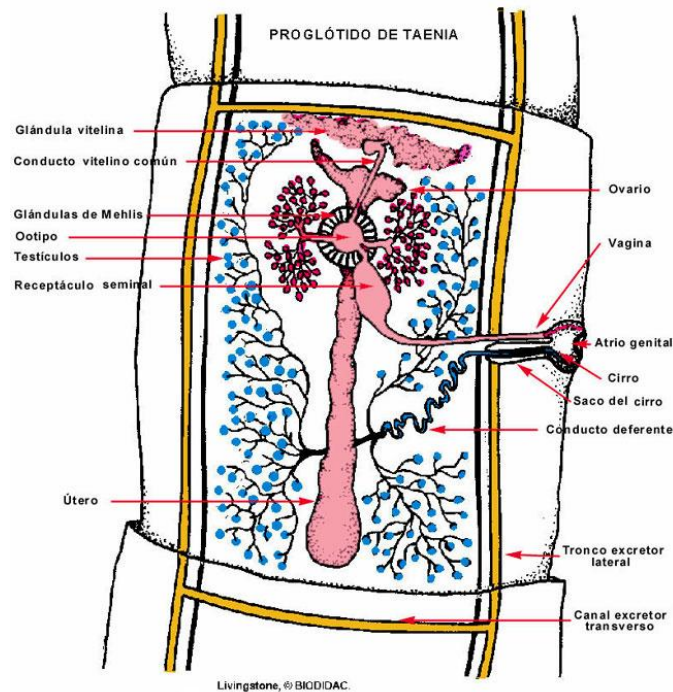
Figura tomada de la página web del CDC (www.dpd.cdc.gov/dpdx)

2.2.2. Reproducción en *T.solium*

Taenia solium es un parásito monoico, es decir posee el aparato reproductor femenino y masculino en el mismo individuo. El aparato genital se desarrolla en los proglótidos inmaduros y sólo en los proglótidos maduros se pueden observar los sistemas reproductivos masculinos y femeninos. El sistema reproductor femenino está compuesto de una vagina, receptáculo seminal, tres ovarios, un oviducto, una

glándula vitelina o vitelarium, glándulas de Mehlis y un útero (Figura 03). El sistema masculino tiene alrededor de 375 a 575 testículos y un conducto espermático.

Figura 03. Sistema reproductor de *Taenia solium*.



Tomado de:

<http://www.facmed.unam.mx/deptos/microbiologia/parasitologia/cestodos.html>

En el proceso de auto-fertilización, el espermatozoide es generado en uno de los 500 testículos en el proglótido (Pawlowski ZS, 2002). Luego, este pasa a través del ducto espermático y se dirige hacia la vagina a través del poro genital o gonoporo. Los huevos sin fertilizar vienen del ovario hacia el oviducto (Sciutto E, 2000). Después de la fertilización, las glándulas vitelinas y de Mehlis secretan sustancias que rodean al cigoto, las cuales formarían el embrión. El cigoto no tiene otra vía de salida por tanto continúa su desarrollo en el útero hasta que el proglótido se desengancha del cuerpo principal y es expulsado con las heces (Sciutto E, 2000). La fertilización

cruzada entre individuos diferentes podría suceder cuando más de una tenia infecta un humano, pero esto ocurre con poca frecuencia (Pawlowski ZS, 2002; Jeri C, 2004; Sciutto E, 2000; Maravilla et al., 2003; Bustos et al., 2012). Además, la fertilización cruzada entre proglótidos es teóricamente posible pero no ha sido reportado para *T.solium* y *Echinococcus granulosus* (Smyth JD y Smyth MM, 1969).

La auto fertilización reduce la variación genética dentro de las poblaciones, pero la variación en regiones no codificantes pueden ocurrir durante la producción de cada oncósfera, y el desarrollo de los clones. Tal como sucede en *Hymenolepis diminuta* en los cuales se ha reportado variación en sus huevos (Pappas y Leiby 1986).

Maravilla et al, 2003 fueron capaces de detectar linajes locales en diferentes estados de México y sugirieron que *T.solium* experimenta reproducción clonal. El mismo estudio sugirió además la existencia de recombinación genética, la cual se debería a una “deriva genética generada por una reproducción sexual o fertilización cruzada”. Posteriormente, se reportó mayor variabilidad genética, sugiriendo que podría ocurrir recombinación genética (Maravilla et al, 2008; Bobes et al., 2010).

En el caso de *E.granulosus*, un céstode cercanamente relacionado, se encontró que tiene una combinación de auto fertilización y reproducción cruzada tal como se demostró en estudios empleando genes nucleares en Brasil (Haag et al., 2011). *Oochoristica javaensis*, un céstode que infecta lagartos como hospederos definitivos, es hermafrodita, experimenta reproducción sexual y tiene un alto porcentaje de endogamia (Detwiler y Criscione 2011)

2.2.3. Marcadores moleculares empleados en el estudio de la diversidad genética

Los parásitos representan una parte significativa de la biodiversidad (De Meeu's y Renaud, 2002) sin embargo, la cantidad de información disponible de estos organismos es insuficiente. (Criscione et al., 2005). Debido a su tamaño pequeño, localización, biología y conducta, no es posible realizar observaciones directas, es por ello que su ecología, modos de reproducción y tamaño poblacional pueden ser analizados a través de "métodos indirectos" (Nadler, 1995; De Meeus et al., 2002 a, b).

Los métodos indirectos son aquellos que emplean a los marcadores moleculares polimórficos y a la variación de dicho polimorfismo dentro y entre sub-conjuntos de poblaciones (referidas como sub poblaciones) (De Meeus, 2007).

Un marcador genético es una porción de ácido nucleico (ADN) o el producto de una porción de ácido nucleico (proteína) del organismo en estudio. Para realizar inferencias biológicas, se estudia la variación de la misma porción de ADN en individuos procedentes de diferentes regiones geográficas. Es importante que dicha porción de ADN se encuentre en el mismo lugar en el genoma de cada individuo (es decir, en el mismo lugar en el mismo cromosoma), por tanto se usa el término locus. Para que un locus sea informativo debe ser variable (es decir, polimórfico). Esto significa que la secuencia de ADN debe ser diferente entre los individuos y dicha diferencia en las secuencias son denominados alelos.

Existen diferentes tipos de marcadores (Barker 2002; Singh 1997; Campbell et al., 2006). Entre los más usados se encuentran los genes mitocondriales, las regiones espaciadoras internas de transcripción de ácidos ribonucleicos ribosomales (rRNA), repeticiones en tándem, los cuales incluyen a los minisatélites y microsátélites, y las mutaciones puntuales de secuencia (SNP).

Genes mitocondriales

La diversidad genética se basa en las diferencias encontradas en los genes mitocondriales específicos siendo la más empleada la citocromo oxidasa (COX) (Hancock *et al.*, 2001). Los genes mitocondriales evolucionan más rápido que los genes nucleares, esta característica los hace más variables y adecuados para diferenciar los organismos cercanamente relacionados (Campbell *et al.*, 2006). Además que el ADN mitocondrial es circular y más resistente a la degradación.

Regiones espaciadoras internas de transcripción (ITS) de ácidos ribonucleicos ribosomales (rRNA)

Son regiones variables distribuidas al azar en el genoma que usualmente son analizadas con la técnica de RFLP, pero además con RAPDs y SCCP. El ITS está compuesto de ITS1 y ITS2, separados por el gen 5.8S que contiene secuencias repetidas en tándem. Dicha característica permite utilizarlas para la diferenciación entre diferentes individuos.

2.2.4. Diversidad genética en *T.solium*

Hasta el momento, la variación genética de *T.solium* no ha sido descrita del todo, a pesar que podría ser un factor importante que explique la heterogeneidad clínica de la

enfermedad. Las presentaciones clínicas de la cisticercosis varían entre diferentes países (García HH 2003), siendo la cisticercosis subcutánea la más común en Asia y África, pero poco común en América.

La neurocisticercosis tiene también múltiples variantes. En occidente, los pacientes presentan pocos quistes en el cerebro mientras que en partes de Asia, la mayoría de pacientes tienen lesiones causadas por un solo quiste y otros por cientos de quistes (García et al., 2003). Las diferentes manifestaciones clínicas así como los distintos estadios del cisticerco encontrados en el paciente han sido atribuidas a la respuesta inmune del paciente (Sotelo y Del Brutto, 2000). Por otra parte, se han reportado diferencias en los inmunoblots entre muestras provenientes de África y América debido principalmente a distintos patrones de glicosilación (Ito et al., 2003). En un trabajo previo, Yakoleff-Greenhouse (1982) mostró variación antigénica en los extractos de cisticercos de cerdos provenientes de México, sugiriendo que no sólo es el sistema inmune del paciente la causa de la variación sino también la diversidad de los antígenos. El entendimiento de la variabilidad genética podría ser la clave para superar las dificultades encontradas en el diseño de las pruebas de diagnóstico y vacunas.

Los primeros trabajos en genotipificación de *T.solium* estuvieron dirigidos hacia la diferenciación entre distintas especies. Es así como, Gasser pudo diferenciar entre 8 especies distintas de *Taenia* en base al polimorfismo de secuencias mitocondriales como la NADH deshidrogenasa I y la subunidad I de la citocromo C oxidasa. Dicho

análisis fue realizado usando la técnica del polimorfismo conformacional de una sola cadena (SSCP) (Gasser et al., 1999).

Posteriormente, Hancock en el 2001 demostró que existían diferencias entre los aislados de *T. solium* provenientes de diferentes países. Ellos evidenciaron poca variación dentro de los mismos individuos usando como marcadores a la citocromo oxidasa I (COI), al espaciador transcrito interno 1 (ITS1) y al marcador antigénico de diagnóstico Ts14. Además, encontraron variación en las secuencias de COI y Ts14 de los parásitos provenientes de Sudamérica y Asia pero no reportaron variabilidad genética dentro de los 15 individuos peruanos con los que trabajaron.

Hasta el momento, los análisis de genética poblacional se han realizado con marcadores RAPDs y regiones ITS del rDNA los cuales han probado que existe una variación genética en *T. solium* (Campbell et al., 2006). El trabajo desarrollado por Maravilla et al. (2003) encontró variación en las regiones ITS1 trabajando con cisticercos provenientes de distintas localidades de México, Honduras y Tanzania. Dicha investigación sugirió que *T. solium* tiene linajes locales o líneas descendientes con diferentes características y ello tendría un rol en la variación de la presentación de la enfermedad debido a una fertilización cruzada o autofertilización del parásito. Adicionalmente, Vega et al. (2003) encontraron un significativo grado de diferenciación genética en el perfil de amplificación de los cisticercos que infectan a cerdos de Madagascar con respecto a los de dos regiones de México. Además encontraron diferenciación genética entre las dos regiones de México. Bobes et al. (2010) evaluaron la diversidad genética en cisticercos aislados a partir de cerdos

infectados usando marcadores RAPDs. Ellos encontraron diferencias genéticas significativas entre individuos de distintas áreas geográficas y además que la diversidad existente entre cisticercos del mismo individuo era menor que la observada entre diferentes individuos, con ello postularon que los cerdos eran raramente infectados por más de una tenia distinta, por lo que la primera infección causaría una inmunidad concomitante hacia otros individuos (Gemmell et al., 1990). Además, demostraron una distribución no al azar de la cisticercosis porcina.

Da Costa et al. (2012) usando marcadores tipo RAPDs encontraron una relación entre los polimorfismos genéticos presente en los cisticercos de *T. solium* de diferentes áreas en Brazil y las diferencias en la detección de anticuerpos en pacientes con neurocisticercosis.

2.2.5. Microsatélites

Los microsatélites son repeticiones en tándem de 1-6 nucleótidos encontrados con alta frecuencia en los genomas nucleares de muchos taxa (Figura 04). Se conocen también como repeticiones de secuencia simple (SSR), repeticiones en tándem en número variable (VNTR) o repeticiones pequeñas en tándem (STR). (Li et al. 2002; Ellegren 2004).

Un locus microsatélite típicamente varía en longitud entre 5 y 40 repeticiones, pero se pueden encontrar cadenas más largas de repeticiones. Las repeticiones de tipo dinucleótidos, trinucleótidos y tetranucleótidos (de dos, tres y cuatro pares de bases de longitud en el motivo de repetición, respectivamente) son los más escogidos para

los estudios de genética molecular, entre ellas las repeticiones de dinucleótidos son las más empleadas en muchas especies por ser de más fácil análisis (Li et al., 2002). Las repeticiones de tipo mononucleótidos son los menos confiables por los problemas en la amplificación y los de tipos de repetición más largas son menos comunes.

Figura 04. Esquema representativo de los marcadores microsatélites

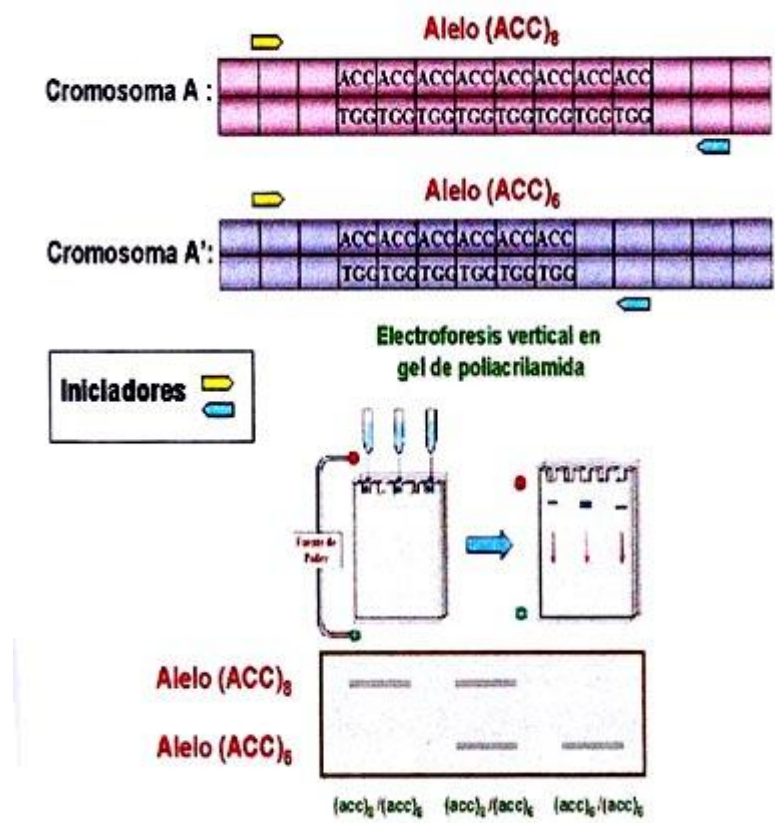


Figura tomada de www.gnis-pedagogie.org

El ADN que rodea un locus microsatélite es denominado región flanqueante. Dichas regiones son muy conservadas entre los individuos de una misma especie y a veces entre diferentes especies, es por ello que un locus microsatélite puede ser identificado por sus regiones flanqueantes. Los oligonucleótidos o cebadores son diseñados a

partir de esta región y guían la amplificación de un locus microsatélite con la reacción en cadena de la polimerasa (PCR).

2.2.5.1.Base genética:

El polimorfismo de los microsatélites resulta de la alta tasa de mutación por generación, más de 10^{-3} mutaciones por locus, en algunos eucariotas (Ellegren 2004). Estas mutaciones son inserciones o deleciones de una o más repeticiones y se cree que resultan esencialmente de un proceso molecular referido como “resbalón de la ADN polimerasa durante la replicación del ADN” o “*DNA replication slippage*”. Este proceso se define como la disociación transitoria de las hebras de ADN durante la replicación, seguido de una re asociación incorrecta. Cuando la cadena naciente se realinea fuera del patrón inicial, la nueva replicación conducirá a la inserción o deleción de una unidad de repetición en relación con el ADN molde inicial (Ellegren 2004).

2.2.5.2.Ventajas:

Los microsatélites han tomado ventaja sobre otros marcadores genéticos como los AFLPs, RAPDs, RFLPs, puesto que: i) tienen el más alto grado de polimorfismo; ii) segregan de manera mendeliana y son codominantes; iii) presentan un solo locus genético por microsatélite y esto hace que la lectura de las bandas sea clara y fácil de interpretar, iv) son selectivamente neutros (Golstein y Pollock, 1994; Vendramin et al., 1996) y v) tienen un modelo de mutación homogénea: en la mayoría de los SSR

las mutaciones son de un paso (una unidad repetitiva) siguiendo el modelo mutacional SSM de Ohta y Kimura (1973).

2.2.5.3. Contenido de información polimórfica (PIC):

Este estadístico permite cuantificar la informatividad y el polimorfismo de un locus génico. Para calcular el PIC de n-alelos de un locus, se multiplica la frecuencia de cada tipo de apareamiento por las fracciones esperadas de la progenie informativa de ese tipo de apareamiento y se añaden los productos (Sanyai et al., 2000):

$$PIC = 1 - \sum_{i=1}^n p_i^2 - 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n p_i^2 p_j^2 = 1 - \sum_{i=1}^n p_i^2 - \left(\sum_{i=1}^n p_i^2 \right)^2 + \sum_{i=1}^n p_i^4$$

Donde pi es la frecuencia poblacional del i-esimo alelo

El PIC se define como la probabilidad de que el genotipo del marcador de un descendiente dado permita deducir al parental afectado del que proviene.

2.2.5.4. Probabilidad de exclusión:

Se define como la probabilidad de excluir a un individuo cualquiera de la población dado los alelos del hijo y de la madre. La probabilidad de exclusión depende del número de alelos y de las frecuencias alélicas.

Existen tres tipos de probabilidad de exclusión definidas por Jamieson A y Taylor SS (1997):

- a) PE1: Probabilidad de excluir a un pariente dado el genotipo de los dos parentales y el descendiente. La fórmula es:

$$P = 1 - 2 \sum_{i=1}^n p_i^2 + \sum_{i=1}^n p_i^3 + 2 \sum_{i=1}^n p_i^4 - 3 \sum_{i=1}^n p_i^5 - 2 (\sum_{i=1}^n p_i^2)^2 + 3 \sum_{i=1}^n p_i^2 \sum_{i=1}^n p_i^3$$

Dónde: p_i es la frecuencia del alelo i y n es el número de alelos en un locus

- b) PE2: Probabilidad de excluir a ambos parentales dado el genotipo de los dos parentales y el descendiente:

$$P = 1 + 4 \sum_{i=1}^n p_i^4 - 4 \sum_{i=1}^n p_i^5 - 3 \sum_{i=1}^n p_i^6 - 8 (\sum_{i=1}^n p_i^2)^2 + 8 (\sum_{i=1}^n p_i^2) (\sum_{i=1}^n p_i^3) + 2 (\sum_{i=1}^n p_i^3)^2$$

Dónde: p_i es la frecuencia del alelo i y n es el número de alelos en un locus

- c) PE3: Probabilidad de excluir la relación filial dado el genotipo de un parental y el descendiente:

$$P = 1 - 4 \sum_{i=1}^n p_i^2 + 2 (\sum_{i=1}^n p_i^2)^2 + 4 \sum_{i=1}^n p_i^3 - 3 \sum_{i=1}^n p_i^4$$

Dónde: p_i es la frecuencia del alelo i y n es el número de alelos en un locus

2.2.5.5. Probabilidad de identidad (Probabilidad de identidad (PID))

Este estimador refleja la probabilidad de que dos individuos compartan el mismo perfil genético. Se basa en la búsqueda de genotipos idénticos asumiendo la asociación de alelos dentro y entre los loci al azar (Mills et al. 2000). Se calcula aplicando la siguiente fórmula:

$$P_{(ID)} = \sum p_i^4 + \sum \sum (2 p_i p_j)^2$$

Dónde: p_i es la frecuencia del alelo I , p_j es la frecuencia del alelo J e $i \neq j$

2.2.6. Estimación de la diversidad genética:

La diversidad genética dentro de una especie puede caracterizarse en tres niveles de complejidad: a) la diversidad genética representada por la cantidad de alelos en un grupo de loci dado; b) la diferenciación genética, que corresponde a la distribución que presentan dichos alelos entre poblaciones; c) la distancia genética, que compara por pares a dichas poblaciones y estima la cantidad de divergencia entre ellas (Lowe et al., 2004).

2.2.6.1. Diversidad Alélica

La frecuencia de un alelo en particular en una población es denominada frecuencia alélica. Este parámetro es fundamental porque cambios genéticos en una población generalmente se expresan como cambios en sus frecuencias alélicas (Frankham et al., 2002). Es por ello que se emplea como un parámetro para caracterizar la diversidad genética o la riqueza de un pool genético en una población. Cuando el alelo más común de un locus presenta una frecuencia menor al 99% en la población bajo estudio (Hartl y Clark, 1997) es denominado locus polimórfico.

a) Heterocigosidad:

La heterocigosidad mide la proporción de individuos heterocigotos existentes para un determinado marcador, o, de forma extendida, su promedio para un conjunto de marcadores. Recordemos que cada célula diploide tiene dos alelos, uno heredado de cada parental. Si un individuo tiene dos alelos en un locus específico, el individuo es

heterocigoto en ese locus; si los dos alelos son iguales, el individuo se denomina homocigote.

La heterocigosis es una buena medida de la diversidad contenida en una población, ya que está íntimamente relacionada con la consanguinidad y la viabilidad de la población.

La heterocigosis se puede medir de dos formas distintas. Una de ellas consiste en calcular directamente la proporción de individuos heterocigotos contando uno a uno los genotipos heterocigotos, o acumulando las frecuencias génicas de éstos. Es lo que se conoce como heterocigosis observada.

La heterocigosis esperada, o diversidad génica (h ; Nei, 1973), es la heterocigosis que se esperaría en condiciones de equilibrio Hardy-Weinberg. Se define como la probabilidad que dos alelos escogidos al azar en una población sean diferentes (Weir, 1996).

$$H_E = 1 - \sum_{i=1}^n p_i^2$$

Dónde: p_i es la frecuencia del alelo i y n es el número de alelos en un locus (Nei et al., 1973)

Las diferencias entre la heterocigosis observada y la esperada pueden ser consecuencia de fenómenos poblacionales como la subdivisión y los cuellos de botella.

b) Índice de Garza-Williamson (G-W)

Parámetro estadístico desarrollado para determinar si una población ha experimentado una reducción en su tamaño efectivo, denominado cuello de botella.

Se define como

$$G-W = k/R + 1$$

Donde k es el número de alelos para un locus dado en una muestra de la población, y R es el rango de alelos o la diferencia entre el máximo y mínimo número de repeticiones de un motivo del microsatélite (Excoffier et al. 2005).

Cuando sucede un cuello de botella el número de alelos disminuye en relación al rango de alelos debido a una reducción en el tamaño de la población. Por tanto, el estadístico de $G-W$ se supone que es muy pequeño en poblaciones que hayan experimentado un cuello de botella y cercano a 1 en poblaciones estacionarias (Garza y Williamson, 2001).

c) Equilibrio de Hardy-Weinberg (HWE)

Este principio es usado para comparar las frecuencias alélicas en una población dada a través de un periodo de tiempo. Una población debe cumplir cinco reglas para ser considerada en equilibrio:

- 1) Ausencia de mutaciones génicas y por tanto los cambios alélicos no ocurren.
- 2) Mínima tasa de migración de los individuos
- 3) Apareamientos al azar.

- 4) Ausencia de deriva génica, una oportunidad de cambio en las frecuencias alélicas.
- 5) Ausencia de selección natural
- 6) Población suficientemente grande

El equilibrio de Hardy-Weinberg nunca ocurre en la naturaleza debido a que ninguna población llega a cumplir todas las reglas pero es un estado ideal que provee una línea de base contra la cual los científicos miden la evolución de los genes en una población dada.

Un muestreo inapropiado, tamaños de muestra pequeños (menores de 20 individuos) así como baja variabilidad de los loci caracterizados son factores comunes que originan una desviación del HWE (Freeland, 2005).

2.2.6.2. Estructura y diferenciación genética

La estructura poblacional constituye la cantidad y distribución de la variabilidad genética dentro y entre subpoblaciones locales e individuos dentro de una misma especie (Templeton, 2006).

Para determinar la estructura genética, es necesario comprender el patrón de variación genética de la especie, es decir, evaluar los genotipos de los diferentes individuos (Slatkin, 1995).

a) Estadística F

Los estadísticos F, propuestos por Wright (1965), permiten evaluar el efecto que tiene la subdivisión en las poblaciones. Los coeficientes se encuentran relacionados directamente con los tres niveles que presenta una población subdividida, es decir, distribuyendo la variación genética a nivel poblacional (T), de subpoblaciones (S) y a nivel individual (I) (Hedrick, 2000) conduciendo a los tradicionales estadísticos F: F_{IS} , F_{IT} y F_{ST} (Balloux y Lugon, 2002).

El F_{IS} , también llamado coeficiente de endogamia, mide la reducción en la heterocigosis de un individuo debido al apareamiento no azaroso dentro de su subpoblación. Este indicador refleja el nivel de consanguinidad. Se calcula de la siguiente manera

$$F_{IS} = (H_s - H_o) / H_s$$

Dónde: H_o es la heterocigosis observada dentro de una población y H_s es la heterocigosis esperada dentro de las subpoblaciones.

Los valores de F_{IS} pueden variar entre -1 (todos los individuos son heterocigotos con el mismo par de alelos), 0 (distribución azarosa de los alelos entre los individuos) hasta +1 (todos los individuos son homocigotos)

El F_{IT} es denominado como coeficiente de endogamia total de un individuo y mide la reducción en la heterocigosis de ese individuo con respecto a la población total (Hartl y Clark, 1989); si los valores son positivos indican una deficiencia de heterocigotos y

si son negativos un exceso de heterocigotos (Hedrick, 2000). Se calcula de la siguiente manera:

$$F_{IT} = (H_T - H_O) / H_T$$

Dónde: H_O es la heterocigosidad observada dentro de una población y H_T es la heterocigosidad esperada total.

El F_{ST} , conocido como índice de fijación subpoblacional, es la correlación entre dos alelos escogidos al azar dentro de las subpoblaciones con relación a los alelos muestreados al azar del total de la población (Wright 1951, 1965). Por tanto, el F_{ST} mide la diferenciación genética entre las subpoblaciones. Los valores de F_{ST} van desde cero cuando no hay diferenciación entre las subpoblaciones, es decir se encuentran en equilibrio de Hardy-Weinberg, hasta uno, cuando se han fijado los diferentes alelos en las subpoblaciones (Wright, 1969; Hartl y Clark, 1997). Se calcula de la siguiente manera:

$$F_{ST} = (H_T - H_S) / H_T$$

Dónde: H_S es la heterocigosidad esperada dentro de una población y H_T es la heterocigosidad esperada total.

Además, el F_{st} representa una medida del efecto de Walhund (Walhund 1928), esto es, una deficiencia de heterocigotos debido a una subdivisión poblacional no detectada. Por lo tanto, el F_{st} mide el déficit de heterocigotos en relación con lo esperado bajo equilibrio de Hardy-Weinberg (Hartl & Clark 1997).

La G_{ST} propuesta por Nei (1973), se aplica para múltiples alelos y se define como la proporción de diversidad genética que se encuentra entre poblaciones y se considera análogo al F_{ST} en el contexto de loci multialélicos.

En el caso de los marcadores microsatélites se ha propuesto un indicador específico para cuantificar la diferenciación genética, el R_{st} , debido que se asume un modelo de mutación “stepwise” (Slatkin 1995).

Otro parámetro que se ha utilizado para medir la diferenciación genética entre poblaciones es el llamado coeficiente de coancestría (Θ), propuesto por Cockerham (1969), el cual representa la probabilidad de que dos genes tomados al azar provenientes de diferentes individuos sean idénticos por descendencia y es equivalente al F_{ST} ; la coancestría se incrementa con el tiempo, de tal manera que indica que tan diferenciadas están las poblaciones (Weir, 1996).

b) Análisis de la Variación Molecular (AMOVA)

La estadística F de Wright está basada en la comparación de las frecuencias génicas entre los *demes* o subpoblaciones, sin embargo, la base de datos molecular revela no sólo la frecuencia de los marcadores moleculares, sino que también puede brindar información acerca de la cantidad de diferencias mutacionales entre los diferentes genes.

El AMOVA se basa en el análisis directo de las diferencias entre las secuencias moleculares para estimar la diferenciación poblacional, en vez de asumir frecuencias

genéticas siguiendo el modelo de Hardy-Weinberg (tal como lo hace la estadística F) y se recomienda su cálculo para poblaciones que no están en equilibrio de HW. Esta técnica permite analizar una gran variedad de base de datos moleculares usando marcadores moleculares como RFLP, AFLP, SSR; secuencias directas o árboles filogenéticos (Excoffier, et al. 1992). Además, proporciona muchas facilidades para analizar la estructura jerárquica de la población, donde los individuos son agrupados en subpoblaciones y estos en grupos (*clusters*) de mayor jerarquía. Sin embargo, el uso de AMOVA en este sentido requiere un conocimiento a priori de dicha jerarquía (Meirmans, 2012).

2.2.6.3. Distancia Genética:

Los índices de distancia genética se calculan entre pares de poblaciones y tiene como objetivo describir el grado de diferencia genética entre estas dos poblaciones. Es así como se puede calcular directamente la F_{st} entre dos poblaciones y usarse como distancia genética, o utilizar la estimación de Nm (Número efectivo de migrantes).

Entre las medidas de distancia genética más usadas tenemos a la distancia genética estándar de Nei (1973, 1977) y la Distancia de Nei (1983), D_A .

La distancia genética estándar estima el número de mutaciones que a nivel nucleotídico se han acumulado en las secuencias de dos linajes a partir del momento de su divergencia original. A partir de datos de frecuencias alélicas es posible calcular estadísticamente el número promedio de las diferencias en las bases nucleotídicas por

locus. Dado que este número es una medida directa de las diferencias genéticas entre dos poblaciones, se le considera como una medida de distancia genética (Nei 1973).

Dada las poblaciones X e Y con r loci y M alelos por locus, la distancia genética es definida como:

$$D_S = -\ln\left(J_{XY} / \sqrt{J_{XX}J_{YY}} \right)$$

donde $J_{XY} = \sum_{i=1}^m \sum_{j=1}^r x_{ij}y_{ij}/r, J_{XX} = \sum_{i=1}^m \sum_{j=1}^r x_{ij}^2/r, J_{YY} = \sum_{i=1}^m \sum_{j=1}^r y_{ij}^2/r,$

X_{ij} es la frecuencia del alelo -i en el -j locus en la población X, y Y_{ij} es la frecuencia del alelo i- en el j- locus en la población Y. El valor paramétrico de D_s entre dos poblaciones que han sido separadas hace t generaciones en el pasado es aproximadamente:

$$D_s \approx 2\mu t$$

Donde μ es la tasa de mutación de alelos infinitos en el locus de estudio. Esta expresión asume que la fragmentación de la población ancestral fue instantánea y completa, y que cada población ha tenido un tamaño efectivo igual al tamaño efectivo de la población ancestral.

Nótese que D_s se incrementa de manera lineal con el tiempo desde cero hasta el infinito y tendrá un valor proporcional a la tasa de mutación. Por ejemplo si dos poblaciones tienen similares frecuencias alélicas, la similitud genética (I) podría ser 1 por lo tanto, la distancia genética (D_s) sería 0 (Freeland, 2005).

La Distancia de Nei (Nei et al., 1983) es una modificación de la distancia geométrica evaluada por Cavalli-Sforza (1967), permite obtener un estimador no sesgado de la distancia genética y su error estándar a partir de frecuencias alélicas obtenidas de muestras pequeñas (Nei y col., 1983).

Su máximo valor es 1.0 lo que significa que las dos poblaciones no comparten ningún alelo en ningún loci (Kalinowski, 2002).

$$D_A = 1 - \frac{\sum_{i=1}^m \sum_{j=1}^r \sqrt{x_{ij}y_{ij}}}{r}$$

Se ha demostrado que ambas medidas de distancia son útiles para la reconstrucción de las filogenias reflejadas en los dendrogramas (Takezaki & Nei 1996).

a) Método de agrupamiento por pares no ponderados con media aritmética (UPGMA)

Es un algoritmo de agrupamiento que une ramas en un árbol basándose en el criterio de maximizar la similitud entre pares y entre promedios de pares ya unidos. En sentido estricto, no es un método de distancia evolutiva (Li, 1997). UPGMA puede generar una topología acertada con las longitudes de ramas verdaderas sólo si la divergencia entre secuencias cumple con el modelo de reloj molecular (Swofford et al., 1996), condición que raramente se cumple en la práctica.

b) Neighbor joining

Se aplica habitualmente cuando se construyen árboles de distancia, independientemente del criterio de optimización. El árbol completamente resuelto se ‘descompone’ a partir de un árbol ‘estrella’ (completamente no resuelto) por el procedimiento de insertar ramas entre el par de vecinos más cercanos (o más aislados de los demás) y el resto de los nodos terminales del árbol.

El par de vecinos más cercanos queda entonces consolidado dando por resultado un nuevo árbol ‘estrella’ y el proceso se repite hasta obtener un árbol resuelto. Es un método comparativamente rápido (Brinkman y Leipe, 2001) que, al igual que los métodos de evolución mínima, descubre la filogenia correcta si las secuencias analizadas cumplen con el supuesto de homogeneidad de tasas de divergencia.

2.2.7. Factores que influyen la diversidad genética:

La estructura genética se ve afectada por varios factores que establecen diferencias en las frecuencias alélicas y genotípicas de una región a otra (Excoffier, 2001), entre los más comunes se encuentran los procesos genéticos como: el flujo génico y migración, la mutación, deriva genética, el modo de reproducción (endogamia) y la selección natural.

a) Flujo génico y migración

El flujo génico es el movimiento de alelos entre las poblaciones, comprende la inmigración y la emigración de alelos (Lowe et al., 2004). Este proceso genera

cambios en las frecuencias alélicas, cuya magnitud varía de acuerdo a la proporción de alelos contribuidos por los migrantes, y a la diferencia entre la frecuencia de los inmigrantes y la población original (Frankham et al., 2002). Cuando los niveles de flujo génico entre las subpoblaciones es alto, la variación genética de las subpoblaciones tenderá a homogenizarse. Por el contrario, si los niveles de flujo génico son bajos, la selección natural, la deriva génica y la mutación pueden originar una diferenciación genética mayor entre las subpoblaciones (Nei, 1987).

La cuantificación indirecta del flujo génico se basa principalmente en observar la distribución espacial de los alelos en las poblaciones para hacer inferencias de los niveles o patrones del mismo (Slatkin, 1985a).

La mayoría de los modelos teóricos surgen de los conceptos de Wright y suponen que los organismos 1) están formando poblaciones discretas –modelo de islas- que se diferencian por mutación y por deriva genética (Wright, 1943), ó 2) están formando poblaciones con una distribución continua – modelo de aislamiento por distancia- en el cual la probabilidad de flujo génico disminuye al incrementarse la distancia espacial.

El modelo usado comúnmente para estimar flujo génico es el modelo de islas infinitas (*infinite islands model*) de Wright (1951). Este modelo considera condiciones en equilibrio entre un número infinito de islas o subpoblaciones de igual tamaño, que intercambian migrantes entre cualquiera de las islas con igual probabilidad a una tasa constante. Las poblaciones pueden ser tratadas como replicas y el modelo se puede

caracterizar con sólo dos parámetros: tamaño poblacional (N) y tasa de migración (m). (Slatkin, 1985a). En el modelo de stepping-stone, introducido por Kimura (1953), los individuos de cada una de las subpoblaciones sólo pueden moverse entre poblaciones adyacentes, pero los resultados del equilibrio, aunque análogos a los del modelo de Wright, son más complicados y menos generales.

El otro modelo es el de aislamiento por distancia, que examina la relación del flujo génico entre pares de poblaciones (M , equivalente al número efectivo de migrantes (Nm) entre pares de poblaciones) con la distancia geográfica (Slatkin, 1995). Bajo este modelo, los valores de M se estiman a partir de valores de θ entre pares de poblaciones (medida de diferenciación genética entre poblaciones, Weir y Cockerham 1984) y se aplica una regresión de $\log_{10}M$ en $\log_{10}k$ (k =distancia geográfica).

Para determinar si existe una relación lineal significativa entre el flujo génico y la distancia geográfica entre poblaciones se utiliza comúnmente una prueba de Mantel (1967).

Otro método para estimar Nm (multiplicación del tamaño efectivo (N_e) por la tasa de migración ' m ') es el de alelos privados (Slatkin 1981, 1985b), en el cual calcula la frecuencia promedio de alelos que está en una sola muestra en base a otras de distintas localidades. Si Nm es mayor que 1, teóricamente el flujo génico supera los efectos de la deriva génica y previene la diferenciación local. Si Nm es menor que 1 entonces se puede decir que la deriva actúa independientemente en cada una de las

poblaciones, y si es mayor de 4 entonces las poblaciones se comportan como una gran población más o menos panmíctica (Whitlock y McCauley 1999).

b) Mutación

La mutación es cualquier cambio en la secuencia de bases de un genoma, dando origen de esta manera a la diversidad genética. Se puede manifestar en distintos niveles, cuando incluyen cambios en los alelos de un locus o en el número de repeticiones (microsatélites). La mayoría de estos cambios ocurre por errores de la ADN polimerasa durante la replicación del ADN, principalmente la adición, delección o sustitución de las bases nitrogenadas (Frankham et al., 2002).

En el modelo de alelos infinitos o IAM, cada mutación transforma un alelo existente en uno nuevo que no estuvo presente previamente en la población. En este modelo, la homoplasia no es posible y todos los alelos idénticos son idénticos por descendencia (heredados por un ancestro en común).

En el modelo de un solo paso SMM cada mutación crea un nuevo alelo ya sea por la adición o la delección de una sola unidad de repetición de un microsatélite, con una probabilidad igual en ambos escenarios. En consecuencia, los alelos que difieren en muchos tamaños serán más distantes en relación con los que tienen el tamaño similar (Kimura & Ohta 1978).

c) **Deriva genética**

Se define como fluctuaciones debidas al azar en una población, tanto en las frecuencias génicas como las genotípicas, de generación en generación (Freeland, 2005; Templeton, 2006).

Las variaciones estocásticas de las frecuencias alélicas y genotípicas dependen del tamaño poblacional. Las variaciones esperadas de las frecuencias en las poblaciones descendientes se incrementan con la disminución del tamaño de la población parental. En poblaciones pequeñas, puede haber fluctuaciones amplias e impredecibles de las frecuencias génicas debido a eventos aleatorios, de una generación a otra.

En la naturaleza, la mayoría de las poblaciones son relativamente grandes, por lo que los efectos de la deriva pueden no ser importantes en comparación con los efectos de la selección o el flujo génico. El efecto fundador es la deriva génica que ocurre cuando un grupo reducido se separa de una población para fundar una población nueva. La deriva génica en estos casos es el resultado de muestreo de la población de origen, y una cierta cantidad de generaciones durante las cuales el tamaño de la nueva población permanece reducido.

En poblaciones que atraviesan una reducción drástica en su tamaño debido a una fragmentación del hábitat (cuello de botella), pueden incrementarse las fluctuaciones aleatorias en las frecuencias alélicas. Aún cuando las poblaciones pueden recuperar su tamaño original, el efecto de la deriva durante el cuello de botella permanece.

d) Endogamia

La endogamia se refiere al apareamiento no aleatorio en donde la probabilidad de que un individuo sea autóctigo (con alelos idénticos por descendencia) es mayor que para una población panmíctica. Por lo tanto, la endogamia no cumple con el supuesto de la panmixia, y reduce el número real de individuos con capacidad reproductiva. Además, una endogamia intensa, producida por ejemplo por autofertilización, también reduce la heterocigosis, y el tamaño poblacional es más parecido al de una población haploide (Hedrick, 2000).

A mayor endogamia, más probabilidad hay en una población de hallar individuos homocigotos para un locus en el cual ambos alelos sean idénticos por ascendencia. La identidad por ascendencia consiste en que dos copias de un gen provengan del mismo gen antecesor.

e) Selección natural

La selección es una fuerza evolutiva importante que puede afectar las frecuencias alélicas y genotípicas. En cada generación, los genotipos que promueven la supervivencia o la reproducción son los que predominarán entre los individuos reproductores. Asimismo, los alelos que promuevan dicha supervivencia o éxito reproductivo serán los que se encuentren con mayor frecuencia en la población. Por tanto, estos genotipos contribuirán de manera desigual a la progenie de la siguiente generación. Este es el proceso de adaptación por selección natural.

Se consideran dos tipos de selección natural: la selección direccional y la selección balanceadora. En genética de poblaciones, la selección direccional se refiere a los casos en los cuales un alelo (o un genotipo homocigoto) siempre tiene el éxito reproductivo más alto. La selección balanceadora se refiere a los casos en los cuales:

- a) una combinación de diferentes alelos (por ejemplo un heterocigoto) tiene el más alto éxito reproductivo relativo o
- 2) cuando el éxito reproductivo de diferentes alelos o genotipos depende de las condiciones medio ambientales o de la composición genotípica de una población.

III. JUSTIFICACION

La predicción de la estructura genética dentro de las poblaciones de parásitos es complicada debido a la diversidad en sus ciclos de vida (Blouin *et al.*, 1999). Sin embargo, ésta es importante para obtener mayor información de la transmisión y de los procesos epidemiológicos involucrados en las enfermedades, debido a que las variantes genéticas pueden estar asociadas a diferencias en infectividad y patogenicidad (Shrivastava *et al.*, 2005). En el caso de *T. solium* ayudaría a identificar las fuentes o el foco de infección en una población; es decir, al individuo teniásico responsable de la dispersión de la enfermedad, así como a determinar la distribución del parásito alrededor de la misma comunidad e identificar rebrotes de las mismas clonas luego de un tratamiento. Todo ello con la finalidad de elaborar mejores estrategias para el control y eventual eliminación del parásito. Además, la variabilidad genética encontrada podría explicar las dificultades halladas en el diseño de herramientas fiables de inmunodiagnóstico y vacunas puesto que sería una de los factores involucrados en la heterogeneidad presente en la neurocisticercosis, además del género y la edad de los pacientes (Vega *et al.*, 2003, Sciutto *et al.*, 2008).

Los microsatélites o repeticiones de secuencia simple (SSRs) son secuencias de ADN que están repetidas en “tandem” en un locus específico (Van Belkum *et al.*, 1998). Los marcadores microsatélites tienen la ventaja de ser marcadores multi-loci y teóricamente neutrales, esto último es muy importante puesto que brinda una visión más real de la estructura poblacional y la dinámica de transmisión del parásito. Además porque se encuentran distribuidos ampliamente por todo el genoma, son

altamente polimórficos, codominantes y pueden ser detectados por una amplificación por PCR convencional. Por todas estas razones, los microsatélites se han convertido en el método de elección recomendado en estudios que establecen diferencias entre individuos (Oura *et al.*, 2003).

Los análisis de variación genética dentro y entre las poblaciones pueden brindar información de cambios evolutivos, diferenciación genética y especiación del parásito (Thompson y Lymbery, 1995). Es por ello, que han sido ampliamente utilizados en estudios de variabilidad genética de otros parásitos como *Leishmania spp.* (Rodríguez *et al.*, 1997), *Plasmodium falciparum* (Su y Wellems, 1996), *Schistosoma japonicum* (Shrivastava *et al.*, 2005) y *Echinococcus multilocularis* (Xiao *et al.*, 2011). En *Leishmania spp.*, los microsatélites fueron los marcadores ideales para la discriminación entre todas las especies dentro del subgénero *Viannia* (Russell *et al.*, 1999; Oddone *et al.*, 2009). En *Schistosoma japonicum* demostraron su utilidad en el establecimiento de distintas subpoblaciones de China. Además, los marcadores microsatélites permitieron detectar mayor variación en especies en las cuales se había reportado una poca diversidad genética como es el caso de *Echinococcus multilocularis* (Nakao *et al.*, 2003).

En ese sentido, este estudio sería pionero en su campo, puesto que utilizaría marcadores microsatélites para la evaluación de poblaciones naturales peruanas de *T.solium*.

IV. OBJETIVOS:

OBJETIVO GENERAL:

Diseñar, evaluar y aplicar un conjunto de marcadores microsatélites para evaluar la diversidad y estructura genética de aislados de *Taenia solium* provenientes de diferentes regiones del Perú.

OBJETIVOS ESPECÍFICOS:

- Determinar la abundancia, distribución y composición de microsatélites o repeticiones de secuencia simple en el genoma de *Taenia solium*
- Diseñar cebadores específicos para un conjunto seleccionado de microsatélites presentes en el genoma de *Taenia solium*.
- Estandarizar la técnica y evaluar el polimorfismo de los marcadores microsatélites presentes en *Taenia solium*.
- Evaluar la diversidad genética de parásitos adultos provenientes de diferentes zonas del país usando los microsatélites polimórficos.

V. MATERIALES Y METODOS:

MATERIAL BIOLÓGICO:

Se emplearon 156 muestras de tejido (proglótidos) de *Taenias* adultas almacenadas en el banco de parásitos del Grupo de Neurocisticercosis del Perú ubicado en el Instituto Nacional de Ciencias Neurológicas. Los proglótidos fueron resuspendidos en un 1mL de etanol 70% en microtubos (Eppendorf) libres de nucleasas y conservados a 4°C hasta su procesamiento.

Las muestras pertenecían a diferentes comunidades de 7 provincias del país: Ancash (n=3), Tumbes (n=60), Piura (n=28), Cajamarca (n=7), Huancayo (n=2), Ayacucho

(n=2), Puno (n=53) y Cusco (n=1) (Tabla 01). La distribución geográfica de las comunidades se muestra en las figuras 05, 06 y 07.

La mayoría de las muestras provienen de campañas realizadas durante el periodo del 2004-2009 con excepción de las muestras de Piura, las cuales fueron colectadas durante este año 2013.

Tabla 01. Origen geográfico de las muestras de *T.solium*, número de parásitos genotipificados por cada comunidad y región, así como las subpoblaciones asignadas

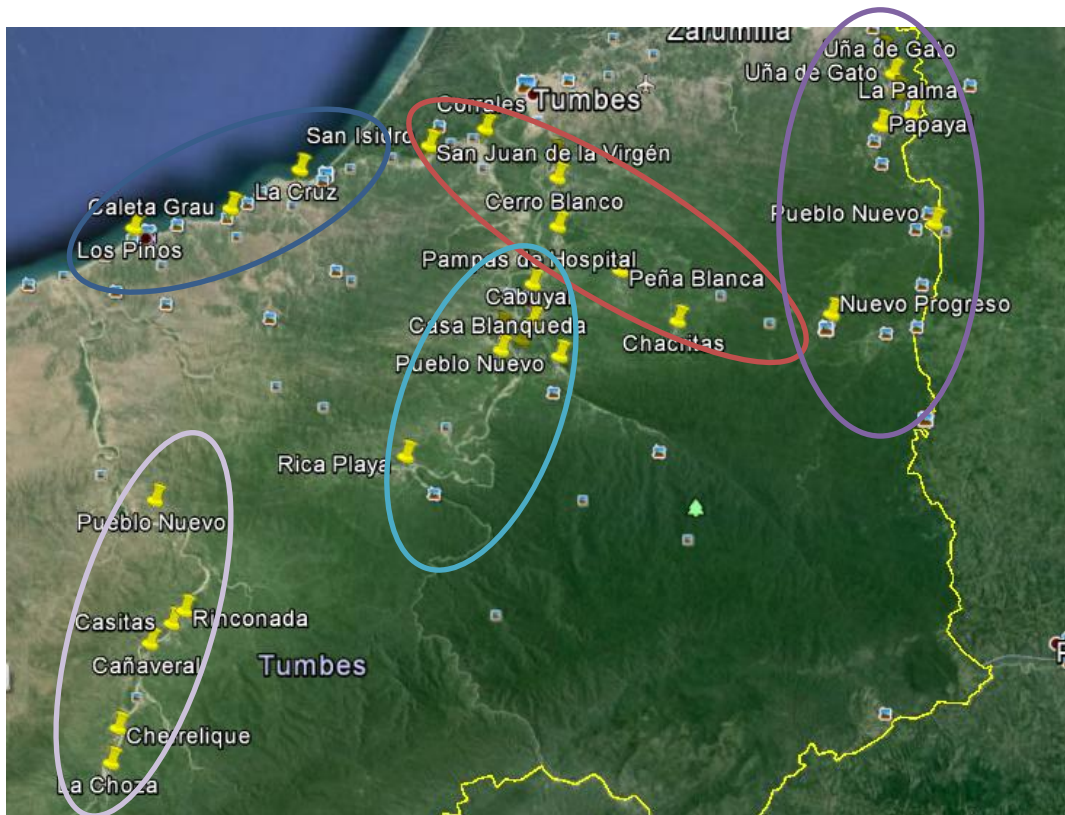
N° de Taenias	Comunidad/provincia	Región	Población asignada en este estudio
7	Cajamarca	Cajamarca	pob_Cajamarca
3	Huaraz	Ancash	otras_poblaciones
1	Cusco	Cusco	otras_poblaciones
2	Ayacucho	Ayacucho	otras_poblaciones
2	Huancayo	Junin	otras_poblaciones
7	AUL	Piura	pob_Piura 1
4	Pampa Elera Baja	Piura	pob_Piura 1
5	Piura	Piura	pob_Piura 1
1	Ayabaca	Piura	pob_Piura 1
7	Joras Centro	Piura	pob_Piura 2
4	Minas de Joras	Piura	pob_Piura 2
8	Fharata	Puno	subpopPuno_1
7	Callata	Puno	subpopPuno_1
1	Tuturuma	Puno	subpopPuno_2

3	Conchaca	Puno	subpopPuno_2
2	Achacuni	Puno	subpopPuno_2
6	Acora	Puno	subpopPuno_3
1	Jayu Jayu	Puno	subpopPuno_3
1	Chulluni	Puno	subpopPuno_3
13	Capachica	Puno	subpopPuno_4
8	Camicachi	Puno	subpopPuno_5
3	Calacota	Puno	subpopPuno_5
1	San Juan	Tumbes	subpopTumbes_1
11	Corrales	Tumbes	subpopTumbes_1
8	Tumbes	Tumbes	subpopTumbes_1
1	Peña Blanca	Tumbes	subpopTumbes_1
1	Cerro blanco	Tumbes	subpopTumbes_1
1	Garbanzal	Tumbes	subpopTumbes_1
1	Fuerte coloma	Tumbes	subpopTumbes_1
1	Chacritas	Tumbes	subpopTumbes_1
2	San isidro	Tumbes	subpopTumbes_2
1	Caleta Grau	Tumbes	subpopTumbes_2
1	Los Pinos	Tumbes	subpopTumbes_2
1	La Cruz	Tumbes	subpopTumbes_2
1	Pampas de hospital	Tumbes	subpopTumbes_3
1	Rica playa	Tumbes	subpopTumbes_3
1	Cabuyal	Tumbes	subpopTumbes_3
1	El limon	Tumbes	subpopTumbes_3
2	El oidor	Tumbes	subpopTumbes_3

1	Casa blanqueada	Tumbes	subpopTumbes_3
1	Cuchareta	Tumbes	subpopTumbes_4
5	Nuevo progreso	Tumbes	subpopTumbes_4
1	Uña de gato	Tumbes	subpopTumbes_4
4	Pueblo nuevo	Tumbes	subpopTumbes_4
2	El porvenir	Tumbes	subpopTumbes_4
5	La palma	Tumbes	subpopTumbes_4
1	Cañaverl	Tumbes	subpopTumbes_5
4	Casitas	Tumbes	subpopTumbes_5

Las provincias o comunidades más cercanas geográficamente fueron agrupadas en una misma subpoblación para aumentar el número total de parásitos y evitar posibles sesgos debido al tamaño poblacional (Figura 05, 06 y 07). Las provincias de Huancayo, Ayacucho, Cusco y Ancash fueron agrupadas como una misma subpoblación debido al bajo número de muestras obtenidas por cada región.

Figura 05. Localización de las comunidades de Tumbes empleadas en este estudio.



Cada una de las subpoblaciones está señalada con los círculos en el mapa. Círculo rojo: subpopTu_1, círculo azul: subpopTu_2, círculo celeste: subpopTu_3, círculo plomo: subpopTu_4, círculo rojo: subpopTu_5.

Figura 06. Localización de las comunidades de Piura empleadas en este estudio



Cada una de las subpoblaciones está señalada con los círculos en el mapa. Círculo amarillo: popPiura_1, círculo rojo: popPiura_2.

Figura 07. Localización de las comunidades de Puno empleadas en este estudio



Cada una de las subpoblaciones está señalada en los círculos en el mapa. Círculo azul: subpopPuno_1, círculo rojo: subpopPuno_2, círculo verde: subpopPuno_3, círculo amarillo: subpopPuno_4, círculo morado: subpopPuno_5.

PROCEDIMIENTO EXPERIMENTAL:

5.1. Identificación de marcadores microsatélites a partir del genoma de *Taenia solium*

Para la identificación de los microsatélites se utilizaron las secuencias parciales de dos genomas obtenidos a través del secuenciamiento *de-novo*, uno de ellos proveniente del estadio adulto de una *Taenia* de Puno (denominado Genoma 1) y otro proveniente del estadio cisticerco (denominado Genoma 2). Ambos genomas fueron ensamblados y anotados parcialmente en el laboratorio de Bioinformática.

Se identificaron las secuencias microsatélites usando el programa *Blast* y el programa desarrollado por *Gurie et al*, el cual puede ser obtenido del sitio <ftp://ftp.technion.ac.il/pub/supported/biotech/ssr.exe>; que reporta el motivo, la longitud del motivo, el número de repeticiones y la localización genómica de los microsatélites. Se consideraron sólo aquellos que presentaban una longitud del motivo repetido de entre 2 – 10 bp y de tres repeticiones como mínimo.

Teniendo como premisa que los marcadores microsatélites polimórficos son los más informativos en los estudios de genotipificación, se identificó un primer grupo de marcadores teniendo en cuenta los siguientes criterios:

- 1) Se eliminó la redundancia en las secuencias a través del análisis de la base de datos de Secuencias Cortas de Expresión (denominadas *ESTs* por sus siglas en inglés) de *T.solium* disponible en el sitio <ftp://bioinformatica.biomedicas.unam.mx/> del Instituto Sanger.
- 2) Se corroboró que las secuencias de microsatélites estén presentes en el mismo locus tanto en los Genoma 1 y 2 como en los ESTs empleando el programa *Blast*.
- 3) Se consideraron como secuencias microsatélites candidatas a aquellas que mostraran variación en el tamaño del motivo presente en las secuencias de los Genomas 1 y 2 y en los ESTs, lo cual indicaría presencia de polimorfismo. Además, se verificó que las regiones flanqueantes de los microsatélites sean conservadas en ambos, tanto en el Genoma 1 y 2 como en los ESTs. Adicionalmente, se seleccionaron otros microsatélites de mayor longitud que

estaban presentes en ambos genomas pero que no se lograron ubicar en las secuencias de ESTs.

Posteriormente, con la disponibilidad del genoma anotado de *T.solium*, se procedió a identificar un segundo grupo de marcadores microsatélites aplicando los siguientes criterios:

- 1) Se hizo una comparación de los *contigs* (secuencias largas de nucleótidos productos del secuenciamiento) de los Genomas 1 y 2. Tomando como referencia a los *contigs* del Genoma 1 se hizo la búsqueda de los SSRs en los *contigs* homólogos del Genoma 2 para la confirmación del polimorfismo.
- 2) Se evaluaron un total de 200 SSRs (100 dinucleótidos y 100 trinucleótidos) con la mayor longitud del motivo de repetición presentes en el Genoma 1 y se consideraron sólo a aquellos cuyo polimorfismo en el fragmento de repetición del motivo es consecuencia de la variación del motivo SSR.

5.2. Diseño de cebadores:

Para ambos grupos de marcadores se diseñaron cebadores a partir de las secuencias que flanqueaban a los microsatélites presentes tanto en regiones codificantes y no codificantes. Para realizar el diseño se utilizó el programa “Primer 3” disponible en Internet teniendo como parámetros una longitud de los iniciadores de 20 bp, un %GC de 50 y un Tm aproximado de 50°C. Luego del diseño de los cebadores, se comprobó que la región de amplificación (identificada *in silico*) tenga la misma longitud tanto

en el Genoma 1 como en el 2. Una vez diseñados los cebadores fueron enviados a la empresa Invitrogen para su síntesis.

5.3. Metodología de los marcadores microsatélites:

5.3.1. Extracción de ADN genómico a partir de proglótidos del estadio adulto de *Taenia solium*.

Se extrajo el ADN genómico de los 156 proglótidos de *Taenia* usando el QIAamp DNA Mini Kit (Qiagen) de acuerdo al protocolo del fabricante. Se resuspendió el ADN genómico en un volumen de 100 ul de buffer de elución del kit, se los colocó en tubos de microcentrífuga de 1.5 mL de capacidad y almacenó a – 20°C hasta su utilización.

5.3.2. Cuantificación y dilución de ADN genómico:

Se estableció la concentración de ADN extraído mediante el uso del espectrofotómetro de amplio espectro NANODROP. Se evaluó la relación entre las absorbancias 260/280 para determinar la contaminación protéica, y la relación entre las absorbancias 260/230 para determinar la contaminación salina.

Se diluyeron las muestras con cantidad suficiente de agua libre de nucleasas (NFW) hasta alcanzar la concentración final adecuada.

5.3.3. Amplificación de los microsatélites por PCR:

5.3.3.1. Estandarización de las Condiciones de Reacción de PCR:

La estandarización del protocolo del PCR, consiste en definir las condiciones óptimas de amplificación para cada par de cebadores. Los parámetros evaluados para la estandarización de la reacción fueron la temperatura de hibridación o Ta (*Annealing temperature*), concentración de ADN, cebadores y de componentes como: Agua libre de nucleasas o NFW (*nucleases free water*) y MgCl₂ (cloruro de magnesio).

En la Tabla 02, se muestra las concentraciones finales para cada uno de los componentes de la mezcla (*master mix*) para la amplificación por PCR de cada cebador.

Tabla 02. Componentes del master mix del PCR

Componentes del <i>master mix</i>		
Reactivo	C _o	C _f
NFW (SIGMA)	--	--
Buffer Taq (Promega)	10X	1X
MgCl ₂	25mM	1.5mM/2.5mM
NTPs	10mM	0.2mM
Iniciador F	10uM	1uM
Iniciador R	10uM	1uM
Taq Pol	0.1U	0,05U
ADN	10ng/ul	20 ng
Volumen total del <i>master mix</i>	25 µl por rxn	

NFW: agua libre de nucleasas. Buffer Taq Promega=Solución tampón, MgCl₂=cloruro de magnesio. dNTPs = dideoxinucleótidos. Taq Pol = enzima Taq polimerasa. Co= concentración inicial. Cf= concentración final. rxn = reacción.

Las temperaturas de hibridación se determinaron a través de ensayos de gradiente de temperatura, las cuales variaron entre 55 °C - 65°C.

Las reacciones de PCR se llevaron a cabo para cada par de SSR en un termociclador modelo MJ Research utilizando los cebadores detallados en la Tabla 05. Las condiciones de amplificación fueron las siguientes:

- a) INICIO: 1 ciclo de 94 °C por 4 minutos
 - b) DESNATURALIZACIÓN DEL ADN: 1 minuto a 94 °C.
 - c) ALINEAMIENTO: Aquí se realizaron las pruebas de gradiente de temperatura entre 55 °C a 65 °C, esto duró 1 minuto.
 - d) EXTENSIÓN: 72 °C por 1 minuto.
- b), c) y d) se realizaron durante 35 ciclos.
- e) ULTIMA EXTENSION: a 72 °C por 10 minutos.

Y finalmente una etapa para conservar la reacción a 12 °C.

5.3.3.2. Evaluación del polimorfismo de los marcadores microsátélites:

Inicialmente, se evaluó el nivel de polimorfismo en 12 individuos de la población, los seis primeros pertenecientes al norte del país (Tumbes) y los 6 siguientes al sur del país (Puno), como paso previo a la amplificación en la población completa. Aquellos SSRs que mostraran una variación en la longitud de los motivos reflejado en una diferencia en el tamaño de las bandas por electroforesis capilar fueron amplificados en la población total.

Una vez determinado el polimorfismo se procedió a la amplificación de los microsatélites polimórficos en toda la población de *Taenias*. (Los 156 individuos).

5.3.3.3. Electroforesis:

Los productos de la amplificación fueron separados por electroforesis en capilar en el equipo QIAxcel Advanced. La separación de los ácidos nucleicos en base a su tamaño es realizada en los capilares de un cartucho pre ensamblado de gel que forma parte del kit de Alta resolución o Qiaxcel High resolution kit que requiere el equipo QIAxcel Advanced para realizar la corrida electroforética.

Como paso previo al análisis de las muestras, se realizó una calibración de la intensidad del cartucho. Con ello, las intensidades de cada capilar son normalizadas y se calcula un factor de corrección el cual es aplicado a cada nueva corrida para evitar cualquier variación en la lectura de la intensidad de cada capilar del cartucho.

Luego, se procedió a la configuración del software del equipo. Siguiendo las recomendaciones del fabricante, se eligió como método de corrida al OM800 (Figura 08). De acuerdo a ello, el equipo ajusta el tiempo de inyección de la muestra en el capilar, el tiempo de separación y el voltaje aplicado. Adicionalmente, para calcular el tamaño del amplificado, el software del equipo requiere tener como referentes dos marcadores: uno de peso molecular comercial y otro de alineamiento interno. Siguiendo las recomendaciones del fabricante, se empleó el marcador de peso molecular de 25pb -1Kb (Invitrogen) y el marcador de alineamiento interno (*QX*

alignment marker) de 15pb-1Kb (Figura 08). Finalmente, se ubica la placa de 96 pocillos en el equipo y se deja correr por 1 hora aproximadamente.

5.3.3.4. Genotipificación:

El genotipaje se realizó en el sistema avanzado Qiaxcel utilizando del software QIAxcel ScreenGel acoplado al sistema. Una vez terminada la corrida, el software brinda un patrón de bandas en gel así como un electroferograma.

El software calcula y asigna el tamaño del fragmento de ADN visualizado como pico o banda, basado en el tiempo de migración de dicho fragmento comparado con el marcador de peso molecular y el marcador de alineamiento interno del sistema.

Figura 08. Recomendaciones especificadas en el manual del fabricante (Qiagen)

Método	Tamaño del fragmento			
	100–500 bp	500 bp – 1 kb	1–5 kb	5–10 kb
OM400* OL400† OH400‡	20 bp	100 bp	500 bp	N/A
OM500* OL500† OH500‡	10 bp	50 bp	200 bp	N/A
OM800* OL800† OH800‡	3–5 bp	N/A	N/A	N/A
OM1200* OL1200† OH1200‡	N/A	N/A	500 bp – 1 kb	1–1.5 kb

* Se recomiendan los métodos OM400, OM500, OM700 y OM1200 para las concentraciones de ADN de 10-100 ng/ul (por ejemplo, productos de PCR amplificados a partir de ADN genómico con 30-40 ciclos)

† Se recomiendan los métodos OL400, OL500, OL700 y OL1200 para las concentraciones de ADN menores de 10 ng/ul.

‡ Se recomiendan los métodos OH400, OH500, OH700 y OL1200 para las concentraciones de ADN mayores a 100 ng/ul

N/A No aplica

5.3.4. Ensayo de reproducibilidad

Con la finalidad de verificar la precisión de nuestros procedimientos, se amplificó el marcador TS_SSR_01 cuatro veces de manera separada en 9 muestras distintas de ADN. Los productos fueron analizados usando el sistema de Qiaxcel.

5.3.5. Elaboración del archivo *Input* para el análisis de diversidad

Los tamaños (en pares de bases) de los marcadores microsatélites polimórficos para cada uno de los individuos de la población fueron registrados en una hoja de Excel. Posteriormente, se hizo un ajuste en el tamaño de las bandas teniendo en cuenta el genotipo de la *Taenia adulta* (Genoma 1) secuenciado. Asimismo, se eliminaron a los individuos que no tenían más del 50% de información genotípica. Dicho archivo se utilizó como *input file* del programa Microsatellite Analyzer (MSA) (Dieringer, D. & Schlötterer, C., 2001) el cual lo transforma este archivo en diferentes formatos requeridos por otros programas de genética de poblaciones.

5.4. Análisis de diversidad genética dentro de las subpoblaciones de *T. solium*

El número de alelos por locus, la heterocigosidad observada y esperada (Nei, 1973), el contenido de información polimórfica (PIC) (Botstein y col., 1980) fueron calculados para cada uno de los marcadores microsatélites polimórficos utilizando el programa Microsatellite Analyzer (MSA) (Dieringer, D. & Schlötterer, C., 2001) y el programa Cervus v3.0.

Las frecuencias alélicas, el número de alelos por locus, la heterocigosidad observada y esperada (Nei, 1973) e índice de Garza-Williamson (Garza y Williamson, 2001) de cada una de las subpoblaciones fueron determinados utilizando el programa Arlequín v3.1 (Excoffier, Laval y Schneider, 2005).

Las desviaciones del equilibrio de Hardy-Weinberg (deficiencia o exceso de heterocigotos) para cada loci y cada subpoblación, así como los análisis de

diferenciación génica y genotípica entre subpoblaciones fueron estimados mediante la prueba U (Score test) y el método de Fisher, utilizando el algoritmo de Cadena de Markov (MC) [Dememorization: 10 000, batches: 100, itineration per batch: 10000] para estimar el valor exacto de probabilidad (Guo y Thompson, 1992), disponible en el programa Genepop v4.0 (Raymond y Rousset, 1995a).

Se determinó el desequilibrio de ligamiento entre pares de loci mediante la prueba G exacta (exact G test) y el método de Fisher, utilizando el algoritmo de cadena de Markov (MC) [Dememorization: 10 000, batches: 100, itineration per batch: 10 000] para estimar el valor exacto de probabilidad (Guo y Thompson, 1992), utilizando el programa Genepop v4.0.

5.5. Análisis de diversidad genética entre subpoblaciones de *T.solium*

El grado de diferenciación genética entre las 7 subpoblaciones de *Taenias* fue cuantificada mediante los estadísticos F (F_{ST} , F_{IS} , y F_{IT}) (Weir y Cockerham, 1984). Se determinó el valor del coeficiente de consanguinidad F_{IS} e índice de fijación F_{ST} para cada subpoblación y para la totalidad de individuos analizados utilizando el programa Genepop v4.0.

Se calcularon los valores de F_{ST} , R_{st} (Slatkin, 1995; Michalakis y Excoffier, 1996) y coeficiente de coancestría (Reynolds y col., 1983) entre cada par de subpoblaciones en forma de matriz pareada utilizando el programa Arlequín v3.1. Además se calculó el número efectivo de migrantes para cada subpoblación en base a los valores del F_{ST} . Los dendrogramas se obtuvieron mediante el algoritmo matemático Neighbor-

Joining (NJ; Saitou y Nei 1987) y el unweighted pair group method with arithmetic mean - UPGMA utilizando el programa DISPAN (Ota, 1993) a partir de la matriz de distancias genéticas (D_A y D_S) entre los pares de poblaciones correspondientes a los distintos grupos muestrales considerados. La robustez de la topología del dendograma fue evaluado con un test de 1000 bootstraps.

5.6. Análisis de estructura poblacional

Se evaluó la estructura presente en las poblaciones de *Taenias* a partir de sus genotipos mediante una análisis de Agrupamiento Bayesiano bajo un modelo de apareamientos no azarosos (no admixture) y el algoritmo MCMC (MonteCarlo-Cadena de Markov) [Burnin period: 30 000, number of MCMC reps after Burnin: 100000] empleando el programa Structure v2.2 (Pritchard y col., 2000).

5.7. Análisis de varianza molecular (AMOVA)

El análisis de la varianza molecular (AMOVA) (Weir & Cockerham 1984; Excoffier et al. 1992; Weir 1996) basado en la varianza de las frecuencias alélicas fue calculada utilizando el programa Arlequin v3.

5.8. Análisis de asociación –Test de Mantel

En base a los valores del coeficiente de ancestría y a los de distancia geográfica se evaluó la correlación entre ambas variables mediante el test de Mantel disponible en el programa Arlequin v3.1.

VI. RESULTADOS

6.1. Determinación de la abundancia, distribución y composición de microsatélites o repeticiones de secuencia simple en el genoma de *Taenia solium*

El número total de microsatélites identificados en el Genoma 1 (parásito adulto) de *T.solium* fue de 9129 y en el Genoma 2 (cisticerco), 9936. En ambos casos, los microsatélites se encontraron distribuidos a lo largo de todo el genoma (Tabla 03).

Los ESTs usados en este estudio fueron normalizados. Por consiguiente, se minimizó la redundancia en la base de ESTs y se pudo encontrar una gran cantidad de secuencias únicas de cDNA. Se hallaron un total de 400 secuencias presentes en los ESTs reportados en la base de datos del GenBank, a partir de las cuales se escogieron 5 secuencias (1 penta-, 2 tetra y 2 tri- nucleótidos) que cumplieran con los criterios de presencia en el mismo locus en todos los genomas analizados y evidencia de polimorfismo en las secuencias. Además se seleccionaron 8 microsatélites de mayor longitud (2 tetra-, 2 tri-, 4 di-nucleótidos) los cuales no se ubicaron en los ESTs pero que estaban presentes en los genomas de *Taenia*.

Para el segundo grupo de microsatélites, de las 200 secuencias microsatélites analizadas, sólo 23 SSRs cumplieran con los criterios anteriormente descritos (11 tri- y 12 di- nucleótidos).

Tabla 03. Distribución de los microsatélites con respecto al motivo de repetición y tamaño del genoma en *T.solium*

Tipo de genoma	Tamaño analizado(pb)	Tipo del motivo de repetición						Total
		Mono	Di	Tri	Tetra	Penta	Hexa	
Genoma 1	109,898,809	2,298	3,393	2,393	889	123	33	9,129
Genoma 2	114,605,177	2,737	3,537	2,464	1,044	122	32	9,936

En ambos grupos, las repeticiones dinucleotídicas constituyeron la mayoría de los microsatélites, de éstos la mayoría tienen una longitud total ≤ 5 pb. Los microsatélites mononucleotídicos, trinucleotídicos y tetranucleotídicos se encontraron en menor cantidad (Tabla 04).

Tabla 04. Frecuencia relativa de los diferentes tipos de motivos en cada clase de microsatélites con repeticiones tipo mono-, di- y tri- nucleotidos analizados en los genomas de *T.solium*

Tipo de repetición	Tipo	
	Genoma 1	Genoma 2
Mono-nucleotido		
A/T	0,85	0,78
C/G	0,15	0,22
Di-nucleotido		
AC/GT	0,69	0,67
AG/CT	0,24	0,26
AT/AT	0,07	0,07
CG/CG	0,00	0,00
Tri-nucleotido		
AAC/GTT	0,09	0,09
AAG/CTT	0,11	0,11
AAT/ATT	0,05	0,05
ACC/GGT	0,18	0,19
ACG/CGT	0,02	0,02
ACT/AGT	0,03	0,03
AGC/CTG	0,10	0,08

AGG/CCT	0,34	0,35
ATC/ATG	0,06	0,06
CCG/CGG	0,02	0,01

Asimismo, existió una alta representación de microsatélites dinucleotídicos ≥ 6 bp de longitud del tipo AC/GT y AG/CT y una baja representación del tipo AT/AT y CG/CG mientras que en el caso de microsatélites mononucleotídicos ≥ 6 bp de longitud existe una mayor abundancia del tipo A + T y una baja representación del tipo C + G

6.2. Diseño de cebadores específicos de los microsatélites presentes en el genoma de *Taenia solium*.

A partir de las secuencias microsatélites identificadas en el genoma y siguiendo el procedimiento previamente descrito se diseñaron 36 pares de cebadores (Tabla 05).

Tabla 05. Lista de cebadores diseñados a partir de las secuencias microsatélites usando el programa Primer 3.

Cebador	Secuencia del cebador (5'-3')		Tamaño esperado (pb)	Tipo de repetición	Temperatura de hibridación
	Cebador de avance	Cebador de retorno			
TS_SSR_01	ACCGGTGGTCGGAATTATTA	GTTCTTGCTGAGGTGGTTCC	222	(CCATT)	60°C
TS_SSR_02	CTCCGTGTCTTGACAGCAAA	TGACGAAATGGAACAGTGG A	188	(ATGA)	60°C
TS_SSR_03	TTTCAAGCACGTGTCAGCAT	GCTGGCAGACAGTGAGTAG G	152	(CATT)	60°C
TS_SSR_04	CAGATGAGGGGATGATGCTT	GAACGATCCCAACCTCCATA	175	(GTT)	60°C
TS_SSR_05	GGGAAAATGCAGTTCAGAGC	GGTCTGATGCGAGGTCTAGG CAAGAATGAACGGGAGCAA	194	(TAA)	60°C
TS_SSR_06	GACCAAGCCCAACACCTCTA	C	233	(GGTA)	62°C
TS_SSR_07	GCACACAACTGGTCACTCG	TGCTATGCGTTTGCTTGT	217	(CAAT)	60°C
TS_SSR_08	TCGTCAGTGTGGGAGAGTGA	TGGTTGGATTTGTGCTTTGA	176	(ACG)	58°C

TS_SSR_0	9	AAGCCAATGGTGACCAAGAG	GCCAGCATAGAAGAGCCTGT	172	(GGT)	60°C
TS_SSR_1	0	CGACTCACGGCATTTCATCTA	TCCAAGACCCTGTGAAATCC	210	(GT)	60°C
TS_SSR_1	1	TCATCTTCCCCGTAAGGCTA	ACACTCGAAGCGCAGTGTTT	178	(GA)	60°C
TS_SSR_1	2	ATCTCGACAGGCTCGAGTTC	TCCGAACAGCTTCGAGTTTT	210	(TG)	60°C
TS_SSR_1	3	GTAGCGGTAACGGAGTGAGG	TCAGGCTGGTAACGTGTCAG	178	(GT)	60°C
TS_SSR_1	4	AGCCGGTTCTCAGTTGATTG	AATGCACTCATGCCATCTCA	157	(TG)	60°C
TS_SSR_1	5	GAAAAGAACGGCATGCAAAT	GTTTGGCCATTTTGCCTCTA	161	(AT)	60°C
TS_SSR_1	6	CGCTGGACTAGGGTCGAATA	CAGCAGAACAACAGCACCA	157	(GT)	60°C
TS_SSR_1	7	GCATTCCGAGGATGAATGAT	CGTTTTTCTGCACACTTGGA	155	(CA)	60°C
TS_SSR_1	8	AGTTAGCGTGCTTGCTTGGT	ATTCTGTTGCAACCTCCAC	161	(GT)	60°C
TS_SSR_1	9	TCCCTTACACCCTTCACGTC	AAAGGCGGTAGATTGTGTGC	159	(TG)	60°C
TS_SSR_2	0	GGCCATTTCAGTACCAACCAT	TGTGCATGCCATTGTATGTG	148	(CT)	60°C
TS_SSR_2	1	CTATGCCACACCCAACAATG	GGCCTTCAAGATCACTCGTC	182	(GT)	60°C
TS_SSR_2	2	CCTATTCCACTGGGGTGATG	TCGATGAGCTTGCTGTATGT	168	(TG)	60°C
TS_SSR_2	3	CCTTTTTCGGTGAAGTCGAT	GCCTCCTTACACACATGCAA	198	(CA)	60°C
TS_SSR_2	4	CCCCATTTCTGTTTCCTCT	GCGGTGGCAATATAAGCATT	138	(CT)	60°C
TS_SSR_2	5	AGGTGGCGTTATGAATCAGC	GCAAACCATCGGATAAAGG	165	(AC)	60°C
TS_SSR_2	6	CGGTTTGCTTTTATGCCAAT	AAATGGTCGCCTGAAATGAC	159	(GAA)	60°C
TS_SSR_2	7	GAGGTCTCGCCTCATCAAAG	TTTCCACTCCAAAAACTCG	165	(GAA)	60°C
TS_SSR_2	8	TGACGCTGGTAAGCTGTTG	GGAACCTTGGCACGAGATA	200	(GTA)	60°C
TS_SSR_2	9	AAAGATGGACGGAAACAGG	GTTGGACGGAGATGTGTGTG	180	(AGG)	60°C
TS_SSR_3	0	TGACGTGTCGTCAGGTAGGA	CGCATAGCCAGTACTTGTTT	183	(TCC)	60°C
TS_SSR_3	1	GGTTGCTTTTGCTTGCTCCTC	CACTCTCCACGAGTCCACAA	190	(TGA)	60°C
TS_SSR_3	2	TGACGTTAACGAGGGTGTG	AGATCTCGCCTTGCAACAAT	182	(AGC)	60°C
TS_SSR_3	3	CCAGCGGCATATTACAAAGG	ACTCAAAAGCGCCGAAATTA	123	(AGG)	60°C
TS_SSR_3	4	ATCACTCCTGTCCCAACTGC	GGGTCGATTGGTCAGAGAA	178	(CCT)	60°C
TS_SSR_3	5	GGGCGTGAACCTCGAATAAAA	GGGGCAGACAAGTGAAAAA	162	(CCA)	60°C
TS_SSR_3	6	GCCCTGATTGTTGCTTTTGT	AACGACACGCGGAAAATAT	177	(TCT)	60°C

6.3. Amplificación de los marcadores microsatélites de *Taenia solium*.

6.3.1. Extracción y cuantificación de ADN genómico:

La concentración de ADN obtenido a partir de los proglótidos de *Taenia* fue alta en todas las muestras. Los valores oscilaron entre 50 ng/ul y 200 ng/ul.

La pureza y calidad del ADN fue analizada a través de la proporción de absorbancia a 260 nm / 280 nm y 260 nm / 230 nm. En el análisis de la relación de absorbancia A260/A280, los valores menores de 1.8 son indicadores de contaminación proteica del ADN, los valores entre 1.8 y 2.0 son indicadores de pureza del ADN en cambio valores mayores a 2.0 indican contaminación con fenol. Todas las extracciones de ADN obtenidas presentaron una relación de absorbancia A260/280 entre 1.97 y 2.07, indicando que el ADN no se encontraba contaminado por proteínas que pudieran interferir con las reacciones de amplificación de PCR.

En cuanto al análisis de la relación de absorbancia A260/A230, en todas las muestras fue alrededor de 1.9 en todas las muestras lo que indica que el ADN estaba libre de contaminantes.

6.3.2. Estandarización de las Condiciones de Reacción de PCR

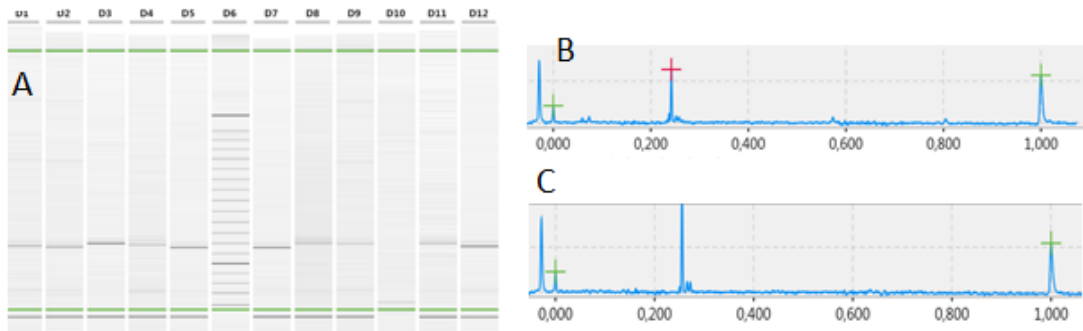
La temperatura de hibridización ideal para la mayoría de los marcadores fue de 60°C, excepto para los marcadores TS_SSR06 y TS_SSR08, los cuales tuvieron una temperatura de hibridización de 62°C y 58°C, respectivamente.

6.3.3. Evaluación del Polimorfismo en los marcadores Microsatélites

Los 36 pares de cebadores fueron amplificados en una muestra de 12 individuos y mostraron un tamaño similar al calculado por herramientas bioinformáticas. Solo 7 loci de microsatélites identificados resultaron ser polimórficos: TS_SSR_01, TS_SSR_09, TS_SSR_16, TS_SSR_18, TS_SSR_27, TS_SSR_28, TS_SSR_32 y fueron elegidos para su amplificación en la población total. Se hizo la selección en base a la variación en la longitud de los motivos reflejado en una diferencia en el tamaño de las bandas o de los picos del Qiaxcel (Figura 09).

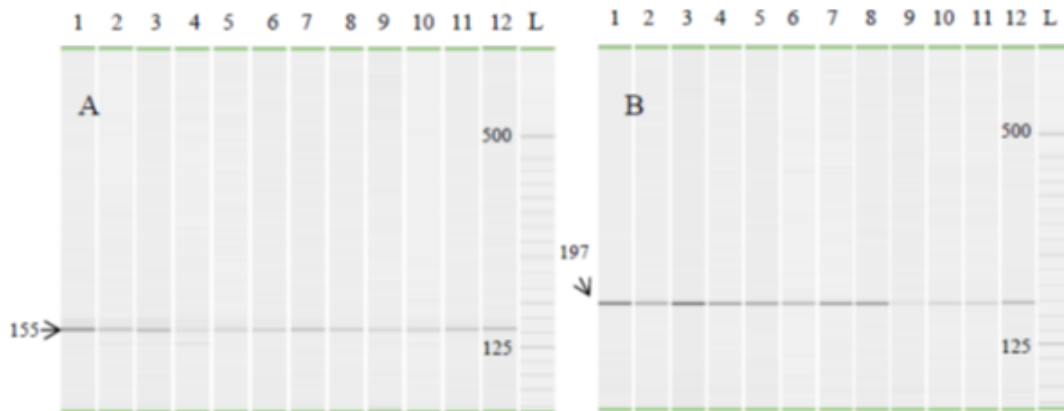
El resto de marcadores no mostraron variación en el tamaño del amplificado y fueron considerados monomórficos, es decir no presentaron variación o polimorfismo (Figura 10). El microsatélite TS_SSR07 amplificó sólo en las muestras procedentes de Puno (Figura 11), es por ello que no fue incluido en el análisis total. El microsatélite TS_SSR08 no logró amplificar, a pesar de probar varias condiciones del PCR.

Figura 09. Gel de electroforesis y electroferograma con los productos de PCR de los microsatélites TS_SSR09 evaluados en las *Taenias* adultas.



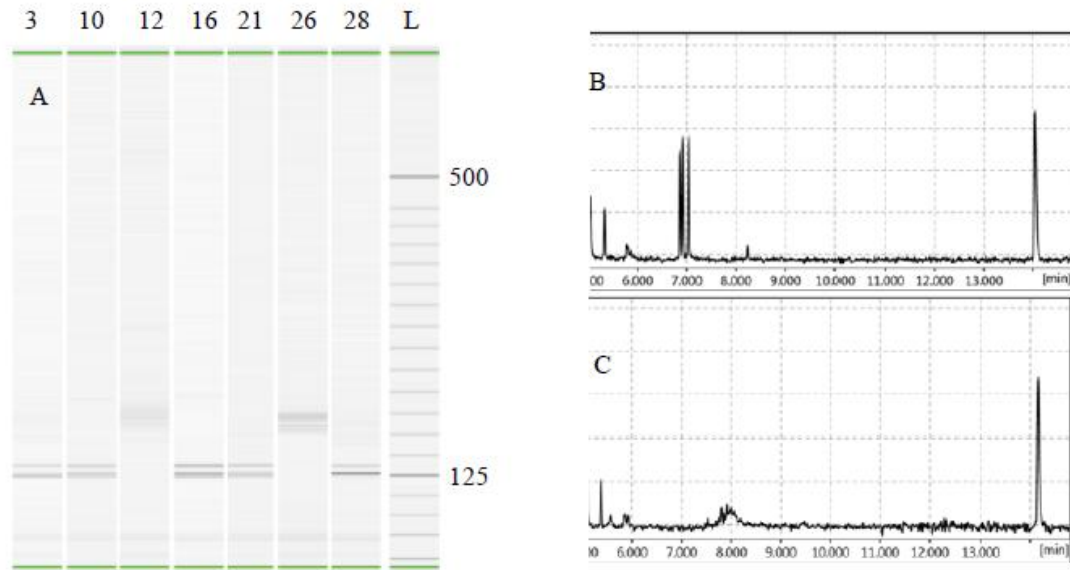
Gel brindado por el software del equipo Qiaxcel. Las *Taenias* están identificadas por sus números ID genotipificadas con el TS_SSR09. A: Gel de electroforesis con los amplificados de 11 muestras. L: marcador de peso molecular de 25pb-500pb (Invitrogen). B: electroferograma de la muestra D2 que muestra un solo pico de 166pb. C: electroferograma de la muestra D3 que muestra un pico de 169pb.

Figura 10. Gel de electroforesis con los productos de PCR de los microsatélites TS_SSR03 y TS_SSR05 evaluados en las 12 *Taenias* adultas.



Gel brindado por el software del equipo Qiaxcel. Las *Taenias* son identificadas por sus números ID, genotipificadas por A: TS_SSR03 y B: TS_SSR05. L: marcador de peso molecular de 25pb-500pb (Invitrogen).

Figura 11. Gel de electroforesis y electroferogramas de los productos de PCR de los individuos amplificados con el marcador TS_SSR07.



Gel brindado por el software del equipo Qiaxcel. Las *Taenias* son identificadas por sus números ID, genotipificadas por marcador TS_SSR07. A: Imagen del gel que muestra los amplificados de 7 muestras. L: marcador de peso molecular de 25pb-500pb (Invitrogen). B: electroferograma de la muestra de Puno que muestra la presencia de dos picos bien definidos. C: electroferograma de la muestra de Tumbes en el cual no se observa ningún pico. Las bandas de color verde corresponden al marcador de alineamiento interno. Las supuestas bandas que aparecen en ambas muestras corresponden a inespecificidades del PCR pues no coinciden con los tamaños esperados.

6.4. Ensayo de reproducibilidad:

Todas las cuatro repeticiones dieron los mismos resultados. El coeficiente de variación fue menor de 0.7%. (Tabla 06).

Tabla 06. Ensayo de reproducibilidad

Muestra	Repetición				CV*(%)
	1	2	3	4	
2	217	217	217	216	0.230681
	222	222	222	221	0.225479
3	212	212	212	213	0.235571
	229	228	228	231	0.617561
4	207	207	207	208	0.241255
5	207	207	207	--	0
6	207	207	207	207	0
7	207	207	207	207	0
8	207	207	207	207	0
9	217	217	217	217	0
	222	221	221	222	0.260655
10	212	212	212	212	0
	222	222	222	222	0

CV*: coeficiente de variación.

Las bandas fueron obtenidas a partir de 4 diferentes PCR usando el marcador TS_SSR01 en las muestras 2 -10. La electroforesis fue realizada en el equipo avanzado Qiaxcel.

6.5. Análisis de la diversidad genética dentro de las subpoblaciones de *T. solium*

Del total de individuos inicialmente genotificados (156) se eliminaron 3 individuos pertenecientes a las subpoblaciones 1, 3 y 4 de la provincia de Tumbes por presentar ausencia de bandas en el 60% de los marcadores polimórficos, es decir, de los siete marcadores polimórficos sólo amplificaron tres marcadores en estos individuos.

Se identificaron un total de 58 alelos presentes en la nueva población total. El número de alelos varió entre 5 alelos para el locus TS_SSR01 y 11 alelos para el locus TS_SSR32 con un valor promedio de 7 alelos por locus.

La heterocigosidad observada (H_o) varió entre 0 para los marcadores TS_SSR16, TS_SSR18, TS_SSR28 y 0.274 para el marcador TS_SSR32, mientras que la heterocigosidad esperada (H_E) varió entre 0.46 para el marcador TS_SSR01 y 0.84 para el marcador TS_SSR28.

El contenido de información polimórfica (PIC) varió entre 0.4305 para el locus TS_SSR01 y 0.821 para el marcador TS_SSR28 con un valor promedio de 0.646 por locus (Tabla 07). La probabilidad de exclusión (PE2) varió entre 0.314 para el locus TS_SSR_28 y 0.7332 para el locus TS_SSR_01. La probabilidad de identidad varió entre 0.044 para el locus TS_SSR_28 y 0.3211 para el locus TS_SSR_01.

El número de alelos, heterocigosidad observada, heterocigosidad esperada, contenido de información polimórfica, probabilidad de identidad y de exclusión para cada loci se detallan en la Tabla 07.

Tabla 07. Número de alelos, heterocigosidad observada (H_o), heterocigosidad esperada (H_E), contenido de información polimórfica (PIC), Probabilidad de exclusión (PE) en la población total de *Taenias* peruanas (N=153)

Locus	N° alelos	H_o	H_e	PIC	PE1	PE2	PE3	P(ID)
TS_SSR_01	5	0.2026	0.4618	0.4304	0.8859	0.7332	0.5671	0.3211
TS_SSR_09	7	0.0130	0.6257	0.5887	0.7727	0.5943	0.3979	0.1766
TS_SSR_16	8	0	0.6804	0.6419	0.7241	0.5428	0.3466	0.1398
TS_SSR_18	10	0	0.7981	0.7676	0.5707	0.3935	0.2082	0.0696
TS_SSR_27	7	0.0130	0.7010	0.6478	0.7129	0.5441	0.3602	0.1415
TS_SSR_28	10	0	0.8422	0.8210	0.482	0.3147	0.1408	0.0442
TS_SSR_32	11	0.2745	0.5221	0.4952	0.8435	0.6726	0.48	0.2554

PE1: Probabilidad de exclusión para un candidato a pariente.
PE2: Probabilidad de exclusión para candidato a pariente dado el genotipo de un pariente conocido del sexo opuesto.
PE3: Probabilidad de exclusión para un par de candidatos a parientes.
P_(ID): Probabilidad de identidad

Todos los marcadores microsatélites mostraron una desviación significativa del equilibrio de Hardy-Weinberg (HWE) debido a un déficit de heterocigotos ($p < 0.01$). Sólo se encontró presencia de desequilibrio de ligamiento ($p < 0.01$) entre los pares de loci TS_SSR09 y TS_SSR18 cuando se analizaron en la población total.

Todas las subpoblaciones presentaron un valor de F_{IS} , F_{IT} y F_{ST} igual a 0.8704 ($Rho_{IS} = 0.7234$), 0.8882 ($Rho_{IT} = 0.7952$) y 0.1374 ($Rho_{ST} = 0.1374$) respectivamente (Tabla 08) con una frecuencia de alelos privados de 0.063.

El menor número de alelos promedio (2.25) se encontró en la subpoblación 2 de Tumbes, mientras que la población 4 de Puno presentó el mayor número de alelos promedio (5.2).

El valor mínimo de heterocigosidad esperada fue calculada para la población 2 de Tumbes (0.27937) y el valor máximo fue para la población 4 de Puno (0.7365). Los principales indicadores de variabilidad genética para cada subpoblación analizada están detallados en la Tabla 09.

Las diferentes subpoblaciones mostraron tener diferentes distribuciones alélicas (diferenciación génica) y genotípicas (diferenciación genotípica) cuando fueron comparadas entre pares para cada uno de los microsatélites ($p < 0.01$).

Tabla 08. Valores FIS, FST, FIT (Weir y Cockerham, 1984) y RhoIS, RhoST , RhoIT (Michalakis y Excoffier, 1996) para cada locus microsatélite analizado.

Locus	Fis	Fst	Fit	Rhois	Rhost	Rhoit
TS_SSR_01	0.4177	0.2447	0.5477	0.3217	0.3785	0.5784
TS_SSR_09	0.9768	0.0240	0.9788	0.9720	0.1536	0.9763
TS_SSR_16	1	0.1719	1	1	0.3082	1
TS_SSR_18	1	0.1975	1	1	0.3750	1
TS_SSR_27	0.9737	0.1639	0.9794	0.9387	0.2729	0.9554
TS_SSR_28	1	0.0854	1	1	0.1654	1
TS_SSR_32	0.4205	0.0908	0.4731	-0.1548	0.2845	0.1738
Total	0.8704	0.1374	0.8882	0.7234	0.2596	0.7952

Tabla 09. Número promedio de alelos, heterocigosidad observada (HO), heterocigosidad esperada (HE) e índice de Garza-Williamson (G-W) en las 13 subpoblaciones de *Taenias*.

Población	N° individuos	N° promedio de alelos	Ho	He	G-W
Cajamarca	7	2.714	0.1667	0.3181	0.82143
pop_Piura 1	8	3.714	0.02679	0.40970	0.9000
Pop_Piura 2	28	2.571	0.02778	0.48576	0.95833
Otras poblaciones	8	4	0.07143	0.68929	0.80952
subpop_Puno_1	8	4.286	0.19643	0.60519	0.74069
subpop_Puno_2	4	3.143	0.17857	0.68367	0.66587
subpop_Puno_3	7	3.857	0.16327	0.66719	0.70873
subpop_Puno_4	14	5.286	0.1978	0.73657	0.80998
subpop_Puno_5	20	5	0.13271	0.66664	0.89683
subpop_Tumbes_1	24	5.143	0.02461	0.5875	0.76429
subpop_Tumbes_2	5	2.25	0	0.27937	1
subpop_Tumbes_3	7	2.6	0	0.39101	0.80833
subpop_Tumbes_4	16	4	0.02941	0.48863	0.77718
subpop_Tumbes_5	5	2.429	0	0.57143	0.96429

6.6. Análisis de la diversidad entre subpoblaciones de *T. solium*

Las subpoblaciones 2, 3 y 5 de la provincia de Tumbes presentaron el mayor valor de endogamia ($F_{IS} = 0.9484$) y la subpoblación 2 de la provincia de Puno presentó el mayor valor de diferenciación genética con respecto a la población total ($F_{ST} = 0.773$). Los valores de F_{IS} y F_{ST} para todas las subpoblaciones se aprecian en la Tabla 10.

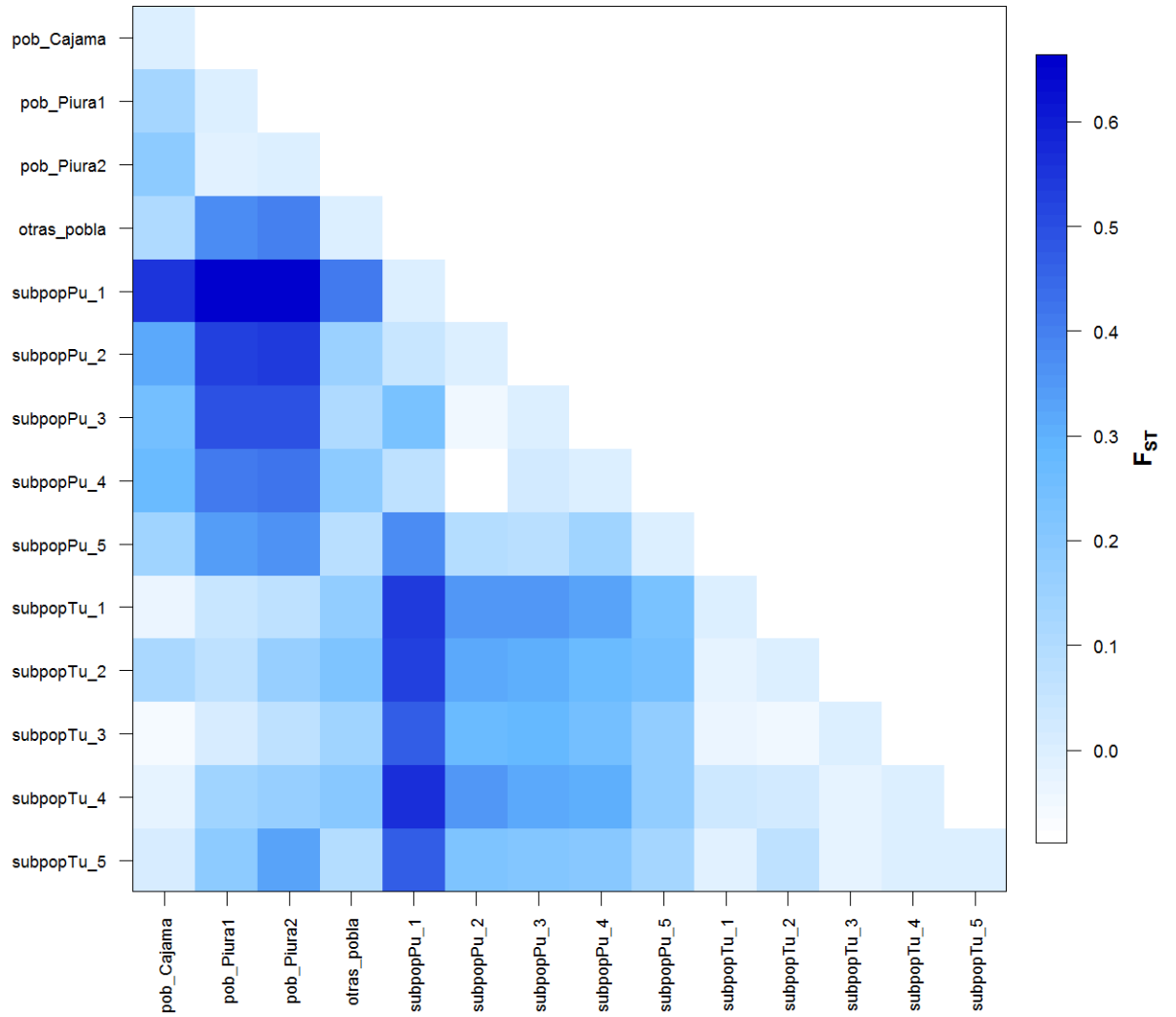
La matriz pareada de valores de F_{ST} y R_{ST} para cada una de las subpoblaciones mostró resultados similares entre estos dos indicadores y valores estadísticamente significativos para algunas subpoblaciones ($p < 0.01$). Los valores más elevados de diferenciación genética se encontraron entre las subpoblaciones de Tumbes y las de Puno, así como entre las subpoblaciones de Piura y las de Puno, tal como se aprecia en la Tablas 11 y 12 y en la Figura 12.

Tabla 10. Valores de F_{IS} y F_{ST} calculada para cada una de las 12 subpoblaciones de *T. solium*

	Fst	Fis
pob_Cajamarca	0.4735	0.904
pop_Piura1	0.4222	0.936
pop_Piura2	0.4328	0.9436
Otras poblaciones	0.7334	0.9026
subpopPuno_1	0.6468	0.6851
subpopPuno_2	0.7679	0.7674
subpopPuno_3	0.7092	0.7698
subpopPuno_4	0.7578	0.7418
subpopPuno_5	0.6817	0.8073
subpop_Tumbes_1	0.6002	0.9594
subpop_Tumbes_2	0.3143	1
subpop_Tumbes_3	0.4281	1
subpop_Tumbes_4	0.5051	0.9484
subpop_Tumbes_5	0.6515	1

La matriz pareada del número de migrantes entre las catorce subpoblaciones mostró que la subpoblación 2 de Piura y la subpoblación 1 de Puno presentan el menor número de migrantes con un valor de 0.1328, mientras que el mayor número de migrantes estuvo entre las subpoblaciones 4 y 5 de Tumbes con un valor de 111.361 tal como se aprecia en la Tabla 13.

Figura 12. Matriz pareada en base a los valores de Fst



La matriz es calculada entre cada par de subpoblaciones de *T. solium*. A la derecha la escala de valores de F_{ST} . A mayor valor de F_{ST} mayor intensidad del color.

Tabla 11. Matriz pareada con los valores de F_{ST} usando el programa Arlequin v3.1

	Cajamarc a	Piura_1	Piura_2	Otras poblac	Puno_1	Puno_2	Puno_3	Puno_4	Puno_5	Tumbes_ 1	Tumbes_ 2	Tumbes_ 3	Tumbes_ 4
Piura_1	0.19312*	0											
Piura_2	0.1586	-0.0162	0										
Otras poblac	0.11669	0.22535 *	0.32736 *	0									
Puno_1	0.19851*	0.21746 *	0.32138 *	0.13613 *	0								
Puno_2	0.18488	0.31569 *	0.32298 *	0.04747	0.02541	0							
Puno_3	0.26932*	0.37448 *	0.37542 *	0.11851	0.12444	0.01061	0						
Puno_4	0.18299*	0.30441 *	0.30391 *	0.07471	0.03223	-0.0006	0.03105	0					
Puno_5	0.13515	0.23068 *	0.21742 *	0.05697	0.04532	0.04536	0.1238*	0.05228	0				
Tumbes_ 1	0.09899	0.16652 *	0.10574 *	0.10162	0.19386 *	0.17809 *	0.22293 *	0.16264 *	0.09306 *	0			
Tumbes_ 2	0.11075	0.37874 *	0.28284 *	0.2587*	0.3185*	0.3696*	0.39094 *	0.2908*	0.22589 *	0.11059	0		
Tumbes_ 3	0.02819	0.18429	0.11382	0.1415	0.25967 *	0.26532 *	0.3371*	0.24239 *	0.14535 *	0.03006	0.14152	0	
Tumbes_ 4	0.06449	0.22869 *	0.14818 *	0.15876 *	0.20884 *	0.248*	0.30746 *	0.21508 *	0.10903 *	0.0279	0.08553	0.06428	0
Tumbes_ 5	0.0818	0.22107 *	0.21615	0.07816	0.17385	0.1413	0.20364 *	0.11692	0.09345	0.07869	0.24554	0.18011	0.13606

Las diferencias son significativas ($p < 0.01$) para el caso de los valores marcados con asteriscos (*). Los otros valores de F_{ST} a pesar de ser bastante altos no logran rechazar la hipótesis nula de diferenciación igual a cero. Los valores negativos son equivalentes a cero, es decir el grado de diferenciación es nula entre estas dos subpoblaciones.

Tabla 12. Matriz pareada con los valores de R_{ST} usando el programa Arlequin v3.

	Cajamarc a	Piura_1	Piura_2	Otras poblac	Puno_1	Puno_2	Puno_3	Puno_4	Puno_5	Tumbes_ 1	Tumbes_ 2	Tumbes_ 3	Tumbes_ 4
Piura_1	0.12693												
Piura_2	0.19209	-0.0162											
Otras poblac	0.10435	0.37223 *	0.39749 *										
Puno_1	0.55096*	0.65298 *	0.66346 *	0.40635 *									
Puno_2	0.31389	0.52675 *	0.54398 *	0.14644	0.05037								
Puno_3	0.24106	0.49443 *	0.49227 *	0.10943	0.2298	0.04881							
Puno_4	0.27133	0.40868 *	0.41797 *	0.18209	0.07526	0.08896	0.02292						
Puno_5	0.13515	0.34545 *	0.36414 *	0.07595	0.37441 *	0.09847	0.08657	0.14443 *					
Tumbes_ 1	-0.03109	0.05016	0.06765	0.17447	0.54017 *	0.35725	0.35661 *	0.32351 *	0.23128 *				
Tumbes_ 2	0.1217	0.06583	0.16761	0.2248	0.52329 *	0.32164	0.30394	0.26722	0.24188 *	-0.02655			
Tumbes_ 3	-0.03241	0.00638	0.06999	0.13478	0.46985 *	0.27196	0.28049 *	0.25113 *	0.18076	-0.03606	-0.04285		
Tumbes_ 4	-0.02646	0.14493 *	0.15871 *	0.19461	0.56565 *	0.35635 *	0.31784 *	0.30654 *	0.17322 *	0.03037	0.01965	-0.02255	
Tumbes_ 5	0.01503	0.19228	0.32605 *	0.08999	0.47105 *	0.22196	0.20907	0.20056	0.12667	-0.01131	0.07262	-0.02192	-0.00225

Las diferencias son significativas ($p < 0.01$) para el caso de los valores marcados con asteriscos (*). Los otros valores de R_{ST} a pesar de ser bastante altos no logran rechazar la hipótesis nula de diferenciación igual a cero. Los valores negativos son equivalentes a cero, es decir el grado de diferenciación es nula entre estas dos subpoblaciones.

Tabla 13. Matriz pareada del número de migrantes¹ calculado entre las catorce subpoblaciones de *T.solium*

	Cajamarca	Piura_1	Piura_2	Otras poblac	Puno_1	Puno_2	Puno_3	Puno_4	Puno_5	Tumbes_1	Tumbes_2	Tumbes_3	Tumbes_4	Tumbes_5
Piura_1	1.7196													
Piura_2	1.0515	15.7299												
Otras poblac	2.1458	0.4216	0.3789											
Puno_1	0.2038	0.1329	0.1268	0.3652										
Puno_2	0.5465	0.2246	0.2096	1.4572	4.7133									
Puno_3	0.7871	0.2556	0.2579	2.0346	0.8379	5.3719								
Puno_4	0.6714	0.3617	0.3481	1.1229	3.0718	3.0603	10.6575							
Puno_5	1.5998	0.4737	0.4365	3.0416	0.4177	2.2888	2.637836	1.4809						
Tumbes_1	8.2912	4.7341	3.4455	1.1829	0.2128	0.4498	0.451046	0.5228	0.8309					
Tumbes_2	1.8042	3.5477	1.2416	0.8621	0.2277	0.5273	0.572531	0.6856	0.7836	9.6662				
Tumbes_3	4.2558	38.9350	3.3219	1.6049	0.2821	0.6693	0.641297	0.7455	1.1330	7.1829	6.0843			
Tumbes_4	9.6982	1.4750	1.3252	1.0346	0.1920	0.4516	0.536559	0.5656	1.1933	7.9818	12.4726	11.3365		
Tumbes_5	16.3834	1.0502	0.5168	2.5281	0.2807	0.8763	0.945772	0.9965	1.7236	22.3543	3.1926	11.6551	111.3	

1. Wright (1969) [$N_m = (1 - F_{ST}) / 4 F_{ST}$]

Tabla 14. Matriz pareada de distancias genéticas D_A encima de la diagonal y D_S debajo de la diagonal entre las catorce subpoblaciones de *T.solium*

	Cajamarca	Piura_1	Piura_2	Otras poblac	Puno_1	Puno_2	Puno_3	Puno_4	Puno_5	Tumbes_1	Tumbes_2	Tumbes_3	Tumbes_4	Tumbes_5
Cajamarca		0.2157	0.1439	0.2743	0.3865	0.3919	0.5028	0.3703	0.2299	0.2072	0.1252	0.1781	0.1587	0.2327
Piura_1	0.1914		0.1153	0.3274	0.5087	0.5149	0.6097	0.4644	0.3523	0.1835	0.2810	0.2153	0.1870	0.2984
Piura_2	0.0300	0.1104		0.3069	0.5101	0.5247	0.6328	0.4718	0.3223	0.1593	0.1815	0.1803	0.1271	0.2291
Otras poblac	0.2386	0.3862	0.2983		0.3172	0.2957	0.2992	0.2204	0.1440	0.2367	0.3655	0.3616	0.2826	0.2927
Puno_1	0.4097	0.8356	0.6057	0.4467		0.2091	0.2668	0.1692	0.1867	0.3958	0.4687	0.4594	0.4360	0.4035
Puno_2	0.3235	0.6509	0.5420	0.2043	0.0541		0.1746	0.2098	0.2242	0.4028	0.5146	0.4810	0.4787	0.3888
Puno_3	0.6210	1.0136	0.9410	0.3565	0.3030	-0.0063		0.1589	0.249	0.4247	0.5664	0.5989	0.5347	0.4612
Puno_4	0.4100	0.7911	0.6259	0.2273	0.0870	-0.0138	0.0506		0.111	0.2800	0.4535	0.4603	0.3652	0.3479
Puno_5	0.2194	0.5301	0.3297	0.2059	0.0822	0.0846	0.2933	0.1085		0.1933	0.3334	0.3431	0.2267	0.2396
Tumbes_1	0.0772	0.2609	0.0986	0.2369	0.4096	0.3377	0.5145	0.3451	0.1593		0.1947	0.1866	0.1190	0.1597
Tumbes_2	0.0306	0.3672	0.1109	0.3702	0.5264	0.5260	0.7621	0.5063	0.3524	0.0982		0.1198	0.1446	0.2703
Tumbes_3	0.0191	0.1881	0.0601	0.2601	0.4336	0.3890	0.7343	0.4738	0.2513	0.0491	0.0420		0.1920	0.3259
Tumbes_4	0.0524	0.2367	0.0543	0.3103	0.5286	0.5529	0.8519	0.5399	0.2693	0.0702	0.0536	0.0766		0.2328
Tumbes_5	0.1689	0.4016	0.2234	0.2575	0.3448	0.2395	0.4622	0.2249	0.1566	0.1166	0.2575	0.2230	0.2854	0

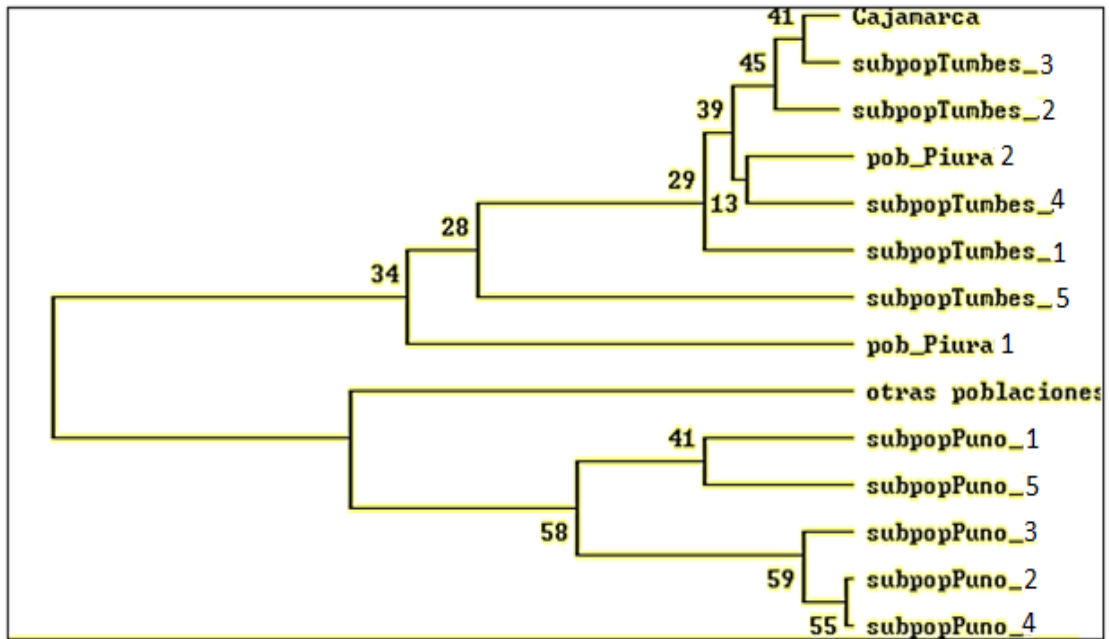
DS: Distancia genética estándar de Nei (Nei y col., 1972).

DA: Distancia genética de Nei (Nei y col., 1983).

La distancia genética entre las subpoblaciones determinada mediante la distancia genética estándar de Nei (D_S) muestra una mayor distancia genética entre las subpoblaciones 1 de Piura y 3 de Puno con un valor de 1.0136 y la menor distancia genética entre las subpoblaciones 2 y 3 de Puno con un valor de -0.0063. La distancia genética de Nei (D_A) indica una mayor distancia genética entre las subpoblaciones 3 de Puno y 2 de Piura mientras que la menor distancia se encuentra entre las subpoblaciones 1 y 4 de Tumbes con un valor de 0.1190. La lista de valores de distancias genéticas se encuentra detallada en la Tabla 14.

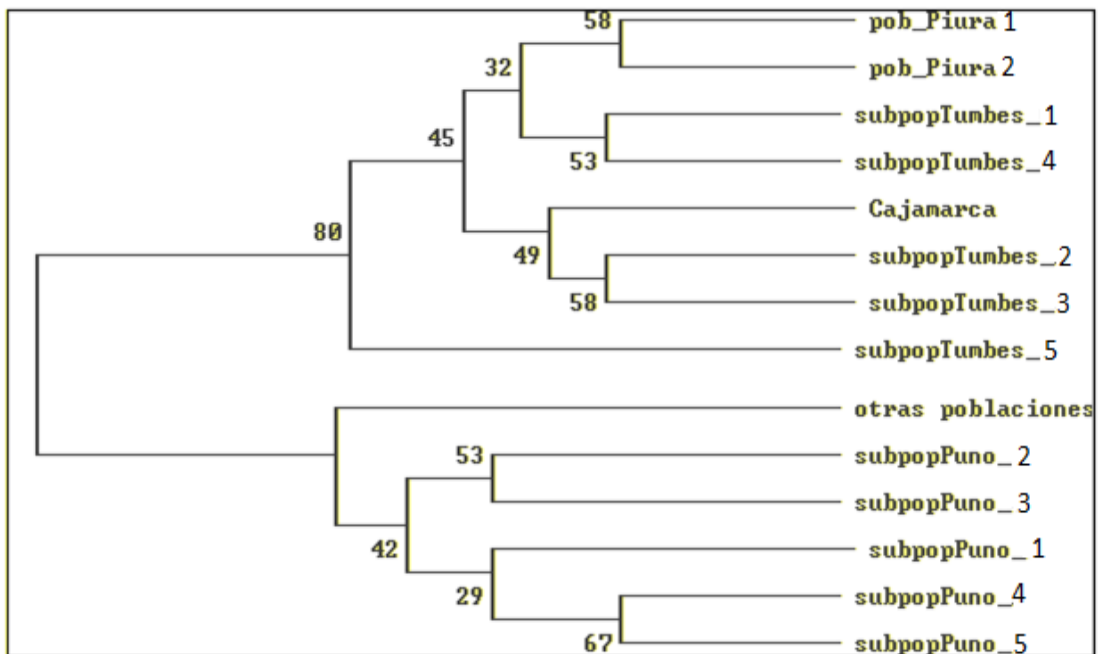
Las 14 subpoblaciones de *Taenias* fueron agrupadas en 3 grupos genéticos mediante la topología de los dendrogramas generados en base a D_A y D_S utilizando el método de Neighbour-joining (NJ) y Unweighted Pair Group Method with Arithmetic Mean (UPGMA). Los tres grupos genéticos identificados corresponden a: Norte (Tumbes, Piura y Cajamarca), Sur (Puno_1, Puno_2, Puno_3, Puno_4 y Puno_5) y en la parte central se ubican las otras poblaciones (Figuras 13, 14, 15 y 16). Sin embargo, la robustez del dendrograma no asegura la validez de los grupos identificados.

Figura 13. Dendrograma generado a partir del método de UPGMA en base a la distancia genética estándar (Ds)



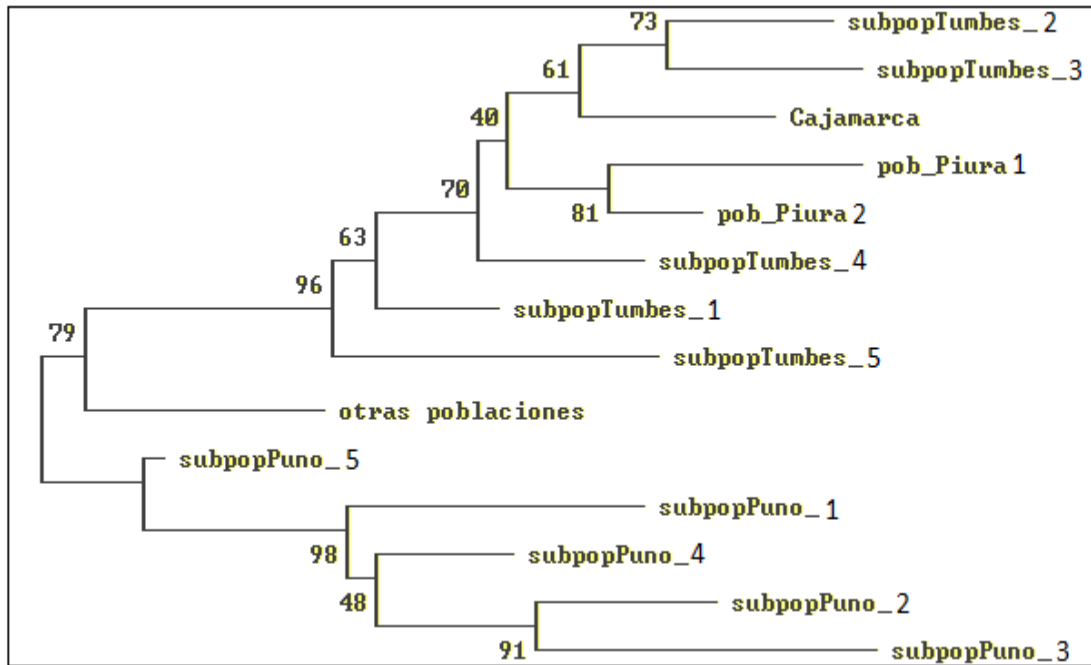
El dendrograma es obtenido usando el programa Dispan. Los números de los nodos indican el porcentaje de ocurrencia por cada 1000 bootstraps.

Figura 14. Dendrograma generado a partir del método de UPGMA en base a la distancia genética de Nei (Da)



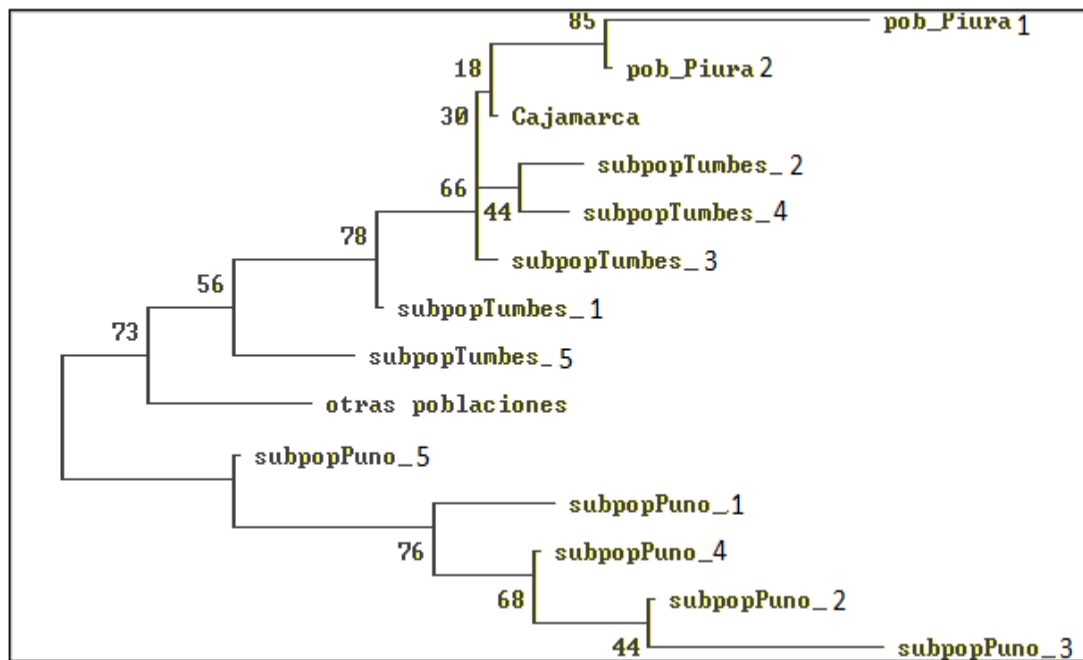
El dendrograma es obtenido usando el programa Dispan. Los números de los nodos indican el porcentaje de ocurrencia por cada 1000 bootstraps.

Figura 15. Dendrograma generado a partir del método de Neighbor joining (NJ) calculado en base a la distancia genética de Nei (Da)



El dendrograma es obtenido usando el programa Dispan. Los números de los nodos indican el porcentaje de ocurrencia por cada 1000 bootstraps.

Figura 16. Dendrograma generado a partir del método de Neighbor joining (NJ) calculado en base a la distancia genética estandar (Ds)



El dendrograma es obtenido usando el programa Dispan. Los números de los nodos indican el porcentaje de ocurrencia por cada 1000 bootstraps.

6.7. Análisis de estructura poblacional

El análisis de estructura mostró la presencia de tres grupos genéticos [K=3, Pr(X/K)=1.0] (Tabla 15) inferidos a partir de los datos de las 14 subpoblaciones de *Taenias*.

El análisis de estructura (K=3) muestra un patrón de ancestría específico (Grupo 1 y 2) para todas las subpoblaciones de Tumbes (subpoblacion1, 2, 3, 4 y 5) Piura (subpoblacion1 y 2) y Cajamarca. Las subpoblaciones de Puno muestran también un patrón de ancestría específico pero para el Grupo 3 (Tabla 16).

La subpoblación que agrupa otras poblaciones mostró un coeficiente de ancestría mixto (Grupo 1, 2 y 3). La proporción estimada de contribución de cada grupo en cada una de las siete poblaciones de *Taenias* se detalla en la Tabla 16.

Tabla 15. Inferencia del valor de K (número de poblaciones) a partir de los 7 loci microsatélites en la población total de *Taenias*.

K	ln Pr(X/K) ¹	Pr(K) ²
1	-3024.8	4.036x10 ⁻²⁶⁹
2	-2575.1	8.095x10 ⁻⁷⁴
3	-2406.8	1.000
4	-2494.4	9.032x10 ⁻³⁹
5	-2556.1	1.445x10 ⁻⁶⁵

¹ Ln de la distribución de probabilidades de los genotipos individuales (X) dado un número de poblaciones inferidas (K) obtenidos mediante el programa Structure v2.2.

² Distribución de probabilidad de K dados los genotipos individuales (X) determinado mediante regla bayesiana (Pritchard y col., 2000).

Tabla 16. Proporción estimada de contribución de cada grupo genético inferido (K=3) en las 14 subpoblaciones de *Taenias*

	Grupo1	Grupo 2	Grupo 3	Nº de individuos
Cajamarca	0.278	0.677	0.044	7
otras poblaciones	0.252	0.237	0.511	8
Piura_1	0.080	0.920	0.000	17
Piura_2	0.003	0.997	0.000	11
Tumbes_1	0.351	0.507	0.142	24
Tumbes_2	0.036	0.964	0.000	5
Tumbes_3	0.135	0.747	0.118	6
Tumbes_4	0.375	0.624	0.001	17
Tumbes_5	0.401	0.406	0.193	5
Puno_1	0.054	0.000	0.947	8
Puno_2	0.002	0.000	0.998	4
Puno_3	0.047	0.000	0.953	7
Puno_4	0.195	0.000	0.804	14
Puno_5	0.429	0.103	0.468	20

6.8. Análisis de varianza molecular (AMOVA)

Teniendo en cuenta los resultados brindados por la estadística F y dado que las subpoblaciones no se encontraron en equilibrio de Hardy-Weinberg, se realizó una

prueba de AMOVA empleando el programa Arlequin v3.1. Toda la población fue dividida en los dos grandes grupos (población Norte y Sur). Los resultados obtenidos para este estudio indicaron que la mayor variación (83.62%) se encuentra dentro de las subpoblaciones y que existe una variación moderada (16%) entre los grupos del norte y sur del país (Tabla 17).

Tabla 17. Promedio de los resultados del AMOVA para los 7 loci analizados en la población total de *T.solium* .

Fuente de variación	Suma de cuadrados	Componentes de varianza	Porcentajes de variación
Entre poblaciones	56.276	0.41329	16.37785
Dentro de las poblaciones	600.599	2.11017	83.62215
Total	656.875	2.52346	

El test de Mantel mostró una correlación positiva entre la distancia geográfica y la distancia genética para todas las subpoblaciones combinadas (distancia genética por aislamiento geográfico = 0.43 indicando que el 43% de la distancia genética fue determinada por la distancia geográfica entre las subpoblaciones, coeficiente de correlación = 0.65; $p = 0.00000$). Una correlación positiva significativa indica que la distancia genética se incrementa a medida que la distancia geográfica aumenta.

VII. DISCUSION

Los marcadores moleculares son esenciales para la investigación de la variabilidad genética y el entendimiento de la dinámica de los genomas. En el caso de *T. solium*, el desarrollo de marcadores moleculares ha sido bastante lento y se ha centrado básicamente en marcadores mitocondriales y sólo algunos marcadores de ADN. Los microsatélites o SSRs han probado ser marcadores muy útiles en estudios de genética de poblaciones de muchos parásitos (Shrivastava et al., 2005; Xiao et al., 2011); sin embargo, la principal limitante es su identificación en el genoma, dado que se realiza principalmente a partir de librerías genómicas. Esta metodología es muy costosa, laboriosa y extensa. Hoy en día, se plantea como estrategia alternativa la minería de datos (*data mining*) a partir de secuencias anotadas, ESTs o genomas completos, lo cual supera las limitaciones antes mencionadas (Sharma et al., 2007).

Con la disponibilidad del genoma de *T. solium* y usando herramientas bioinformáticas, se pudo en este estudio identificar secuencias microsatélites en el genoma y diseñar cebadores en las regiones flanqueantes conservadas, para su amplificación y posterior análisis. Además, se pudo realizar una comparación *in silico* de dos genomas de *T. solium*, uno a partir del estadio adulto y otro del estadio cisticerco encontrándose diferencias en la longitud de las secuencias microsatélites.

En este estudio, por primera vez, se evidencia la variabilidad genética de *T. solium* en el Perú utilizando microsatélites como marcadores moleculares. Los estudios previos

habían demostrado una baja variabilidad genética en este parásito inclusive analizando muestras alrededor del mundo (Gasser, R.B., 1995; Nakao, M. 2002; Ito, A. 2003). Anteriormente, se habían evaluado 15 *Taenias* adultas provenientes de diferentes localidades del Perú con marcadores mitocondriales pero no se encontró ningún tipo de polimorfismo (Hancock *et al.*, 2001). Recientemente, se ha evaluado la diversidad genética de varios linajes locales en México, los cuales plantean que *T. solium* podría tener un tipo de reproducción sexual lo cual explicaría el grado de diversidad genética encontrado (Maravilla P. 2003, Vega R. 2003, Maravilla P. 2008, Bobes, R.J. 2010).

Las secuencias microsatélites son abundantes y se encuentran distribuidas a lo largo de todo el genoma de *T. solium*. Si bien es cierto, la frecuencia de los microsatélites y su cobertura relativa en el genoma varía considerablemente entre organismos (Sharma *et al.*, 2007), el número de secuencias encontrado en *T. solium* coincide con la cantidad de SSRs identificada en otros parásitos como *Caenorhabditis elegans*, *Trypanosoma cruzi*, *Plasmodium falciparum* (Sharma *et al.*, 2007). En este último organismo, el número de secuencias microsatélites fue excepcionalmente elevado. El número elevado de secuencias SSRs se justifica puesto que se cree desempeñan un rol importante en la evolución de los genomas, creando y manteniendo cuantitativamente la variación genética (Tautz *et al.* 1986; Kashi *et al.* 1997).

En *T. solium*, los SSRs fueron identificados tanto en regiones codificantes como no codificantes, siendo estas últimas las que contenían mayor cantidad de secuencias

microsatélites. Nuestros resultados son consistentes con los estudios previos los cuales reportan que en organismos eucariontes, los SSRs se encuentran más abundantes en las regiones no codificantes comparadas con las distribuciones al azar (Metzgar et al., 2000). La relativa baja frecuencia de SSRs en regiones codificantes puede ser explicada por una selección negativa en contra de las mutaciones que cambian el marco de lectura en las secuencias que se transcriben ([Metzgar et al., 2000](#) y [Li et al., 2004](#)). Además, los microsatélites identificados en las regiones no codificantes podrían brindar mayor variabilidad y eventualmente podrían ayudar a diferenciar parásitos provenientes de localidades cercanas. Esto ya ha sido corroborado en otros organismos como *C. elegans* (Haber, M. 2005), en el cual empleando microsatélites fue posible identificar una alta variabilidad genética no reportada previamente.

Un locus microsatélite típicamente varía entre 5 y 40 repeticiones, pero son posibles cadenas más largas también. En ambos genomas, los microsatélites muestran una disminución en su número a medida que se aumenta el tipo de repetición (Temnykh, S. et al. 2001, Grover, A. et al. 2007). Nuestros resultados coincidieron totalmente con ello puesto que los SSRs con motivos más pequeños, mononucleótidos, dinucleótidos y trinucleótidos, estuvieron en mayor porcentaje con respecto a los tetra-, penta- y hexa- nucleótidos. Los motivos de repetición de dinucleótidos fueron los más abundantes en *T.solium* entre todas las repeticiones. Este resultado no coincide con resultados reportados en otros parásitos como *Caenorhabditis elegans*, *Trypanosoma cruzi* y *Plasmodium falciparum* (Sharma et al., 2007), los cuales

indican que son los microsatélites mononucleotídicos los más abundantes. Sin embargo, coincide con otras investigaciones que reportan que en organismos eucariontes, los SSRs dinucleotídicos son los más abundantes (Li, Y.C. 2002; Karaoglu et al., 2005). Esta divergencia se debería a las herramientas y criterios establecidos en la búsqueda de los microsatélites en el genoma de *T.solium*, puesto que se tomaron motivos mayores de 5pb de longitud (Hamarshah O. y Amor A., 2011). Además, los motivos diméricos largos son inestables en las secuencias de expresión, mientras que en las regiones no codificantes las repeticiones dinucleotídicas pueden ser de larga longitud, probablemente debido a la alta tolerancia del ADN no codificante hacia las mutaciones y a la mayor probabilidad de producirse por el mecanismo del resbalón de la polimerasa durante la replicación (Dokholyan et al., 2000).

El motivo de dinucleótidos AC/GT fue el más abundante en el genoma de *T. solium*. Este tipo de repetición es muy frecuente en primates, roedores, mamíferos, vertebrados y hongos. En el caso de parásitos, el estudio realizado por Castagnone-Sereno et al. 2010 analizando genomas de distintos nemátodos reveló que (AT)_n es el tipo de repetición más común. Sin embargo, en el caso de céstodes no se han analizado ni secuenciado muchos genomas, motivo por el cual no se puede hacer una interpretación fehaciente y especular sobre algún tipo de significancia biológica. Adicionalmente, es importante recalcar que los patrones de distribución de los diferentes tipos de repetición (desde mono- hasta hexanucleótidos) son específicos de

cada taxón y esto se debería a la interacción de los mecanismos de mutación y la selección diferencial (Li et al., 2002).

Los microsatélites candidatos para la evaluación del polimorfismo fueron elegidos teniendo en cuenta en lo posible, la longitud del motivo y el número de repeticiones. Se sugiere como una estrategia seleccionar motivos di- y tri- nucleótidos y al menos 12 repeticiones para encontrar variabilidad (Haber M. 2005). El tamaño del motivo y el número de repeticiones de los microsatélites probados fueron diversos desde 2 hasta 5 nucleótidos. Muchos de los dinucleótidos no fueron polimórficos, a pesar que ellos tuvieron un alto número de repeticiones (16-19). Por otro lado, a pesar del tamaño grande del motivo de los marcadores TS_SSR02, TS_SSR03, TS_SSR04 y TS_SSR05, ellos tuvieron pocas repeticiones y fueron monomórficos. Sivasundar encontró baja variabilidad a nivel mundial en *Caenorhabditis elegans* cuando empleó microsatélites con bajo número de repeticiones (Sivasundar, A. 2003). TS_SSR06 y TS_SSR07 tienen motivos tetranucleótidos y tuvieron un alto número de repeticiones (23 repeticiones); sin embargo, no fueron empleados por presentar problemas en la amplificación. El bajo nivel de polimorfismo es frecuente en individuos con altos niveles de endocruzamientos, bajo tamaño poblacional y cuellos de botella frecuentes o severos, además es común encontrar polimorfismo en microsatélites más cortos en la longitud del motivo de repetición, por ejemplo en los dinucleótidos (De Woody & Avise, 2000; Neff & Gross 2001).

De un total de 36 cebadores diseñados a partir de regiones microsátélites, 7 de ellos mostraron variación dentro de las subpoblaciones, es decir fueron polimórficos y tuvieron alto valores de índice de información polimórfica. Además, la mayoría de ellos mostraron tener altas probabilidades de exclusión (hasta 0.73) que indican la gran capacidad que tienen estos marcadores para asignar a los parentales a pesar de ser individuos con altos niveles de consanguinidad. Adicionalmente, mostraron bajos valores de probabilidades de identidad (hasta 0.32) con lo cual aumentan el poder de diferenciar entre dos individuos de una población. Un marcador (TS_SSR01) fue identificado a partir de secuencias codificantes (ESTs), esto indicaría una posible función de este marcador en el parásito. El conjunto de microsátélites permitieron diferenciar *Taenias* adultas provenientes de diferentes regiones del norte y sur del país. Nuestros resultados confirman de esta manera que los microsátélites son los marcadores ideales para caracterizar la estructura poblacional de *T. solium*.

Existen varios indicadores de variabilidad genética en poblaciones de parásitos. El número de alelos, la H_E y el coeficiente de consanguinidad (Fis) constituyen los principales indicadores de la variabilidad genética intrapoblacional. El análisis de las 14 subpoblaciones de *Taenias* provenientes de Tumbes, Piura, Huancayo, Ayacucho, Ancash y Puno reveló altos niveles de heterocigosidad esperada (0.73) y variabilidad presente a nivel intra- e inter- poblacional contradiciendo los bajos niveles de variación genética frecuentemente reportados para esta especie. Esta diferencia se debería básicamente al tipo de marcador empleado, en los estudios previos se emplearon RAPDs los cuales no permiten evidenciar la variación intrapoblacional o

sino también emplearon marcadores mitocondriales los cuales no sufren ningún proceso de recombinación. Por todo ello, los marcadores microsatélites son más idóneos para el análisis de la diversidad genética de una población.

A pesar que el número de individuos en muchas de las subpoblaciones fue bastante bajo ($n= 5$), la estimación de la heterocigocidad esperada es válida puesto que generalmente no está afectada por el tamaño de la muestra y aun teniendo pocos individuos se puede estimar este parámetro siempre y cuando se cuente con un gran número de loci (Gorman and Renzi, 1979). Es importante subrayar la presencia de individuos heterocigotos en algunos marcadores lo cual apoyaría la hipótesis de la posibilidad de reproducción sexual en el parásito diploide *T.solium*. Sin embargo, los bajos niveles de heterocigocidad observada para la mayoría de las subpoblaciones (Tabla 09) en concordancia con los altos niveles de consanguinidad ($F_{is} = 1$) y la gran diferencia entre la heterocigocidad esperada y observada se justifica más bien cuando el tipo de reproducción predominante en ese organismo es la autofertilización. Tal como sucede en *E.multilocularis*, en el cual los altos niveles de homocigocidad estarían asociados a una proliferación asexual de los protoescoleces en la fase larval (Knapp et al., 2007). Existen otros factores que podrían estar determinando el déficit de heterocigotos: 1) una subestructuración poblacional no detectada (efecto Walhund) o 2) un proceso de cuello de botella. Este último según los resultados obtenidos por el índice de G-W (valores cercanos a 1) no habría sucedido en las subpoblaciones mientras que la subestructuración poblacional es una posibilidad no descartada.

Entre todas las subpoblaciones de *T.solium*, fueron las de Puno las que mostraron valores más altos de heterocigosidad observada y esperada. Esto último indica que presentan una mayor diversidad genética en comparación con las demás subpoblaciones (Tabla 10). Una hipótesis que explicaría esto sería la tasa alta de migración de poblaciones de Bolivia hacia la región de Puno por el lago Titicaca, lo cual aumentaría el pool genético presente en esta zona y con ello la presencia de mayor cantidad de genotipos distintos en comparación con las subpoblaciones del norte. Otra explicación estaría centrada en las condiciones medioambientales presentes en esta región que favorecen la viabilidad del huevo cuando es expulsado al medio ambiente, puesto que se ha comprobado que las bajas temperaturas así como una humedad relativa alta favorecen la subsistencia del huevo de *T.solium* en el ambiente (Lawson & Gemmell, 1983; Coman 1975). Ambas condiciones son características de las provincias de Puno muestreadas en este estudio.

El alto nivel de consanguinidad de *T.solium* fue confirmado por la presencia de desviaciones del equilibrio de Hardy-Weinberg debido a una deficiencia significativa de heterocigotos ($p < 0.01$). Un exceso de homocigotos detectado puede ser el resultado de la presencia de alelos selectivos no-neutrales, o alelos nulos, o de una subestructuración poblacional no detectada [efecto Walhund (Hartl & Clark, 1997)]. Dado que es poco probable observar alelos nulos en todos los loci y las muestras, el efecto del apareamiento no al azar, la endogamia y la subestructuración poblacional

parecen ser los factores más trascendentes que estén influyendo en la deficiencia de heterocigotos.

El desequilibrio de ligamiento se encuentra de manera típica en los individuos de auto-fecundación o poblaciones endogámicas cuando el apareamiento no es al azar (Hartl & Clark 1997). Sin embargo, la mayoría de los marcadores microsatélites aquí evaluados, exceptuando el TS_SSR09 y TS_SSR16, no presentaron desequilibrio de ligamiento ($p < 0.01$). Esto último estaría sugiriendo eventos de recombinación que podrían ser una fuente menor de variabilidad genética en *T.solium* (Richard and Pâques 2000; Kelkar et al. 2008).

La estadística F, R y los estimadores de distancia genética han mostrado ser herramientas útiles para esclarecer los patrones y la extensión de la variación genética dentro y entre poblaciones. Los valores de F_{ST} y R_{ST} en cada una de las matrices mostraron valores de diferenciación genética similar para ambos indicadores, sin embargo no todas las diferencias entre las subpoblaciones fueron significativas, esto se debe a la fuerte influencia que tiene el tamaño de la muestra y la variabilidad de la misma en el cálculo del p-value. En nuestro caso, las muestras que mostraron cierto grado de diferenciación pero no significativo se debería al pequeño número de individuos por muestra. El valor promedio de F_{ST} entre las subpoblaciones mostraron que un 13.7% de la variación genética total puede ser explicada por diferencias a nivel poblacional mientras que un 86.3% de la variación genética total se debe a diferencias entre los individuos de cada población (Tabla 08). Sin embargo, la

variación intra poblacional podría estar siendo sobre estimada por el bajo número de individuos provenientes de cada subpoblación.

El valor de diferenciación genética moderada ($F_{st} = 0.1374$) es mayor al reportado en estudios en parásitos cercanos como *Echinococcus multilocularis* (Knapp et al., 2012) pero son congruentes con estudios poblacionales realizados en otros parásitos un poco más lejanos evolutivamente como *Schistosoma japonicum* (Shrivastava et al., 2005), *Plasmodium falciparum* (Larrañaga et al., 2013; **Abdi Khaireh et al., 2013**), *Trypanosoma brucei* (**Simo et al., 2009**), *Trypanosoma cruzi* (**Oliveira et al., 1998**).

Las medidas de distancia genética permiten evaluar la cantidad de variación compartida entre grupos. La topología de los dendrogramas generados mediante NJ y UPGMA en base a D_A y D_S (Figuras 13, 14, 15 y 16), así como los valores pareados de F_{ST} y R_{ST} (Tablas 11 y 12, Figura 12) mostraron dos grupos principales: i) Norte (Tumbes, Piura, Cajamarca), ii) Sur (Puno). Sin embargo, la asociación de las subpoblaciones en cada uno de los dendrogramas no fue exactamente la misma. Por ejemplo, en el caso de las dos subpoblaciones de Piura no se agrupan juntas en el dendrograma UPGMA basado en D_S , mientras que usando el método de NJ basado en D_S y D_A , ambas subpoblaciones se encuentran juntas. Estas diferencias se deberían al modelo de mutación de cada metodología. En el caso del UPGMA asume un modelo de reloj molecular mientras que el NJ no lo hace y si bien algunos marcadores SSRs han demostrado seguir este modelo, esto depende del organismo en estudio (Sun et al., 2009). Dado que en *T.solium* no se cuenta con este tipo de información, el

dendrograma usando el método de NJ sería el más adecuado. Además, entre los dos tipos de distancia, D_a y D_s , la primera es la más recomendada en estudios en los cuales el tamaño muestral es pequeño (Nei, 1973).

En el caso de la subpoblación denominada “otras poblaciones” mostró diferente ubicación en el dendrograma según el método de UPGMA y el de NJ. En el primer caso, se agrupó con las subpoblaciones del sur mientras que en el segundo se agrupa con las subpoblaciones del norte. (Figura 13, 14, 15 y 16). Teniendo en cuenta lo mencionado anteriormente acerca del método UPGMA, se tomarían los resultados brindados por el NJ y las subpoblaciones estarían más relacionadas con las subpoblaciones del norte, lo cual concuerda con la ubicación geográfica de estas subpoblaciones. Adicionalmente, se hizo un análisis sin incluir a esta subpoblación y de igual manera todas las subpoblaciones del norte se separaban de todas las subpoblaciones del sur.

El análisis de estructura genética para un modelo de reproducciones no al azar (*no-admixture*) evidenció la presencia de tres grupos genéticos, los cuales denominaremos: Norte: Tumbes, Piura y Cajamarca; Centro: otras poblaciones y Sur: Puno (Tablas 14 y 15). Sin embargo, dado el bajo número de individuos obtenidos en las provincias de Huaraz, Cusco, Ayacucho y Huancayo agrupadas como “otras poblaciones” (Tabla 01) no se puede aceptar de manera concluyente estos resultados, además porque el análisis de estructura parte de un supuesto de equilibrio H-W de las poblaciones analizadas.

Dado que las poblaciones no se encontraron en equilibrio de H-W, los resultados brindados por la estadística F y el análisis de estructura son referenciales. En base a esos resultados toda las subpoblaciones se agruparon en dos poblaciones más grandes (subpoblación norte y subpoblación sur excluyendo a la subpoblación “otras subpoblaciones”) para el análisis de AMOVA (Tabla 17), el cual es el más apropiado e idóneo en poblaciones en desequilibrio. Dicho análisis mostró resultados similares, puesto que indica que la mayor fuente variación se encuentra dentro de las poblaciones (83.62%) y no entre las poblaciones (16.38%). Nuestros resultados contrastan con aquellos encontrados analizando otros parásitos con marcadores SSRs en los cuales las diferencias entre las poblaciones son la mayor fuente de variación de las poblaciones (Sumo et al., 2010; Shrivastava et al., 2005). El bajo número de individuos empleados en este estudio no permiten aseverar de manera fehaciente estos últimos resultados.

Los niveles más bajos de diferenciación genética ($F_{ST} = -0.02$, $R_{hoST} = -0.006$ y $DA = 0.006$) se encontraron entre todas las subpoblaciones que pertenecían a la misma provincia, ya sea en Tumbes, Piura, y Puno. Esto se vio reflejado en la distribución alélica y genotípica similar (Tablas 11, 12 y 14) presente entre ellas. Asimismo, el nivel de diferenciación entre las subpoblaciones ubicadas al norte (Tumbes, Piura, Cajamarca) fue bajo ($F_{st} = 0.144$, $p < 0.01$). El que dos poblaciones sean similares puede deberse a una separación reciente, un elevado flujo génico entre ellas, a su gran tamaño (hay poca deriva génica) o a presiones de selección semejantes que

hayan afectado a loci similares en ambas poblaciones (Hedrick, 2000). En nuestro caso, el alto grado de similitud genética encontrada se explicaría por la proximidad geográfica entre las subpoblaciones, lo cual incrementa el flujo génico entre ellas y ello tiende a homogenizar la variación genética presente. El alto número de migrantes (Tabla 13) que intercambian estas subpoblaciones corrobora la hipótesis de un flujo genético elevado.

Caso contrario es lo que sucede cuando se compara las subpoblaciones de Puno con las subpoblaciones del norte (Tumbes, Piura y Cajamarca) en los cuales se encontraron niveles moderados significativos ($p < 0.01$) de diferenciación genética ($F_{ST} = 0.663$, $Rho_{ST} = 0.39$ y $DA = 0.632$) (Tablas 11, 12 y 14), los cuales indican una distribución alélica y genotípica particular para dichas poblaciones. Dichas diferencias dependen principalmente de los niveles de flujo génico que exista entre las subpoblaciones, es decir si los niveles de flujo son bajos, la selección natural, la deriva génica y la mutación pueden originar una diferenciación genética mayor entre las subpoblaciones (Nei, 1987). Ambas poblaciones pudieron haberse separado hace mucho tiempo y ya no presentar flujo génico dada su lejanía o también puede deberse a que la deriva génica ha generado grandes diferencias o posiblemente a que haya diferentes presiones selectivas entre las dos poblaciones (Hedrick, 2000). Esto último podría ser un factor determinante puesto que se ha comprobado que las diferencias en las condiciones medioambientales afectan los estilos de vida y alimentación del hospedero. Todo ello podría inducir una adaptación por parte del parásito reflejado en la presencia de genotipos distintos para las subpoblaciones del Norte (Tumbes, Piura,

Cajamarca) con respecto a las del Sur (Puno). Además, otra razón sería que hoy en día el transporte de cerdos entre provincias es muy restringido puesto que se requiere autorizaciones especiales y constancias de vacunación; por tanto, se estaría limitando la dispersión del parásito por parte del hospedero intermediario.

Existen dos modelos teóricos para estimar el flujo génico entre poblaciones, uno de ellos es el modelo de isla y el otro es el modelo de aislamiento por distancia (Wright, 1951). En nuestro caso, el flujo génico entre las subpoblaciones estaría mejor explicado por el modelo de aislamiento por distancia, lo cual es afianzado por los resultados del test de Mantel que indican una alta correlación (86%) entre la matriz de distancia geográfica y coeficiente de ancestría entre las subpoblaciones. Las subpoblaciones en estudio no cumplen con tener el mismo número de individuos cada una, intercambiar migrantes con la misma probabilidad y no tener mutación (por la misma naturaleza de las secuencias microsatélites analizadas), los cuales son los supuestos del modelo de islas de la estadística F. (Wright, 1931).

Idealmente, los individuos muestreados para la estimación de la estructura genética poblacional deben pertenecer a la misma generación, puesto que las frecuencias alélicas varían no sólo a través del espacio sino también a través del tiempo dado que las poblaciones son de tamaño finito (Waples 1989a, b). Esto es particularmente importante después de algún efecto fundador o cuellos de botella (Hansson et al., 2000). En nuestro caso, las subpoblaciones pertenecían a diferentes años puesto que provienen de distintas campañas de tratamiento realizado por el grupo de estudio de

cisticercosis en el Perú. Las campañas se hacían en distintos departamentos en tiempos distintos, por tanto si se las agrupaba por año, el número muestral hubiese sido muy pequeño. Sin embargo, esto podría estar afectando la estructuración genética observada.

La mayor limitación del presente estudio fue el bajo número de individuos obtenidos en ciertas comunidades. Sin embargo, se agruparon varias comunidades según su cercanía geográfica para superar esta valla. A pesar de esta limitante y la diferencia en la temporalidad de las muestras, los marcadores microsatélites desarrollados a partir del genoma de *T. solium* permitieron especular acerca de la distribución genética de *T. solium*. Las subpoblaciones del sur se diferenciaron claramente de las del norte y además mostraron una variabilidad genética mayor con respecto a los del norte. Con ello, se corroboró que los microsatélites son los marcadores ideales para reflejar la variabilidad genética presente en esta especie. En ausencia de estudios sobre diversidad y caracterización genética en *T. solium*, los resultados aquí generados constituyen una primera aproximación hacia la elucidación de la biología del parásito y eventualmente podrían ser aplicados para el diseño de mejores estrategias de control, para analizar posibles rebrotes y rutas de re-infección. Como resultado, las poblaciones humanas endémicas se verían beneficiadas a largo plazo, debido que esto contribuye a los esfuerzos de control del parásito.

VIII. CONCLUSIONES

- Se observó la presencia de dos grupos genéticamente distintos de *T.solium* en el Perú: el grupo norte compuesto por todas las subpoblaciones de Tumbes, Piura y Cajamarca, y el grupo sur compuesto por todas las subpoblaciones de Puno.
- Se comprobó que existe variación genética en las poblaciones de *Taenias*, tanto dentro de las poblaciones (intra poblacional) como entre las poblaciones (inter poblacional).
- Se ha observado un alto nivel de diversidad genética (73%), un alto índice de consanguinidad (100%) y una diferenciación genética moderada significativa entre las poblaciones del norte (Tumbes, Piura, Cajamarca).

IX. RECOMENDACIONES

- Ampliar el estudio genético poblacional con la adición de muestras de otras comunidades endémicas del Perú.
- Obtener mayor cantidad de muestras en regiones endémicas del sur para corroborar la alta diversidad genética presente en esa zona.
- Realizar un análisis de regiones con alta variabilidad del genoma mitocondrial para correlacionarlo con los resultados obtenidos en este estudio.
- Evaluar la variabilidad de estos marcadores microsatélites en los cisticercos presentes en el cerdo.

X. BIBLIOGRAFIA

1. Balloux F, Lugon-Moulin N. The estimation of population differentiation with microsatellite markers. *Mol Ecol* 2002; 11:155–165.
2. Barker GC. Microsatellite DNA: a tool for population genetic analysis. *Trans R Soc Trop Med Hyg* 2002 Apr;96 Suppl 1:S21-4.
3. Blouin M., Liu J., Berry R. Life cycle variation and the genetic structure of nematode populations. *Heredity. Trends in parasitology* 1999, 83: 253-259.
4. Bobes RJ, Fragoso G, Reyes-Montes Mdel R, Duarte-Escalante E, Vega R, de Aluja AS, et al. Genetic diversity of *Taenia solium* cysticerci from naturally infected pigs of central Mexico. *Vet Parasitol* 2010 Feb 26;168(1-2):130-135.
5. Botstein D, White RL, Skolnick M, and Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 1980; 32:314–331.
6. Brinkman FSL, Leipe DD. Phylogenetic analysis. In: *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, 2nd edition (eds Baxevanis AD & Ouellete BFF), John Wiley & Sons, USA; 2001. pp. 323–358.
7. Bustos JA, Rodriguez S, Jimenez JA, Moyano LM, Castillo Y, Ayvar V, et al. Detection of *Taenia solium* taeniasis coproantigen is an early indicator of treatment failure for taeniasis. *Clin Vaccine Immunol* 2012 Apr; 19(4):570-573.
8. Butler JM. Forensic DNA typing. Biology, technology and genetics of STR markers. 2da. Ed. Elsevier Academic Press 2005; p. 455-539.

9. Campbell G, Garcia HH, Nakao M, Ito A, Craig PS. Genetic variation in *Taenia solium*. *Parasitol Int.* 2006; 55 (Suppl.): S121-S126.
10. Castagnone-Sereno P, Danchin GJ, Deleury E, Guillemaud T, Malausa T and Abad P. Genome-wide survey and analysis of microsatellites in nematodes, with a focus on the plant-parasitic species *Meloidogyne incognita*. *BMC Genomics* 2010, 11:598.
11. Cavalli-Sforza LL and Edwards AFW. Phylogenetic analysis models and estimation procedures.” *American Journal of Human Genetics* 1967; **19**: 223-257.
12. Cockerham CC. Variance of gene frequencies. *Evolution* 1969; 23: 72-84.
13. Coman BJ. The survival of *Taenia pisiformis* eggs under laboratory conditions and in the field environment. *Australian Veterinary Journal* 1975: 51.
14. Correa D, Morales Z, Medina Y, García C, Medina E, Mandujano A, Ortiz D y Meza A. Teniasis y cisticercosis por *Taenia solium*, Una revisión de viejos y nuevos descubrimientos. Instituto Nacional de Diagnóstico y Referencia Epidemiológicos “Dr. Manuel Martínez Baez”. 1991. Publicación Técnica del Indre # 4.
15. Criscione CD, Poulin R, y Blouin MS. Molecular ecology of parasites: Elucidating ecological and microevolutionary processes. *Molecular Ecology* 2005; 14:2247-2257 (Invited review).
16. Da Costa I, De Oliveira SJ. et al. Genetic polymorphism in *Taenia solium* metacestodes from different Brazilian geographic areas. *Mem Inst Oswaldo Cruz*, Rio de Janeiro 2012; *107*(1): 24-30.

17. De Meeûs T & Renaud F. Parasites within the new phylogeny of eukaryotes. **Trends in Parasitology** 2002; 18: 247-25.
18. De Meeus T, McCoy KD, Prugnolle FF, Chevillon C, Durand P, Hurtrez-Bousse S, Renaud F. Population genetics and molecular epidemiology or how to “débusquer la bete”. *Infection, Genetics and Evolution* 2007; (7) :308–332.
19. De Woody JA & Avise JC. Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *J. Fish Biol.* 2000; 56, 461–473.
20. Detwiler JT, Criscione CD. Testing Mendelian inheritance from field-collected parasites: Revealing duplicated loci enables correct inference of reproductive mode and mating system. *Int J Parasitol.* 2011 Sep; 41(11):1185-1195.
21. Dieringer D, and C. Schlotterer. Microsatellite analyzer (MSA)—a platform independent analysis tool for large microsatellite data sets. *Mol. Ecol.* 2003; Notes 3: 167–169.
22. Ellegren H. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* 2004; 5(6):435–445.
23. Excoffier L, Hofer T, Foll M. Detecting loci under selection in a hierarchically structured population. *Heredity* 2009; 103, 285–298.
24. Excoffier L, Laval G, and Schneider S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 2005; 1:47-50.
25. Excoffier L, Smouse PE, and Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992; 131: 479-491.

26. Frankham R, Ballou JD, Briscoe DA. Introduction to conservation genetics. Cambridge, UK: Cambridge University Press 2002; p. 617.
27. Freeland JR. Molecular Ecology. John Wiley y Sons Inc 2005; pag. 110-111.
28. Garcia HH, Del Brutto OH, Cysticercosis Working Group in Peru. Neurocysticercosis: updated concepts about an old disease. Lancet Neurol 2005 Oct;4(10):653-661.
29. Garcia HH, Gilman RH, Gonzales AE, Verastegui M, Cysticercosis Working Group in Peru. Epidemiology of *Taenia solium* Infection in Peru. In: Garcia HH, Martinez M, editors. *Taenia solium*. Taeniasis/Cysticercosis. Second ed. Lima, Peru: Editorial Universo; 1999 p. 297-305.
30. Garcia HH, Gilman RH, Gonzalez AE, Verastegui M, Rodriguez S, Gavidia C, et al. Hyperendemic human and porcine *Taenia solium* infection in Peru. Am J Trop Med Hyg 2003 Mar; 68(3):268-275.
31. Garcia HH, Gonzalez AE, Evans CA, Gilman RH, Cysticercosis Working Group in Peru. *Taenia solium* cysticercosis. Lancet 2003 Aug 16; 362(9383):547-556.
32. Garza JC, Williamson EG. Detection of reduction in population size using data from microsatellite loci. Molecular Ecology 2001; 10:305-318.
33. Gasser RB, Zhu X, Woods W. Genotyping *Taenia* tapeworms by single-strand conformation polymorphism of mitochondrial DNA. Electrophoresis 1999 Oct; 20(14):2834-2837.

34. Gemmell MA, Lawson JR, Roberts MG, Griffin JF. Population dynamics in echinococcosis and cysticercosis: regulation of *Taenia hydatigena* and *T. ovis* in lambs through passively transferred immunity. *Parasitology* 1990; 101:145–151.
35. Gilman RH. 1999. Short cuts to development: Methods to control the transmission of cisticercosis in developing countries. *In Taenia solium Taeniasis/cisticercosis* (2nd ed., p. 313-326). Edited by Garcia, H.H., Martinez, S.M. Editorial Universo, Lima-Peru
36. Goldstein DB, Pollock DD. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* 1997; 88: 335-342.
37. Gorman GC y Renzi JJ. Genetic Distance and Heterozygosity Estimates in Electrophoretic Studies: Effects of Sample Size. *Copeia* 1979(2):242-249.
38. Grover, A. et al. Biased distribution of microsatellites in rice genome. *Mol. Genet. Genomics* 2007; 277, 469–480.
39. Guo SW and Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. 1992; *Biometrics*. 48. 361-72.
40. Haag KL, Marin PB, Graichen DA, De La Rue ML. Reappraising the theme of breeding systems in *Echinococcus*: is outcrossing a rare phenomenon? *Parasitology* 2011 Mar;138(3):298-302.
41. Haber M, Schungel M, Putz A, Muller S, Hasert B, Schulenburg H. Evolutionary history of *Caenorhabditis elegans* inferred from microsatellites: evidence for spatial and temporal genetic differentiation and the occurrence of outbreeding. *Mol Biol Evol* 2005 Jan; 22(1):160-173.

42. **Hamarsheh O. y Amro A.** Characterization of simple sequence repeats (SSRs) from *Phlebotomus papatasi* (Diptera: Psychodidae) expressed sequence tags (ESTs). *Parasites & Vectors* 2011, 4:189.
43. Hancock K, Broughel DE, Moura IN, Khan A, Pieniazek NJ, Gonzalez AE, et al. Sequence variation in the cytochrome oxidase I, internal transcribed spacer 1, and Ts14 diagnostic antigen sequences of *Taenia solium* isolates from South and Central America, India, and Asia. *Int J Parasitol.* 2001 Dec;31(14):1601-1607.
44. Hartl, D. L. and A. G. Clark. *Principles of Population Genetics*. 2nd ed. Sinauer Associates, Sunderland, Massachusetts; 1989.p. 682.
45. Hedrick PW y Kalinowski ST. Inbreeding depression in conservation biology. *Annual Review of Ecology System* 2000; 31:139–62.
46. Hedrick, P. W. *Genetics of Populations*. Jones & Bartlett Learning. Third Edition, Boston, MA; 2005. p. 400.
47. Ito A, Yamasaki H, Nakao M, Sako Y, Okamoto M, Sato MO, et al. Multiple genotypes of *Taenia solium*--ramifications for diagnosis, treatment and control. *Acta Trop* 2003 Jun; 87(1):95-101.
48. Jamieson A, Taylor SS (1997) Comparisons of three probability formulae for parentage exclusion. *Animal Genetics* 28: 397-400.
49. Jeri C, Gilman RH, Lescano AG, Mayta H, Ramirez ME, Gonzalez AE, et al. Species identification after treatment for human taeniasis. *Lancet* 2004 Mar 20;363(9413):949-950.
50. Jurka J, Pethiyagoda C. Simple repetitive DNA sequences from primates: compilation and analysis. *J. Mol. Evol.* 1995; 40:120–126

51. Kalinowski ST. Evolutionary and statistical properties of three genetic distances. *Molecular Ecology* 2002; 11, 1263–1273.
52. **Karaoglu H, Lee KMY, Meyer W: Survey of simple sequence repeats in completed fungal genomes. *Mol Biol Evol* 2005, 22:639-649.**
53. Kashi Y, King D, and Soller M. Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet* 1997; 13: 74-78.
54. Kelkar YD, Tyekucheva S, Chiaromonte F, Makova KD. The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Res* 2008; 18(1):30–38.
55. **Khairah BA, Ashenafi A, Hassan H et al.** Population genetics analysis during the elimination process of *Plasmodium falciparum* in Djibouti. *Malaria Journal* 2013, **12**:201
56. Kimura M and Ohta T. Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc Natl Acad Sci USA* 1978; 75(6): 2868–2872.
57. Knapp J, Bart JM, Glowatzki ML, Ito A, Gerard S, Maillard S, Piarroux R, and Gottstein B. Assessment of Use of Microsatellite Polymorphism Analysis for Improving Spatial Distribution tracking of *Echinococcus multilocularis*. *J Clin Microbiol* 2007 September; 45(9): 2943–2950.
58. Knapp J, Staebler S, Bart JM, Stien A., Yoccoz NG., Drögemüller C., Gottstein B, Deplazes P. *Echinococcus multilocularis* in Svalbard, Norway: microsatellite genotyping to investigate the origin of a highly focal contamination. *Infect Genet Evol.* 2012 Aug;12(6):1270-4.

59. Kohn MH, York E, Kamradt DA, Haught G, Sauvajot R, y Wayne RK. Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London, Series B* 1999; 266:657–663.
60. Larrañaga N, Mejía R, Hormaza J, Montoya A, Soto A and Fontecha G. Genetic structure of *Plasmodium falciparum* populations across the Honduras-Nicaragua border. *Malaria Journal* 2013, 12:354.
61. Lawson JR., Gemmell MA. Hydatidosis and cysticercosis: the dynamics of transmission. In: Baker J.R., Muller R (eds) *Advances in parasitology*. Academic Press, London 1953; 22: 262-308.
62. Li YC, Korol AB, Fahima T, Beiles A y Nevo E. Microsatellites: genomic distribution, putative functions, and mutational mechanisms: a review. *Mol. Ecol.* 2002; 11, 2453–2465.
63. Li, WH. 1997. *Molecular Evolution* (Sunderland, MA: Sinauer Associates).
64. Lowe A, Harris S y Ashton P. *Ecological Genetics: design, analysis, and application*. Blackwell Publishing 2004; 326pp.
65. Maravilla P, Gonzalez-Guzman R, Zuniga G, Peniche A, Dominguez-Alpizar JL, Reyes-Montes R, et al. Genetic polymorphism in *Taenia solium* cysticerci recovered from experimental infections in pigs. *Infect Genet Evol* 2008 Mar;8(2):213-216.
66. Maravilla P, Souza V, Valera A, Romero-Valdovinos M, Lopez-Vidal Y, Dominguez-Alpizar JL, et al. Detection of genetic variation in *Taenia solium*. *J Parasitol* 2003 Dec;89(6):1250-1254.

67. Meirmans PG. Amova-Based Clustering of Population Genetic Data. *Journal of Heredity* 2012;103(5):744–750.
68. Metzgar D, Bytof J, et al. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res* 2000; **10**(1): 72-80.
69. Michalakis, Y., and L. Excoffier. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* 142: 1061-1064.
70. Mills L, Citta J, Lair K, Schwartz M, Tallmon D. Estimating animal abundance using non-invasive DNA sampling: promise and pitfalls. *Ecological Applications* 2000; 10, 283–294.
71. Moran C. Microsatellite repeats in pig (*Sus domestica*) and chicken (*Gallus domesticus*) genomes. *J. Heredity* 1999; 84: 274–28.
72. Nadler, SA. Microevolution and the genetic structure of parasite populations. *J. Parasitol* 1995; 81: 395–403.
73. Nakao M, Okamoto M, Sako Y, Yamasaki H, Nakaya K, Ito A. A phylogenetic hypothesis for the distribution of two genotypes of the pig tapeworm *Taenia solium* worldwide. *Parasitology* 2002; 124, 657–662.
74. Nakao M, Sako Y, Ito A. Isolation of polymorphic microsatellite loci from the tapeworm *Echinococcus multilocularis*. *Infect, Genet, Evol* 2003;3: 159–63.
75. Neff BD, Gross MR. Dynamic adjustment of parental care in response to perceived paternity. *Proc R Soc Lond B* 2001; 268: 1559-1565.
76. Nei M, Tajima F, Tateno Y. Accuracy of estimated phylogenetic trees from molecular data. *Journal of molecular evolution* 1983; 19:153–170.

77. Nei M. Analysis of gene diversity in subdivided populations. Proceedings of the national academy of sciences USA 1973; 70:3321-3323.
78. Nei M. Genetic distances between populations. *American naturalist* 1972; 106:283-292.
79. Nei M. Molecular Evolutionary Genetics. Columbia University Press, New York. 1987.
80. Nei M. F-statistics and analysis of gene diversity in subdivided populations. *Ann. Hum. Genet* 1977; 41: 225-233.
81. Nicoletti A, Bartoloni A, Reggio A, et al. Epilepsy, cysticercosis and toxocariosis: a population-based control study in rural Bolivia. *Neurology* 2002; 58: 1256-61.
82. Ota T (1993) *Dispan: Genetic Distance and Phylogenetic Analysis*. Pennsylvania State University, University Park.
83. Otha T y Kimura M. The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. *Genetic Research* 1973; 22: 201-204.
84. Ouraa B, Odongoc DO, Lubega et al. A panel of microsatellite and minisatellite markers for the characterization of field isolates of *Theileria parva*. *International Journal for Parasitology* 2003, 33: 1641–1653.
85. Paetkau D, Strobeck C. Microsatellite analysis of genetic variation in black bear populations. *Molecular Ecology* 1994; 3, 489–495.

86. Pappas PW y Leiby DA. Variation in the sizes of eggs and oncospheres and the numbers and distributions of testes in the tapeworm, *Hymenolepis diminuta*. J Parasitol 1986 Jun;72(3):383-391.
87. Pawlowski ZS. *Taenia solium*: Basic Biology and Transmission. In: Singh G, Prabhakar S, editors. *Taenia solium* Cysticercosis. From Basic to Clinical Science. First ed. London, UK: CAB International; 2002. p. 1-13
88. Pritchard JK., Stephens M, et al. Inference of population structure using multilocus genotype data. Genetics 2000; 155: 945-959.
89. Raymond M and Rousset F. Genepop (version 1.2): Population genetics software for exact test and ecumenicism. J. Hered 1995; 86: 248-249.
90. Reed JZ, Tollit D, Thompson P, Amos W. Molecular scatology: the use of molecular genetic analysis to assign species, sex, and individual identity to seal faeces. Molecular Ecology 1997; 6, 225–234.
91. Reynolds, J., B.S. Weir, and C.C. Cockerham. 1983. Estimation of the coancestry coefficient: Basis for a short-term genetic distance. Genetics 105: 767-779.
92. Richard GF, Pâques F. Mini- and microsatellite expansions: the recombination connection. EMBO Rep 2000; 1: 122–126.
93. Riva P. Oliveira, Broude NE, Macedo A. M, Cantor CR., Smith CL., Pena SDJ. Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. Proc. Natl. Acad. Sci. USA Genetics 1998; 95: 3776–3780.

94. Rodríguez N, De Lima H, Rodríguez, A, Brewster S y Barker DC. Genomic DNA repeat from *Leishmania (Viannia) braziliensis* (Venezuelan strain) containing simple repeats and microsatellites. *Parasitology* 1997; 115: 349-358.
95. Rodríguez-Hidalgo R, Benitez-Ortiz W, Praet N, Saa LR, Vercruysse J, Brandt J, et al. Taeniasis-cysticercosis in Southern Ecuador: assessment of infection status using multiple laboratory diagnostic tools. *Mem Inst Oswaldo Cruz* 2006 Nov; 101(7):779-782.
96. Russell R, Iribar M, Lambson B, Brewster S, Blackwell J, Dye, C y Ajioka, J. Intra and inter-specific microsatellite variation in the *Leishmania* subgenus *Viannia*. *Molecular and Biochemical Parasitology* 1999, 103: 71-77.
97. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.
98. Sanjay S, Tiwari H, Elston R. On estimating the heterozygosity and polymorphism in information content value. *Theoretical Population Biology* 2000; 57 (3): 265–271.
99. Schenone H, Villaroel F, Rojas A, Ramirez R. Epidemiology of human cysticercosis in Latin America. In: Flisser, A., Willms, K., Lacleste, J.P., Larralde, C., Ridaura, C., Beltran, F. (Eds.), *Cysticercosis. Present State of Knowledge and Perspectives*. Academic Press, New York 1982; p. 25-38.
100. Sciutto E, Fragoso G, Fleury A, Lacleste JP, Sotelo J, Aluja A, et al. *Taenia solium* disease in humans and pigs: an ancient parasitosis disease rooted in developing countries and emerging as a major health problem of global dimensions. *Microbes Infect* 2000 Dec; 2(15):1875-1890.

101. Shrivastava J, Qian BZ, Mcvean G, Webster JP. An insight into the genetic variation of *Schistosoma japonicum* in mainland China using DNA microsatellite markers. *Mol Ecol* 2005 Mar;14(3):839-849.
102. Simo G., Njiokou F., Tume C., Lueong S., De Meeu T., Cuny G., Asonganyi T. Population genetic structure of Central African *Trypanosoma brucei gambiense* isolates using microsatellite DNA markers. *Infection, Genetics and Evolution* 2010; 10: 68–76
103. Singh B. Molecular methods for diagnosis and epidemiological studies of parasitic infections. *Int J Parasitol* 1997 Oct;27(10):1135-1145.
104. Sivasundar A, Hey J. Population genetics of *Caenorhabditis elegans*: the paradox of low polymorphism in a widespread species. *Genetics* 2003 Jan; 163 (1):147-157.
105. Slatkin M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 1995; 139:457-462.
106. Slatkin M. Gene flow in natural populations. *Annual Review of Ecology and Systematics* 1985a; 16, 393-430.
107. Slatkin M. Rare alleles as indicators of gene flow. *Evolution* 1985b; 39, 53-65.
108. Smyth JD, Smyth MM. Self insemination in *Echinococcus granulosus* in vivo. *J Helminthol* 1969;43(3):383-387.
109. Su X, Wellems T. Toward a high-resolution *Plasmodium falciparum* linkage map: polymorphic markers from hundreds of simple sequence repeats. *Genomics*. 1996; 33: 436-444.

110. Swofford DL, Olsen GL, Waddell PJ, Hillis DM. Phylogenetic inference. In: Molecular Systematics (eds Hillis DM & Mortiz C) Sinauer Assocs, Sunderland, MA; 1996. p. 407–514.
111. Takezaki N, Nei M. Empirical test of the reliability of phylogenetic tree constructed with microsatellite DNA. *Genetics* 2008; 178, 385–392.
112. Tautz D, Trick M, and Dover G. Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 1986; 322: 652-656.
113. Temnykh S. et al. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Research* 2001; 11: 1441-1452.
114. Templeton A. Population Genetics and Microevolutionary Theory. John Wiley y Sons, New Jersey 2006: 82-246.
115. Thompson RCA, Lymbery AJ. Echinococcus and hydatid disease. Wallingford' CAB International 1995; p. 300-315.
116. Toth G, Gaspari Z & Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res* 2000; 10, 967– 981.
117. Van Lith HA, Van Zutphen LF. Characterization of rabbit DNA microsatellites extracted from the EMBL nucleotide sequence database. *Anim. Genet.* 1996; 27: 387-395.
118. Varshney RK, Thiel T, Stein N, Langridge P, Graner A. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol. Biol. Lett.* 2002; 7: 537–546.

119. Vega R, Piñero D, Ramanankandrasana B, Dumas M, Bouteille B, Fleury A, Sciutto E, Larralde C, Frago G. Population genetic structure of *Taenia solium* from Madagascar and Mexico: implications for clinical profile diversity and immunological technology. *Int J Parasitology* 2003; 33: 1479-1485.
120. Vendramin GG, Lelli L, Rossi P, Morgante M. A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. *Mol Ecol* 1996; 5:595–598.
121. Waits J, Leberg P. Biases associated with population estimation using molecular tagging. *Animal Conservation* 2000; 3:191–199.
122. Waits L, Taberlet P, Swenson J, Sandegren F, Franzén R. Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear (*Ursus arctos*). *Molecular Ecology* 2000; 9, 421–431.
123. Waits LP, Luikart G, y Taberlet P. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Mol.Ecol* 2001; 10:249–256.
124. Wang Z, Weber JL, Zhong G, Tanksley SD. Survey of plant short tandem repeats *Theor. Appl. Genet.* 1994; 881–6.
125. Weir B. Genetic data análisis II. Methods for discrete population genetic data. Sinauer Associates Inc Publishers. Massachusetts; 2006. p. 50-156; 209-211.
126. Weir BS and Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution* 1984; 38:1358-1370.
127. White C. Neurocysticercosis: A major cause of neurological disease worldwide. *Clinical Infectious Diseases* 1997; 24: 101-115.

128. Whitlock MC and McCauley DE. Indirect measures of gene flow and migration: F_{ST} doesn't equal $1/(4Nm+1)$. *Heredity* 1999; 82: 117-125.
129. Wright S. Evolution and the Genetics of Populations: The Theory of Gene Frequencies. University of Chicago. 1969; vol 2.
130. Wright S. The genetical structure of populations. *Ann. Eugenics* 1951; 15,323–354.
131. Wright S. The interpretation of population structure by F-statistics with special regard to system of mating. *Evolution* 1965; 19, 395–420.
132. Xiao N, Remais J, Brindley P, Qiu D, Spear R, Lei Y, Blair D. Polymorphic microsatellites in the human bloodfluke, *Schistosoma japonicum*, identified using a genomic resource. *Parasites & Vectors*. 2011, 4: 13.
133. Yakoleff-Greenhouse V, Flisser A, Sierra A, Larralde C. Analysis of antigenic variation in cysticerci of *Taenia solium*. *J Parasitol* 1982 Feb; 68(1):39-47.
134. Zhang XD, Hewitt MG. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* 2003; 12:563-584.