

UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE CIENCIAS

ANÁLISIS DE VARIANTES POLIMÓRFICAS
DEL PROMOTOR, EXÓN 1 Y EXÓN 3 DEL GEN
RECEPTOR A DOPAMINA D4 EN
PACIENTES CON PRIMER EPISODIO PSICÓTICO

TESIS

QUE PARA OPTENER EL GRADO ACADÉMICO DE

DOCTORA EN CIENCIAS

P R E S E N T A

M. en C. ANA JULIA AGUIRRE SAMUDIO

DIRECTOR DE TESIS: DR. CARLOS SABÁS CRUZ FUENTES

México, D.F.

2008



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Dr. Isidro Ávila Martínez
Director General de Administración Escolar, UNAM
Presente

Me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día 26 de Mayo de 2008, se aprobó el siguiente jurado para el examen de grado de **DOCTORA EN CIENCIAS** de la alumna **ANA JULIA AGUIRRE SAMUDIO** con número de cuenta **98801622** con la tesis titulada: **"Análisis de variantes polimórficas en el promotor, exón 1 y exón 3 del gen receptor a dopamina D4, en pacientes con primer episodio psicótico"**, realizada bajo la dirección del **DR. CARLOS SABÁS CRUZ FUENTES**:

Presidente:	DR. FÉLIX RECILLAS TARGA
Vocal:	DR. LUIS MEDRANO GONZÁLEZ
Vocal:	DRA. MARÍA TERESA TUSIÉ LUNA
Vocal:	DR. ALICIA ELIZABETH HERNÁNDEZ ECHEAGARAY
Secretario:	DR. CARLOS SABÁS CRUZ FUENTES
Suplente:	DRA. ARIADNA ESTELA GONZÁLEZ DEL ANGEL
Suplente	DR. EDGAR ARTURO ZENTENO GALINDO

Sin otro particular, me es grato enviarle un cordial saludo.

Atentamente
"POR MI RAZA HABLARA EL ESPIRITU"
Cd. Universitaria, D.F. a, 29 de Septiembre de 2008.


Dr. Juan Núñez Farfán
Coordinador del Programa

c.c.p. Expediente de la interesada.

AGRADECIMIENTOS

Sea extensivo mi agradecimiento al Posgrado de Ciencias Biológicas, sede de inscripción para realizar este doctorado.

A la Dirección general (DGEP) por la beca otorgada durante el período de estudios de doctorado.

También agradezco al Posgrado de Ciencias Biológicas por aceptado la solicitud de apoyo financiero para asistir a los congresos de la organización HUGO “HGM 2003 Workshop” y al XII World Congreso on Psychiatric Genetics 2004, Dublín, Irlanda.

Al Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz, en donde se realizó el trabajo de investigación. Y quien me otorgó una beca para terminar la tesis.

Al jurado designado para evaluar este trabajo de tesis:

Dr. Félix Recillas Targa por haberme aceptado en su laboratorio para hacer una estancia y aprender de epigenética. Al Dr. Luis Medrano González por todas las recomendaciones, por dar su punto de vista en el análisis de genética de poblaciones y por los análisis filogenéticos. A la Dra. María Teresa Tusie Luna, Dra. Alicia Elizabeth Hernández Echeagaray, Dra. Ariadna Estela González del Ángel y al Dr. Edgar Arturo Zenteno Galindo por sus comentarios y observaciones atinadas en el manuscrito para el refinamiento de este trabajo de tesis.

Al Dr. Carlos Sabás Cruz Fuentes por todas las horas de dedicación, interés y explicación durante el trabajo experimental; así como en la perfección de la redacción de este trabajo.

Agradecimientos personales

A mis hijas Ana María y Rosana, quienes han sido y serán los motores y fuente de inspiración a lo largo de toda mi vida académica y personal. Siempre tratando de compensar el tiempo de dedicación a su ternura y candidez.

A mi esposo, gracias por todo el apoyo durante la escritura de este trabajo. A mis padres por haber creído en mis ansias de conocimiento, por su apoyo infinito y que me implantaron la semilla de la educación. Esta cultura de saber trascenderá a todas las personas que en un futuro llegase a enseñar. Gracias por haberme enviado a este maravilloso país.

RESUMEN

El gen receptor a dopamina D4 (DRD4) es uno de los más polimórficos identificado en el genoma humano. A lo largo de todo el gen se encuentran varios SNPs (polimorfismo único de nucleótidos), y en especial, en el exón 3 se presenta una región muy variable tipo minisatélite o VNTR (número variable de repeticiones en tandem) con una unidad repetida de 48 pares de bases (pb), y con polimorfismos de 2 a 11 unidades repetidas. Los alelos más comunes en la población mundial son 2, 4 y 7 repetidas (R). La variación del exón 3 ha sido objeto de estudio en diferentes comportamientos y trastornos psiquiátricos como esquizofrenia.

Dada la importancia de este gen en los trastornos psicóticos llevamos a cabo estudios de asociación con los tres polimorfismos del gen DRD4 más estudiados en la población mundial. Estos son: el SNP -521 (un cambio de C/T) ubicado en el extremo 5' del promotor, un VNTR de 12 pb en el exón 1, y el VNTR de 48pb en el exón 3. Se encontró una asociación entre los alelos poco comunes o raros (incluyen los alelos 3R, 5R, 6R y 8R) del polimorfismo del exón 3 y los trastornos psicóticos, lo cual no había sido reportado antes. Los haplotipos formados por los polimorfismos del exón 1 y 3 fueron inferidos a partir de los genotipos y se encontró una asociación de los haplotipos de las formas raras con los pacientes con diagnóstico de psicosis. También se encontraron diferencias estadísticamente significativas entre la distribución alélica del SNP -521 con los pacientes psicóticos a diferencia del grupo control.

El análisis de la secuencia del polimorfismo del exón 3 del gen DRD4, se ha estudiado poco en población mundial y mucho menos en población de origen mestizo mexicana. Uno de los objetivos de esta tesis fue caracterizar la región del exón 3 del gen DRD4, mediante análisis de polimorfismo, de desequilibrio de ligamiento (DL) y de regiones conservadas (evaluado con el análisis de entropía). Nosotros reportamos nuevas variaciones nucleotídicas y haplotipos formados por el VNTR del exón 3 de este gen. La mayoría son mutaciones que contribuyen al conocimiento de la variación genética. La diversidad nucleotídica del grupo de secuencias de los alelos fue mayor en los de 7R. La variación que se ubicaba dentro de los patrones de polimorfismo, fueron evaluados con la Teoría Molecular de la Neutralidad. Las secuencias de los alelos de 2R y 4 R presentan variaciones de tipo neutral, y las secuencias de los alelos de 7R están sujetas a la selección direccional o positiva. El análisis de DL y conservación en los nucleótidos de los alelos de 7R reveló sitios fuertemente asociados que se conservan en el tiempo. Ambos análisis DL y entropía apoyada con la comparación de secuencias de 7R en especies de gibones (género *Hylobates*), nos ayudó a identificar sitios SNP fuertemente ligados y conservados que se sugiere pueden ser susceptibles a enfermedades o fenotipos asociadas al alelo de 7R.

La reconstrucción filogenética relacionó varias especies de primates no humanas con las variaciones encontradas en este estudio de los alelos 7R (región del VNTR del exón 3). El análisis ayudó a visualizar sitios susceptibles de la esquizofrenia, cuyas variaciones reportadas en este estudio, son de aparición reciente. Sin embargo, las formas comunes y el control⁷⁰ tienen variaciones cuyo nodo ancestral refiere a los platirinos con el *S. boliviensis*, que se separa del nodo del *Homo sapiens* unos 26 millones de años.

Nuestros resultados sugieren que las variantes genéticas situadas a lo largo del gen DRD4 pueden jugar un papel en el desarrollo de los trastornos psicóticos. Además, los análisis de evolución molecular nos sugieren una dinámica poblacional en donde el alelo de 7R puede tener una ventaja adaptativa sobre los demás alelos 2R y 4R, y en cuya secuencia nucleotídica se distinguen sitios que podrían actuar como marcadores en el mapeo de la característica del comportamiento o en especial en enfermedades complejas asociadas al alelo de 7R.

ABSTRACT

The receptor D4 dopamine gene (DRD4) has been described as one of most polymorphic in the human genome. This gene have several SNPs (single nucleotide polymorphism), and especially, the exon 3 has a region very variable, a VNTR (variable number of tandem repeat) of 48 pairs of bases (pb). The polymorphism has 2 to 11 alleles, but the alleles most common in the world-wide population are 2, 4 and 7 repeat (R). This variation has been analyzed in different behaviors and psychiatric diseases such as schizophrenia.

On the other hand, given the importance of this gene in psychotic disorders, we carried out association studies with the three polymorphisms of the DRD4 gene. These are: SNP -521 (C/T substitution) located in end 5' flanking of the promoter, a VNTR-12pb in exon 1, and the VNTR-48pb in exon 3. There was an association between the rare alleles (3R, 5R, 6R and 8R) of the polymorphism of exon 3 and psychotic patients, which had not been reported before. The haplotypes formed by the exon 1 and 3 polymorphisms were inferred from the genotypes and it was an association with the patients with diagnosis of psychosis.

Also, we found significant allele frequency differences of -521 T between the case and the control groups.

The analysis of the VNTR-exon 3 sequence of the DRD4 gene has been little studied in world-wide population. One of the objectives of this thesis was to characterize the region of VNTR-exon 3. Here we analyzed by DNA sequencing of 2R, 4R and 7R alleles, the polymorphism sites, the nucleotide diversity, the Neutral test of three alleles, the linkage disequilibrium (LD) and entropy (to evaluate conserved regions). We reported new variations and haplotypes of the VNTR of this gene, and the majority is mutations that contribute significantly to the genetic variation. The nucleotide diversity of the 2R and 4R sequences displayed neutrality. Nevertheless, we found that the diversity was greater in 7R alleles. This variation was evaluated with the Neutral Theory, which showed that the sequences of the alleles of 7R are subject to the positive selection.

Moreover, strong LD and region conserved were found in the nucleotides of the 7R allele. Both analyses (LD and entropy) and under the comparison of sequences of 7R with species of gibones (genus *Hylobates*), were enable to identify sites SNP strongly linkage and conserved, which suggest they can be susceptible to diseases or phenotypes associated to the allele of 7R of DRD4 gene.

Phylogenetic reconstruction related the non-human primate species with the variations find of this study in 7R sequences (VNTR-exon3 region). The analysis was able of identify sites or haplotypes that it suggest is susceptibility to schizophrenia in our sample. The nucleotide variations reported here can be of recent apparition. However, the common form and control 70 had polymorphisms with the ancestral node in the Platyrrhini, *S. boloviensis* that it separate of *Homo sapiens* past 26 Myr.

Our results suggest that genetic variants in DRD4 gene can play a role in the development of the psychotic disorders. The analysis of molecular evolution suggest a population dynamics where the alleles of 7R can have an adaptive advantage on the other alleles (2R and 4R), and in whose nucleotide sites are distinguished strong LD. These sites can especially to function like markers in the mapping of characteristic of the behavior or in complex diseases associated to the 7R allele.

INDICE

PRESENTACIÓN	11
INTRODUCCIÓN	14
1. El uso de genes como marcadores en el Mapeo de enfermedades complejas	15
1.1 Uso de marcadores polimórficos y haplotipos	16
1.1.1 Marcadores moleculares en el Genoma Humano.....	17
1.1.2 Haplotipos.....	19
1.1.3 Análisis de desequilibrio de ligamiento.....	20
1.2 Análisis de asociación	22
1.2.1 Estudio de casos-control.....	23
2. Diversidad Nucleotídica y Neutralidad	24
2.1 Estimadores.....	24
2.2 Neutralidad y predicciones de variación molecular.....	27
2.3 Pruebas de neutralidad	27
2.4 Patrones de diversidad humana.....	29
3. Factores que forman patrones de la diversidad	31
3.1 Tasas de Mutación y Parámetros de mutación Neutral.....	31
3.2 Selección Molecular.....	32
4. Dinámica, historia y estructura poblacional	35
5. Aplicación de la variación genética	37
6. Estructura del gen que codifica al receptor tipo 4 a dopamina (DRD4) y su variación polimórfica	39
6.1 Caracterización molecular del DRD4	39
6.1.1 VNTR en el exón 3.....	40
6.1.2 Variabilidad en la secuencia interna de los repetidos VNTR exón3.....	41
6.1.3 Polimorfismo -521 en la región promotora.....	42
6.1.4 Polimorfismo del exón 1.....	44
6.2 Variación en diferentes taxas	45
7. Proteína receptora a dopamina D4	46
7.1 Estructura de la proteína D4.....	46
7.2 Predicción mediante modelaje molecular.....	48
7.3 Distribución anatómica y expresión.....	48
7.4 Farmacología del receptor D4.....	47
8. Características clínicas de la psicosis	53
8.1 Factores asociados a la enfermedad.....	54
8.2 Hipótesis dopaminérgica de la esquizofrenia.....	57
II. OBJETIVOS DEL ESTUDIO	60
III. MATERIALES Y MÉTODOS	62

1. Sujetos	63
2. Determinación de genotipos.....	64
3. Análisis estadístico.....	65
4. Análisis de la secuencia del VNTR del exón 3	66
5. Análisis polimórfico y selección.....	66
6. Prueba de Desequilibrio de Ligamiento.....	67
7. Entropía	68
8. Modelaje molecular	68
9. Reconstrucción Filogenética.....	69
IV. RESULTADOS.....	70
1. Artículo I.....	71
2. Artículo II.....	73
3. Polimorfismo del sitio -521 C/T.....	79
V. DISCUSIÓN.....	108
VI. CONCLUSIONES.....	118
VII.REFERENCIAS.....	120
VIII. APENDICE.....	131
• Geles de electroforesis de los polimorfismos estudiados	
• Secuencias del VNTR del exón 3	
• Secuencias del VNTR del exón 1	
• Secuencia del polimorfismo -521, extremo 5' flanqueante	
• Alineamiento de las secuencias para el análisis de entropía	

PRESENTACIÓN

LISTA DE PUBLICACIONES

La tesis está basada en los siguientes artículos, el cual se refieren en el texto por sus números Romanos.

I Ana Julia Aguirre, Rogelio Apiquian, Ana Fresan, Carlos Cruz-Fuentes. Association analysis of exon III and exon I polymorphisms of the dopamine D4 receptor locus in Mexican psychotic patients. *Psychiatry Research* 153:209-215, 2007. PREMIO PAUL JANSSEN DE INVESTIGACIÓN 2007

II Ana Julia Aguirre Samudio¹, Carlos Cruz Fuentes¹, Mario Cásarez Espinosa, Luis Rosales-León. Haplotypes and nucleotide diversity analysis in the polymorphic sequence of the 48bp-VNTR-DRD4 gene in Mexican subjects. Manuscrito.

Este trabajo de investigación tiene cómo tema central al gen que codifica a la proteína receptora a dopamina tipo 4 (DRD4), el cual es uno de los múltiples genes candidatos en relación a diferentes trastornos mentales. Se tipificaron variantes en diferentes regiones del gen lo que permitió desarrollar un estudio de asociación con los trastornos del espectro psicótico. Asimismo, se caracterizó parte de su secuencia nucleotídica, lo que permitió llevar a cabo el análisis de la diversidad nucleotídica en una región con relevancia fisiológica. Además, se evaluó si esta variación está sujeta a un proceso de mutación neutral o selección direccional. Se determinó el posible desequilibrio de ligamiento entre las variaciones además de calcular valores de entropía nucleotídica, datos que permitieron intentar establecer relaciones filogenéticas entre especies (Esquema 1). En conjunto, todos estos análisis pretendieron evaluar el comportamiento del gen DRD4 en la población tanto normal como aquella afectada por el trastorno. Finalmente intentamos responder a la siguiente pregunta ¿El gen DRD4 y sus variaciones pueden actuar como un marcador molecular en el diagnóstico de las enfermedades complejas como lo es la psicosis?

Por otra parte, este estudio intentó contribuir, aunque de forma modesta al análisis de la diversidad en el genoma humano. Recordemos que los 1.2 millones de variaciones entre las secuencias hasta ahora reportadas ha proporcionado material valioso para evaluar la importancia antropológica, evolutiva y médica de la variabilidad genómica. El análisis de la secuencia del polimorfismo del exón 3 del gen DRD4, se ha estudiado poco en la población mundial. Este estudio caracterizó y analizó la variabilidad de esta región en sujetos de origen mexicano, desde una perspectiva de la genética de poblaciones, a partir del cual hemos tratado de hacer inferencias aplicables a la historia de una población humana. En un sentido de dinámica poblacional, hemos tomado desde el punto de vista de la evolución molecular para tratar de entender las causas del espectro de la variación obtenida. Para ello, hemos analizado el papel de fenómenos como la mutación, selección, sitios conservados bajo un desequilibrio de ligamiento y filogenia. Para tener una apreciación de cómo se comportan las variaciones entre especies, nosotros comparamos la secuencia del VNTR-exón 3 con especies de primates no humanas, lo cual pretendió ayudarnos a entender el papel de la variación en la relación fenotipo-genotipo, es decir qué tanto ha variado o conservado entre los taxones cercanos a la filogenia del hombre y si esas mutaciones pueden guardar una distancia genética entre ellos.

En dicha región hemos logrado detectar alelos bajo selección positiva lo mismo

que sitios conservados bajo presión selectiva. Esta información podría ser útil en el manejo de marcadores moleculares de genes que contribuyen de alguna forma al desarrollo de psicosis como la esquizofrenia.

Por otra parte, dada la importancia de este gen en los trastornos psicóticos llevamos a cabo estudios de asociación con los tres polimorfismos más estudiados en la población mundial. Cabe resaltar el hallazgo de la asociación de alelos poco comunes o raros del polimorfismo del exón 3 y los trastornos psicóticos, el cual no había sido reportado antes. Finalmente, el disponer de datos de variabilidad situadas a lo largo del gen DRD4 nos ha permitido comparar frecuencias poblacionales que han contribuido al estudio de enfermedades complejas o trastornos mentales como la psicosis.

I. INTRODUCCIÓN

Conducir ordenadamente mis pensamientos, empezando por los objetos más simples.

René Descartes.

1. EL USO DE GENES COMO MARCADORES EN EL MAPEO DE ENFERMEDADES COMPLEJAS

Durante muchos años se ha planteado la hipótesis sobre la relación entre genes y enfermedades. No obstante con el avance tecnológico y la completa secuenciación del genoma Humano, los científicos tienen la oportunidad de encontrar la relación entre el conjunto de variaciones genéticas específicas y una entidad clínica particular. Esto involucra a una genética compleja con relación a los principios de Mendel, en el que los médicos deben apreciar su complejidad para el desarrollo de un tratamiento para el padecimiento (Kelley, 2007). La genética, desde esta visión compleja, estudia las contribuciones de múltiples loci para exhibir un fenotipo específico. Dichos fenotipos pueden estar influenciados por un estímulo ambiental y la ruta para examinar los loci es mediante la genómica.

Anteriormente, lo más conocido por la medicina eran las enfermedades con herencia mendeliana. Actualmente se sabe que los genes identificados como los responsables de enfermedades Mendelianas, presentan en la mayoría de los casos una variación por mutación que aparecen raramente (Chakravarti, 1999). A la fecha, existe un compendio de información de enfermedades genéticas y los genes relacionados, el cual está editado desde 1966, y su versión en internet está disponible en el sitio <http://www.ncbi.nlm.nih.gov/omim/>. El OMIM cuyas siglas significan Online Mendelian Inheritance in Man, inicia su base de datos en la Escuela de Medicina de la Universidad de Johns Hopkins con los algoritmos desarrollados por el NCBI (National Center for Biotechnology Information). En la actualidad, esta base registró 12000 genes con secuencia conocida y la relación genotipo fenotipo (Hamosh *et al*, 2005).

Sin embargo, la mayoría de los fenotipos humanos en el que se incluyen enfermedades, son complejos en donde los patrones mendelianos no son los responsables del trastorno. Los genetistas saben que el origen de una enfermedad no es exclusivo de los genes ni del medio ambiente, sino de ambos. La variación genética encontrada en los organismos puede conferir un efecto susceptible o protector para que se manifieste una enfermedad en un individuo que forma parte de una población.

La predisposición dada por una variación genética está sujeta a factores del medio ambiente como agentes infecciosos en la etapa adulta o embrionaria, productos químicos,

tabaquismo, dietas, estrés que ante la variante, se suman para que realmente se manifieste la enfermedad (Kelley, 2007). Un ejemplo claro de la complejidad genética y ambiental es la diabetes de tipo 2.

La forma para determinar los factores genéticos en un fenotipo complejo puede llegar a ser muy elaborada, lo cual requiere buscar o mapear genes, donde cada uno proporciona un efecto sutil, aditivo sobre el padecimiento. Sin embargo muchas veces se desconocen las variantes genéticas que originan dicho fenotipo.

La estrategia para tratar de solucionar en parte este problema, va dirigida a explorar la estructura de los haplotipos en el genoma (este tema se ampliará en la sección 1.1.2), el cual se define como la combinación de alelos marcadores en un cromosoma y al desequilibrio de ligamiento en extensión o la asociación de alelos no al azar (el desequilibrio de ligamiento se aborda con amplitud en la sección 1.1.3). El análisis de desequilibrio de ligamiento tiene como objetivo mapear detalladamente las variaciones que podrían estar asociadas a enfermedades.

1.1 USO DE MARCADORES POLIMÓRFICOS Y HAPLOTIPOS

Un gen es polimórfico cuando presenta diferencias en la secuencia entre individuos. En una población, el polimorfismo se presenta cuando la frecuencia de los alelos de un gen varía al menos 0.01, de manera que los individuos heterocigotos portadores de uno de los alelos, pueden detectarse en el 2% o más de la población (Oliva, 1996). Los grados de variación van desde simples diferencias de pares de bases hasta números diferentes de repetidas de segmentos de DNA específicos. A dicha variación le denominan polimorfismo molecular. La presencia de variantes en la secuencia del DNA (variación alélica) dentro del genoma humano explica en parte las diferencias entre individuos. Esta variación consta de alrededor del 0.1% en la secuencia del DNA entre dos individuos. La identificación de polimorfismos en el genoma humano y el estudio de su herencia aportan marcadores genéticos útiles para estudios de mapeo físico, mapeo genético y análisis de ligamiento (Aguirre y Nicolini, 2005).

En esta tesis la herramienta principal de trabajo fue el uso de marcadores moleculares y su variación o polimorfismo.

1.1.1 Marcadores moleculares en el genoma humano

Un marcador molecular es un elemento ubicado en el mapa del genoma que puede ser: un gen, una secuencia marcada in situ (STS), un fragmento de DNA, un oligonucleótido o un sitio de restricción (Oliva, 1996).

A lo largo de la evolución, el genoma humano presenta evidencias de cambio que se manifiestan como duplicaciones, inserciones, deleciones, reorganizaciones, mutaciones puntuales y otro tipo de cambio. Las frecuencias alélicas de estas variaciones están determinadas por fuerzas evolutivas como la migración, la selección natural y la deriva genética, lo cual puede variar dentro y entre poblaciones (Oliva, 1996). Las fuerzas evolutivas determinan los patrones de una dinámica poblacional, que pueden ser útiles al estudio de las frecuencias poblacionales de un polimorfismo en una enfermedad. La identificación de polimorfismos en el genoma humano y el estudio de su herencia aportan marcadores genéticos útiles que son independientes del fenotipo expresado.

Hay diferentes tipos de marcadores genéticos, entre ellos se pueden mencionar por su importancia y porque fueron tema central en este estudio, a los marcadores de número variable de repeticiones en tandem (VNTR) y el polimorfismo de nucleótido único (SNP).

Marcadores de número variable de repeticiones en tandem (VNTR) o Minisatélites

Los minisatélites o VNTR se definen como secuencias repetidas en tandem (con una longitud entre 10 a 100 pares de bases), cuya unidad de repetición se puede expandir de 0.5 a varias kilobases. Esta secuencia repetida generalmente es rica en dinucleótidos CG.

Los minisatélites son altamente polimórficos con heterocigocidad de $H \geq 0.5$, y son marcadores multialélicos apropiados para el estudio de ligamiento y de asociación. Estos marcadores, se han asociado con algunas características de la biología y evolución del genoma humano, así como con patologías de origen genético (Vergnaud y Denoeud, 2000).

Los VNTR se consideran hipermutable debido a que su proporción de mutación promedio en las células germinales es alta y alcanza 0.5% (la frecuencia de mutación varían tanto en el macho como en la hembra); además posee una variación interna en su secuencia (Vergnaud y Denoeud, 2000). La distribución cromosómica de este tipo de

polimorfismo se encuentra hacia las regiones teloméricas y regiones ancestralmente teloméricas, y como son altamente polimórficos son aptos en la identificación de microdeleciones en los extremos de los cromosomas (Denoeud *et al.*, 2003).

Los minisatélites tienen una participación activa en el funcionamiento del genoma humano, debido a que:

- 1 Algunos son parte del marco de lectura y varían entre los individuos de una población.
- 2 Los que interfieren en la codificación de proteínas. Por ejemplo, los minisatélites ubicados en el extremo 5' de los genes que participan en la regulación de la transcripción (Kennedy *et al.*, 1995); los que se ubican en la región intergénica porque pueden interferir con el splicing o empalmamiento (Turri *et al.*, 1995). Algunos estudios indican que los minisatélites pueden funcionar como biomarcadores en la exposición de agentes genotóxicos como la radiación ionizante en animales (Dubrova *et al.*, 1998) y en humanos crónicamente expuestos a la liberación de material radioactivo como en la explosión de la estación nuclear de Chernobil (Dubrova *et al.*, 1997).

Polimorfismo de nucleótidos único (SNP)

Los polimorfismos de nucleótidos único (SNPs), son variaciones de un nucleótido y pueden tener una frecuencia de 0.01 o mayor en la población en general. Los SNP representan aproximadamente el 90% de todos los polimorfismos del DNA humano (Collins *et al.*, 1998).

Existen 4 tipos de SNP bialélicos; siendo la transición C/T (G/A) la más común. Los SNP no sinónimos son los que se presentan en regiones codificantes que cambian en el triplete la secuencia de la proteína. Los SNP sinónimos son los que no cambian la secuencia de codificación.

Los genotipos de SNPs pueden ser identificados con alta fidelidad por lo que se dice que llegan a ser buenos marcadores para el mapeo de enfermedades complejas; tienen una tasa baja de mutación y los eventos de cambio de bases permiten que los alelos del SNP sean muy estables (Gray *et al.*, 2000). La mayoría (aproximadamente >80%) son comunes en todas las poblaciones, pero con diferentes frecuencias alélicas (Sachidanandam *et al.*, 2001).

Los SNPs se utilizan ampliamente en los estudios de asociación de enfermedades complejas en humanos (Risch y Merikangas, 1996), o en el análisis de desequilibrio de ligamiento para el mapeo de enfermedades. A la fecha no hay consenso sobre la cuál es la densidad de SNPs requeridos para el mapeo de un fenotipo en poblaciones con mezcla al azar.

En la actualidad, con el avance tecnológico en el desarrollo de plataformas de tipificación masiva y la integración de consorcios internacionales se ha logrado desarrollar un mapa de haplotipos tipo SNP (The Internacional HapMap Consortium, <http://www.hapmap.org>), el cual constituirá una base de información muy útil para los investigadores en el campo de la medicina, antropología y biología.

1.1.2 Haplotipos

Es la combinación de alelos marcadores ligados en un cromosoma. Los haplotipos son “bloques” físicos de polimorfismos que son heredados juntos con una frecuencia mayor a la esperada.

La información que arroja un haplotipo es de gran importancia en el análisis de datos genéticos moleculares. Una ventaja del estudio de haplotipos es que son más polimórficos que el uso de un solo locus marcador. El ejemplo 1 describe cómo se puede presentar un haplotipo. Experimentalmente obtenemos los genotipos, sin embargo, este método no provee información acerca de los haplotipos. Éstos pueden ser inferidos a partir de los genotipos.

Si un individuo tiene genotipos en tres loci (bialélicos),
 Aa, Bb, CC,
 Entonces es posible tener los siguientes haplotipos:
 ABC. abC o AbC. aBC.

Ejemplo 1.

Existen varias razones para intentar explorar la estructura de haplotipos en el genoma humano, entre ellas se pueden mencionar las siguientes:

1. La estructura y formación de los haplotipos hacen referencia a la historia evolutiva y a la identificación de variantes genéticas, que pueden influir en las diferencias entre los fenotipos y las características humanas. Dichos haplotipo pueden ser

evaluados con el análisis de desequilibrio de ligamiento (ver sección 1.1.3). Dentro de los haplotipos se incluyen marcadores fuertemente ligados, por ejemplo aquellos que están en un mismo gen con alelos que muestran dependencia estadística o están en desequilibrio de ligamiento (DL) o asociación alélica (Zhao *et al.*, 2003).

2. También pueden proporcionar información acerca de los eventos de recombinación, y de la estructura de una población. Asimismo, nos permite intentar entender las presiones de selección evolutivas sobre una región específica del genoma.

Un aspecto de gran relevancia para este trabajo, se basó en el análisis de DL entre haplotipos para identificar variantes genéticas asociadas a enfermedades complejas; ya que el DL permite tener un diseño más eficiente de los estudios de asociación del tipo casos-control (Zhao *et al.*, 2003).

Una manera de formar haplotipos informativos es conociendo las variaciones nucleotídicas de los SNPs. Las combinaciones de SNP que forman haplotipos son usadas para el estudio de enfermedades relacionadas a genes. El SNP *Consortium* por ejemplo tiene mapeado más de 1.4 millones de SNPs (Internacional SNP map Working group). Sin embargo, debido a la abundancia de los SNPs en el genoma, es más complejo identificar qué SNPs dentro de un haplotipo permiten la caracterización de todas las posibles variaciones alélicas en una región, sin tener que tipificar todos los sitios. A este grupo de SNPs que forma estos haplotipos se le denomina Haplotype “Tagging” SNP (htSNPs), que son marcadores que capturan las combinaciones en una región en desequilibrio de ligamiento, y su identificación puede ahorrar tiempo y recursos en los estudios. El DL es uno de los métodos más informativos en el mapeo de enfermedades (Zhang *et al.*, 2004; Fallin y Schork, 2000).

1.1.3 Análisis de desequilibrio de ligamiento (DL)

El desequilibrio de ligamiento, se define como los patrones de asociación de los alelos no al azar en diferentes loci, que forman haplotipos dentro de una población, y cuyos individuos no tienen parentesco (Nordorg y Tavaré, 2002). Estas asociaciones dadas en una población son fuente de información para el mapeo fino de loci de

enfermedades humanas.

El concepto de DL fue usado a partir de principios del siglo XX, el cual fue desarrollado por Lewontin y Kojima (1960), quienes lo llamaron como una asociación no al azar de alelos en un par de loci.

En promedio, dos marcadores que están cercanos físicamente tienen más probabilidad de estar asociados con relación a aquellos que están lejanamente espaciados (ver ejemplo 2). Así el DL disminuye con la distancia física cromosómica, sin embargo hay una gran variación entre diferentes regiones cromosómicas. Es importante mencionar que el poder y la eficiencia de la estadística del DL se ve afectado por eventos de recombinación, mutación, eventos de conversión de genes, deriva génica, mestizaje, y selección natural, el cual puede distorsionar el valor calculado del DL y la distancia física interlocus (Jorde, 2000).

Un par de loci dialélicos A y B en el mismo cromosoma
Con Alelos A_1 y A_2 B_1 y B_2

Haplotipos A_1-B_1 , A_1-B_2 , A_2-B_1 , A_2-B_2

Si las frecuencias observadas de los alelos A_1 , A_2 , B_1 y B_2 en la población son en total de 0.5...Entonces esperaríamos que las frecuencias de cada haplotipo en la población fuera de 0.25, ver ecuación E.1

Ejemplo 2

Denominamos P_{ij} a la frecuencia observada del gameto A_iB_j , y p_{ij} a la frecuencia esperada de ese gameto si hay equilibrio de ligamiento. En un sistema de dos genes ligados con dos alelos cada uno (A_1, A_2 y B_1, B_2), tendremos:

Gametos	Frecuencia Observada	Frecuencia en el equilibrio
A_1B_1	P_{11}	p_1q_1
A_1B_2	P_{12}	p_1q_2
A_2B_1	P_{21}	p_2q_1
A_2B_2	P_{22}	p_2q_2

En general, un gameto constituido por los alelos A_i y B_j tendrá la frecuencia observada de P_{ij} y la frecuencia esperada en el equilibrio $p_i q_j$.

El desequilibrio de ligamiento vendrá definido por la diferencia entre la frecuencia observada del gameto P_{ij} y la esperada si hubiera equilibrio $p_i q_j$, la relación está dada por:

$$D = P_{ij} - p_i q_j \quad (\text{Ecuación E.1})$$

Si el valor de D es diferente de cero, entonces se puede decir que hay un DL. El grado de DL entre 2 loci es dependiente de la fracción de recombinación, θ , y del tiempo generacional, t , (es decir, el tiempo 0 a partir de que una nueva mutación causa una enfermedad): $D_t = D_0 (1-\theta)^t$. Así, la D nos indica la frecuencia de la recombinación y la distancia física entre dos loci (Jorde, 2000).

1.2 ANÁLISIS DE ASOCIACIÓN

El análisis de asociación es una herramienta efectiva en el estudio de las enfermedades complejas, ya que tiene un mayor poder estadístico que los análisis de ligamiento, en especial cuando hay una heterogeneidad de locus (Risch y Merikangas, 1996). Este análisis compara las frecuencias de la variable a estudiar, es decir son los alelos de los individuos afectados y los que no lo están.

Una ventaja de los estudios de asociación particularmente en el diseño casos control (sección 1.2.1) está en que los individuos seleccionados pueden ser confrontados con otros que tengan características ambientales similares, como el estilo de vida, y esto puede ser importante en el desarrollo de la enfermedad (Goldstein y Weale, 2001)

Actualmente, los estudios de asociación se emplean frecuentemente en la determinación del papel de la genética de las enfermedades complejas, en donde se utilizan los genes candidatos. Estos a su vez son aquellos genes que se seleccionan debido a su posible papel biológico en la patología de la enfermedad y sus variantes son analizadas en poblaciones que pueden compartir condiciones similares del medio

ambiente.

Al aplicar en el estudio de asociación el análisis de ligamiento, dicho estudio puede alcanzar mayor fuerza estadística (mediante pruebas no paramétricas) y mayor eficiencia en el mapeo de genes asociados a enfermedades complejas, ya que aumenta la posibilidad de encontrar una relación entre loci y enfermedad. Porque los alelos en diferentes loci se encuentran juntos y proporcionan mayor información basado en las frecuencias alélicas (Jorde, 2000).

1.2.1 Estudios de casos-control

El diseño de casos-control es la estrategia más usada en los estudios de asociación en la genética epidemiológica (Khoury y Beaty, 1994). La ventaja es la relativa facilidad de los sujetos de estudio y donde es importante que tanto los casos como los controles sean cuidadosamente seleccionados. En especial el grupo control, debe estar relacionado en edad, género y grupo étnico similares, para que la asociación estadísticamente significativa encontrada refleje la influencia alélica estudiada.

En un estudio de casos-controles el objetivo es comparar la exposición de factores de riesgo (ambientales y genéticos) entre individuos afectados y controles no afectados para encontrar asociación entre factores de riesgo y la enfermedad (Zonder van y Cardon, 2004).

En estudios de población, el método de casos-control se usa para 1) estudios en donde el objetivo es determinar mutaciones humanas, 2) evaluar el papel de indicadores genéticos no-específicos, como endogamia o mezcla racial, 3) y medir el papel de características genéticas en la etiología de las enfermedades (Khoury y Beaty, 1994).

2. DIVERSIDAD NUCLEOTÍDICA Y NEUTRALIDAD

La diversidad nucleotídica hace referencia a la variación en la secuencia del DNA, que está basado principalmente en la diversidad de SNPs y los patrones que forman. Esta información es importante ya que permite explorar los patrones de la evolución molecular de las especies. En especial la humana porque puede intentar identificar el origen de enfermedades o fenotipos de relevancia médica.

Además de los SNP, las secuencias repetidas pueden formar parte de otras formas de variación, que se pueden tomar de referencia para medir la diversidad nucleotídica, y representan del 35% a 50% de las secuencias en el Genoma humano.

La diversidad nucleotídica puede ser evaluada a partir de estimadores que nos indican como es la variación en determinada región, entre especies.

2.1 ESTIMADORES DE DIVERSIDAD

La diversidad genética en las especies pueden proveer información valiosa que ayuda a deducir la evolución de las especies, el cual incluye cambios en el tamaño de la población en el pasado, efectos de estructura social sobre la diversidad, patrones de migración en el pasado y señales de selección de eventos pasados. Por esta razón numerosos estudios de la diversidad genética son realizados en humanos, ya que es el fundamento para entender las bases genéticas de las enfermedades, tanto en poblaciones ancestrales como en especies emparentadas cercanas al humano (como los gorilas) (Yu *et al.*, 2004). Por ejemplo, la hipótesis del origen humano en África, ha sido apoyada por el estudio de la diversidad nucleotídica en poblaciones humanas y en donde se ha logrado identificar un efecto fundador que se origina en África en la región del Sahara, con frecuencias ancestrales que reflejan la variación en la dinámica poblacional entre regiones geográficas, la relación entre haplotipo y heterocigosidad y la distribución geográfica (Li *et al.*, 2008). Datos como estos nos dan una caracterización más amplia y completa de la variación genética humana. Aunque se sabe muy poco de la diversidad relacionados con los trastornos mentales.

Un comparación de la diversidad nucleotídica en diferentes tipos de genes muestra el alto grado de variabilidad que puede existir inter-poblaciones e inter-especies (ver tabla 1).

Una manera muy sencilla de medir la cantidad de variación nucleotídica es determinando el número de sitios que cambian en una muestra. La medida del polimorfismo de DNA no dependerá del tamaño de la muestra sino del número promedio de diferencias nucleotídicas por sitio entre dos secuencias escogidas al azar. Es decir, la diversidad nucleotídica puede ser medida mediante la relación de π (ver ecuación E.2),

$$\pi = \sum_{ij} x_i x_j d_{ij} \quad E.2$$

en donde q es el número total de alelos (secuencias alélicas diferentes), x_i es la frecuencia poblacional del alelo i , y d_{ij} es el número de diferencias nucleotídicas o sustituciones por sitio entre los alelos i y j en la secuencia del DNA.

Estos son los parámetros centrales de la teoría neutral de la Evolución molecular, propuesta por Motoo Kimura (1968). La teoría Neutral señala que la mayoría de las sustitución de nucleótidos en el transcurso de la evolución son el resultado de la fijación al azar de mutantes neutros, o que tienden a la neutralidad, por deriva génica, en lugar de fijarse por la selección natural (Kimura, 1980). Kimura asume que la mayoría de las variantes moleculares que son polimórficas son afectadas por la interacción de la mutación, generando nuevas variantes y la deriva genética eventualmente fija las variantes (Hedrick, 1999). En el curso de esta deriva aleatoria, la mayoría de los alelos mutantes se pierden por azar, pero el resto se fija en la población. El tiempo promedio necesario para fijarse es cuatro veces el tamaño efectivo de la población, $4N_e$. El tamaño efectivo es aproximadamente igual al número de individuos que se reproducen de una generación y es menor al total de individuos de una especie (Kimura, 1980).

Tabla 1. Comparación del estimado de la diversidad nucleotídica en grupos de genes de diferentes especies.

	$\pi \times 10^{-4}$	$\theta \times 10^{-4}$	Estadística D de Tajima	Referencia
<i>MC1R</i>	10.1	64.2	-2.08	Savage <i>et al.</i> , (2008)
Genes del Telomero (n = 7)	3.0	5.5	-1.365	Savage <i>et al.</i> [2005].
Genes receptores de Hormonas Nucleares (n = 40)	4.1	7.5	ND	Pungliya <i>et al.</i> [2004]
Genes del procesamiento y presentación de Antígenos (n = 72)	4.7	8.3	ND	Pungliya <i>et al.</i> [2004]
Genes Referencia (n = 4950)	4.8	7.9	ND	Pungliya <i>et al.</i> [2004]
Genes de la respuesta farmacológica (n = 1598)	5.6	ND	Negativo	Schneider et al [2003]
Genes de respuesta al ambiente (n = 213)	6.7	ND	ND	Livingston <i>et al.</i> [2004].
Gen ADH1 en humano (n = 20)	0			
-en chimpancé (n = 22)	10.1			Jensen-Seaman <i>et al.</i> , 2001
-en bonobo (n = 10)	0			
-en gorila (n=28)	8.29			
Gen DRD2 en humano (n = 190)	0			
-en chimpancé (n = 71)	24.68			Jensen- Seaman <i>et al.</i> , 2001
-en bonobo (n = 33)	11.33			
-en gorila (n = 27)	0			
Gen DRD4 en humanos (n = 110)	9.68			
-en chimpancé (n = 32)	33.21			Jensen- Seaman <i>et al.</i> , 2001
-en bonobo (n = 12)	21.2			
-en gorila (n = 57)	14.03			
Gen HOXB en humanos (n = 210)	6.22			
-en chimpancé (n = 82)	20.39			Jensen- Seaman <i>et al.</i> , 2001
-en bonobo (n = 36)	21.15			
-en gorila (n = 58)	20.69			

ND: no determinada. Los valores de π del gen *MC1R* de la melanocortina son altos en relación a otro grupo de genes (Savage *et al.*, 2008). También, entre especies de primates no humanos, se ha reportado que el gen *DRD4* tiene altos valores de π .

2.2 NEUTRALIDAD Y PREDICCIONES DE VARIACIÓN MOLECULAR

El problema que ha causado controversia por años, ha sido describir qué factores son los responsables de la variación en los loci. De aquí divergen dos fuertes tendencias teóricas, una que dice que la mayoría de las variaciones fueron mantenidas por selección de diferentes tipos como la balanceadora; y la de neutralidad, que sostiene que la gran mayoría de variantes moleculares son fijadas aleatoriamente (Fontdevila, 1978).

Como se mencionó en la sección anterior, Kimura (1968) propone un modelo en donde la selección tiene un papel menor en la determinación del mantenimiento de las variaciones y en donde los diferentes genotipos moleculares tienen casi la misma adecuación (Hedrick, 1999). De esta manera, la mayoría de las sustituciones de nucleótidos que se producen en el transcurso de la evolución debe ser el resultado de la fijación al azar de mutantes neutros, más que el resultado de una selección. Así la persistencia de la variación neutra en una población se debería al equilibrio existente entre la aportación del polimorfismo por mutación y su eliminación al azar, provocando cambios aleatorios en sus frecuencias génicas causadas por el tamaño finito de las poblaciones (deriva genética).

En este trabajo de tesis para tratar de entender los patrones de variación entre las poblaciones, se midieron la diversidad nucleotídica y examinamos si la variación es consistente con las predicciones de la Teoría Neutral para distinguir los efectos de la selección de los efectos del desequilibrio demográfico (panmixia).

2.3 PRUEBAS DE NEUTRALIDAD

El Modelo Neutral está basado en condiciones de panmixia, tamaño constante y el Modelo de mutación de Sitio Infinito (ISM). El ISM señala que las mutaciones recurrentes no ocurren en el mismo sitio. Tajima introdujo un estadístico que es ampliamente usado para probar la hipótesis nula de una población en equilibrio y tamaño constante (Teoría Neutral). Tajima consideró dos estadísticos: (π) el número promedio de diferencia de

nucleótidos por sitio entre dos secuencias y el número de sitios que segregan (S). Bajo la hipótesis nula, los valores esperados de ese estadístico son:

$$\begin{aligned} E[\pi] &= \theta \\ E[S] &= a_1 \theta \end{aligned}$$

donde $\theta = 4Nu$, N es el tamaño efectivo de la población, $2N$ es para el tamaño de población haploide, u es la tasa de mutación por generación, y a_1

$$a_1 = \sum_{i=1}^{n-1} \frac{1}{i}$$

en donde n es el tamaño de la muestra (el número de secuencias de DNA estudiadas).

Bajo la hipótesis nula, π y S/a_1 ambos estiman a θ , de forma que ambos pueden tener valores similares. Si estos estadísticos tienen valores iguales entonces no se puede rechazar la hipótesis. Por el contrario, si tienen valores diferentes se rechaza la hipótesis. Tajima formula la varianza de una muestra con los estadísticos descritos que está definida por D:

$$D = \frac{\pi - \theta}{\sqrt{\text{var}(\pi - \theta)}}$$

el denominador es la raíz cuadrada de la varianza de la diferencia de π y θ .

Una D de Tajima con valores negativos indica un exceso de polimorfismos de baja frecuencia, sugiriendo una expansión poblacional y/o selección disruptiva o purificadora (descrita en la sección 3.2). Una D positiva definiría bajos niveles de polimorfismos de baja y alta frecuencia, indicando una disminución del tamaño de la población y/o selección balanceadora.

Otra manera frecuente de estimar la neutralidad es a través del estadístico de D de Fu y Li (1993). Este estadístico considera que las mutaciones se distribuyen en el ramaje interno y externo de un árbol genealógico. Las mutaciones más viejas están genealógicamente en la parte interna del árbol y las nuevas en la parte externa. Bajo la hipótesis de neutralidad no debería haber grandes diferencias entre π y θ para el número

de *singletons* (polimorfismos que aparecen solo una vez), en relación al número total de polimorfismos.

El cálculo de valores de D nos permite determinar si las mutaciones están sujetas a selección direccional o purificadora (también llamada negativa porque elimina alelos), cómo consecuencia de un exceso de mutaciones deletéreas y cuyas frecuencias son bajas, o a una selección positiva. De esta forma al ser eliminados los sitios que se encuentran en desequilibrio de ligamiento, podríamos esperar un incremento de las variantes raras *versus* los polimorfismos de frecuencias más altas, lo cual es consistente con el crecimiento poblacional o selección positiva (Thomas *et al.*, 2006). La selección positiva crea mutaciones sobre el ramaje externo de un árbol genealógico, la mayoría de estas variaciones se espera que sean jóvenes en aparición, y si estas mutaciones llegan a tener alguna ventaja podrían llegar también a fijarse en la población.

Valores negativos de la D de Tajima (y de manera similar de la prueba de F_u y L_i), son interpretados cómo evidencia de una selección positiva.

Por otra parte, en el caso de la selección balanceadora habría pocas mutaciones externas, las mutaciones viejas serían reemplazadas por nuevos alelos con frecuencias intermedias. Este escenario se relacionaría con una reducción poblacional de las frecuencias poblacionales. Estas fuerzas alternadamente influenciarán a π pero afectarán muy poco a θ , en la prueba D de Tajima. Los valores de D en este caso serían positivos.

2.4 PATRONES DE DIVERSIDAD HUMANA

El estudio de la variación genómica se ha desarrollado mediante el diseño de plataformas que proveen bases de datos dirigidos a entender la estructura del genoma y su evolución. Una fuente de información está contenida en el *GenBank* en la base de dbSNP, que es la primera fuente de variación a pequeña escala (www.ncbi.nlm.nih.gov/snp), este es el banco en donde se depositan los datos de todos los proyectos públicos, el cual lleva detalles de la información tanto individual como poblacional de las variaciones genómicas.

En promedio hay cuatro posibles variaciones SNP por gene, con una aparición esperada de 120000 a 219000 a lo largo del genoma (Kruglyak y Nickerson, 2001), y la frecuencia de SNP no sinónimos está entre 50,000 y 876,000 (40%), además de que la diversidad aumenta en las regiones intergénicas. Hasta abril 17 del 2008, *Homo sapiens*

tiene 3,706,296 SNPs que han sido ingresados en el *NCBI variation Database* (dbSNP) (www.hapmap.org).

La distribución de la variación nucleotídica a lo largo del genoma no es al azar, sino que se debe a la acción de las fuerzas evolutivas y a la dinámica poblacional (Reich *et al.*, 2002). En este sentido, algunas regiones del genoma muestran altos niveles de diversidad, el cual parece ser consistente con características estructurales del genoma humano, como es el caso de las secuencias repetidas, secuencias ricas en GC, islas CpG, distancia de los telómeros y centrómeros, regiones de alta variación (Hellmann *et al.*, 2005). Por ejemplo, los datos reportados en el HapMap, para el gen DRD4, ubicado en la región subtelo mérica, con dos islas CpG, está identificado en una región rica en SNPs (Figura 1).

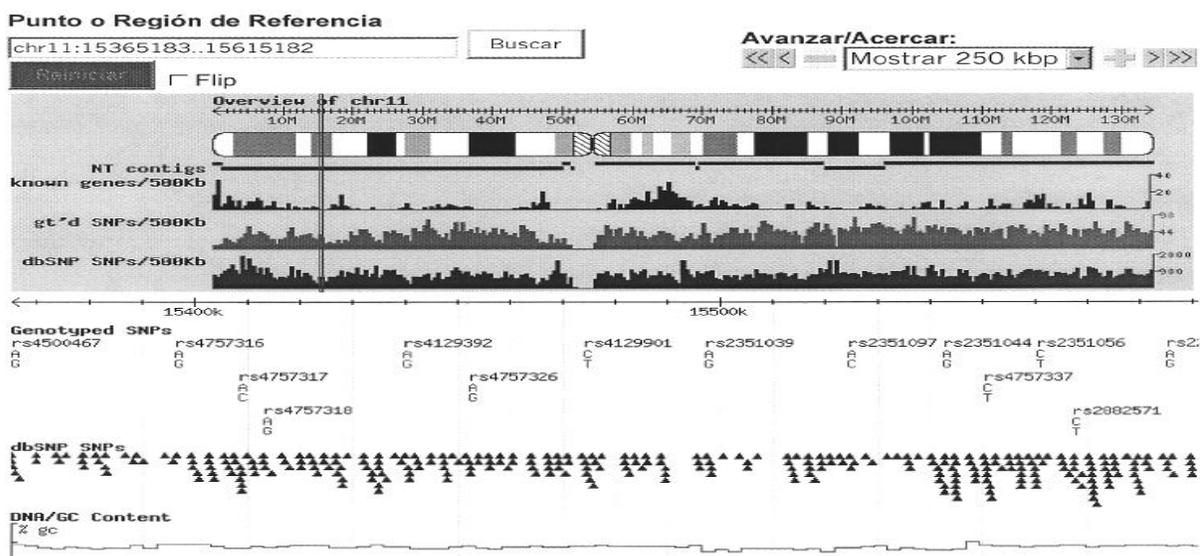


Figura 1. Región que detalla gráficamente los SNP que están dentro del brazo corto del cromosoma 11 (15.5), en donde se mapea el gen DRD4. Internacional HapMap Proyecto.

3. FACTORES QUE FORMAN PATRONES DE LA DIVERSIDAD

Los patrones de la variabilidad se forman y son sensibles a factores como mutación, que a su vez serán fijados por el efecto de la selección o neutralidad y puede observarse a lo largo de la evolución del gen. Sin embargo, hay correlaciones en los niveles de diversidad sobre las distancias, las cuales pueden ser muy cortas tanto como de 100 pares de bases hasta 1000 kb que son consistentes con los patrones de desequilibrio de ligamiento (Reich *et al.*, 2002).

3.1 TASAS DE MUTACIÓN Y PARÁMETROS DE MUTACIÓN NEUTRAL

Según la estimación de Nei (1987), la tasa de mutación espontánea, es difícil de medir adecuadamente debido a que es muy baja, además de ser crucial para estudios de evolución. Por ejemplo, la tasa de mutación que producen efectos letales que afectan el fenotipo es del orden de 10^{-5} por locus por generación en organismos como el hombre, *Drosophila* y maíz, aunque esas mutaciones contribuyen muy poco a la evolución.

La tasa de sustitución nucleotídica de pseudogenes es más alta que la de genes funcionales, y bajo la hipótesis nula de mutación neutral la tasa de mutación se estima en 4.7×10^{-9} por sitio por año (Nei, 1987). Como ejemplo, los genes de alfa globina de mamíferos (423 nucleótidos y 141 codones) tienen una tasa de mutación por gen (solo codones) estimada de 2.0×10^{-6} por año (Nei, 1987).

Recientes estudios han encontrado que la tasa de mutación neutral está entre 1.3×10^{-8} y 3.4×10^{-8} con un promedio de 2.5×10^{-8} (Nachman y Crowell, 2000). En general los genes muestran diferencias en su tasa de mutación y se puede relacionar con su ubicación en el gen, es decir, se asocia a si se encuentra en una región codificante, en intrones, o regiones CpG y repetidas.

La tasa de mutación puede afectar la densidad de SNPs y por consiguiente los valores de π y θ . Si se asume que la región no es funcional y que la tasa de mutación es

alta, el número de mutaciones con alelos de baja frecuencia, podría ser grande afectando a θ más que a π , de ser así la D de Tajima (descrita anteriormente en la sección 2.3) obtenida indicaría un exceso de variantes raras. En cambio, regiones con alta tasa de mutación promueve las mutaciones raras y crea débil correlación entre los SNPs cercanos.

3.2 SELECCIÓN MOLECULAR

La genética evolutiva tiene el papel de profundizar en los mecanismos que han ocurrido a través del tiempo en las estructuras genómicas actuales. La selección natural es el más importante de los factores que determinan cambios evolutivos valederos (Mettler y Gregg, 1982).

La base de la variabilidad se concentra en las mutaciones, en especial las que provocan sustituciones en la secuencia de aminoácidos, las cuales podrían tener un efecto sobre la *adecuación* de sus formas. La *adecuación* de un genotipo mide su relativa capacidad para reproducirse a sí mismo comparado a otros genotipos, el cual es favorecido por la selección. Por ejemplo, las sustituciones deletéreas de aminoácidos pueden originar deformaciones congénitas o la aparición temprana de una enfermedad, lo cual reduciría la oportunidad de que el acarreador pase como mutación a la siguiente generación.

No obstante, la mutación es selectivamente negativa y se mantiene pero en baja frecuencia. Todo lo contrario sucede con las mutaciones que tienen un papel beneficioso, dichas mutaciones mejoran la supervivencia del portador, lo cual serán seleccionadas en la edad reproductiva y aumentan positivamente la probabilidad de que llegue a ser polimórfica o a ser fijada en la población. Otro suceso sería que las mutaciones no tuvieran un efecto sobre la *adecuación*, lo cual llegarían a ser ligeramente deleterias o a tener un efecto pequeño en la *adecuación*. Dichas mutaciones serán consideradas neutrales y su frecuencia en la población está sujeta al efecto de la deriva génica (Sunyaev *et al.*, 2003).

La deriva genética es un factor que disminuye la variación hereditaria en las poblaciones. Implica fluctuaciones fortuitas en las frecuencias de los alelos debido a errores de muestreo, y existe una tendencia a la fijación de un alelo u otro, especialmente en poblaciones muy pequeñas. Es un factor que puede actuar junto con la selección en la

reformación de las variaciones existentes, pero a diferencia de la selección, la deriva es indeterminada (Mettler y Gregg, 1982).

La selección dentro de las poblaciones se divide en balanceadora, direccional y disruptiva. Estos términos son los actos que discriminan, conservando o rechazando fenotipos como progenitores de una generación.

En la balanceadora, la selección actúa de modo constante en toda la historia de la población para conservar los genotipos que dirigen la formación de genotipos mejor adaptados al conjunto de condiciones que haya en un tiempo dado. Si una población vive en un medio constante durante varias generaciones puede tener oportunidad de alcanzar un alto grado de adaptabilidad (que se mide con el valor de W). Si esto ocurre es probable que la variación genética se haga estable y las frecuencias génicas de la mayoría de los loci estén cerca de los valores de equilibrio. La selección opera y adquiere un papel estabilizador.

De acuerdo a la selección balanceadora, dos ó más alelos en un locus particular pueden tener una selección ventajosa, así los alelos serán mantenidos en frecuencias intermedias. La selección se puede categorizar más a fondo en la selección dependiente de la frecuencia negativa, donde la *adecuación* del alelo disminuye cuando llega a ser el más común, y donde los heterocigotos tienen una ventaja selectiva sobre homocigotos (Banshad y Wooding, 2003). Los polimorfismos que son selectivamente balanceados son influidos por el desarrollo y adaptación de los humanos a diferentes ambientes, inmunidad y protección contra agentes infecciosos. Un ejemplo puede ser el loci HLA (Hughes y Yeager, 1998), la Glucosa -6-fosfato deshidrogenasa y la protección contra la malaria, CCR5 polimorfismo en el promotor y la protección contra HIV (Bamshad *et al.*, 2002). Muchos de los polimorfismos que ocurren en las regiones codificantes de los genes son el resultado de la selección. En este sentido, muchas de las regiones que codifican a lo largo del genoma humano no tienen exceso de alelos raros, lo que podría indicar que lo más común es la selección balanceadora.

La selección direccional o positiva. Aquí las variantes más raras que estuvieron relativamente mal adaptadas pueden así volverse las elegidas y preferirse algún carácter fenotípico óptimo distinto de la norma ancestral. Se favorece una variante con respecto a otras en un locus, dicha variante aumentará su frecuencia en la población hasta fijarse.

Los alelos que son seleccionados positivamente incrementan su frecuencia junto

con el polimorfismo neutral que lo rodea (Nachman, 2001). En el proceso habrá una disminución en el número de polimorfismos con alta frecuencia que ya existían y un incremento de variantes raras, esto también lo indica una D de Tajima (ver la sección 2.3) negativa, en donde se espera que el DL incremente con los nuevos sitios ligados. El grado de la selección alrededor de las mutaciones estará sujeto a la recombinación. En este caso el DL se puede usar como un indicador de selección positiva, debido a que el DL se rompe rápidamente después de eventos de recombinación y de un barrido selectivo (Przeworski, 2002).

Selección disruptiva. Esta actúa de forma contraria a la selección estabilizadora por cuanto atañe a múltiples fuerzas que actúan sobre diversas formas de una sola población. Cuando la selección favorece más de una condición fenotípica óptima, es decir, genotipos muy diferentes, y discrimina los intermedios, se le llama selección disruptiva. Los resultados pueden tener una de dos formas básicas: 1) polimorfismo, 2) divergencia y aislamiento, en que las condiciones óptimas están formadas por individuos de subpoblaciones en adaptación local.

4. DINÁMICA, HISTORIA Y ESTRUCTURA POBLACIONAL

Las dinámicas de poblaciones forman la diversidad a través del genoma, y la selección puede formar la diversidad a un nivel local y afectar el DL.

El crecimiento poblacional reduce la deriva génica, como resultado del balance entre la mutación y la deriva se inclina hacia la mutación, incrementando el número de alelos con baja frecuencia y reduciendo el DL. Para que este efecto se extienda va depender de la tasa de crecimiento, lo cual interrumpe el DL si la población crece rápidamente (Pritchard y Przeworski, 2001).

Por el contrario, una disminución del tamaño de la población seguido por crecimiento (un cuello de botella), reduce la diversidad genética e incrementa el DL entre los marcadores que no han sido eliminados en el tiempo.

También la distribución poblacional puede tener un efecto similar al de cuello de botella, provocando una disminución de la diversidad y un incremento en el DL en la nueva población fundadora.

La estructura de población es un aspecto muy estudiado en la genética de las poblaciones, refiere a cualquier desviación del apareamiento al azar y está influenciado por fenómenos como endogamia, apareamiento asociativo (donde la reproducción es estratificado entre genotipos) y por subdivisión geográfica. En la estructura geográfica, los organismos no se dispersan al azar, en su lugar ellos tienden a mantenerse cerca del lugar de donde nacieron. Bajo estas circunstancias, las diferencias genéticas y fenotípicas pueden acumularse entre las poblaciones, se presentan entonces apareamientos no al azar por la cercanía local.

Dentro de las poblaciones también hay sub-estructuras que pueden ser afectadas por la deriva génica en la diversidad. La dispersión demográfica entre sub-poblaciones (mestizaje), podría originar DL, pero estos dependen del tiempo del mestizaje. El modelo de la estructura de la población asume que las sub-poblaciones seguían siendo distintas y proporcionan la relación entre el genotipo y las frecuencias de alelos en la población entera. Estos acontecimientos resumen el modelo de la adición, que asume la modificación de algunas sub-poblaciones por la afluencia de alelos de otras sub-poblaciones, dicho efecto de las subdivisiones se podría reducir en las siguientes generaciones.

En estudios de asociación casos-control, este problema de mezcla poblacional o estratificación podría originar falsos positivos en la asociación y ausencia del equilibrio de

Hardy-Weimberg.

Cuando una población es subdividida, puede haber eventos genéticos que logran hacer grupos, como la cantidad de flujo génico entre los subgrupos. El flujo génico indica que el movimiento entre grupos da como resultado un intercambio genético. Cuando el flujo génico es alto, la variación genética tiene un efecto homogenizador entre los grupos. Por el contrario, cuando el flujo es bajo, la deriva génica, la selección y la mutación en ambos grupos pueden provocar una diferenciación genética (Hedrick, 2000).

Una consideración que estima la cantidad de diferencia en la subdivisión de una población consiste en el calculo del coeficiente F (desarrollado por Wright, 1951), y se usa para localizar la variabilidad genética de la población total (T), las subdivisión (S), y la individual (I), que son los valores de F_{ST} , F_{IT} y F_{IS} , que se Interrelacionan,

$$F_{ST} = \frac{F_{IT} - F_{IS}}{1 - F_{IS}}$$

F_{ST} es una medida de diferenciación genética entre poblaciones. F_{IS} y F_{IT} son medidas de desviación de la proporción de Hardy-Weinberg dentro de las subpoblaciones y en la población total. Los valores positivos indican una disminución de los heterocigotos y los negativos indican un exceso de estos (Hedrick, 2000).

5. APLICACIÓN DE LA VARIACIÓN GENÉTICA.

Para tratar de entender la contribución de las diferentes fuerzas que actúan sobre el genoma de los organismos, es necesario llevar a cabo la descripción de los patrones de la variación genética dentro y entre las poblaciones naturales. En este sentido, una alternativa de particular relevancia para humanos y organismos no-modelos es analizar la variación genética dentro y entre especies, con el objeto de identificar regiones genómicas que parecen haber evolucionado bajo selección natural más que por deriva genética solamente. Por ejemplo, un método es contrastar el polimorfismo y la divergencia en sitios sinónimos y no sinónimos para detectar evolución no neutral en las secuencias de los aminoácidos (Yan y Bielawski, 2000). Una sustitución selectivamente benéfica puede ser detectada en los datos polimórficos, lo cual pudo haber ocurrido recientemente. La fijación de este alelo favorable en la población distorsiona patrones de variación en sitios ligados; por lo tanto, conlleva a distinguir señales o marcas en al menos 10^4 generaciones en humanos, y a distinguir datos polimórficos bajo el efecto de selección positiva, aumentando la frecuencia alélica que puede alcanzar un valor de la *adecuación* (Przeworki *et al.*, 2005, Sabeti *et al.*, 2002).

Por otra parte, las marcas o señales de barridos selectivos indican la presencia de la variación genética que puede tener algún efecto significativo sobre la variación fenotípica humana. Aunque las marcas de la naturaleza de la selección no son muy claras, muchas de esas variantes afectan fenotipos complejos de importancia médica (Voight *et al.*, 2006). A este respecto hay varios reportes que señalan que hay un determinado número de genes que muestran señales de fuerte y reciente selección que pueden favorecer nuevos alelos. Un ejemplo es el caso de la respuesta a la malaria (Tishkoff *et al.*, 2001), el gen de la deficiencia a la lactasa y su implicación en un error innato del metabolismo (Bersaglieri *et al.*, 2004), una variante que sensible a la concentración al NaCl en respuesta al clima (Thompson *et al.*, 2004), y en los genes implicados en el desarrollo del cerebro (Evans *et al.*, 2005).

A la fecha, los mejores ejemplos de selección reciente en seres humanos han sido descubiertos en estudios de genes candidatos a enfermedades donde se había planteado una hipótesis previa de selección natural.

En este sentido, ya se han construido mapas de alelos (SNP) con reciente selección positiva y con diferente *adecuación* en el genoma humano. Estos loci de SNPs pueden ser utilizados en estudios de mapeo de enfermedades o características complejas

tales como la diabetes, cáncer, enfermedades cardíacas y psiquiátricas, depresión, asma etc. (Voight *et al.*, 2006).

A esto se suma la identificación de genes en enfermedades complejas a partir del uso de metodologías como el desequilibrio de ligamiento, análisis propuesto como una poderosa herramienta para encontrar genes de modesto efecto que contribuyen a la susceptibilidad de enfermedades y revela la historia y orígenes humanos debido en parte a la historia poblacional (Pritchard y Przeworski, 2001).

Además los patrones de conservación nucleotídica, un parámetro cuantificado por la entropía (basada en la teoría matemática de Shannon, 1948), pueden también llegar a ser puntos clave en la identificación de mutaciones asociadas a enfermedades.

La entropía de una variable al azar se define en términos de su distribución de probabilidad y se puede demostrar para ser una buena medida de aleatoriedad o de incertidumbre (Parkhomchuk, 2006)

En la actualidad, se han determinado patrones de conservación para 6185 mutaciones no sinónimas heredables asociadas a enfermedades en más de 230 genes identificados con la información filogenética (Mooney y Klein, 2002).

6. ESTRUCTURA DEL GEN QUE CODIFICA AL RECEPTOR TIPO 4 A DOPAMINA (DRD4) Y SU VARIACION POLIMÓRFICA

6.1 CARACTERIZACIÓN MOLECULAR DEL DRD4

La familia de receptores a dopamina originalmente fue agrupada en dos tipos de receptores (D1 y D2), e incluye 5 receptores. Los receptores de la familia D1 agrupan a D1 y D5, cuya diferencia principal consiste en su distribución, porque su farmacología y funcionalidad es indistinguible. En tanto que la familia de receptores D2 incluye dos formas originadas por *splicing* (corte y empalme) alternativo del receptor D2, el D3 y el D4. Estos receptores tipo D2 tienen la característica común de inhibir la adenil ciclasa y activar otras vías de traducción de señales (Van Tol, 1996).

En 1990, se empleó el uso de sondas del receptor D2 con el objeto de identificar nuevos subtipos de receptores mediante homología con el receptor D1 y D2. Bajo esta metodología se lograron hibridaciones con dicha sonda en líneas celulares de neuroblastoma de ratón, en células de riñón de hámster y en células de neuroepiteloma humano. A partir del tamizaje de las bibliotecas de cDNA se aislaron clonas de cDNA parciales que tenían homología con el receptor D2 (Van Tol *et al.*, 1991). Los estudios farmacológicos detectaron un subtipo de receptor a dopamina, el cual fue denominado receptor a dopamina D4.

Basados en la homología con los receptores D2 y D3, la estructura completa de la región codificante para el receptor humano D4 fue identificado en una región genómica de 4.5kb de DNA (Van Tol *et al.*, 1991). Estudios paralelos de hibridación in situ localizaron el fragmento hacia la región telomérica del brazo corto del cromosoma 11 (p15.5) (Gelernter *et al.*, 1992). Se organiza en 4 exones y 3 intrones (Figura 2). Por otra parte la región promotora se localiza en el extremo 5' cercano al codón de inicio, entre los nucleótidos -591 y -123 donde inicia la transcripción, siendo una secuencia rica en GC. Adicionalmente, se ha descrito un sitio de modulación negativa, entre los nucleótidos -770 y -679 (Kamakura *et al.*, 1997).

La secuencia nucleotídica presenta variantes o polimorfismos en diferentes regiones a lo largo del gen. Particularmente, son relevantes para este trabajo los descritos en la región promotora, así como en los exones 1 y 3 (Figura 2) (Choy Wong *et*

al., 2000). Las frecuencias alélicas correspondientes a estos SNP y VNTRs, han sido descritas en diferentes grupos étnicos (Jönsson *et al.*, 1999, Chang *et al.*, 1996).

6.1.1 VNTR EN EL EXÓN 3

Posterior a la clonación del gen humano DRD4, Van Tol *et al.* (1991) demostraron que en la región correspondiente al exón 3 se presenta una región minisatélite tipo VNTR con una unidad repetida de 48 pares de bases (pb), y en donde se localizaron alelos desde 2 hasta 10 unidades de repeticiones (R) Lichter *et al.*, 1993; Van Tol *et al.*, 1992). Posteriormente, otros grupos confirmaron estos hallazgos y se llegó a describir hasta un alelo de 11R. Estudios realizados a lo largo de estos años en múltiples grupos de población han mostrado que con contadas excepciones, los alelos más comunes corresponden en orden de frecuencia decreciente al alelo de 4R con una frecuencia promedio del 64.3% en población de origen Caucásico, el segundo alelo más común es el de 7R con una frecuencia promedio de 20.6%, en tanto que el tercero más común es el alelo de 2R con una frecuencia de 8.2%. Cabe mencionar que estas frecuencias son altamente variables, ya que por ejemplo el alelo 4R parece estar aumentado en el continente Americano (frecuencia de 48.3%), en tanto que en poblaciones de Asia el alelo 7R es poco frecuente, y el alelo de 2R tiene una frecuencia promedio del 18.1% (Chen *et al.*, 1999). La tabla 2, modificada del reporte de Lung *et al.* (2002), muestra la relación de los alelos entre 2R a 8R entre sujetos caucásicos y asiáticos.

Estas variaciones han sido empleadas como marcadores del gen DRD4 en asociación con múltiples trastornos psiquiátricos incluida la esquizofrenia.

Tabla 2. Distribución de alelos del VNTR exon 3 del gen DRD4 entre caucásicos y asiáticos (tomado de Lung *et al.*, 2002)

Variación alélica	Caucásicos (%) n= 615	Asiáticos (%) n= 274
2	6	7.2
3	1.5	0.36
4	64.4	83
5	1.6	5.4
6	0.5	2.5
7	25	0.72
8	0.8	0.72

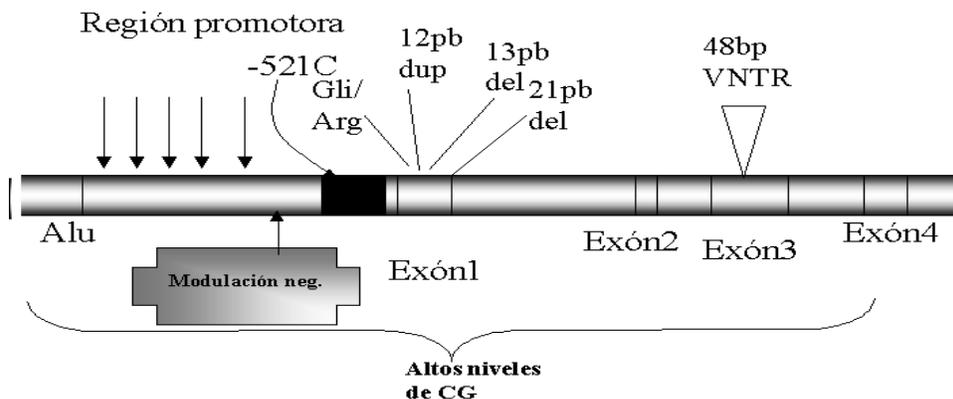
6.1.2 Variabilidad en la secuencia interna de el VNTR exón 3

Una característica importante e interesante de las secuencias repetidas en este polimorfismo, se refiere a la variabilidad en la secuencia interna que se ha venido reportando (Lichter *et al.*, 1993). Un ejemplo de lo anterior se describe en la figura 3.

Lichter *et al.* (1993) con base al análisis de secuencia de diversos alelos humanos, propuso inicialmente una nomenclatura establecida en el alfabeto griego, en donde el denominado alelo alfa (α) o uno, correspondía a la secuencia más frecuentemente encontrada, y sirve de base para codificar a los subsecuentes alelos.

Posteriormente, esta lista original se ha tenido que ampliar dada la caracterización de más alelos, por lo que se adoptó desde entonces una nomenclatura numérica secuencial (Ding *et al.*, 2002), que se emplea a la fecha.

Así, un polimorfismo VNTR en esta región particular podría estar formado por la combinación de diferentes alelos de secuencia, incrementando substancialmente la variabilidad de este sistema alélico. Esta combinación alélica de variantes de secuencia se define de esta manera cómo un haplotipo. En donde la secuencia de un alelo esta constituido por las unidades de 48 pb descritas en la figura 3.



Ejemplo:

Figura 2. Representación de la hebra sentido del gen receptor a dopamina D4, se señalan los exones y los espacios entre ellos, son los intrones. Distancia en pares de bases. Se puede apreciar un modulador negativo, y el rectángulo negro incluye el sitio de inicio de la transcripción. El rectángulo señala las variaciones del polimorfismo del VNTR del exón 3 (Aguirre y Nicolini, 2005).

En años recientes se han publicado algunos reportes sobre esta variabilidad tanto

en relación con trastornos psiquiátricos (e.g. TDAH, autismo, Grady *et al.*, 2003, 2005), cómo el análisis de la misma en función de procesos evolutivos de selección. Aspecto que se discute en forma más amplia en función de los datos obtenidos en este trabajo de tesis.

Cabe resaltar, que el análisis de secuencia de los haplotipos del VNTR del exón 3 tanto en humanos, como la de primates no humanos apoyan la idea de que la variante de 4R puede ser el alelo ancestral y el de 7R de aparición reciente, a partir de eventos de mutación y recombinación (figura 4) que incrementan su frecuencia en poblaciones humanas mediante selección natural (Livak *et al.*, 1995; Ding *et al.*, 2002; Wang *et al.*, 2004).

6.1.3 Polimorfismo -521 en la región promotora

En la región promotora del gen DRD4 está presente un polimorfismo del tipo SNP en el sitio -521 (Figura 2), en el que hay una sustitución de citosina por timina (C-T) (Okuyama *et al.*, 1999).

Los ensayos de expresión en células de retinoblastoma humano demostraron que la transición de C por T puede modificar los niveles de expresión del gen. En el ensayo las células con el alelo T, redujeron la eficiencia de transcripción del gen, (Okuyama *et al.*, 1999). Otros polimorfismos de la región promotora han sido identificados empleando enzimas de restricción y obteniéndose fragmentos de restricción de longitud polimórfica (RFLP) (Okuyama *et al.*, 2000). Sin embargo se desconoce si tiene un papel en la regulación del gen DRD4.

6.1.4 Polimorfismo del exón 1

Otra región polimórfica de importancia por su posible papel en el mecanismo de traducción proteínica corresponde al fragmento del exón 1. Este exón presenta un polimorfismo que codifica para 4 aminoácidos en el extremo amino terminal (región extracelular y dominios I y II), y que funcionalmente pueden ser importantes en la unión a dopamina (Schetz *et al.*, 2000). Se han descrito 3 polimorfismos en este exón que incluyen: un VNTR de 12 pb de 1 a 3 repetidas (Catalano *et al.*, 1993); Seeman *et al.*, 2000), una delección de 13 pb, únicamente descrito por el momento en población caucásica europea (Nothen *et al.*, 1994), y una delección de 21 pb (Cichon *et al.*, 1995) que podría codificar una proteína no funcional. El polimorfismo de 12 pb del exón 1 también ha sido descrito en otros primates de la familia de los póngidos (chimpancés, gorilas y orangutanes) con pocas diferencias interespecíficas (Seeman *et al.*, 2000), lo cual ha permitido establecer una relación filogenética y origen evolutivo en el hombre a partir de una divergencia o secuencia ancestral presente en los póngidos (Seeman *et al.*, 2000).

Además de las regiones polimórficas ya descritas, se han caracterizado otros sitios polimórficos, pero han sido muy poco usados como marcadores de diagnóstico molecular. (Choy Wong *et al.*, 2000).

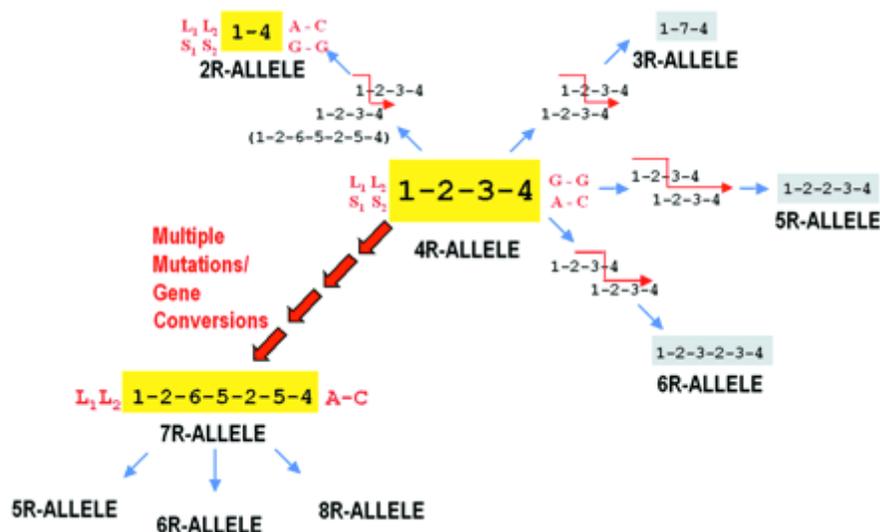


Figura 4. Modelo que explica el origen de las repeticiones en los diferentes alelos del VNTR del exón 3, del gen DRD4 (Ding *et al.*, 2002).

6.2 VARIACIÓN DEL GEN DRD4 EN DIFERENTES TAXONES

El taxón es un grupo de organismos emparentados, que son clasificados para ser agrupados, asignándole al grupo un nombre en latín, una descripción, y un "tipo". En este esquema organizativo, cada grupo de organismos en particular es un taxón.

Evolutivamente, la estructura de secuencias de repeticiones de la región del exón 3 del gen DRD4, se mantiene. Pero difieren en la composición de la secuencia en especies de primates no humanos (Livak *et al.*, 1995). También varía en algunas taxones del orden Carnivora (Inoue-Murayama *et al.*, 1998), así como en perros (Ito *et al.*, 2004), y algunas especies del orden Cetacea (Mogensen *et al.*, 2006). Esto enfatiza su importancia fisiológica y variación genética.

Es importante analizar la estructura y variación del gen DRD4 entre Primates no humanos debido a que son las especies más cercanas en el linaje evolutivo del *Homo sapiens*. La reconstrucción evolutiva de este gen se produce a partir de los eventos de mutación en la región de repeticiones de 48pb del exón 3, entre especies a lo largo de un árbol inferido, considerando la sistemática cladista (o sistemática molecular).

En este sentido, surge la pregunta bajo la consideración cladista de ¿cuál es el Primate no humano más cercano al humano en relación al VNTR de 48pb del gen DRD4? Esto es complicado de responder por la alta variabilidad de los repetidos internos, sin embargo algunas reconstrucciones filogenéticas sugieren que el chimpancé y el gorila (con alelos de 5R) son los más cercanos al hombre (Livak *et al.*, 1995). El chimpancé se separa del hombre aproximadamente 8 millones de años. La estructura moderna del gen DRD4 aparece como repeticiones de unidades de 48pb con sustituciones nucleotídicas. Sin embargo, el mono araña presenta alelos de 9R, y tiene duplicaciones en las estructuras de haplotipos que forman las unidades de repetición. También estas unidades de repetición difieren en mucho a la secuencia de haplotipos en humanos. Livak *et al.* (1995), señala que el mono ardilla (*Saimiri boliviensis*), que pertenece a los Cebidos, con múltiples repeticiones de 48pb no indican necesariamente que la estructura del gen sea primitiva, ya que el mono araña se separa del hombre aproximadamente 30 millones de años.

En cambio, se presume que el Primate ancestral tuvo una unidad de repetición de 48pb y la duplicación ocurrió a nivel de prosimios (excepto tarsioideos, grupo monofilético del orden Primate) (Inoue-Murayama *et al.*, 1998). Las evidencias sugieren que los Primates superiores evolucionaron hacia un incremento de las unidades de repetición

cambiando y conformando una función neurofisiológica. Tal es el caso de los alelos de 7R que son sobre-expresados en sujetos con dependencias a opiodes (Kotler *et al.*, 1997), y se han asociado a la característica de la personalidad de buscadores de la novedad o *novelty seeking* (Ebstein *et al.*, 1996).

7. PROTEÍNA RECEPTORA A DOPAMINA D4

7.1 ESTRUCTURA DE LA PROTEÍNA D4

Se han identificado cinco receptores a dopamina (D1, D2, D3, D4 y D5), de los cuales D1 y D5 forman a la familia D1. Los D2, D3 y D4 se clasifican en la familia D2, y fueron propuestos con base en su estructura/función, su respuesta farmacológica y su distribución anatómica (citado en Aguirre y Nicolini, 2005).

El gen DRD4 codifica una proteína que pertenece a la superfamilia de receptores acoplados a la proteína G (GPCR), y está organizada topográficamente en 7 dominios transmembranales (**TM**). La similitud entre los receptores D2 y D3 es del 40%, y dentro de la secuencia se preservan asas intra e extracelulares y colas, las cuales conforman los dominios transmembranales que forman los sitios de unión a ligandos como se ha observado en los receptores adrenérgicos (Van Tol, 1996). Esto coincide con el hecho de que los receptores dopaminérgicos tienen aminoácidos indispensables en la unión de ligandos como la asparagina en el dominio TM III, y 2 residuos de serina en el dominio TM V, los cuales son cruciales en la unión de dopamina y ligandos en varios receptores catecolaminérgicos (Strader *et al.*, 1989). También, del lado extracelular del dominio TM II y III, 2 aminoácidos son críticos en la unión selectiva de drogas como la piperazina, el raclopride, así como de la dopamina (Schetz *et al.*, 2000).

En analogía a los modelos creados para proteínas GPCR el extremo amino está en el lado extracelular y contiene un sitio N-glicosilado. El extremo carboxilo contiene una cisteína, como se ha descrito en receptores D2 y D3 y puede servir como sustrato para la palmitoilación (O'Dowd *et al.*, 1989), en donde el extremo amino es modificado por la adición de un ácido graso para formar proteínas N-aciladas. La forma más común de tal modificación es la adición de un grupo palmitoílico.

Una característica particular y única al receptor D4 en humano, es la presencia de

16 aminoácidos que se repiten en la tercera asa citoplásmica, la cual no ha sido identificada en el gen de la rata. La localización de esta secuencia repetida corresponde a la ubicación del corte y empalme de 29 aminoácidos de la secuencia del receptor D2 (Van Tol, 1996).

La supuesta secuencia de aminoácidos inferida a partir de la secuencia nucleotídica del D4, muestra que tiene un alto contenido de prolinas en la región de repetidas (de 30 a 40%), las cuales se han identificado como dominios potenciales de unión a SH3 o puentes disulfuro (Ren *et al.*, 1993).

La estructura de dicha proteína en 3D (tercera dimensión) es difícil de aislar debido a que es una proteína transmembranal, y es complejo llevar a cabo un protocolo de cristalización.

Un paso previo obligado a la resolución de la estructura de las moléculas de interés biotecnológico es su cristalización o difracción de rayos X sobre cristales. Esto es así porque es la técnica actual más eficaz para descubrir la estructura molecular de cualquier compuesto. Sin embargo, existen dificultades sobradamente conocidas para obtener cristales suficientemente grandes de buena calidad. Estas dificultades tienen que ver con los movimientos de convección y con la sedimentación, provocados por la gravedad terrestre, que se dan en un fluido con materia en suspensión, como el fluido en el que se precipita el compuesto que se va a cristalizar. El resultado son dislocaciones e imperfecciones en los cristales resultantes. Esto no es un problema insalvable cuando se trata de moléculas pequeñas, ya que se puede descubrir la estructura molecular a pesar de las imperfecciones, pero no cuando se trata de macromoléculas como el DNA o las proteínas

Pese a las dificultades un método apropiado para conocer parte de su estructura, es mediante simulación computacional, el cual predice la estructura en 3D (tercera dimensión) usando la secuencia de aminoácidos, como es el caso del receptor de dopamina D2 humano que ya ha sido visualizado en su estructura terciaria (Kalani *et al.*, 2004). Este tema se tratará en la sección 7.2.

En este sentido, de suma relevancia sería conocer más acerca de la región codificante de las repetidas del exón 3 del gen DRD4, ya que tiene importancia fisiológica debido a que participa en el proceso de activación de la proteína G inhibidora de la formación de AMP cíclico durante la neurotransmisión (Asghari *et al.*, 1995).

La región polimórfica del VNTR del exón 3, subyace entre los residuos citoplásmicos cuya posición de inicio está entre el aminoácido 249 al 312 (varía en

tamaño dependiendo de la longitud de la repetida) del péptido receptor D4 (UniProtKB n° de acceso P211917).

7.2 PREDICCIÓN MEDIANTE MODELAJE MOLECULAR

Dado el papel fisiológico y evolutivo de la región polimórfica sería conveniente observar como varían los aminoácidos dentro de la estructura proteínica. Sin embargo, como es una proteína perteneciente a la superfamilia de receptores GPCR, la visualización de estructuras en 3D resulta difícil debido a que se basa en la construcción de cristales de alta resolución y se requieren en una cantidad suficiente para obtener datos de difracción de rayos X de alta resolución y el uso de la espectroscopía de resonancia magnética nuclear (NMR) para determinar estructuras tales como dicha proteína transmembranal (Trabanino *et al.*, 2004). En la actualidad, no existen estructuras experimentales disponibles para cualquier GPCR humano, y sólo se puede predecir sitios de unión a ligandos y la energía de unión a dichos ligandos. Así los GPCR de interés farmacológico como el receptor a dopamina D4 sólo tienen datos de constantes de unión a ligandos, de agonistas y antagonistas (Tarazi *et al.*, 1997; Tarazi y Baldessarini, 1999).

Las aplicaciones de la construcción de modelos de estructuras proteínicas son varios: estudios de mecanismos catalíticos, diseño e improvisación de ligandos, acoplamiento de macromoléculas, tamizaje virtual y acoplamiento de pequeños ligandos, definición de epitopes de anticuerpos, reemplazo molecular en cristalografía de rayos X, diseño de quimeras, estable y variantes cristalizables; soporte de mutagenesis dirigida, refinamiento de estructuras NMR, colocación dentro de la densidad electrónica de baja resolución, relación funcional de similitud estructural, identificación de sitios de residuos conservados en superficie, identificación de sitios funcionales por motivos 3D (Baker y Sali, 2001).

Las primeras predicciones de la estructura y función de los receptores GPCR fueron a partir de la validación por la estructura cristalizada de la rodopsina bovina (Trabanino *et al.*, 2004). Las predicciones del receptor a dopamina D4 aun no son muy claras y están incompletas.

7.3 DISTRIBUCIÓN ANATÓMICA DEL RECEPTOR DOPAMINÉRGICO D4

El sistema dopaminérgico (SD) está distribuido en gran parte en el sistema nervioso central (SNC), pero se concentra en las neuronas de la vía nigroestriatal de la sustancia

nigra (A9), las cuales se proyectan hacia la región del neostriado. El SD también está ubicado en la vía mesocorticolímbica compuestas por la vía mesocorticolímbica formadas por neuronas del área tegmental ventral (A10), las cuales inervan a estructuras de la corteza límbica (figura 5). La vía dopaminérgica nigroestriada está relacionada con las disfunciones extrapiramidales, el cual se ha demostrado que es parte de la degeneración neuronal que ocurre en los cerebros de pacientes diagnosticados con la enfermedad de Parkinson (revisado en Civelli *et al.*, 1991), lo que ha sido objeto de tratamientos terapéuticos en sujetos con este padecimiento.

La vía mesocorticolímbica ha sido implicada como la principal vía en la etiología de la psicosis. Aquí también se aplica la hipótesis dopaminérgica (ver sección 8.2), que está asociada a la patogénesis de la psicosis, en particular esquizofrenia, o en aquellos estados de excesiva liberación de dopamina similar a la esquizofrenia, ya que los antipsicóticos usados en el tratamiento son antagonista de receptores a dopamina. Estos datos explican el dilema asociado a drogas usadas en el tratamiento de la liberación de dopamina. El bloqueo del sistema dopaminérgico reduce los síntomas psicóticos e induce un efecto extrapiramidal (revisado en Civelli *et al.*, 1991).

Por otra parte, específicamente, el RNA mensajero (RNAm) del receptor D4 ha sido detectado en varias regiones cerebrales de humano, mono y rata. Análisis de *Northern blot* han mostrado una distribución similar a la de los receptores D2 y D3. En el mono, el receptor D4 muestra altos niveles de RNA mensajero en corteza frontal, amígdala y médula espinal (Van Tol *et al.*, 1991). Y por hibridación in situ se observó la presencia del RNAm del D4 en altos niveles en giro dentado (Meador-Woodruff *et al.*, 1991). Aparte también ha sido detectado altos niveles en corazón de rata (O'malley *et al.*, 1992) y retina (Cohen *et al.*, 1992).

El DRD4 se ha localizado en el hipotálamo, tálamo, bulbo olfatorio, hipocampo (partes del sistema límbico), regiones corticales, corteza frontal, lóbulos occipitales y cerebelo (Berger *et al.*, 2001; Mulcrone *et al.*, 1997).

7.4 FARMACOLOGÍA DEL RECEPTOR D4

La familia de los receptores a dopamina acoplados a la proteína G se ubican en los botones postsinápticos del estriado; pero los D2 ayudan a regular la síntesis de dopamina y se acoplan a proteína Gi, la cual está involucrada en muchas funciones, como por ejemplo el movimiento. La activación del subtipo D1 y D5 conduce a la formación de

AMPC por estimulación de la adenilato ciclasa, el cual es un proceso mediado por la proteína G α s. Ambos receptores son activados por SKF 38393 y su antagonista es el SCH-23390.

La acción farmacológica del receptor D4 se basa en una cascada de activación proteínica, de este modo el D4 actúa como receptor activado por un ligando, activa a la proteína G y disocia las subunidades de su estructura conformacional e inhibe la actividad de la enzima adenilciclasa (AC) y del segundo mensajero AMPC (Longstaff, 2000). Estimula el transportador antiport Na⁺/H⁺, activa canales de potasio, produce un influjo de Cl⁻, y altera la morfología neuronal. Además, tiene un efecto intracelular como lo es la liberación del ácido araquidónico, hidrólisis fosfoinositídica, inhibición de canales de Ca²⁺ y apoptosis (Oak *et al.*, 2001).

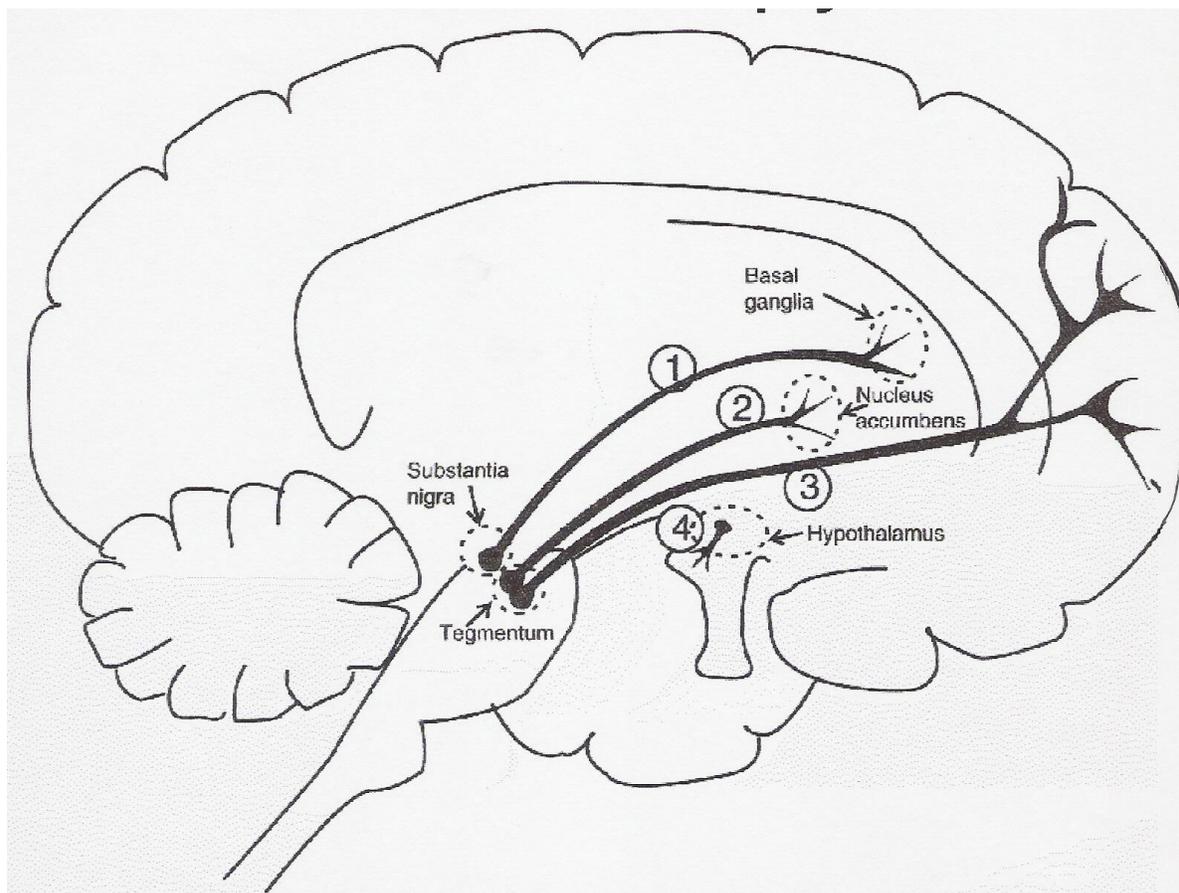


Figura 5. Principales vías dopaminérgicas (DA) en el sistema nervioso central en donde tienen efecto los antipsicóticos. 1) Vía DA nigroestriada relacionada con el movimiento. 2) Vía DA mesolímbica con función en el control del movimiento y emociones. 3) Vía DA mesocortical tiene un papel en la cognición y la planeación. 4) Vía DA tuberoinfundibular relacionada con la secreción de prolactina.

Estos receptores tipo D2 de forma independiente estimulan la actividad de cinasas como MAPK ERK1/2° c-Jun NH2-terminal (Luo *et al.*, 1998; Oldeh Hof *et al.*, 1998). Las MAPK son cinasas tipo serina/treonina que mediante cascadas de activación celular inducen la proliferación, diferenciación y proliferación a largo plazo, generalmente atribuidas a receptores de factores de crecimiento.

La activación de las MAP-cinasas no parece afectarse por la longitud del tercer asa citoplásmicas del receptor con alguno de los 3 alelos más comunes en el polimorfismo del exón 3 (Oak *et al.*, 2001).

La caracterización farmacológica del receptor D4 se ha llevado a cabo mediante la tecnología del DNA recombinante, por expresión de las células COS-7, sin diferencias aparentes, excepto por el hecho que pueden ser obtenido altos niveles de expresión del gen clonado en estos vectores de expresión (Van Tol, 1996).

Las características de los ligandos de unión al receptor D4 han sido determinadas por varios antagonistas y agonistas de los D2, sugiriendo que los perfiles farmacológicos son similares con algunas excepciones. Como la alta afinidad de la clozapina por el D4 (con una KI de 10 a 20nM) es diez veces más alta en comparación con el D2 (Ki de 100 a 200nM) y la pobre afinidad del antagonista de D2 el raclopride por el receptor D4 de 1 a 2µM (Ashgari *et al.*, 1994).

Un aspecto relevante del gen DRD4 radica en que la proteína que codifica tiene sitios de unión a fármacos, que pueden ser útiles en el tratamiento de las enfermedades mentales. Y es que el receptor D4 se localiza en las áreas límbicas cerebrales, y es un blanco para el diseño de drogas selectivas como los antipsicóticos (Aguirre y Nicolini, 2005).

Experimentalmente, se ha demostrado que el exón 3 con el polimorfismo de 2 repetidas, puede bloquear la actividad de la AC (Bouvier *et al.*, 1995). Por el contrario, la variante larga de 7R puede dar una respuesta reducida en cuanto al bloqueo de la AC y en la sensibilidad de unión del antipsicótico clozapina, lo que podría indicar una respuesta farmacológica diferencial de acuerdo a los polimorfismos del exón 3 (Van Tol *et al.*, 1992; Asghari *et al.*, 1995). Algunas evidencias sugieren la hipótesis de que la esquizofrenia se debe a un incremento del funcionamiento del sistema dopaminérgico, y esto es consistente con la acción de los antipsicóticos, como las fenotiacinas (clorpromacina). A pesar del efecto regulador de los síntomas, las fenotiacinas recetadas por períodos largos a esquizofrénicos, provocan reacciones secundarias como: efectos extrapiramidales

(acinesia, rigidez muscular y finos temblores de reposo), distonías, acatisia, discinesia tardía, efectos atropínicos (es decir, boca seca, variabilidad en la frecuencia cardíaca), hipotensión y reacciones alérgicas (Aguirre y Nicolini, 2005).

8. CARACTERÍSTICAS CLÍNICAS DE LA PSICOSIS EN ESQUIZOFRENIA

Según el Manual de trastornos mentales y estadístico (DSM) IV (American Psychiatric Association, 1994) los síntomas psicóticos hacen referencia a una conducta irracional, inapropiada y poco predecible, además de que puede ser potencialmente peligrosa.

Dentro de los síntomas psicóticos se puede distinguir un primer episodio psicótico, caracterizado por alucinaciones, delirios, alteraciones conductuales y deterioro en el funcionamiento global del individuo. El primer episodio psicótico es el inicio de la esquizofrenia. Lieberman lo describe en la historia natural de la esquizofrenia, como el deterioro de la capacidad funcional del individuo y el brote de los síntomas psicóticos, que aumenta en cada nuevo episodio y tiende a estabilizarse hacia la fase final de la edad adulta. Sin embargo desde antes de este episodio en un estado llamado prodromos ya se manifiestan los síntomas psicóticos y algunos síntomas negativos, así como cognitivos (citado en López Mato *et al.*, 2000).

No existe un cuadro clínico único, sino que hay múltiples síntomas característicos; como síntomas emocionales, cognitivos, de personalidad y de actividad motora. La sintomatología debe estar presentes durante no menos 1 mes y persistir durante al menos 6 meses.

Síntomas positivos: aumento o distorsión de las funciones normales como:

- Alucinaciones: definidas como percepciones irreales, que pueden ser de tipo auditivas, visuales, táctiles, olfativas o gustativas, siendo más frecuentes las 2 primeras.
- Ideas delirantes: refiere a las alteraciones del pensamiento como ideas falsas al razonamiento argumental.
- Lenguaje desorganizado e incoherente. Suelen ser ideas de persecución, de grandeza, religiosos, de celos e hipocondríacos.
- Comportamiento gravemente desorganizado como agitación, incapacidad de organizarse y de mantener la higiene personal o catatónico (con una disminución de la actividad psíquica y motora hasta llegar a una falta total de atención y rigidez).

Síntomas negativos: parecen reflejar una disminución o pérdida de las funciones normales. Los síntomas negativos comprenden restricciones como:

- Embotamiento afectivo: no reaccionan ante estímulos emocionales.
- Pobreza del habla (alogia).
- Abulia o apatía: falta de voluntad, incapacidad para persistir o para iniciar una actividad.
- Anhedonia: incapacidad para disfrutar de los placeres.

En México existen muy pocos estudios epidemiológicos de las enfermedades mentales. Caraveo *et al.* (1996) reportan que la prevalencia de la esquizofrenia es del 0.7%, en población urbana adulta, es decir afecta entre 500,000 a 700,000 personas en México. De acuerdo a los reportes de la Organización mundial de la Salud (OMS), a nivel mundial esta enfermedad está representada por el 1% de la población, entre 15 a 54 años de edad. Es una de las enfermedades más discapacitantes dentro de los trastornos mentales.

8.1 FACTORES ASOCIADOS A LA ESQUIZOFRENIA

Los factores de origen de la enfermedad son diversos, pero se aceptan los factores ambientales y los genéticos, ambos en combinación.

En este sentido, en la esquizofrenia se han propuesto varias hipótesis de origen como traumas intrauterinos y del nacimiento, problemas del neurodesarrollo (Lieberman *et al.*, 1997), exposición prenatal a infecciones virales (Fatemi *et al.*, 2002; Chen *et al.*, 1999) y agentes retrovirales (Karlsson *et al.*, 2001). Pero, debido a la complejidad y a lo limitante de la metodología para investigar estas hipótesis, poco se ha logrado para entender la etiología. Algunos modelos de esquizofrenia referente al neurodesarrollo e infecciones virales cerebrales en neonatos han tratado de explicar el fenómeno (Aitta-aho, 2003). Sin embargo, cabe señalar que las evidencias de los efectos ambientales representan sólo una parte del origen del problema. La mayoría de los expertos están de acuerdo en que una serie de factores combinados con un riesgo genético contribuyeron al desarrollo de la enfermedad (Aguirre y Nicolini, 2005).

Una nueva área de investigación como la epigenética propone que los genes implicados en la esquizofrenia pueden estar alterados en su expresión fenotípica. La epigenética es la rama de la biología del desarrollo que investiga el mecanismo por el cual cambios heredables en la expresión del gen ocurren sin que cambie el material genético. La variación epigenética puede estar dada por la adición de grupos metilo y acetilos en la estructura del DNA o en la cromatina impidiendo que se expresen los genes. Pero se desconoce cuánto puede influir el ambiente sobre este fenómeno. Se ha sugerido que mecanismos epigenéticos pueden tener un papel importante en el desarrollo de la esquizofrenia y psicosis (Petronis, 2000). Sin embargo en la psiquiatría el análisis epigenético todavía está muy pobremente estudiado. Recientemente, Petronis *et al.* (2003) en un estudio piloto con 2 parejas de gemelos monocigóticos, describieron una variación epigenética (en metilación de citosinas) en el promotor del gen receptor D2. Los gemelos con esquizofrenia presentaban una disminución de los patrones de metilación en comparación con su par no afectado. Lo que podría sugerir una susceptibilidad a padecer la enfermedad asociada a los eventos epigenéticos (revisado en Aguirre y Nicolini, 2005).

Tabla 3. Factores de riesgo ambiental biológico y Esquizofrenia

Situación ambiental biológica	Riesgo
Preclampsia	9%
Trauma obstétrico	7%
Incompatibilidad Rh o grupo	3%
Desnutrición (2do. Trimestre)	2%
Infección viral materna (2do trimestre)	2%

Secretaría de Salud de México, Lijeja 7, Col. Juárez, 06696-México D.F.

Los primeros estudios en genética psiquiátrica fueron de tipo epidemiológicos realizados en familias y en gemelos. Los estudios epidemiológicos mostraron que el riesgo de padecer esquizofrenia aumenta con el grado de parentesco (tabla 4).

Tabla 4. Relación de parentesco asociado con el riesgo de padecer esquizofrenia

Parentesco con esquizofrenia	Riesgo de padecer la enfermedad
Gemelo monocigótico	48 %
Gemelo dicigótico	18 %
Ambos padres	35 %
Un padre	12 %
Hermano	12 %
Tíos o primos	4%
Población general	1%

Los parientes de sujetos con esquizofrenia presentan un mayor riesgo que aquellos sin historia familiar con la enfermedad, empezando con el primer grado de consanguinidad y la estimación se observa en la tabla 4. Estos estudios en familia comenzaron en 1916, sin embargo en 1987 algunas investigaciones fueron realizados por Gottesman *et al* (1987). Dichos estudios han mostrado diferencias en el criterio de diagnóstico tanto en familias como en la población en general, lo que dificulta la interpretación. Recientes estudios realizados en la transmisión de la enfermedad en grupos familiares son más rigurosos y los criterios se han unificado, empleando análisis neurológicos y clínicos, cuyos resultados indican bajos índices de riesgo en comparación con los reportados por Gottesman. Aunque los estudios en familias presentan baja sensibilidad de predicción, en la actualidad se recopila información diagnóstica más directa y confiable, que consiste en entrevistar a cada uno de los familiares del paciente con esquizofrenia, en lugar de entrevistar solo a unos cuantos familiares, este método se denomina estudio familiar. Este tipo de estudio refleja que la probabilidad de que alguien padezca de esquizofrenia si tiene un familiar de primer grado es del 3 al 8% y varía el riesgo en función de la edad en que inicia el padecimiento y el predominio de los síntomas (Citado por Ortega Soto y Valencia Collazo, 2001).

8.2 HIPÓTESIS DE LA ESQUIZOFRENIA

El ácido gammaaminobutírico (GABA) es un neurotransmisor con función inhibitoria en el sistema nervioso central que se ha relacionado a la etiopatogenia de la esquizofrenia. La hipótesis de una disregulación GABAérgica en la esquizofrenia fue formulada por Roberts en 1972, considerando que una alteración en este sistema llevaría a una disminución del control inhibitorio que ejerce el GABA sobre las vías dopaminérgicas. Desde entonces, se han realizado numerosos estudios intentando verificar dicha hipótesis, con resultados contradictorios. En los últimos años y con el auge de la teoría dopaminérgica y del posible papel desempeñado por la serotonina, la hipótesis GABAérgica no ha estado tan vigente, aunque algunos resultados publicados recientemente vuelven a demostrar la implicación de este neurotransmisor en la etiopatogenia de este trastorno (revisado en la revisión bibliográfica de Baeza Pertegaz *et al.*, 2000).

Un polimorfismo descrito en el receptor GABA B que consta de repetidas (AC)_n del tipo dinucleótidos se encontró ser diferente en esquizofrénicos y en pacientes con historia familiar de esquizofrenia, sin embargo la corrección de Bonferroni diluyó la significancia (Imai *et al.*, 2000). Algunos estudios con fármacos agonistas GABAérgicos como el Baclofén, el muscimol y el ácido valproico administrados por separado a un número de pacientes esquizofrénicos, no mejoraron los síntomas psicóticos, algunos empeoraron. La mayoría de los estudios apuntan a que hay una disminución del GABA en sus estructuras, lo cual estaría relacionada directamente con la hipótesis dopaminérgica, una disminución del GABA en la corteza temporal aumentaría la dopamina y los síntomas positivos. Aunque los fármacos GABAérgicos no mejoran los síntomas, posiblemente porque no es el único sistema de neurotransmisores afectado (Baeza Pertegaz y cols, 2000).

La serotonina es un neurotransmisor involucrado en algunos comportamientos como el sueño, el humor, alucinaciones y depresión. Se han identificado 7 receptores: 5HT-1, 5HT-2, 5HT-3, 5HT-4, 5HT-5, 5HT-6, 5HT-7, con 5 subtipos de 5HT-2 y 3 distintos 5HT-3 (Barnes y Sharp, 1999).

Algunas drogas como reserpina reducen las concentraciones de serotonina en el cerebro, lo que podría corroborar su papel etiológico en la esquizofrenia, ya que se ha reportado que en estos pacientes los receptores de serotonina y sus sitios de recapturas están alterados en el sistema límbico. También se han estudiado polimorfismos en varios genes de esta vía serotoninérgica, mostrando una asociación positiva con el 5HT-2A en la respuesta con clozapina (revisado en Prasad *et al.*, 2002).

La mayoría de los estudios señalan una disminución en la concentración de GABA, sus receptores o enzima GAD en cerebros posmortem en la corteza frontal y temporal. Estos resultados también se han confirmado en dos únicos estudios realizados con resonancia magnética espectroscópica. Esta disminución estaría directamente relacionada con la hipótesis dopaminérgica. Así, una disminución del GABA a nivel temporal aumentaría la dopamina en esta zona, relacionándose con la producción de síntomas positivos como las alucinaciones auditivas o trastornos formales del pensamiento. Además, los estudios funcionales también sugieren una alteración en el sistema GABAérgico en esquizofrénicos (respuesta aplanada de la GH y prolactina después de administrar fármacos GABAérgicos), aunque la administración de fármacos que facilitan la transmisión GABAérgica en general no disminuyen la sintomatología psicótica, a excepción de una mejoría en el grado de ansiedad. A pesar de esto, la mayoría de los estudios indican que el sistema GABAérgico puede ser otro de los sistemas de neurotransmisión alterados en la esquizofrenia, sobre todo debido a la respuesta parcial de algunos pacientes a antipsicóticos que inhiben la vía dopaminérgica e incluso a los que afectan a otros neurotransmisores como la serotonina. Por ello, para poder conseguir un mayor conocimiento de la implicación del GABA en la esquizofrenia se necesitan más estudios, considerando primordialmente la neuroimagen funcional y los receptores *post mortem* como posibles áreas de investigación en el futuro (citado en la revisión bibliográfica de Baeza Pertegaz *et al.*, 2000).

Por otra parte, en 1963 se propuso la hipótesis dopaminérgica (Carlsson y Lindqvist, 1963, citada en Laruelle *et al.*, 1999), la cual predice una hiperactividad en la transmisión dopaminérgica, que era la responsable de los síntomas de la esquizofrenia. Una de las hipótesis más estudiadas.

Dentro de los sistemas de neurotransmisión, el dopaminérgico es uno de los más importantes, ya que está involucrado en los procesos de la cognición, la conducta y las

emociones. La hipótesis dopaminérgica comenzó a ser una de las principales teorías en la patología de la esquizofrenia, y se empezaron a desarrollar drogas antipsicóticas dirigidas hacia los receptores antagonistas dopaminérgicos. Se incluyen las drogas como el haloperidol, así como las fenotiacinas (tioridacina, flufenacina, trifluoperacina). Dichas drogas fueron las principales en el tratamiento de la esquizofrenia, a pesar de los serios efectos colaterales (mencionados en la sección 7.4).

En especial, los antagonistas de los receptores D2 mostraron una eficacia en el tratamiento de la esquizofrenia sustentando así la hipótesis de que este trastorno puede deberse en parte a una hiperactividad de la transmisión dopaminérgica. La característica farmacológica de la familia de receptores D2 es que inhiben la adenilciclasa y en consecuencia la formación de AMPc por activación de la proteína Gai, aumentando la acidificación extracelular por intercambiador Na^+/H^+ .

II.OBJETIVOS

Todo lo que una persona puede imaginar, otros pueden hacerlo realidad

Julio Verne

GENERAL

Analizar la diversidad nucleotídica formando haplotipos así como la distribución de las frecuencias de alelos y genotipos en distintas regiones polimórficas del gen DRD4.

PARTICULARES

A) Estudiar en la secuencia nucleotídica del exón 3 su estructura de haplotipos. Así como analizar la diversidad nucleotídica del VNTR-48pb.

B) Describir posibles evidencias de evolución molecular en los patrones de variación, que contribuyen en la reconstrucción de la historia de las poblaciones humanas y con el conocimiento de la diversidad genética.

C) Identificar sitios bajo presión selectiva, con desequilibrio de ligamiento y regiones conservadas para entender mejor la compleja relación entre fenotipo-genotipo y que podrían ser usados como marcadores en la identificación de enfermedades psiquiátricas o comportamientos

D) Analizar en un grupo de pacientes psicóticos con esquizofrenia y controles en una muestra de individuos de origen mestizo mexicano:

- Las variantes genéticas del VNTR de 48 pb ubicado en el exón 3
- El polimorfismo del VNTR de 12 pb del exón 1
- El polimorfismo del SNP –521 C/T, que se encuentra en la región del promotor del gen DRD4

III. MATERIALES Y MÉTODOS

The importance of the great principle of Selection mainly lies in the power of selecting scarcely appreciable differences...which can be accumulated until the result is made manifest to the eyes of every holder

Charles Darwin (1859)

1. SUJETOS

Este estudio se realizó considerando la Declaración de Helsinki, y bajo las normas del Comité Científico y de Ética del Instituto Nacional de Psiquiatría *Ramón de la Fuente*, ciudad de México. Todos los sujetos que participaron en este estudio dieron su consentimiento por escrito, y fueron incluidos los que tuvieran abuelos de origen mestizo mexicano, no relacionados. El grupo afectado fue de 149 (79 hombres y 70 mujeres); entre los 14 a los 58 años (promedio = 29 ± 8) los cuales experimentaban su primer episodio psicótico. Fueron evaluados mediante la prueba psicométrica "Schedule for the Clinical Assessment in Neuropsychiatry" (SCAN), y el diagnóstico cumplía con los criterios del DSM III-R [American Psychiatric Association, 1987].

La muestra consistió en un grupo de pacientes con primer episodio psicótico. Los estudios con pacientes de un primer episodio proporcionan una metodología adecuada para homogenizar la variabilidad que existe en el curso del desarrollo de la esquizofrenia. También permite ser más preciso en el diagnóstico, la fisiopatología y el tratamiento, en especial en la esquizofrenia. Además de eliminar el efecto de la cronicidad, incluye a pacientes sin tratamiento farmacológico, lo que permite una mejor evaluación clínica e identificar aspectos biológicos que se relacionan con la etiología del trastorno mental (Apiquián *et al.*, 1999).

El grupo de pacientes presentaron los siguientes desórdenes psiquiátricos: esquizofrenia paranoide, esquizofrenia indiferenciada, esquizofrenia desorganizada, trastorno esquizoafectivo, trastorno esquizofreniforme, depresión mayor con síntomas psicóticos y trastorno bipolar también con síntomas psicóticos, trastorno psicótico breve, trastorno de delirio, y trastorno psicótico no específico.

El grupo control estaba formado por 169 individuos de origen mestizo mexicanos (86 hombres y 83 mujeres) no relacionados, entre los 19 a los 70 años de edad (promedio = 40 ± 14 años de edad). El criterio de inclusión consideraba que los sujetos no tuvieran algún síntoma psicopatológico o historial con enfermedades psiquiátricas, el cual fue evaluado mediante Symptom Check List-90 revised (SCL90-R) (Derogatis *et al.* 1973).

2. DETERMINACIÓN DE GENOTIPOS

El DNA genómico fue obtenido de sangre venosa, mediante el método de extracción de fenol-cloroformo. El proceso de obtención de genotipos del VNTR de 48-pares de bases (pb) del exón 3 fue realizado utilizando los oligonucleótidos D4-3 y D4-42 reportados por Lichter *et al.* (1993). Las condiciones de PCR y termociclaje se describen detalladamente en Aguirre *et al.* (2007) (este parte del trabajo de tesis fue presentado en el Artículo II). Los fragmentos amplificados (tamaños entre 350 a 700pb), fueron separados por electroforesis en geles de agarosa Metaphor® al 2.2%, y visualizados con bromuro de etidio. Algunos genotipos de sujetos homocigotos fueron secuenciados para corroborar la identidad y longitud de los alelos específicos.

Otra parte del trabajo de tesis se analizó en el Artículo II. Se secuenciaron 66 cromosomas para el análisis de los haplotipos, los cuales fueron previamente descritos (Ding *et al.*, 2000; Wang *et al.*, 2004). Detalles de las condiciones de amplificación se observan en el Artículo II.

La amplificación mediante PCR de la región polimórfica del VNTR de 12pb en el exón 1, fue realizada usando los oligonucleótidos D4EX1F y D4EX1R reportados por Catalano *et al.* (1993) (Artículo II). Los amplicones para la forma duplicada del VNTR de 12pb (alelo A1: 275 pb) y la forma sencilla o de una copia (alelo A2: 263 pb), fueron observados en geles de poliacrilamida al 6% desnaturalizados con 10% formamida y teñidos con SYBR Green I. La electroforesis vertical se realizó durante toda la noche en una cámara Dcode™ System (BioRad), con placas de 22cm de largo. Los fragmentos de esta región fueron verificados mediante secuenciación.

El análisis del fragmento de DNA entre -627 y -343 pb, que contiene el polimorfismo -521 SNP, fue evaluado mediante PCR, el volumen final fue de 25 µl, y contenía 100ng de DNA, Buffer de PCR 1x, DMSO al 10%, MgCl₂ (2mM), 200µM de cada dNTP, 150µM de cada oligonucleótido (secuencia descrita en tabla1 de esta sección de materiales y métodos) y 0.8 unidades de Taq-Gold polimerasa. Las condiciones de la reacción de PCR incluyó desnaturalización a 98°C por 1min y 95°C por 10 min, 35 ciclos a 95°C/ 30seg, 57°C/ 30 seg, 72°C/ 40 s, y la extensión a 72°C /10 min. El producto de amplificación fue verificado en un gel de agarosa al 1.5%, digerido con la endonucleasa de restricción Fsp 1, del cual se obtuvo fragmentos de 285, 175 y 110 pb, y se observaron en geles de agarosa al 2.5%.

Este fragmento del promotor del gen DRD4 fue clonado en dos sistemas de expresión TOPO TA cloning (Invitrogen) y PGem Teasy (Promega). Brevemente, el

fragmento de PCR amplificado con los oligonucleótidos descritos (tabla 1), fue verificado y cortado del gel de agarosa. Posteriormente, dicho fragmento fue adenilado y ligado al vector plasmídico durante 16 horas a 4°C, la transformación se efectuó en bacterias competentes JM109 a 42°C, y se incubaron en medio LB agar más ampicilina durante 16 horas. El tamizaje molecular se llevó a cabo seleccionando las clonas positivas. Posteriormente fueron vistos por secuenciación automatizada con oligonucleótidos M13 (*Forward*), se usó el sistema de *BygDye Terminator v3.1* y fue electroforazo en un *ABI Prism 3100 Genetic Analyzer* (Applied Biosystems).

Tabla 1. Relación de oligonucleótidos utilizados en este estudio.

Nombre Oligonucleótido	Región	Cita	
D4EX1F	CGCCATGGGGAACCGCAG	VNTR-12pb exón 1, y secuenciación	Seaman <i>et al.</i> , 2000
D4EX1R	CGGCTCACCTCGGAGTAGA		
Drd4-1	CGGGGGCTGAGCACCAGAGGCT GCT	SNP-521 región promotora	Okuyama <i>et al.</i> , 2000
Drd4-2	GCATCGACGCCAGCGCCATCCTA CC		
EX35Q	CGTACTGTGCGGCCTCAACGA	Secuenciación haplotipos VNTR exón 3	Ding <i>et al.</i> , 2002
EX35QR	GACACAGCGCCTGCGTGATGT		
D4-3	GCGACTACGTGGTCTACTCG	VNTR-48pb exón 3	Lichter <i>et al.</i> 1993
D4-42	AGGACCCTCATGGCCTTG		

3. ANÁLISIS ESTADÍSTICO

Las frecuencias de los genotipos se obtuvieron por conteo de acuerdo a un patrón de bandeo establecido. La estimación del equilibrio de Hardy-Weinberg de los diferentes polimorfismos fue realizado utilizando el apartado HWE del programa LINKAGE (Ott, 1988).

La comparación de las frecuencias alélicas entre los pacientes y los controles fue analizada con una Chi cuadrada, dada por el programa CLUMP v 2.2 (Sham y Curtis, 1995) (ver Artículo I), en el cual la p empírica obtenida se basa en la consideración de Monte Carlo, de donde se derivan simulaciones que generan tablas al azar que contienen el mismo total final de los valores observados, evitando así el uso de la Corrección de Bonferroni y múltiples casillas con valores pequeños como es el caso del polimorfismo multialélico del VNTR de 48pb.

Para calcular la inferencia de la distribución individual y frecuencia de haplotipos se utilizó el algoritmo EM (expectation maximization algorithm) que es parte de Arlequin V. 2.0 (Schneider *et al.*, 2000). La comparación de frecuencias entre grupos fue hecho por el programa EH+ PM+ (Zhao *et al.*, 2000). En este análisis también se midió la fuerza de asociación mediante el odds ratio (OR) para estimar el posible riesgo, donde el valor de OR y de confianza debe ser significativamente > 1.0 que indica una asociación positiva entre un alelo y la enfermedad (casos-contrales, Artículo I).

4. ANÁLISIS DE LAS SECUENCIAS DEL VNTR DEL EXÓN 3

En este estudio nosotros usamos la nomenclatura reportada por Ding *et al.* (2002) para clasificar los haplotipos y fueron comparados por los anteriormente descritos (Ding *et al.*, 2000; Wang *et al.*, 2004). Los alineamientos y comparaciones de la secuencia nucleotídica fueron realizados usando los programas BLAST y BioEdit.

4.1 ANÁLISIS POLIMÓRFICO Y NEUTRALIDAD

Los sitios polimórficos, la prueba de Tajima (Tajima, 1989) para evaluar la teoría Neutral de la Evolución Molecular, la prueba de Fu y Li (Fu y Li, 1989), así como la estadística descriptiva y el desequilibrio de ligamiento fueron evaluados con la ayuda del software DnaSP ver. 4.10.9 (Rozas *et al.*, 2003) y Arlequin v 2.0 (Schneider *et al.*, 2000). La diferencia entre poblaciones dada por la F_{ST} convencional y la F_{ST} de Kimura 2 parámetros fueron calculados. En el resultado fue corregido el porcentaje de nucleótidos en la cual dos haplotipos son diferentes. Dicha corrección considera las sustituciones por sitio, toma en cuenta la tasa de sustituciones entre transiciones y transversiones (Kimura, 1980). Se realizó un análisis de varianza molecular (AMOVA, Analysis of Molecular

Variance) con en el programa Arlequin versión 2.001 (Schneider *et al.*, 2000), con el objetivo de establecer el nivel de diferenciación entre grupos de poblaciones. Este es un método para estimar la diferenciación poblacional directamente de los datos moleculares y probar la hipótesis acerca de tal diferenciación (Excoffier *et al.*, 1992). En este estudio los datos que se usaron corresponden a las secuencias del VNTR del exón 3 del gen DRD4.

Además, se calculó la distribución teórica de la D de Tajima (95% de confianza con intervalos entre -2 y +2). Esta asume que la presencia de polimorfismos es independiente de la frecuencia de alelos. Altos valores de la D de Tajima sugieren un exceso de variación nucleotídica común en una región del DNA, la cual es consistente con una selección balanceadora. Los valores negativos de D indican un exceso de variación rara, consistente con el crecimiento poblacional o selección direccional.

4.2 PRUEBA DE DESEQUILIBRIO DE LIGAMIENTO

También se calculó el valor de D' para el análisis de desequilibrio de ligamiento entre los nucleótidos de los tres grupos estudiados (2R, 4R y 7R), se usó tablas de contingencia de 2x2 y la significancia estadística se calculó con la prueba exacta de Fisher, y la corrección de Bonferroni para análisis múltiples. El valor de desequilibrio de ligamiento está dado por D', cuya expresión muestra el máximo posible valor de D', con un intervalo de 0 a 1, dadas las frecuencias alélicas. La presencia de desequilibrio de ligamiento en casos y controles entre dos loci estuvo bajo la hipótesis nula de no asociación entre los loci evaluados y fue analizados con DnaSP (Rozas *et al.*, 2003). Una $p = 0.01$ fue considerada como significativa.

La prueba exacta de Fisher permite analizar si dos variables dicotómicas están asociadas cuando la muestra a estudiar es demasiado pequeña y no se cumplen las condiciones necesarias para que la aplicación del test X^2 sea adecuada. Estas condiciones exigen que los valores esperados de al menos el 80% de las celdas en una tabla de contingencia sean mayores de 5. Así, en una tabla 2x2 será necesario que todas las celdas verifiquen esta condición, si bien en la práctica suele permitirse que una de ellas muestre frecuencias esperadas ligeramente por debajo de este valor. Otra característica relevante de esta prueba es que permite establecer con exactitud el correspondiente P - valor, y no de manera aproximada como es el caso cuando se recurre

a la distribución χ^2 . Cada una de estas probabilidades se obtiene bajo la hipótesis nula de independencia de las dos variables que se están considerando.

4.3 ENTROPÍA

El grado de conservación en las posiciones de las secuencias de los alelos de 7R fue calculado con el promedio de la entropía. Shannon (1948) lo define como una medida de incertidumbre en predecir valores de una variable al azar x con valores posibles $\{x_1, \dots, x_n\}$, es una medida de variabilidad de variables discretas, definida por la ecuación E.3:

$$H(X) = - \sum P(x_i) \log(P(x_i)), \quad \text{Ecuación E.3}$$

Donde una variable al azar es distribuido en los valores de x_i con probabilidad de $P(x_i)$ de la sumatoria de todos los diferentes genotipos i en x .

Alta entropía indica que es muy variable, y baja entropía altamente conservado. Una de las interpretaciones de entropía esta basada en el contenido de información del mensaje, el cual es subjetiva a la secuencia genómica, en donde se asume que la entropía refleja el grado de funcionalidad. Por ejemplo, el contenido de CG se desvía de un 50%, que baja la entropía de la secuencia, por lo que deja de ser aleatoria por mutaciones como se espera. También están las secuencias repetitivas (las cuales tienen baja variabilidad) que se forman por numerosos procesos como replicación por slippage, retrotransposición, etc., estas regiones genómicas pudieran no estar bajo un proceso azaroso (Parkhomchuk, 2006).

El análisis de entropía fue hecho con el software BioEdit, y en el alineamiento se compararon las secuencias de humanos con las de primates no humanos con secuencias de 7R, obtenidas del GenBank (*Hylobates muelleri*, clave de acceso: AB032910), *Hylobates la*: AB032907, *Hylobates pileatus*: AB032908, y el *Hylobates agilis*: AB032908).

5. MODELAJE MOLECULAR

Se evaluó los cambios estructurales asociados a la variación en el número de repeticiones (VNTR de 48 pb) presentes en la región del exón 3 de la proteína DRD4, se llevaron a cabo varios ensayos de modelaje molecular, con el fin de tener modelos para las variantes 2R, 4R y 7R. Las proteínas utilizadas como moldes para la simulación o modelaje fueron obtenidas de la base de datos de receptores acoplados a proteínas G

creada por Zhang *et al.* (2006) (Artículo II).

La predicción de estructuras proteínicas mediante el modelaje de una secuencia basada en estructuras conocidas consiste en 4 pasos: encontrar una estructura conocida relacionada a la secuencia de interés (templado), alineamiento de la secuencia con el templado, construir y medir el modelo (Baker y Sali, 2001). Estos análisis se llevaron a cabo en colaboración con el Depto. de Visualización Científica de la Dirección General de Servicios de Cómputo Académico. UNAM.

Reconstrucción filogenética

Se reconstruyeron varios árboles filogenéticos basado en el método bayesiano (MB), con la finalidad de encontrar la relación de los sujetos de la muestra tanto controles como esquizofrénicos, los únicos primates no humanos (gibones) que tienen alelo 7R. En la primera reconstrucción filogenética, se usó como grupo externo de referencia el haplotipo compuesto generado de la secuencia ortóloga de *Macaca arctoides* (acceso al GenBank n° EU744540). La muestra de sangre periférica de *M. arctoides* fue donada de la colonia de macacos del Instituto Nacional de Psiquiatría Ramón de la Fuente.

Los otros árboles filogenéticos usando el MB se reconstruyó con otras especies de primates de la región del VNTR del exón 3 del gen DRD4. Claves de acceso a los taxones usados en el estudio: *Macaca mullata*, L27757; *Papio hamadryas*, L27803; *Clorocebus sabaesus*, DQ666090; *Pan troglodytes*, L27804; *Gorilla gorilla*, L27750; *Pongo pygmaeus*, L27806; *Pongo pygmaeus*, L27752; *Papio hamadryas*, L27747; *Saimiri boliviensis*, L27755; *Chlorocebus sabaesus*, DQ666090; *Tarsius bancanus*, AB016203; *Nycticebus coucang*, AB016201; *N. coucang*, AB016200; *Lemur catta*, AB013202.

El método bayesiano determina para cada árbol, la probabilidad de que sea la hipótesis correcta (en función de los datos de secuencias de 7R del gen DRD4). A la probabilidad de cada árbol dados unos datos, $P(H|D)$, se denomina *probabilidad posterior* porque se calcula la probabilidad de una hipótesis a partir de los resultados que produciría dicha hipótesis. La aproximación para estimar la distribución de probabilidad posterior (i.e. la probabilidad de cada árbol) se realiza mediante la consideración de Markov Chain Monte Carlo (MCMC).

IV. RESULTADOS

Una gran obra es el producto de una gran pasión al servicio de una gran idea.

Goethe

ARTÍCULO I

La identificación de los genotipos del polimorfismo del exón 3 fue larga debido a que tiene un alto contenido de CG en su secuencia y no se observaba con claridad las bandas a pesar de que ya existía un protocolo previo en nuestro laboratorio. Se logró establecer un protocolo que modificaba el tiempo de desnaturalización de 95°C/20seg. a 95°C/40seg. También la temperatura de alineamiento fue cambiada de 55°C/20seg. a 56°C/40seg. Además se aumentó el tiempo de elongación de 7 a 10 min y mayor cantidad de DNA, lo cual favoreció a la definición de los fragmentos de DNA de la región de interés. La separación de los fragmentos en los geles de agarosa Metaphor® se muestra en la figura 1 de la sección del apéndice.

Los alelos más comunes del VNTR-exón 3 en la población objeto de estudio fueron las formas 4R y 7R repetidas. Encontramos una diferencia estadísticamente significativa entre los pacientes con respecto a los controles (CLUMP T4 $X^2 = 21.0$, $p < 0.01$). Las diferencias se relacionaban a las frecuencias de alelos poco frecuentes como los alelos 3R, 5R, 6R y 8R. A partir de este resultado se realizó una comparación entre las frecuencias de alelos menos frecuentes o “raros” contra los alelos más comunes del 4 y 7 repetidas entre la muestra de estudio. Hubo una diferencia estadísticamente significativa entre los pacientes psicóticos ($X^2 = 19.3$, 1df, $p < 0.01$) en relación a los controles, confirmando así el aumento de alelos raros en la muestra de pacientes.

La muestra de pacientes psicóticos fue reagrupada con el objeto de identificar si la diferencia estadística que encontramos estaba dada por los pacientes esquizofrénicos. El subgrupo fue comparado con los controles y se encontró que las frecuencias de los pacientes eran diferente significativamente (CLUMP, T4 $X^2 = 15.3$, $p < 0.01$). El valor de p fue obtenido mediante la agrupación de los alelos 3R, 5R, 6R y 8R, la cual fue realizada por el programa CLUMP. Al agregar al alelo de 2R dentro del grupo de alelos raros vs. comunes también se observaron diferencias ($X^2 = 13.6$, 1df, $p < 0.01$).

POLIMORFISMO DEL VNTR-120pb DEL EXÓN 1

Para identificar y lograr una separación suficiente entre los genotipos del polimorfismo de 12pb ubicado en el exón 1 del gen DRD4, fue necesario preparar un gel en condiciones naturalizantes que permitieron una migración estable de los fragmentos de DNA. También usamos un colorante sensible como el SYBER Green, que nos ayudó a

observar con claridad los fragmentos de DNA (como se muestra en la figura 2 de la sección del apéndice). La variación fue corroborada por secuenciación (figuras 7 y 8, del apéndice).

La frecuencia alélica del VNTR de 12 pb, del exón 1 no fue diferente significativamente entre los pacientes y los controles ($X^2 = 0.12$, $p = 0.76$).

ANÁLISIS DE HAPLOTIPOS

La estimación de las frecuencias de haplotipos formados de las combinaciones entre el polimorfismo del exón 1 con el exón 3 entre casos y controles, fue diferente significativamente. Detectamos 14 posibles haplotipos, de los cuales todos fueron observados en pacientes psicóticos, 13 en una sub-muestra de sujetos con diagnóstico de esquizofrenia y sólo 12 fueron encontrados en los controles.

La diversidad del gene para los dos loci en pacientes y controles fue moderadamente alta con 0.78 ± 0.02 y 0.68 ± 0.02 , respectivamente. El haplotipo más común fue 4-A1 tanto en pacientes como en el grupo de controles, seguido por la forma 7-A1. Los haplotipos 5-A2 y 8-A2 no estuvieron presentes en el grupo de sujetos controles, mientras que las frecuencias de las formas 5-A1, 6-A2 y 8-A1 fueron tan bajas como de menos del 1%. Los haplotipos 3-A1, 5-A1, 6-A1 y 8-A1 tuvieron frecuencias altas en pacientes con diagnóstico de psicosis o esquizofrenia en relación a los controles. La comparación de las frecuencias de haplotipos entre sujetos psicóticos y los controles mostró una evidencia de asociación (X^2 con un valor de p empírica de 0.012, con 250 permutaciones). La comparación de frecuencias no fue significativa entre el subgrupo de sujetos con esquizofrenia y los controles ($p = 0.08$).

El Odds ratios de los datos de haplotipos fue estadísticamente significativo de entre las formas 3-A1, OR: 4.26 (95% CI: 1.4 a 12.91, $p = 0.005$), 6-A1, OR: 3.14 (95% CI: 1.0 a 9.86, $p = 0.039$), y 4-A1 OR: 0.65 (95% CI: 0.458 a 0.935, $p = 0.019$). Además, un resultado significativo se obtuvo para esos individuos que llevan uno de lo siguiente cuatro haplotipos (es decir 3-A1, 5-A1, 6-A1 y 8-A1), OR: 4.5 (95% CI: 2.2 a 9.2, $p = 0.001$).

ARTICULO II

Análisis de Secuencia

Los electroferogramas de las secuencias de los alelos 4R y 7R se presentan en la figura 4 y 5, así como la unidad de repeticiones en la figura 6 de la sección del apéndice.

Las variaciones internucleotídicas que encontramos en la secuencia de la región polimórfica del exón 3, fueron agregadas como formas nuevas secuencias en la sección *entrez nucleotide* del GenBank. Las nuevas asignaciones encontradas van desde el número 36 al 42 y se puede clasificar como polimorfismos del tipo SNP. Dicha variación se observó en haplotipos contenidos en alelos de 7R, 2R y en menor proporción en alelos de 4R. De las secuencias nuevas encontradas en este estudio algunas mutaciones se agrupan en su mayoría en el último motivo de las unidades de repetición de los alelos como es el caso de los alelos de 2R (haplotipo 1-37), y en el alelo de 7R (por ejemplo; 1-2-27-5-2-5-41 y 1-2-6-5-2-5-41).

La estructuras de haplotipos más frecuente en esta muestra poblacional fueron las formas 1-2-3-4 y 1-2-6-5-2-5-4 en los alelos de 4R y 7R, y esto coincide con datos de otras poblaciones (Ding *et al.*, 2002; Grady *et al.*, 2003). Los grupos de alelos 2R, 4R y 7R, se agruparon por separado como 3 poblaciones y se dividió en sujetos con y sin diagnóstico de esquizofrenia. Se encontró que entre la población de los alelos de 4R no hubo diferencias significativas entre pacientes y controles ($X^2= 1.824$, 2df, $p=0.402$). En cambio, en la población de alelos 7R hubo una tendencia significativa al llevar a cabo 100 simulaciones con el programa CLUMP ($X^2=10.689$, $p 0.049$), al hacer el agrupamiento de columnas la significancia fue de $p= 0.0396$ ($X^2=8.1579$, 100 simulaciones).

Polimorfismo de secuencia

La variación de secuencia fueron analizados bajo el modelo de sitio infinito (ISM), que se desprende de la Teoría Neutral (Kimura, 1983), asume que cada mutación que ocurre en el genoma se da en un nuevo sitio al azar.

Dentro de la unidad repetida (N= 16 secuencias motivos, que forman los haplotipos) encontrada en la muestra de estudio, se localizaron 21 sitios de variación, 27 sitios sin variación, y 13 sitios parsimónicos informativos (2 y 3 variantes). Se entiende por parsimonia a aquellos sitios que contiene el mínimo número de cambios mutacionales y con los cuales se pueden construir árboles filogenéticos e inferir acerca de cuales fueron los sitios ancestrales. Las sustituciones tanto sinónimas como no sinónimas fueron analizadas en las 16 secuencias de las unidades de repetición encontradas en nuestra

población objeto de estudio. El análisis mostró que hay más sustituciones no sinónimas (66%) que sinónimas (34.6%), este es un indicador de la diversidad nucleotídica.

Análisis de desequilibrio de ligamiento (DL)

Este análisis se llevó a cabo en las unidades de repetición (16 secuencias) y solo 2 comparaciones fueron significativas con la prueba de Fisher ($p < 0.01$) (16 y 25; 35 y 38) y 7 comparaciones fueron significativas con la corrección de Bonferroni (Fisher y Chi-cuadrada, $p < 0.01$). Estos sitios fueron clasificados como sitios informativos parsimónicos.

El análisis LD en los grupos de 4R y 2R muestra que la variación internucleotídica es poca o nula, respectivamente. En cambio, en las secuencias de los alelos de 7R hubo 21 sitios polimórficos de 336. El análisis DL en los grupos 7R mostró que de 210 combinaciones pareadas, 43 fueron significativas a un nivel de $p < 0.05$, y 3 combinaciones tuvieron una $p < 0.001$ después de la corrección de Bonferroni. Los sitios con alto LD corresponden a las comparaciones entre los nucleótidos 304, 309, 313 ($D' = 1.0$), los cuales fueron identificados como sitios informativos parsimónicos.

Entropía

En la búsqueda de regiones conservadas en las secuencias de los alelos de 7R y tratando de confirmar la naturaleza de la presión selectiva sobre los sitios con LD. Encontramos 6 regiones conservadas a partir del alineamiento múltiple (Figura 10, del apéndice) con las secuencias del VNTR de 48pb de gibones con alelos 7R y los haplotipos encontrados en la muestra de estudio. Los valores de entropía fueron pequeños (< 0.2).

Diversidad genética y estructura de población

Los grupos de esquizofrénicos (E) y controles (C) con secuencias de 2R presentaron un sesgo debido a su tamaño de muestra. Los valores de π y θ , (estimadores de diversidad) del grupo control fueron altos. En el análisis de estructura de población con F_{ST} tanto convencional como la de Kimura-2P no se puede distinguir diferencia genética entre grupos debido al tamaño de muestra. La D de Tajima calculado para las subpoblaciones E y C no expresa claramente si las sustituciones encontradas son neutrales o están bajo algún tipo de selección.

La población con haplotipos 4R del grupo E muestra un sesgo por el tamaño de muestra. Los valores de F_{ST} tanto convencional como la de Kimura-2P no distinguen la

diferencia genética entre grupos E y C. El grupo C, cuya diversidad nucleotídica es 0.0028, presenta una D de Tajima negativa, que es estadísticamente no significativa, no alcanza la confianza debido al tamaño de la muestra, pero la D negativa podría sugerir que las 3 sustituciones encontradas están sujetas a una selección purificadora.

La diversidad de los grupos con secuencia de 7R fue de 0.003 y 0.001 de esquizofrénicos y controles, respectivamente. Los valores de FST son similares, y aunque no es estadísticamente significativa podría sugerir que hay una pequeña diferencia entre ambos grupos. El valor de D fue de -0.93 ($p=0.21$) para el grupo E y -1.48 ($p=0.0059$) de los controles (Ver tabla 1 de resultados). La diferencia entre los grupos es muy poca. Sin embargo, cuando se suman ambas subpoblaciones E y C, el valor de D y de p aumenta (-2.21, $p<0.01$) y es verificado con la prueba de Fu y Li (-3.35, $p<0.02$), distinguiendo mejor el efecto de la selección sobre las mutaciones. Los valores de D pueden sugerir que la selección purificadora elimina alelos de baja frecuencia debido al crecimiento poblacional de los alelos 7R.

Los resultados del análisis de varianza molecular (AMOVA) mostraron una baja diferenciación genética entre las poblaciones E y C de sujetos con alelos 4R y 7R (1.4% y 2.8%, individualmente). Sin embargo, la población de sujetos con alelos 2R fue mayor con una diferencia genética de 66%, aunque la estimación puede tener un valor sesgado por el tamaño de muestra. Se encontró, una mayor variación genética porcentual dentro de poblaciones para los grupos 2R, 4R y 7R (133, 32.6 y 27.5, correspondientemente).

Tabla 1 de Resultados. Diversidad y estructura genética dadas por las pruebas de F_{ST} y AMOVA de las subpoblaciones de sujetos diagnosticados con esquizofrenia y los controles.

Poblaciones por alelos	N	Nº of Sitios	FST Kimura 2P	FST Conv.	AMOVA (%)	$\pi \pm E$	θ_π	θ_S	D Tajima
2R Esquizofrén.	4	96	0.33 P=0.43	0.44 P=0.14	66	0.0± 0.0	0.0	0.0	0.0 NS
2R Controles	4	96				0.05±0.04			2.21
4R Esquizofrén.	6	192	0.043 P=0.16	0.0008 P=0.25	1.4	0.0055±0.0049	1.07± 0.94	0.88± 0.68	1.03 P=0.87
4R Controles	18	192				0.0028±0.0027	0.53± 0.52	0.87± 0.56	-1.08 P=0.148
7R Esquizofrén.	11	336	0.011 P=0.2	0.011 P=0.21	2.8	0.003±0.0025	1.02± 0.83	1.37± 0.82	-0.93 P=0.21
7R Controles.	23	336				0.001±0.0012	0.34± 0.38	0.81± 0.52	-1.48 P=0.058

Esquizofrénicos (esquizofrénicos), N: número de cromosomas, AMOVA (entre poblaciones), π y (E) su error estándar.

Reconstrucción filogenética

A pesar de que este análisis no es parte de los objetivos de la tesis, el dato es interesante para establecer la relación de parientes cercanos como los de primates no humanos, específicamente gibones con secuencias de 7R; así como encontrar una consideración cladista que detecta sitios susceptibles que forman clades (un clade es un grupo taxonómico que incluye un antepasado común y a todos los descendientes de ese antepasado) en los pacientes diagnosticados con esquizofrenia.

Los árboles filogenéticos obtenidos por el método bayesiano, implementado con el software MrBayes ver 3.0 (Huelsenbeck y Ronquist, 2001), el cual usa la consideración de Monte Carlo (Markov chain Monte Carlo) para calcular la probabilidad de los árboles

filogenéticos.

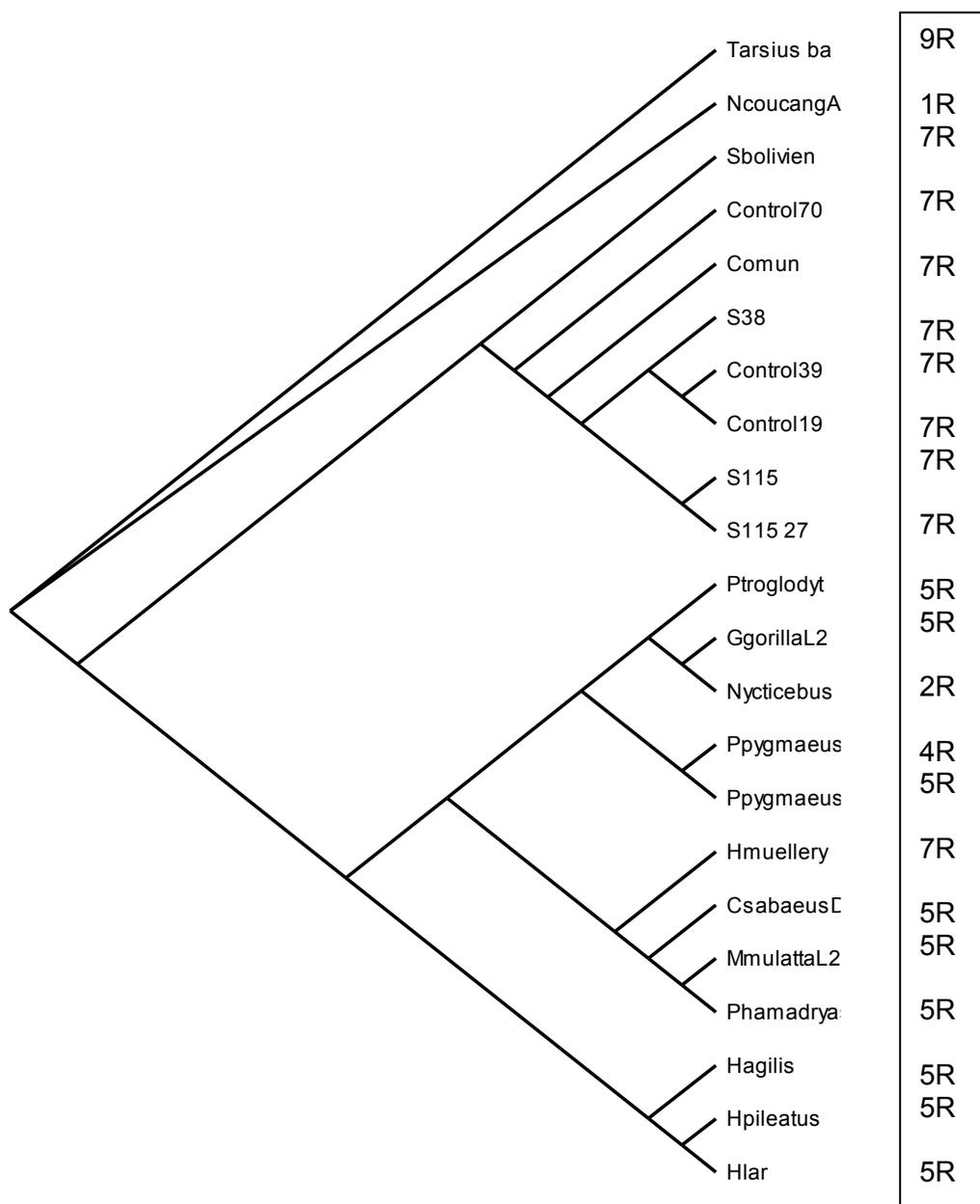


Figura 1 (de resultados). Análisis filogenético de la región VNTR-exón 3 del gen DRD4 en muestras de humanos de este estudio y primates no humanos. Los datos del recuadro representan el número de repeticiones de cada taxón.

Interpretación del árbol filogenético

Los nodos terminales del árbol están constituidos por 22 secuencias que en una lectura de “abajo hacia arriba”, están representados por *Hylobates lar*, hasta *Tarsius bancanus*.- En el recuadro se observan el número de repeticiones correspondiente a cada

taxón.

El grupo externo *Nycticebus coucang* perteneciente al grupo monofilético Lorisoideo, fue escogido como grupo externo. En el Cladograma no aparece como tal, y quien aparece como grupo externo es el *Tarsisus bancanus*. A este respecto, la posición de los Tarsioides en el árbol de primates está a debate. En un análisis de 15 genes nucleares los Tarsioides son colocados como una rama separada entre los prosimios con alta significancia estadística. Otro análisis sugiere que los Tarsios pueden representar un grupo monofilético con los Platiirinos (ceboidea), y catarrinos (Cercophitecoides) (Hipótesis Haplorrhine).

Resultados.

- a.- *Tarsisus bancanus* se ubica como grupo Externo.
- b.- Grupo Interno. Está constituido por las secuencias de *Hylobates lar* hasta *Saimiri boliviensis* (mono ardilla).
- c.- El Grupo Externo representado por *T. bancanus*, así como la secuencia *H. lar*, representan un carácter con valor cero (0), denominado pleisomórfico.
- d.- Las secuencias desde *H. pileatus* hasta *S. boliviensis*, representan un carácter con valor de 1, denominado apomórfico, y las politomías se resuelven quedando *H. lar*-*H.pileatus*, *P. hamadryas*-*M. mulata*. Las 2 variaciones de *P. pygmeus* como secuencias hermanas, *N. coucang*(2R)-*G.gorilla*, los esquizofrénicos S115_27-S115, Control39-control19, y común-control70, como grupos monofiléticos, soportados por la sinapomorfía.
- e.- Los valores de 0 y 1, se explican recurriendo al origen único, por lo tanto, el 0 y 1, son homologías táxicas, para los taxones terminales que los representan.
- f.- Las secuencias correspondientes desde S115 27, hasta *S. boliviensis*, tienen su origen en un nodo del grupo interno, lo cual indicaría la presunción de un origen ancestral reciente, con un ramaje y topología que forma los nodos terminales "clades" implicando los taxones de *Homo sapiens*, y de *S. bolivienses* (Monos del Nuevo Mundo, platiirinos, ceboideos). Así mismo merece una evaluación el alineamiento de las agrupaciones S115 27-S115, y Control19-Control 39. A este respecto se sugiere un análisis más acucioso, dado el criterio de la polaridad de los caracteres.

Análisis del árbol según los criterios de Enraizamiento

- 1.-La secuencia de *T. bancanus*, aparece como grupo hermano de *H. lar*-*N. coucang*, al aplicar la regla operacional
- 2.- Los grupos internos desde *H. lar* a *T. bancanus*, se presuponen monofiléticos.

3.- Aún cuando el grupo externo esté incluido en la matriz de datos, el resultado de la prueba de congruencia, es un árbol todavía sin raíz. El conjunto de estados de carácter para el grupo externo, está codificado en la matriz de datos como cualquier otro taxón, incluido en el análisis. El taxón del grupo externo se usa en la construcción del árbol y No corresponde a un nodo interno, sino a un nodo terminal.

Por lo tanto podemos concluir que el método del grupo externo (MGE), no es un método de enraizamiento, ya que la información que podemos extraer del grupo externo no es suficiente para construir la raíz implícita en la regla operacional. (La raíz del árbol se localiza en el internodo que conecta al grupo interno con el grupo externo).

Sería apropiado que el uso del MGE, en lugar de proveer información sobre caracteres para enraizar, sería de utilidad para orientar información sobre grupos monofiléticos. A este respecto el árbol obtenido, representa la descripción de una tricotomía, en cuyo origen de raíz parcial se involucra al taxón del grupo externo *Tarsisus bancanus*, al taxón terminal mas basal del grupo interno, en este caso representado por *N. coucang*, y a las ramas de los grupos monofiléticos del resto de las taxones terminales.

POLIMORFISMO –521 C/T

Los genotipos fueron identificados en geles de agarosa, como se observa en la figura 3 y en la figura 9 (sección del apéndice) se observa el cambio de nucleótido).

Las frecuencias alélicas y genotípicas del polimorfismo del promotor 521C/T fueron estadísticamente diferente a la de los controles ($p = 0.004$; $p = 0.01$, respectivamente), Tabla 2 de resultados. Los pacientes psicóticos muestran una alta frecuencia de las formas TT y T.

La muestra de pacientes estuvo en equilibrio de Hardy-Weinberg ($X^2 = 2.448$, 1 df, $p = 0.11$). La distribución de la frecuencia en los sujetos controles estuvo en desequilibrio H-W ($X^2 = 7.629$, 1 df, $p = 0.006$). Esto tal vez se deba a una sub-muestra poblacional por el mestizaje o estratificación poblacional.

Tabla 2 de resultados. Distribución de las frecuencias alélicas y de genotipos del polimorfismo -521 SNP en pacientes diagnosticados con psicosis y controles, de origen mexicano

	Frecuencia de Genotipos*			Frecuencia alélica**	
	CC	CT	TT	C	T
Controles	38 (0.3)	76 (0.6)	13 (0.1)	152 (0.6)	102 (0.4)
Psicóticos	32 (0.25)	54 (0.43)	40 (0.32)	118 (0.47)	134 (0.53)

*Chi-squared = 17.988, 2df, 2 sided, $p = 0.000124$, **Chi-squared = 8.098, 1df, $p = 0.0044$.

ARTICULO I

Puede ser consultado en *Psychiatry Res* 2007, 153:209-215.

Association analysis of exon III and exon I polymorphisms of the dopamine D4 receptor locus in Mexican psychotic patients.

Ana Julia Aguirre^a, Rogelio Apiquian^b, Ana Fresan^b, Carlos Cruz-Fuentes*

^aE-mail address: ajua@servidor.unam.mx Departamento de Genética Psiquiátrica, Instituto Nacional de Psiquiatría "Ramón de la Fuente". D.F., México.

^bEpidemiología Clínica, Instituto Nacional de Psiquiatría "Ramón de la Fuente". Cp. 14370, D.F., México.

*Correspondence to: Carlos Cruz Fuentes, Departamento de Genética Psiquiátrica, División de Investigaciones Clínicas, Instituto Nacional de Psiquiatría "Ramón de la Fuente". 101 México-Xochimilco, Col. San Lorenzo Huipulco, cp 14370 D.F., México. E-mail address: cruz@imp.edu.mx (+52) 5655-2811, ext. 210, Telefax (+52) 5513-3722.

Abstract

In this study, we investigated whether polymorphisms of the dopamine D4 receptor (DRD4) gene were associated with psychotic symptomatology rather than with unique diagnosis such as schizophrenia. A number of association studies between the DRD4 gene 48bp-VNTR polymorphism at exon 3 and these disorders have been reported but results have been controversial. 48bp-VNTR as well as the 12bp-VNTR (at exon 1) polymorphisms of this gene, were analyzed a group of 149 unrelated Mexican subjects with a diagnosis for one of the following psychotic disorders: schizophrenia, schizoaffective disorder, schizophreniform disorder, major depression and bipolar disorder, both with psychotic symptoms, brief psychotic disorder, delusional disorder and non-specific psychotic disorder, and in 169 individuals free of psychiatric illnesses. There were no differences in allele or genotype frequencies between groups of comparison for the 12bp-VNTR. However, a significant excess of "rare" alleles (3-, 5-, 6- and 8- 48bp repeats alleles) was found in the group of psychotics. Moreover, haplotypes 3-A1, 5-A1, 6-A1 and 8-A1 were significant most frequently associated with cases. This positive association supports a role of this molecule as a genetic risk factor in psychotic disorders.

Keywords: psychotic disorder, schizophrenia, DRD4, 48bp VNTR exon 3, 12bp VNTR exon 1, haplotypes, Mexicans.

1. Introduction

A disturbance in the neural dopaminergic transmission system has long been implicated in the development of schizophrenia (Winterer and Weinberger 2004). Support of the dopamine hypothesis is partially based on the therapeutic effects of most of the so-called typical and atypical neuroleptics, which in turn are attributed to their well documented occupancy of dopamine receptors sites, particularly to those of the D2-like receptors subtypes (Sanyal and Van Tol, 1997). The dopamine D4 receptor protein has been of particular interest in the research on psychotic disorders, on the basis of its high affinity binding profile for the atypical antipsychotic clozapine (Seeman, 1992). Additional support to a putative role of DRD4 in schizophrenia is given by reports of an increased density of dopamine D4 receptors in the putamen/caudate nucleus of post-mortem brain tissue of patients (Seeman *et al.*, 1993), as well as the relative abundance of D4 mRNA in pre- or orbito- frontal cortex of schizophrenics (Meador-Woodruff *et al.*, 1997; Stefanis *et al.*, 1998;). Nonetheless, negative results have also been reported (Roberts *et al.*, 1996; Lahti *et al.*, 1998).

The human D4 receptor gene (DRD4) located on chromosome 11 (11p.15.5) (Gelernter *et al.*, 1992) is composed of four exons (Van Tol *et al.*, 1991), with a number of polymorphic systems described in its sequence (Wong *et al.*, 2000). A variable number of 48bp tandem repeats (VNTR) in exon 3, encoding the putative third intracellular loop has been of particular interest in psychiatric genetics. Alleles from 2 up to 11 repeats have been reported (Van Tol *et al.*, 1992; Ding *et al.*, 2002), with the 4-, 7- and 2- repeats sequences being the most frequent variants in most populations; although ethnic and population heterogeneity in allele frequencies is acknowledged (Chang *et al.*, 1996; Ding *et al.*, 2002).

The *in vivo* functional significance of these length and sequence variants of the D4 receptor protein is unknown, although *in vitro* studies have showed that in comparison with the commonest 4-repeat allele, the 7-repeat variant expressed in cell transfection systems displays discrete pharmacological differences, both in clozapine competition of (³H) spiperone binding (Van Tol *et al.*, 1992), as in their functional coupling to adenylyl cyclase (Asghari *et al.*, 1995; Jovanovic *et al.*, 1999).

A number of association studies using this allelic system as a genetic marker have been reported for schizophrenia (for recent reviews see Glatt *et al.*, 2003; Jönsson *et al.*, 2003). Overall, the results do not support an important a role for this particular polymorphism and this mental disorder, although positive associations with particular variables (i.e. sex-, clinical symptoms, or clinical subtypes) could not be disregarded.

Our group has previously reported a significantly lower frequency of 4-repeats allele carriers in Mexican patients experiencing their first psychotic episode. This association was sustained mainly by a small group of individuals showing affective

symptoms (Rinetti *et al.*, 2001). The aim of the present study was to reanalyze these data on the VNTR polymorphism in exon 3 in an extended sample of subjects who experienced their first psychotic episode. In addition we also analyzed the 12 base pair tandem repeats polymorphism in exon 1 that codes for a putative four amino acids variant in the extra cellular N- terminal domain of the receptor which has been associated in subjects with delusional disorder (Catalano *et al.*, 1993).

2. Methods

2.1 Subjects

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics and Scientific Committees of the Instituto Nacional de Psiquiatria *Ramon de la Fuente* in Mexico City. All subjects studied were unrelated with parents and grandparents of Mexican origin. Written consent was granted for participation. One hundred and forty-nine individuals (79 men and 70 women); between 14 to 58 years old (mean = 29 ± 8), that experienced their first psychotic episode constituted the affected group. They were evaluated with the Schedule for the Clinical Assessment in Neuropsychiatry (SCAN), and met DSM III-R diagnostic criteria (American Psychiatric Association, 1987) for one the following psychiatric disorders: paranoid schizophrenia (54%), undifferentiated schizophrenia (10.7%), disorganized schizophrenia (5.3%), schizoaffective disorder (6.7%), schizophreniform disorder (6%), major depression (8%) and bipolar disorder (6.7%) both with psychotic symptoms, brief psychotic disorder (0.7%), delusional (1.3%), and non-specific psychotic disorder (0.7%).

The control group was constituted by 169 individuals (86 men and 83 women), between 19 to 70 years old (mean = 40 ± 14 years old); they were included on the basis of clinical interrogation of past or current psychiatric disorders and being free of current psychological distress evaluated by using the Symptom Check List-90 revised (SCL90-R) (Derogatis *et al.*, 1973).

2.2 DNA analysis

Genomic DNA was obtained from venous blood using a conventional phenol-chloroform extraction method. Genotyping of the 48bp VNTR polymorphic system in exon 3 was achieved using the D4-3 and D4-42 primers reported in Lichter *et al.*, (1993) and PCR conditions described previously (Cruz *et al.*, 1997). The amplified fragments (alleles from 350 to 700 bp), were separated by electrophoresis in 2.2% Metaphor®- agarose gels, and visualized after ethidium bromide staining.

PCR amplification of the region containing the 12bp-VNTR at exon 1 was performed using the primers D4EX1F and D4EX1R reported by Catalano *et al.*, (1993). Temperature profile was similar as for exon 3 VNTR, except that the annealing temperature was set up to 57 °C. Amplicons for the duplicated 12-bp repeats (A1 allele:

275 bp) and the single 12-bp repeat (A2 allele: 263 bp), were resolved in 6% denaturalized polyacrilamide gels with 10% formamide, and visualized after SYBR Green I staining. In order to validate the identity and length of specific alleles, amplicons for both polymorphic systems were sequenced in a particular set of homozygous subjects.

2.3 Statistical analyses

Allele and genotype frequencies were obtained by manual counting. Possible deviations of the Hardy-Weinberg equilibrium of the genotype and allele frequencies for both markers were estimated using the HWE subroutine of genetic LINKAGE program (Ott, 1988).

Comparison of allele frequencies between psychotic patients and controls were tested using chi-square tests. Empirical P values were obtained using the CLUMP software v 2.2 (Sham and Curtis, 1995). CLUMP uses a Monte Carlo approach which performs repeated simulations to generate tables having the same marginal totals as the one under consideration, and counting the number of times that a chi-squared value associated with the real table is achieved by the randomly simulated data. It is proposed that by this procedure an estimate of P that yields the correct type I error will be obtained. One hundred simulations were performed in order to evaluate the statistical significance of the chi-squared statistics for a 2 x m contingency table.

The expectation maximization (EM) algorithm included in Arlequin v. 2.0 was used to infer individual haplotypes probability distributions, and estimate their frequencies (Schneider et al., 2000). Haplotype frequencies were compared between groups using the EH+ PM+ software (Zhao *et al.*, 2000). These are a suite of programs designed to perform model-free analysis and permutation tests for marker-disease allelic associations. Chi square values were obtained using a heterogeneity model (T5). This test is advocated to be used when the disease model is unknown, and it is suggested to provide nearly optimal power, especially for a minor-gene model (Zhao *et al.*, 2000).

The strength of the association among specific haplotypes and the disease was evaluated by means of odds ratios (OR). OR with 95% confidence intervals was calculated using the on-line software written by DJR Hutchon (<http://www.hutchon.net>).

3. Results

Tables 1 and 2 show the genotype counts and allele frequencies respectively in patients and controls for the two polymorphic systems analyzed. Genotype distribution at each locus, were consistent with Hardy-Weinberg equilibrium (table 1). As previously reported (Cruz et al., 1997; Rinetti et al., 2001) the most frequent variants for the exon 3 VNTR in the Mexican population were the 4- and 7- 48bp repeat alleles. Moreover; the duplicated 12bp repeat on the exon 1 represented the most common allele in our population.

Allele frequencies for the 12bp polymorphism in exon 1 were not different between the psychotics and controls ($\chi^2 = 0.12$, $P = 0.76$, Table 2); while there was a statistical significant difference between groups for the VNTR system in exon 3 (CLUMP T4 $\chi^2 = 21.0$, $P < 0.01$). A similar result was also observed when the affected group was constituted by the subgroup of patients with schizophrenia (CLUMP, T4 $\chi^2 = 15.3$, $P < 0.01$, Table 2). It is worth to note that either of these differences were the outcome of the CLUMP software through the “clumping” of 3-, 5-, 6-, and 8- repeats alleles. These alleles together with the 2- repeats variant represent the less frequent variants in our population. We therefore decide to compare the frequencies all these “rare” alleles vs. the commonest (4- and 7-) alleles between groups. Significant differences were obtained for the psychotics ($\chi^2 = 19.3$, $df = 1$, $P < 0.01$) or patients with schizophrenia ($\chi^2 = 13.6$, $df = 1$, $P < 0.01$) compared to controls. These results confirm the increase of rare alleles in patients respect to controls.

We estimated the frequencies of haplotypes in cases and control samples. Among 14 possible haplotypes, all were observed in psychotics, 13 were observed in the patients with schizophrenia and 12 in the control group (Table 3). Therefore the gene diversity for these two loci in psychotic patients and control samples were 0.78 ± 0.02 and 0.68 ± 0.02 , respectively. The most common haplotype was the 4-A1 in patients and controls, followed by the 7-A1 haplotype. The 5-A2 and 8-A2 were completely absent in control individuals, whereas the 5-A1, 6-A2 and 8-A1 were detected in very low frequencies (<1%). In psychotics or patients with schizophrenia the following haplotypes 3-A1, 5-A1, 6-A1 and 8-A1 were present in higher frequencies compared to controls (Table 3).

Comparison of haplotype frequencies between psychotic and controls showed evidence for association (χ^2 with an empirical p value of 0.012, 250 permutations). Non significant difference was obtained from the comparison between the patients with schizophrenia and controls ($P = 0.08$).

Odds ratios for the different haplotypes are showed in Table 3. Statistically significant results were obtained for the 3-A1, OR: 4.26 (95% CI: 1.4 to 12.91, $P = 0.005$), 6-A1, OR: 3.14 (95% CI: 1.0 to 9.86, $P = 0.039$), and 4-A1 haplotypes OR: 0.65 (95% CI: 0.458 to 0.935, $P = 0.019$). Moreover, a significant result was obtained for those individuals who bear one of the following four haplotypes (i.e. 3-A1, 5-A1, 6-A1 and 8-A1), OR: 4.5 (95% CI: 2.2 to 9.2, $P = 0.001$).

4. Discussion

In the present study we performed an association study between psychotic disorders and two polymorphisms in the sequence of exon 1 and exon 3 respectively of the DRD4 gene. The main result was the excess of less frequent alleles of the exon 3 VNTR (including 2-, 3-, 5-, 6- and 8- repeat alleles) in psychotic patients (as well as in the

subgroup with schizophrenia) compared with non-affected individuals. In support of this observation, significant odds ratios were obtained for haplotypes bearing these uncommon variants suggesting a discrete role of these variants in association with the psychotic disorders.

The possible association between genetic variants of the DRD4 48bp VNTR in exon 3 and psychotic disorders has been previously explored. In particular, the Italian group of Serretti analyzed an extended group of affected individuals with major psychosis (i.e. schizophrenia, delusional, bipolar and unipolar disorders), but failed to detect a significant contribution of any of the common DRD4 exon 3 alleles to the main diagnoses (Serretti et al., 1999a, 2001). However, an alternative analysis of the data using clinical symptoms as the phenotypic variable detected a significant association between the so-defined delusional factor and the long variants of the gene (a compound of the alleles ≥ 5 repeats) (Serretti et al., 1999b, 2001). This result however could not be further replicated in a different sample of affected individuals of German origin (Serretti et al., 2004). Moreover, there are multiple published studies with respect to the DRD4 VNTR exon 3 and schizophrenia. Two recent meta-analysis (Jönsson et al., 2003; Glatt et al., 2003) did not give positive evidence for an association, while a third one reported a significant association between the long form (≥ 6 repeats) of the gene and schizophrenia, only in Caucasians but not for Asian people (Lung et al., 2002). It is worth to noting that in many of the studies, including some meta-analyses, it has been customary to dichotomize the genetic data on the basis of the number of repeats (i.e. “long” $\geq 6R$ vs. “short” $\leq 5R$). *In vitro* research has shown only minor functional and pharmacological differences properties for the 7 repeat allele in comparison with the other two most common polymorphic variants of the D4 receptor (Ashgari et al., 1995; Jovanovic et al., 1999); although more recent observations suggest that the 7 repeat sequence could modulate differentially the gene expression, via mechanisms involving RNA stability and/or translational efficiency (Schoots and Van Tol, 2003).

It is important to note that the human DRD4 is one of the most polymorphic genes described to date. Much of this variation is the result of SNP changes in the 48-bp tandem repeats. Over 67 different haplotype variants have been reported to date (Ding et al., 2002; Wang et al., 2004), with many more to be uncovered. Sequence analysis of these haplotypes in humans as well as those present in non-human primates support the idea of the 4 repeats variant as the ancestral allele, whereas the 7 repeats sequence seems to be originated more recently as a rare mutational event that increased in frequency in human populations by means of a balanced positive selection genetic mechanism (Livak et al., 1995; Ding et al., 2002; Wang et al., 2004). Moreover, since the sequence motif of the 4-repeat allele is present in most of the non-4 repeats variants, it has been suggested that an unequal recombination between two 4-repeat alleles would raise alleles from the most common 2- to 6-repeats alleles (Ding et al., 2002). All this has led to the suggestion that

characterizing individuals by *DRD4* 4/4 versus non-4 repeats genotypes could be more appropriate for testing gene vs. behavior-disease associations (Wang et al., 2004). However we did not observe any significant difference by comparing the data in this way.

On the other hand, the 12 bp VNTR in exon 1, lies in the region that encodes a four-amino acid duplication, part of extra cellular domain (Seeman et al., 1994). The functional significance of this genetic variation is unknown. *In vitro* experiments have shown only minor pharmacological differences with respect to quinpirole and clozapine binding properties when a recombinant protein constituted by the *DRD4* 48 bp 4-repeats allele plus the A1 allele was compared to the construct containing the A2 variant (Zenner et al., 1998).

Our observation of an absence of association between the exon 1 12-bp polymorphism and psychotic or schizophrenic patients is in agreement with most of the published studies (Ohara et al., 1996; Catalano et al., 1993; Hong et al., 1998; Lee et al., 2000) and a recent meta-analysis (Jönsson et al., 2003). Only Catalano *et al.*, (1993) reported a positive association between this particular genetic marker and patients diagnosed with delusional disorder. Since in our sample only two affected individuals had this particular diagnosis we could not check for this putative association.

Limitations of our study include the population stratification and sample size effects characteristics of case-control genetic association studies. Moreover, although haplotypes are more informative than unique markers in studies of association; they are prone to mechanisms of mutation, selection, genetic drift, sample size sample and admixture that alter their population frequencies.

In this study we showed a novel association between rare alleles and haplotypes of the 48bp- VNTR *DRD4* gene and psychotic disorders in Mexican patients. One must be cautious drawing premature conclusions of our findings, as further investigations into the functional characteristics among the different *DRD4* haplotypes and their physiological consequences are required.

Acknowledgments

This research was supported by CONACYT, grant # 0828P, Mexico. We thank MsC. Beatriz Camarena and Dr. Humberto Nicolini for providing some of DNA samples for the control group. Maria Garcia was helpful in interviewing control subjects.

References

- Asghari, V., Sanyal, S., Buchwaldt, S., Peterson, A., Jovanovic, V., Van Tol, H.H., 1995. Modulation of intracellular cyclic AMP levels by different human dopamine D4 receptor variants. *Journal of Neurochemistry* 65, 1157-1165.

- Catalano, M., Nobile, M., Novelli, E., Nöthen, M.M. , Smeraldi, E., 1993. Distribution of a novel mutation in the first exon of the human dopamine D4 receptor gene in psychotic patients. *Biological Psychiatry* 34, 459-464.
- Chang, F.-M., Kidd, J., Livak, K., Pakstis, A. , Kidd, K., 1996. The world-wide distribution of allele frequencies at the human dopamine D4 receptor locus. *Human Genetics* 98, 91-101.
- Cruz, C., Camarena, B., King, N., Paez, F., Sidenberg, D., de la Fuente, J.R. , Nicolini, H., 1997. Increased prevalence of the seven-repeat variant of the dopamine D4 receptor gene in patients with obsessive compulsive disorder with tics. *Neuroscience Letters* 231, 1.
- Derogatis, L.R., Lipman, R. , Covi, L., 1973. SCL-90: An outpatient psychiatric rating scale-preliminary report. *Psychopharmacology Bulletin* 9, 13-28.
- Ding, Y.C., Chi, H.C., Grady, D.L., Morishima, A., Kidd, J.R., Kidd, K.K., Flodman, P., Spence, M.A., Schuck, S., Swanson, J.M., Zhang, Y.P., Moyzis, R.K., 2002. Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proceedings of the National Academy of Sciences of the United States of America* 99, 309-14.
- Gelernter, J., Kennedy, J., Van Tol, H.H.M., Civelli, O. , Kidd, K.K., 1992. The D4 dopamine receptor (DRD4) maps to distal 11p close to HRAS. *Genomics* 13, 208-210.
- Glatt, S.J., Faraone, S.V. , Tsuang, M.T., 2003. Schizophrenia is not associated with DRD4 48-base-pair-repeat length or individual alleles: results of a meta-analysis. *Biological Psychiatry* 54, 629-635.
- Hong, C.J., Chiu, H.J., Chang, Y.S. , Sim, C.B., 1998. Twelve-nucleotide repeat polymorphism of D4 dopamine receptor gene in Chinese familial schizophrenic patients. *Biological Psychiatry* 43, 432-435.
- Jönsson, E., Sedvall, G., Nöthen, M. , Cichon, S., 2003. Dopamine D4 receptor gene (DRD4) variants and schizophrenia: meta-analyses. *Schizophrenia Research* 61, 111-119.
- Jovanovic, V., Guan, H.C., Van Tol, H.H., 1999. Comparative pharmacological and functional analysis of the human dopamine D4.2 and D4.10 receptor variants. *Pharmacogenetics* 9, 561-568.
- Lahti, R.A., Roberts, R.C., Cochrane, E.V., Primus, R., Gallager, D., Conley, R. , Tamminga, C., 1998. Direct determination of dopamine D4 receptors in normal and schizophrenic postmortem brain tissue – A[3H]NGD 94-1 study. *Molecular Psychiatry* 3, 528-533.
- Lee, H.S., Lee, M.S., Han, D.J., Lee, H.J., 2000. Association of dopamine D4 receptor DRD4 gene polymorphism with Korean schizophrenic patients. *Korean Journal of Biological Psychiatry* 7, 152-158.

- Lichter, J.B., Barr, C.L., Kennedy, J.L., Van Tol, H., Kidd, K., Livak, K.J., 1993. A hypervariable segment in the human dopamine receptor D4 (DRD4) gene. *Human Molecular Genetics* 2, 767-773.
- Livak, K.J., Rogers, J., Lichter, J.B., 1995. Variability of dopamine D4 receptor (DRD4) gene sequence within and among nonhuman primate species. *Proceedings of the National Academy of Sciences of the United States of America* 92, 427-431.
- Lung, F.W., Tzeng, D.S., Shu, B.C., 2002. Ethnic heterogeneity in allele variation in the DRD4 gene in schizophrenia. *Schizophrenia Research* 57, 239-45.
- Meador-Woodruff, J.H., Haroutunian, V., Powchik, P., Davidson, M., Davis, K.L., Watson, S.J., 1997. Dopamine receptor transcript expression in striatum and prefrontal and occipital cortex. Focal abnormalities in orbitofrontal cortex in schizophrenia. *Archives of General Psychiatry* 54, 1089-95.
- Ohara, K., Nakamura, Y., Xie, D.-W., Tatsuya, I., Deng, Z.-L., Tani, K., Zhang, H.-Y., Kondo, N., Liu, J.-Ch., Miyasato, K., Ohara, K., 1996. Polymorphisms of dopamine D2-like (D2, D3 and D4) receptors in schizophrenia. *Biological Psychiatry* 40, 1209-1217.
- Ott, J. 1988-2001. Program HWE©. Utility programs for analysis of genetic linkage. Columbia University. <http://linkage.rockefeller.edu/soft/linkutil/>.
- Rinetti, G., Camarena, B., Cruz, C., Apiquian, R., Fresan, A., Paez, F., Nicolini, H., 2001. Dopamine D4 receptor (DRD4) gene polymorphism in the first psychotic episode. *Archives of Medical Research* 32, 35-38.
- Roberts, D.A., Balderson, D., Pickering-Brown, S.M., Deakin, J.F., Owen, F., 1996. The relative abundance of dopamine D4 receptor mRNA in post mortem brains of schizophrenics and controls. *Schizophrenia Research* 20, 171-174.
- Sanyal, S., VanTol, H.H.M., 1997. Review the role of dopamine D4 receptors in schizophrenia and antipsychotic medication. *Journal of Psychiatry Research* 31, 219-232.
- Schneider, S., Roessli, D., Excoffier, L., 2000. Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- Schoots, O., Van Tol, H.H.M., 2003. The human dopamine D4 receptor repeat sequences modulate expression. *Pharmacogenomics Journal* 3, 343-348.
- Seeman, P., Guan, H., Van Tol, H. H., 1993. Dopamine D4 receptors elevated in schizophrenia. *Nature* 365, 441-445.
- Seeman, P., Ulpian, C., Chouinard, G., Van Tol, H.H.M., Dwosh, H., Lieberman, J., Siminovitch, K, Liu, I.S.C., Wayne, J, Voruganti, P., Hudson, C., Serjeant, G.R., Masibay, A.S., Seeman, M., 1994. Dopamine D4 receptor variant, D4GLYCINE194, in Africans, but not in Caucasians: no association with schizophrenia. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 54, 384-390.

- Seeman, P., 1992. Dopamine receptor sequences therapeutic levels of neuroleptics occupy D2 receptors, clozapine occupies D4. *Neuropsychopharmacology* 7, 261-284.
- Serretti, A., Lilli, R., Di Bella, D., Bertelli, S., Nobile, M., Novelli, E., Catalano, M., Smeraldi, E., 1999a. Dopamine Receptor D4 gene is not associated with major psychoses. *American Journal of Medical Genetics* 88, 486-491.
- Serretti, A., Macciardi, F., Catalano, M., Bellodi, L., Smeraldi, E., 1999b. Genetics variants of dopamine receptor D4 and psychopathology. *Schizophrenia Bulletin* 25, 609-618.
- Serretti, A., Lilli, R., Lorenzi, C., Lattuada, E., Smeraldi, E., 2001. DRD4 exon 3 variants associated with delusional symptomatology in major psychoses: a study on 2,011 affected subjects. *American Journal of Medical Genetics* 105, 283-290.
- Serretti, A., Lorenzi, C., Mandelli, L., Cichon, S., Schumacher, J., Nöthen, M.M., Rietschel, M., Tullius, M., Ohlraun, S., 2004. DRD4 exon 3 variants are not associated with symptomatology of major psychoses in a German population *Neuroscience Letters* 368, 269-273.
- Sham, P.C. , Curtis, D., 1995. Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Annals of Human Genetics* 59, 97-105.
- Stefanis, N.C., Bresnick, J.N., Kerwin, R.W., Schofield, W.N. , McAllister, G. 1998. Elevation of D-4 dopamine receptor mRNA in postmortem schizophrenic brain. *Molecular Brain Research* 53, 112-119.
- Van Tol, H., Bunzow, J.R., Guan, H., Sunahara, R., Seeman, P., Niznik, H.B., Civelli, O., 1991. Cloning of the gene for a human dopamine D4 receptor with high affinity for the antipsychotic clozapine. *Nature* 350, 610-614.
- Van Tol, H., Wu, C.M., Guan, H.-Ch., Ohara, K., Bunzow, J.R., Civelli, O., Kennedy, J., Seeman, P., Niznik, H.B. , Jovanovic, V., 1992. Multiple dopamine D4 receptor variants in the human population. *Nature* 358, 149-152.
- Wang, E., Ding Y.-C., Flodman, P., Kidd, J.R., Kidd, K.K., Grady, D.L., Ryder, O.A., Spence, M.A., Swanson, J.M. , Moyzis, R.K., 2004. The Genetic Architecture of Selection at the Human Dopamine Receptor D4 (DRD4) Gene Locus. *American Journal of Human Genetics* 74, 931-944.
- Winterer, G. , Weinberger, D.R., 2004. Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends in Neurosciences* 27, 683-690.
- Wong, A.H.C., Buckle, C.E. , Van Tol, H.H.M., 2000. Polymorphisms in dopamine receptors: what do they tell us? *European Journal of Pharmacology* 41, 183-203.
- Zenner, M.-T., Nobile, M., Henningsen, R., Smeraldi, E., Civelli, O., Hartman, D., Catalano, M., 1998. Expression and characterization of a dopamine D4R variant associated with delusional disorder. *FEBS letters* 422, 146-150.

Table 3. Pairwise haplotype frequencies estimated for the DRD4 polymorphisms in psychotic or schizophrenics populations and controls.

HAPLOTYPES	PYCHOTICS* Chr = 248 N = 124	OR	SCHIZOPHRENICS Chr = 166 N = 83	OR	CONTROLS Chr = 254 N = 127
2 – A1	0.026	1.19	0.027	1.02	0.025
2 – A2	0.014	1.36	0.009	0.51	0.011
3 – A1	0.062	4.26	0.038	2.34	0.016
3 – A2	0.019	5.14	0.016	4.66	0.004
4 – A1	0.359	0.66	0.390	0.75	0.460
4 – A2	0.137	0.86	0.128	0.80	0.154
5 – A1	0.018	5.14	0.022	6.25	0.004
5 – A2	0.009	-	0.008	-	-
6 – A1	0.047	3.14	0.049	3.17	0.016
6 – A2	0.009	1.01	0.011	1.54	0.008
7 – A1	0.237	0.95	0.220	0.86	0.243
7 – A2	0.029	0.49	0.057	0.98	0.056
8 – A1	0.025	6.2	0.039	9.49	0.004
8 – A2	0.007	-	-	-	-

Chr = number of chromosomes. N= sample size. OR= odds ratio, (-) = no showed haplotype.

*Empirical P -value= 0.012.

ARTÍCULO II

Sometido a la revista *Eur Hum Genet*

HAPLOTYPES AND NUCLEOTIDE DIVERSITY ANALYSIS IN THE POLYMORPHIC
SEQUENCE OF THE 48bp-VNTR-DRD4 GENE IN MEXICAN SUBJECTS

Ana Julia Aguirre-Samudio¹, Carlos Cruz Fuentes¹, Mario Cásarez-Espinosa, Luis
Rosales-León²

¹Departamento de Genética, Instituto Nacional de Psiquiatría Ramón de la Fuente, D.F.
México; ²Departamento de Visualización, Dirección General de Servicios de Cómputo
Académico. Universidad Nacional Autónoma de México.

Corresponding to Carlos Cruz Fuentes, e-mail: cruz@imp.edu.mx

Abstract

Dopamine D4 receptor gene (DRD4) is one the most polymorphic gene described, its variation has been object to study in different behaviours and psychiatric disorders, but the sequence variation in natural population is not many knowing. Our goal was to characterize the polymorphic region of exon 3-DRD4 gene in a sample of Mexican people and construct structure models of its variation. New variation or haplotypes were detected, the majority mutations contributed significantly to the genetic variation. The data revealed that the nucleotide diversity in 7R sequence is higher than the 2 and 4R allele. Interestingly, the variation in 7R is reflects of positive selection, which was estimated with Tajima's test. Also, in this allele was found strong linkage disequilibrium. Susceptibility sites to phenotype associated to 7R were detected using the entropy analysis in the comparison of human with non-human primate species (*Hylobates*). Together, our sequence and structure analysis support a scenario where rare and low variation exists on the tandem unit repeats of the VNTR, while the numbers of repeats in the gene, may have a bigger impact on the phenotype expression of different traits and as a result, on the selection of the variants of the DRD4 gen.

Key words: 48bp VNTR-exon 3, dopamine d4 receptor, nucleotide entropy, linkage disequilibrium, models prediction protein.

Introduction

Dopaminergic transmission by mean of dopamine binding and activation of five G-protein-

coupled receptor molecules exerts an important array of physiological and pharmacological processes in the mammalian CNS. In particular, the dopamine receptor D4, which is a member of the D2-like dopamine receptors family, is of relevance in neurobiology and psychiatry. This receptor has been involved in functions such as movement and motor control (Zhang et al, 2002), behavioral responses to novelty (Dulowa et al, 1999), and target of antipsychotic drugs for the treatment of schizophrenia (Liegeois et al, 1995; Sanyal and Van Tol, 1997).

The human dopamine receptor D4 (DRD4) gene has been mapped near the telomere of chromosome 11p (15.5), and is composed of 4 exons and 3 introns. Several polymorphic systems have been described in its sequence (Van Tol et al, 1991, 1992), which renders it as one of the most polymorphic genes reported to date. Particularly interesting is the 48-bp variable number of tandem repeats region (VNTR), located in exon 3. This polymorphic region encodes the putative third intracellular loop of the receptor, which seems to be involved in the coupling to a Gi protein, therefore mediating changes in the levels of intracellular cAMP (Asghari et al, 1995; Jovanovic et al, 1999; Oak et al, 2000). The VNTR has been reported to have 2 to 11 repeats. Even as one takes into consideration ethnic population heterogeneity (Chang et al, 1996; Ding et al, 2002), the 2, 4 and 7 repeats sequences the most frequent variants reported world-wide (Van Tol et al, 1992; Ding et al, 2002).

In Caucasians the 7R variation has been frequently associated with a wide range of psychiatric disorders and behavioral traits (Ebstein et al, 1996; Swanson et al, 2000). Our group has studied this polymorphic system in the Mexican population, and we have reported positive associations for different alleles with several psychiatric phenotypes (Cruz et al, 1997; Rinetti et al, 2001; Aguirre et al, 2007; Camarena et al, 2007; Fresan et al, 2007).

Moreover, an important level of internal variation of the nucleotide sequence of the 48 bp repeats is due to a single nucleotide polymorphism (SNP). This feature in conjunction with the previously described variation in the number of repeats for a specific allele increases substantially the variability of the gene, leading to the description to date of more than 67 different haplotypes in different world-wide populations (Ding et al. 2002; Grady et al. 2003).

More important the analysis of the variants sequence can provide valuable information for deducing the evolutionary history of the species such as monkeys and human. And the systematic association mapping by SNPs inside of susceptibility regions may lead to the identification of haplotypes that increased disease risk because of linkage disequilibrium, which allowing fine mapping of the localization of disease relevant mutations (Sabeti et al. 2002).

We described the variability of the DRD4 gene from mestizo Mexicans sample; characterized new variants in haplotype structure and their nucleotide diversity. We used

approach computational to evaluate the selection of marker that utilizes both the haplotype information content and the physical distribution of the marker, we employed global entropy as an informatively measure (Hampe et al. 2003), and used cladistic approach to detect susceptibility sites in diseases. This theoretical framework altogether was used to select the SNPs that can to maximize the chance of detecting disease association with the information that can be extracted from the genomic region of the interest.

Moreover, in order to evaluate the structural changes associated with the nucleotides variations by non-synonymous substitution in the number of VNTR repeats, several models of the DRD4 protein were constructed.

MATERIALS AND METHODS

All procedures were reviewed and approved by the independent Research and Ethics Boards of the Instituto Nacional de Psiquiatria Ramon de La Fuente, in Mexico City.

A subset from our original panel of DNA samples obtained from Mexican mestizo subjects, whose genotypes for the exon 3 VNTR DRD4 gene were previously known (Aguirre et al, 2006) were analyzed. Sixty six chromosomes were sequenced, corresponding to 33 homozygous subjects with respect to the genotype size; of this set 18 chromosomes belonged to 9 subjects with diagnostic of schizophrenia (Table 1).

PCR amplification and DNA sequencing

One first batch of samples were amplified using the D4-3 and D4-42 primer set described by Lichter et al, (1993), and using thermal conditions as reported previously (Aguirre et al, 2007). PCR reactions were set up in 50 µl volumes, containing 100ng of genomic DNA. The products were verified in 2% agarose gel and the corresponding fragment was excised of the gel and purified with the QIAquick Gel Extraction Kit® (QIAGEN). Fragments were sequenced using the same D4-3 and D4-42 primer set as above. In a second group of samples, amplification and sequence was obtained employing the additional primers: sense 5'CGTACTGTGCGGCCTCAACGA3' and antisense 5'GACACAGCGCCTG CGTGATGT3' (Ding et al., 2000), under PCR condition: 35 cycles to 95°C by 1min, 55°C/30seg, 72°C/2min and extension 72°C/10min. The sequence reactions were carried out with the BigDye Terminator v3.1 Chemistry (Applied Biosystems) and electrophored in an ABI Prism 3100 Genetic Analyzer (Applied Biosystems). We used the nomenclature reported by Ding et al (2002) and were compared to haplotypes previously described (Ding et al., 2000; Wang et al, 2004). The comparison of the peptide sequences deduced from the nucleotide sequence was done using PSI-BLAST and the BioEdit software. Accession numbers in the Web site, <http://www.ncbi.nlm.nih.gov/Genbank/index.html>, for data presented herein are: EF507590, EF489028-35, DQ846850, DQ846849

Polymorphism and population analysis.

The polymorphism sites, Tajima's test (Tajima, 1989), Fu and Li's D test (Fu and Li, 1993) of the alleles repeats, as well as descriptive statistics and linkage disequilibrium were computed with DnaSP ver. 4.10.9 (Rozas et al, 2003). The genetic structure analysis (F_{ST} and AMOVA) was calculated with Arlequin software (Schneider et al., 2000), with objective of found difference between subpopulations. The variations (neutral or upon natural selection) were evaluated with theoretical distribution of Tajima's D (95% confidence interval of -2 to $+2$), it assumes that polymorphism ascertainment is independent of allele frequency. High values of Tajima's D suggest an excess of common variation in a region, which can be consistent with the balancing selection. Negative values of Tajima's D, on the other hand, indicate an excess of rare variation, or positive selection.

Nucleotide Entropy analyse in repeats alleles

This consideration was based in Shannon entropy (Shannon, 1948) as a measure of genetic diversity and association. Here, Shannon's entropy is applied to different degree of conservation of nucleotide and selection of SNPs. Given a random variable X with probabilities $P(x_i)$ for discrete set of events. Shannon's entropy, also known as Information Entropy, is defined by: $H(X) = -\sum P(x_i) \log P(x_i)$ The entropy, is maximal when the probability of all the events $P(x_i)$ are equal and thus the outcome is most uncertain. The degree of conservation in the positions of the sequences of the alleles of 7R was calculated with the average of the entropy (H_x). High entropy indicates regions very variable and low entropy that it is highly conserved. This analysis was computed with the BioEdit software. The alignment of the sequences of humans with those of nonhuman primates were compared, which they were obtained of the GenBank (*Hylobates muelleri*, access number: AB032910), *Hylobates lar*: AB032907, *Hylobates pileatus*: AB032908, and the *Hylobates agilis*: AB032908 (Inoue-Murayama et al, 2000).

Molecular Modelling of DRD4

In order to evaluate the structural changes associated with the variations in the number of VNTR repeats, several models of the DRD4 protein were constructed. For this purpose, the 3D coordinates for the DRD4 protein were extracted from the database of G-protein coupled receptors (Zhang et al, 2006). From the group of models corresponding to the DRD4 receptor (ID: P21917, Protein Database, <http://www.ncbi.nlm.nih.gov>), the model with the best contacts profile, as evaluated with PROSA2003 (Sippl, 1993), was chosen as the template for the molecular modelling assays (model 3). The final models for the DRD4 with 4 and 2 VNTR repeats were constructed from this 7R template, by deletion of the central repeats. Finally, the modified loops were refined by interactive molecular dynamics (IMD) with NAMD2 (Kale et al, 1999) and VMD (Humphrey et al, 1996 (<http://www.ks.uiuc.edu/Research/vmd/>)).

RESULTS

Association between the haplotypes reported and Mexican variants

From the analysis of sequencing of the 48bp-VNTR-exon 3 of the DRD4 gene in Mexican mestizo subjects, we found 13 different haplotypes, of whom five were described previously. In agreement with the results described by Ding et al. (2002) and Grady et al. (2003) in different world-wide populations, the more frequent haplotypes in our sampling were the 1-4, 1-2-3-4 and 1-2-6-5-2-5-4 for the 2R, 4R and 7R alleles respectively, representing as a whole 79% of the total number of chromosomes analyzed (table 1). The other two haplotypes previously reported were the 1-2-13-4 and 1-2-6-5-2-5-19. Moreover, eight new haplotypes were observed: five in chromosomes with 7R alleles, two corresponding to 2R alleles and one for a new 4R variant. Of these new variants TWO were found in a 7R/7R subject with diagnosis of schizophrenia.

DRD4 nucleotide sequence variation

The fragment of DNA of VNTR in exon 3 was examined. Because the number of the bases varies per allele (Table 2), these numbers are used to calculate all measures of nucleotide diversity. The sequence variability per repeat unit is summarized in figure 1. We found 16 variants, which had synonymous (SS) and non-synonymous (NS) substitutions (with 16 codons). Of which the 1 to 27 sequences were described. Here, we reported 6 repeats units news.

All 15 codons of each one of the repeats units, there were 15 SS, and 30 NS. Of the NS sites found in our sample, 2 replacement SNPs were found at the sites 12 and 13 (Gly→Ile change at amino acid codon 4). Also, more replacements were in the sites 38-40 (Asp→Thr change at amino acid codon 13), at sites 41-43 (Cys→Val change at amino acid codon 14), at sites 44-46 (Ala→Leu change at amino acid codon 15), although, these NS are rare and it belongs to haplotype 1-37 (figure 1).

On the basis of all sites for each repeats allele sample, our two nucleotide-variability measures were π and θ (see table 2), and our data are consistent with other polymorphism studies (Przeworski et al, 2000; Verrelli et al, 2002). The description of the data of the three alleles (2, 4 and 7R) is summary in table 2. The genetic variation was major in the sequence of 7R with 21 nucleotide substitutions, and invariable sites detected were 315.

We analysed the genetic structure between the subgroups of the schizophrenics (S) and controls (C) with AMOVA analysis, which showed a little difference of subjects with 4R and 7R sequences (1.4% and 2.8%, respectively, not significant). The F_{ST} test showed that the subject with 7R alleles in both subsamples S and C (0.003 and 0.001), too have a little difference among groups.

The complete sample was evaluated with the D' Tajima. The value obtained was statistically significant (-2.21, $p < 0.01$) in the 7R chromosomes. A similar result was

obtained when employing the Fu and Li's D statistic test (-3.35, $p < 0.02$) (Fu and Li, 1993). The 2R sequences data showed a bias due to the size of the sub-samples.

Linkage disequilibrium

Linkage disequilibrium (LD) was studied by calculating the P value of Fisher's exact test in all pairwise associations of substitution polymorphisms. The LD analysis showed that the sequences of 4R alleles had 3 of 192 sites polymorphic. The D obtained was significant in one pairwise comparison (nucleotides 107 vs. 127) (Bonferroni's corrected $X^2: 12.0$, $p < 0.001$). The group of 2R alleles was tested and was not significant.

In the case of the 7R alleles, the numbers of polymorphic sites were 21 of 336 sites. In this analysis, we find that our sample size is sufficiently large to detect significant LD with SNPs that have frequencies of $\geq 3\%$. Of 210 possible pairwise combinations between informative sites, 43 were significant at the 0.05 level, of which 3 maintained significance at the < 0.0001 level after the Bonferroni procedure (Figure 2). All three pairwise comparisons highly significant ($D' = 1.0$, $p < 0.0001$). These sites with high LD (figure 3) were the 304, 309, 313 nucleotides (with frequencies of 12%), which were identified as parsimony informative sites.

Entropy

A way of evaluate the mapping of LD upon haplotypes SNPs were to use the strategy of calculate the mean entropy of the 7R sequences. We find 6 regions conserved from the multiple alignment with sequences of the VNTR of 48pb of gibbons and humans with sequences of 7R (Table 3). The values of entropy were small (< 0.2) suggesting high conservation. The sites such as 309 and 313 with significant LD within each allele class lies inside of entropy region, which maximize the use of these sites as markers in phenotypes associated to 7R.

Molecular Modelling of DRD4.

We constructed several models from the nucleotide sequence information for the 7R, 4R and 2R variants of the D4 dopamine receptor protein, as coded by the VNTR of exon 3. One of the most important steps during the models construction was the quality evaluation of the several available models from group P21917. In this regard, the use of Prosa2003 allowed us to choose among very similar models, the one with the best combined energy for C β interactions.

The use of IMD as a refining technique, allowed us to refine the VNTR region in a simple and fast way, without the need of a simulation over the whole membrane-receptor model. Also, by using IMD we were able to monitor carefully the position of the loops corresponding to this region, in order to keep it on the "cytoplasmic" side of the receptor, as there is an important body of evidence that is consistent with the VNTR region as a cytoplasmatic important region for the interaction with G-proteins.

The final models are shown on figure 3. As expected, the model for the 2R shows the biggest structural changes on the VNTR loop, caused mainly by the deletion of a bigger number (five) of repeats, when taking as reference the 7R. Our analysis of the sequence variability of the VNTR showed that the tandem repeat unit is highly conserved in all variants analyzed (Figure 1), with an average value of 88%. On this respect, half of the residues encompassing the tandem unit show 100% conservation, with the lowest conservation value being 50% for position 7 of the aminoacid sequence (in Figure 1). Similarly, when the average variability of the tandem repeat unit are mapped on the spatial structure of the DRD4 models (Figure 4), is not possible to locate regions of high variability.

DISCUSSION

In this study, we characterized the sequence of the region of the 48bp-VNTR-exon 3 of the DRD4 gene, identified new haplotypes, analyzed the diversity nucleotide and detected sites in linkage disequilibrium. Besides, we built one model for each variants studied as prediction of the drd4 protein.

The 48bp-VNTR sequencing has not been reported before in the mestizo Mexican people, however some allelic variations have been described in other populations (Ding et al. 2002; Grady et al, 2003). From genotype screening which included 2, 4 and 7R alleles by size, taken to intention because they are most common, we found common haplotypes or more frequents in world population, which they are consistent with the previously described by Ding et al. (2002) and Grady et al. (2003). Too, news forms inside of haplotype sequence of VNTR-exon 3 were observed. Although, it is not possible to be discarded that there are other variants not even identified in our population study, its ethnic variability is given, because it is possible that our sample is not representative of all mestizo population. Although, subjects with 2R alleles were a few, because its population frequency this on was represented. Nevertheless we did one on selection of 7R alleles, since there are reports of clinical samples in where 7R have rare haplotypes (Grady et al, 2003).

The nucleotide variability study of the DRD4 gene could help to understand the functionality of the 48bp-VNTR polymorphism in relation to the psychiatric disorders or traits of the personality, which have been associated to this polymorphism. Thus, the nucleotide diversity was calculated by each sequence of 2, 4 and 7 alleles due to difference in size. The 4 and 7 alleles presented range of π between 0.34% and 0.54%, respectively. On report of DRD4 gene in human and chimpanzee, bonobo and gorilla show nucleotide diversity a pattern similar to that observed to π of 4R allele (Jensen- Seaman et al, 2001). But the nucleotide diversity in 7R is high, and this result could be subject to variation by its recent origin (Ding et al, 2002).

To assess whether the nucleotide variation of DRD4 gene characterized in our sample are neutral or by natural selection, we investigated the patterns of polymorphism by means of ISM way, derived of Neutral Model. One common measure is Tajima's test (D). The 2R and 4R groups had neutral variation. However, the D of 7R groups suggest that positive selection may be present in our sample, which is based on the presence of negative Tajima D and significant p values. Simulations show that this test does possess the statistical power to detect alterations in the frequency of polymorphisms (Braverman et al, 1995). The selection effect on 7R is in agreement with the positive selection model proposed by Ding et al. (2002). At that time, that 7R evolution is relative recent apparition, which increases its frequency around the world. It should be noted the D-value is subject to the same population dynamic where 7R follows fixing.

Alternatively, has been speculated that the correlation between 7R and its positive selection may also provide a corresponding migration patterns in prehistoric times as an adaptive value in migratory societies (Chen et al, 1999), due that had been associated to novelty seeking, hyperactivity, and exploration behaviour and the high proportion of 7R allele. The extent world geographical distribution linkages to novelty-seeking or attention deficit /hyperactivity disorder (ADHD) and some authors think that this allele benefits to certain styles of society rather than being merely genetic coincidences with effects individual (Williams and Taylor, 2006).

The analyses to detail of the nucleotide sequence showed that 4R allele had a few variation and LD sites, which may appear to be older in origin than 7R. We found that 7R allele had variation and strong LD between parsimony informative sites or regions conserved of the sequence, suggesting that this allele could be associated by evolutionary mechanisms that fixed for functionally with strong selective pressures. The identifying regions of the human genome that have experienced substantial selective pressure can provide insights into the location of functionally important polymorphisms and may help prioritize targets for association mapping (Sabeti et al, 2002; Clark et al, 2003).

The abundant polymorphism of 7R compared to 2 and 4R is likely the result of a relatively recent change in selective pressures in humans. Our comparison of human and gibbons (*Hylobates*) of the exon 3-DRD4 nucleotide sequences implies that has been no protein change for this receptor possible for the past 28 Myr. The entropy analysis confirms the loci inside of regions conserved in all 7R sequence of both genus *Homo* and *Hylobates*, been it the unique non human primate that present 7R allele.

Our results from the sequence analysis of the residue conservation on the VNTR unit show that this is indeed a highly conserved region. Even as the VNTR-exon 3 of DRD4 gene shows nucleotide diversity, most of the present mutations are conservative, without any effect on the aminoacid sequence of the protein. Likewise, the residue variation on the VNTR is very low and, in most of the reported variants, the physicochemical properties of the original aminoacid are conserved.

All of these results, point towards a scenario where nucleotide/residue variation of the VNTR has low impact on the selection of the possible variants of this gene, while the numbers of repeats in the gene, may have a bigger impact on the phenotype expression of different traits and as a result, on the selection of the variants of the DRD4 gene.

In this work it was not tried to study association with some clinical phenotype, nevertheless within the sample subjects with schizophrenia diagnosis were included and was there a new variation (Table 1). Interestingly and of particular clinical relevance is the association with ADHD, where over 10% of novel or rare haplotypes 7R has been described (Grady et al, 2003). Characteristic of patients with ADHD diagnostic due that in autism probands do not found susceptibilities variants (Grady et al, 2005). Nevertheless, our next studies go directed to evaluate the sample with ADHD.

In conclusion, our study suggest that the nucleotide diversity in the VNTR-exon 3 in 7R alleles of DRD4 gene have not evolved according to the neutral model and one possibility is that it has evolved under positive selection, which has been fixed. LD analysis between nucleotides was strong in three different sites of 7R allele, and this can be explained by the idea that natural selection works on the function of conserved exons to final protein products, and it confirms that the discovered mutations no could to alter the amino acid sequence of the protein.

ACKNOWLEDGEMENTS

All the Molecular Dynamics simulations were done on the facilities of the Visualization Department at the DGSCA, UNAM. The authors wish to thank the Postgrade in Biology, UNAM, and especially Lic. Héctor Cuevas Vázquez del Mercado from the Visualization Department of DGSCA for the support granted to the Ph.D. student Ana Julia Aguirre Samudio.

REFERENCES

- Asghari V, Sanyal S, Buchwaldt S, Peterson A, Jovanovic V, Van Tol H.H: Modulation of intracellular cyclic AMP levels by different human dopamine D4 receptor variants. *J Neurochem* 1995; **65**: 1157-1165.
- Braverman JM, Hudson RR, Kaplan NL, Langley CH, Sthepan W: the hitchhiking effect on the site frequency spectrum of DNA polymorphism. *Genetics* 1995; **140**: 783-796.
- Camarena B, Loyzaga C, Aguilar A, Weissbecker K, Nicolini H: Association study between the dopamine receptor D(4) gene and obsessive-compulsive disorder. *Eur Neuropsychopharmacol* 2007; **17**: 406-9.

- Chen C, Burton M, Greenberger E, Dmitrieva J: Population migration and the variation of dopamine D4 receptor (DRD4) allele frequencies around the globe. *Evol Human Behavior* 1999; **20**: 309–324.
- Clark AG, Glanowski S, Nielsen R, Thomas P, Kejariwal A, Todd MJ, Tanenbaum DM, Civello D, Lu F, Murphy B, et al: Inferring nonneutral evolution from human–chimp–mouse orthologous gene trios. *Science* 2003; **302**: 1960–1963.
- Cruz C, Camarena B, King N, Paez F, Sidenberg D, de la Fuente JR, Nicolini H: Increased prevalence of the seven-repeat variant of the dopamine D4 receptor gene in patients with obsessive-compulsive disorder with tics. *Neurosci Lett*. 1997; **231**:1-4.
- Ding YC, Chi HC, Grady DL, Morishima A, Kidd JR, Kidd KK, Flodman P, Spence MA, et al: Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proc Natl Acad Sci USA* 2002; **99**: 309-14.
- Dulowa S, Grandy DK, Low MJ, Paulus M, Geyer MA: Dopamine D₂ receptor-knock-out mice exhibit reduced exploration of novel stimuli. *J Neuroscience* 1999; **19**: 9550-9556.
- Fresan A, Camarena B, Apiquian R, Aguilar A, Urraca N, Nicolini H: Association study of MAO-A and DRD4 genes in schizophrenic patients with aggressive behavior. *Neuropsychobiology* 2007; **55**: 171-5.
- Fu Y-X, Li W-H: Statistical tests of neutrality of mutations. *Genetics* 1993; **133**: 693-709.
- Grady DL, Chi H-C, Ding Y-C, Smith M, Wang E, Schuck S, Flodman P et al: High prevalence of rare dopamine receptor D4 alleles in children diagnosed with attention-deficit hyperactivity disorder. *Mol Psychiatry* 2003; **8**: 536-545.
- Grady DL, Harxhi A, Smith M, Flodman P, Spence MA, Swanson JM, Moyzis RK: Sequence Variants of the DRD4 Gene in Autism: Further Evidence That Rare DRD4 7R Haplotypes Are ADHD Specific. *Am J Med Genet* 2005; **136B**:33–35.
- Gelernter J, Kennedy J, Van Tol HHM, Civelli O, Kidd KK: The D4 dopamine receptor (DRD4) maps to distal 11p close to HRAS. *Genomics* 1992; **13**: 208-210.
- Hampe J, Schreiber S, Krawczak M: Entropy-based SNP selection for genetic association studies. *Hum Genet* 2003; **114**: 36-43.
- Humphrey W, Dalke A, Schulten K: VMD – visual molecular dynamics. *J Mol Graphics*, 1996; **14**: 33-38 (<http://www.ks.uiuc.edu/Research/vmd/>).
- Inoue-Murayama, M., Niimi, Y., Takenaka, O. and Murayama, Y: Allelic variation of the dopamine receptor D4 gene polymorphic region in gibbons. *Primates* 2000; **41**, 383-392.
- Jensen-Seaman MI, Deinart AS, Kidd KK. Modern African ape populations as genetic and demographic models of the last common ancestor of humans, chimpanzees, and gorillas. *J Heredity* 2001, **92**: 475-480.

- Jovanovic V, Guan HC, Van Tol HH: Comparative pharmacological and functional analysis of the human dopamine D4.2 and D4.10 receptor variants. *Pharmacogen* 1999; **9**: 561-568.
- Kimura M: *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, MA. 1983, pp
- Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozak Ai, Varadarajan K, Schulten K: NAMD2, Greater scalability for parallel molecular dynamics, *J Comp Phys* 1999; **151**: 283–312.
- Kato M, Sekine A, Ohnishi Y, Johnson TA, Tanaka T, Nakamura Y, Tsunoda T: Linkage disequilibrium of evolutionarily conserved regions in the human genome. *BMC Genomics* 2006, **7**:326.
- Lichter JB, Barr CL, Kennedy JL, Van Tol H, Kidd K, Livak KJ: A hypervariable segment in the human dopamine receptor D4 (DRD4) gene. *Hum Mol Genet* 1993; **2**: 767-773.
- Liegeois JF, Bruhwyler J, Damas J, Rogister F, Masereel B, et al: Modulation of the clozapine structure increases its selectivity for the dopamine D4 receptor. *Eur J Pharmacol* 1995; **273**: R1-3.
- Li H, Zhang Y, Zhang Y-P, Fu Y-X: Neutrality test using DNA polymorphism from multiple samples. *Genetics* 2003; **163**: 1147-1151.
- Przeworski M, Hudson R, Di Rienzo A. Adjusting the focus on human variation *TIG* 2000; **7**: 296-302.
- Rinetti G, Camarena B, Cruz C, Apiquián R, Fresán A, Páez F, Nicolini H: Dopamine D4 receptor (DRD4) gene polymorphism in the first psychotic episode. *Arch Med Res*. 2001; **32**:35-8.
- Rozas J, Sánchez-Del Barrio J C, Messeguer X, Rozas R: DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 2003; **19**: 2496-2497.
- Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko, JV, Patterson NJ, McDonald GJ, et al: Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002; **419**: 832–837.
- Sanyal S, Van Tol HHM: Review: the role of dopamine D4 receptors in schizophrenia and antipsychotic action. *J Psychiat Res* 1997; **31**: 219-232.
- Schneider, S., Roessli, D. , Excoffier, L. Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, 2000. University of Geneva, Switzerland.
- Sippl, M J: Recognition of Errors in Three-Dimensional Structures of Proteins. *Proteins*, 1993; **17**, 355-362.
- Swanson JM, Flodman P, Kennedy J, Spence MA, Moyzis R, Schuck S, et al: Dopamine

- genes and ADH. *Neurosci Biobehav Rev* 2000; **24**: 21-25.
- Tajima, F: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989; **123**: 585-595.
- Van Tol H, Bunzow JR, Guan H, Sunahara R, Seeman P, Niznik HB, Civelli O: Cloning of the gene for a human dopamine D4 receptor with high affinity for the antipsychotic clozapine. *Nature* 1991; **350**: 610-614.
- Van Tol H, Wu CM, Guan H-Ch, Ohara K, Bunzow JR, Civelli O, Kennedy J, et al: Multiple dopamine D4 receptor variants in the human population. *Nature* 1992; **358**: 149-152.
- Verrelli B, McDonald J, Argyropoulos G, Destro-Bisol G, Froment A, Drousiotou A, Lefranc G et al: Evidence for balancing selection from nucleotide sequence analyses of human G6PD. *Am J Hum Genet* 2002; **71**: 1112-1128.
- Wang E, Ding Y-C, Flodman P, Kidd JR, Kidd K, Grady DL, Ryder O, Spence MA, Swanson JM, Moyzis RK. The genetic architecture of selection at the human dopamine receptor D4 (DRD4) gene locus. *Am J Hum Genet* 2004; **74**: 931-944.
- Williams J, Taylor E: The evolution of hyperactivity, impulsivity and cognitive diversity. *J R Soc Interface* 2006; **3**: 399-413.
- Zhang K, Tarazi F, Davids E, Baldessarini RJ: Plasticity of dopamine D4 receptors in rat forebrain: temporal association with motor hyperactivity following neonatal 6-hydroxydopamine lesioning. *Neuropsychopharmacol* 2002; **26**: 625-633.

Table 1. Haplotypes frequencies of VNTR exon 3 of DRD4 gene

Repeats	Haplotypes	Controls	S	χ^2 , P	N (%)
2	1-4	1	4	4.8 0.091	5 (63)
	1-37	2			2 (25)
	1-36	1			1 (12)
4	1-2-3-4	15	4	1.824 0.402	19 (79)
	1-2-13-4	2	2		4 (17)
	1-2-40-4	1			1 (4)
7	1-2-6-5-2-5-4	20	8	7.89 0.246	28 (82)
	1-2-6-5-2-38-4		1		1 (3)
	1-2-6-39-2-5-19	1			1 (3)
	1-2-6-5-2-5-19	1			1 (3)
	42-2-6-5-2-5-4	1			1 (3)
	1-2-6-5-2-5-41		1		1 (3)
	1-2-27-5-2-5-41		1		1 (3)

N (Number of chromosomes). Sub-sample with subjects had diagnostic of schizophrenia (S). χ^2 was calculated with CLUMP. Haplotypes (i.e. 1-2-13-4 and 1-2-6-5-2-5-19) were previously reported in low frequency and ADHD subjects (Ding et al, 2001, Grady et al, 2003).

Table 2. Descriptive and estimators of nucleotide diversity in sequences of three alleles of 48bp-VNTR exon 3-DRD4 gene in the complete sample.

Allels	N	N° of Sites	Sustitutions	$\pi \pm SE$	θ Per site	D Tajima	D Fu and Li
2R	8	96	11	4.35± 2.7	4.42	-0.075 NS	0.578 NS
4R	24	192	3	0.34± 0.3	0.44	-0.43 NS	-0.188 NS
7R	34	336	21	0.54 ± 0.3	1.53	-2.21 p<0.01	-3.35 p<0.02

Repeat Variant	Nucleotide Sequence Protein Sequence
----------------	---

```

REPEAT
1 1 A CCC GCG CCC CGC CTC CCC CAG GAC CCC TGC GGC CCC GAC TGT GCG CC 48
    P A P R L P Q D P C G P D C A P
2 1 C ... .. G.. ..T ... .G. .GT ... .. 48
    P A P G L P R G P C G P D C A P
3 1 C G.. ... .. A.. ... ..T ... .. 48
    A A P S L P Q D P C G P D C A P
4 1 C ... .. G.. ... ..C. ... ..T.. A.. ... ..T .. 48
    P A P G L P P D P C G S N C A P
5 1 C ... .. G.. ... ..C. ... .. 48
    P A P G L P Q D P C G P D C A P
6 1 C G.. ... .. G.. ... ..C. ... ..C .GC .. 48
    A A P G L P P D P C G P D C G P
13 1 C G.. ... .. G.. ... ..C. ... ..T.. A.. ... ..T .. 48
    A A P G L P Q D P C G P D C A P
19 1 C ... ..T G.. ... ..C. ... ..T.. A.. ... ..T .. 48
    P A P G L P P D P C G S N C A P
27 1 C ... .. G.. ... ..C. ... ..C .GC .. 48
    P A P G L P P D P C G P D C G P
36 1 C ... .. G.. ..T ... .G. ..T ... ..T.. A.. ... ..T .. 48
    P A P G L P R D P C G S N C A P
37 1 C ... .. G.. ... ..C. ... ..T.. ACT GTG CTC .. 48
    P A P G L P P D P C G S T V L P
38 1 C ... .. G.. ... ..T ... .. 48
    P A P G L P Q D P C G P D C A P
39 1 C ... .. G.. ... ..A.. ... .. 48
    P A P G L P Q D P C G P N C A P
40 1 C G.. ... .. AT. ... ..T ... .. 48
    A A P I L P Q D P C G P D C A P
41 1 C ... .. G.A ... ..C. ... ..T.. A.. ... ..T .. 48
    P A P G L P P D P C G S N C A P
42 1 . ... ..G ... .. 48
    P A P G L P Q D P C G P D C A P

```

Figure 1. Alignment of 48-VNTR repeats in a sample of Mexican origin. The number of unit followed the nomenclature report by Ding et al, (2001). Observe that from 36 to 42 alleles, were found in this sample.

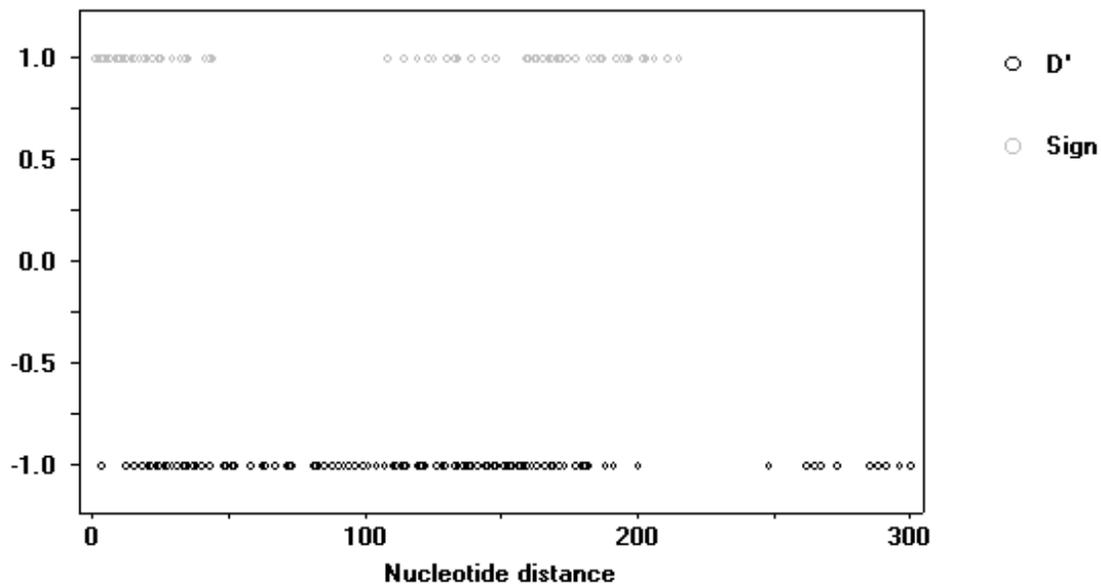


Figure 2. Measurements and plots of LD. Relationship between the LD measure as D' and the distance (in bp). The SNP found had frequencies >3% in a sample of 7R allele of Mexican population. The Number of pairwise comparisons was 210. Of which, 43 were statistic significant with Fisher' test.

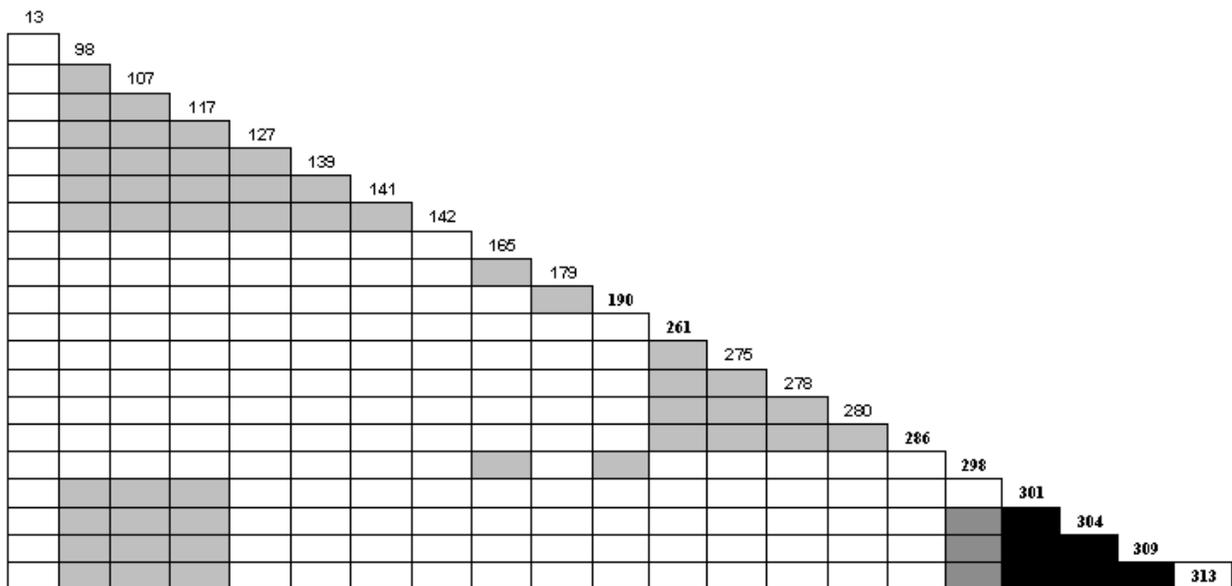


Figure 3. Plot of pairwise comparison among SNP found for 34 chromosomes. Blackened boxes indicate significant LD at the 0.01% level for 3 of the possible 53 associations, with frequency of 12%.

Table 3. Mean entropy of the nucleotides of 7R allele between human and some species of genus *Hylobates*.

Region	Position	Secuencias	Mean Entropy
1	115-131	CAGGACCCCTGCGGCC	0.0072
2	133-163	GACTGTGCGCCCCCGCGCCCGCCTTCCCC	0.0196
3	240-254	CCCCGCGCCCGGCCT	0.0325
4	290-303	CCGCGCCCGGCCTTCC	0.0152
5	307-336	CGGGTCCCTGCGGCCCGACTGTGCGCCC	0.0203
6	409-428	CCCTGCGGCTCCAACCTGTGC	0.0183

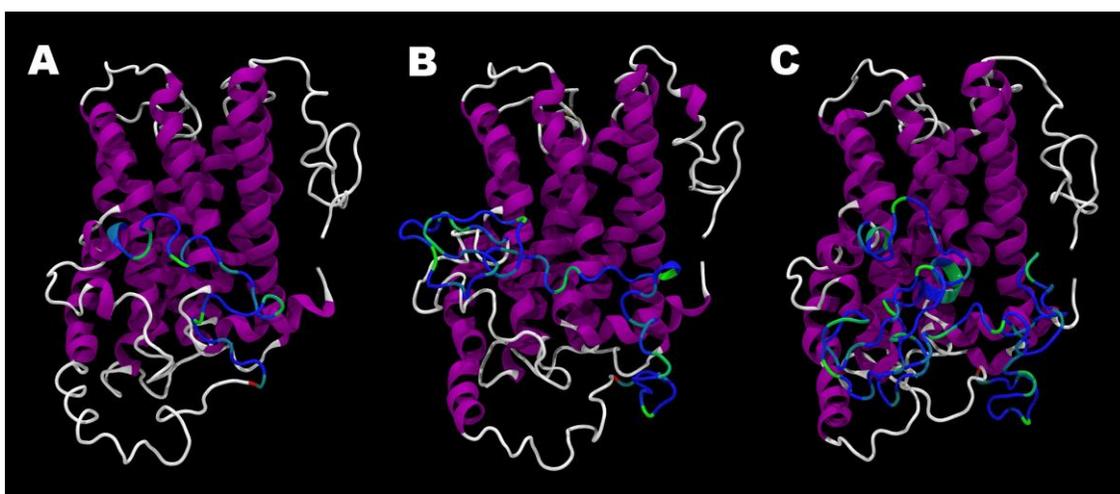


Figure 4. Structure of the modelled variants of the human dopamine receptor D4. Cartoon representation of DRD4-2R, showing the secondary structure of the receptor (alpha helix in purple and coil regions in white), the VNTR is painting according to the sequence variability (highly conserved residues in blue and least conserved residues in green) found in all reported variants of the tandem repeat; (B) DRD4-4R; (C) DRD4-7R.

V. DISCUSIÓN

En este trabajo hemos analizado la variabilidad de varias regiones polimórficas del gen receptor a dopamina D4 (DRD4). Los principales resultados han sido preparados en artículos para publicación por separado. Cada uno centrado en un aspecto distinto de los patrones o las implicaciones que conlleva la variación y dinámica poblacional representada en las frecuencias de sus alelos, y que hemos encontrado en la población estudiada de origen mestiza mexicana tanto de grupos con diagnóstico de psicosis como sin el padecimiento. En cada uno de los manuscritos se han presentado tanto resultados como la discusión correspondiente a cada uno de los artículos.

A continuación trataremos de discutir los aspectos más importantes de este trabajo y de integrar todos los resultados obtenidos.

OBTENCION DE GENOTIPOS Y HAPLOTIPOS

La determinación de los genotipos de la región polimórfica del VNTR del exón 3 fue una parte laboriosa de este trabajo. A pesar de que la técnica ya había sido probada con anterioridad, los resultados no eran convincentes. La metodología fue modificada porque no se podían visualizar claramente los genotipos, y la electroforesis se realizó cambiando a geles en donde la separación de las bandas fue un punto crítico en la identificación del genotipo.

Dentro de los objetivos de este trabajo, uno de ellos fue el de llevar a cabo un estudio de asociación entre sujetos con diagnóstico de psicosis y varios polimorfismos ubicados en la región promotora, exón 1 y exón 3 del gen receptor a dopamina DRD4. El principal resultado, no mencionado con anticipación en la literatura, fue el hallazgo de la asociación entre alelos poco comunes o raros del VNTR del exón 3. Los alelos raros incluyen a 2R, 3R, 5R, 6R y 8R en los pacientes con psicosis (así como en un subgrupo de sujetos diagnosticados con esquizofrenia) en comparación con el grupo de sujetos no afectados. Estos resultados fueron apoyados por los valores de *Odds ratio* (ver sección 3 de materiales y métodos) obtenidos a partir de la agrupación de haplotipos formados entre los alelos de los polimorfismos del VNTR 48pb, el de 12 pb.

La posible asociación entre las variantes del VNTR 48 pb y los trastornos psicóticos ya habían sido estudiadas en un grupo de sujetos afectados de origen italiano con psicosis principal (el cual incluía esquizofrenia, trastorno unipolar y bipolar y trastorno de delirio), en los que no encontraron asociación entre los alelos comunes del VNTR 48pb y la psicosis (Serretti *et al.*, 1999, 2001). Sin embargo en otro análisis en donde la clasificación refería a los síntomas como un fenotipo variable, Serretti *et al.*, (1999b, 2001)

detectaron una asociación significativa entre los síntomas de delirio y las variantes largas del VNTR 48pb (es decir, alelos entre 5R a 8R ó 10R). No obstante dicho resultado no fue replicado en población de sujetos afectados de origen alemán (Serretti *et al.*, 2004). Además, dos estudios del tipo meta-análisis (Jönsson *et al.*, 2003; Glatt *et al.*, 2003) no han mostrado alguna evidencia de asociación de este polimorfismo con esquizofrenia; mientras que un reporte encontró asociación entre los alelos largos (6 repetidos en adelante) y sujetos diagnosticados con esquizofrenia, únicamente en caucásicos pero no se encontró asociación en población de origen asiática (Lung *et al.*, 2002).

Cabe señalar que en muchos estudios del polimorfismo VNTR 48 pb y el padecimiento, incluyendo los de meta-análisis, tienden a agrupar los alelos con base al número de repeticiones como en alelos largos vs. cortos (=6R vs. = 5R). Tal vez la razón de concentrar unos u otras formas, se sustente en los ensayos *in vitro*, los cuales han demostrado pequeñas diferencias funcionales y farmacológicas relacionadas al alelo de 7R en comparación con las otras dos variantes más comunes en la población mundial (2 y 4R) (Ashgari *et al.*, 1995; Jovanovic *et al.*, 1999). Aunque estudios recientes sugieren que la secuencia del alelo 7R, podría modular la expresión del gen, vía mecanismos que envuelven la estabilidad del RNA así como su eficiencia de traducción (Schoots y Van Tol, 2003).

Por otra parte, el VNTR de 12 pb del exón 1 del gen DRD4, el cual codifica una duplicación de 4 aminoácidos y forma parte del dominio extracelular del receptor tiene una importancia funcional; aunque se desconoce con exactitud. Para realizar la genotipificación de este polimorfismo se llevo a cabo el montaje de la técnica.

El estudio de asociación con el VNTR de 12 pb dio resultados de ausencia de asociación entre este polimorfismo y los pacientes con síntomas psicóticos o esquizofrénicos, lo cual coincide con la mayoría de los estudios publicados (Ohara *et al.*, 1996; Catalano *et al.*, 1993; Hong *et al.*, 1998; Lee *et al.*, 2000) y con un meta-análisis (Jönsson *et al.*, 2003). Solamente, Catalano *et al.*, (1993) reporta una asociación entre este marcador genético y los pacientes con diagnóstico de trastorno de delirio. En nuestra muestra objeto de estudio, sólo dos individuos afectados tenían este diagnóstico en particular, de manera que nosotros no podríamos verificar esta supuesta asociación. Estudios posteriores podrían aumentar el tamaño de muestra debido a la baja frecuencia del alelo A2.

Un muestreo subóptimo y la sobre-interpretación de los datos son los motivos más comunes por los cuales no se logra obtener resultados positivos. Los parámetros

representativos en enfermedades multifactoriales son complejos, esto incluye interacciones del loci de la enfermedad y el tamaño de muestra, o que la frecuencia alélica de la enfermedad podría reflejar un rango de mutaciones en el mismo locus. Estos puntos resumen los objetivos básicos de los diseños de estudios de asociación (Zonder van y Cardon, 2004).

Por otra parte, en este apartado del trabajo de investigación también se llevaron a cabo análisis de haplotipos por inferencia. La inferencia de alelos ha sido la técnica más ampliamente usada en los estudios de genética de poblaciones humanas. Nosotros usamos el programa Pm, en el que lleva a cabo permutaciones, es decir lleva a cabo todas las posibles combinaciones a partir de los genotipos observados de cada individuo. También, usamos otro programa (Arlequin de autoría Schneider *et al*, 1995) que estima las frecuencias haplotípicas en la población (esto evita desestimar el estudio, debido al gran número de casillas que se agrupan por haplotipo). El Arlequín utiliza el algoritmo de “maximización de la esperanza” (expectation-maximization) para la inferencia de haplotipos. Este método resuelve la incapacidad de estimar con precisión la frecuencia de haplotipos raros. En el análisis, la muestra de pacientes tuvo más haplotipos inclusive raros en comparación con el grupo control. En este estimado de haplotipos se logró obtener una asociación entre los haplotipos raros de la población psicótica en relación con la población sin el trastorno.

En este trabajo también se analizó el SNP ubicado en el sitio -521 (cambio de C por T) de la región promotora, el cual es importante en la señal de traducción del RNA mensajero del gen DRD4 (Kamakura *et al*, 1997; Okuyama *et al*, 1999). Para el análisis de este polimorfismo, no fue requerido un gran tamaño de muestra (Gordon *et al*, 2002, 2003), con una n de 127. Las frecuencias poblacionales de los alelos C y T de los sujetos controles (0.6 y 0.4, respectivamente) fueron parecidas a las de otras poblaciones (C= 0.55, T= 0.45) (Bookman *et al*, 2002; Ronai *et al*, 2001; Jönsson *et al*, 2001). Nosotros encontramos una diferencia estadísticamente significativa entre el alelo T y genotipo TT y los pacientes psicóticos ($p < 0.01$). Sin embargo, los resultados no son replicados en una muestra de pacientes esquizofrénicos suizos (Jönsson *et al.*, 2001) y japoneses (Mitsuyasu *et al.*, 2001; Okuyama *et al.*, 1999).

Cabe señalar que aunque los datos no son consistentes con otros estudios de asociación, su importancia no puede desestimarse, debido a que este sitio polimórfico se

encuentra dentro de la región regulatoria o isla CpG y es parte del dinucleótido CG. Cualquier alteración en esta región como la modificación en los patrones de metilación, podrían interferir con la transcripción del gen (Kamakura *et al.*, 1997; Bird, 2002). Los estudios *in vitro* con el alelo T muestran que tiene una actividad transcripcional menor en comparación con el alelo C (Okuyama *et al.*, 1999, 2000). De ser así, el riesgo de poseer un alelo T daría un receptor D4 que disminuiría su función como inhibidor de la formación del segundo mensajero AMP cíclico.

El análisis de asociación con el SNP -521, presenta algunas limitaciones, que incluyen la mezcla poblacional, y el efecto del tamaño de la muestra control. Sin embargo, en estos estudios, aunque los haplotipos son más informativos en comparación del uso de un solo marcador, estos están expuestos a mecanismos de mutación, selección natural, deriva génica, efecto del tamaño de la muestra y la mezcla poblacional, los que alteran la frecuencia poblacional.

DIVERSIDAD NUCLEÓTIDICA

El gen humano DRD4 es uno de los más polimórficos y la mayor parte de la variación se debe a cambios en los SNP, en especial dentro de las repetidas de 48pb del exón 3. Aproximadamente 67 diferentes variantes de haplotipos de nucleótidos han sido reportados a la fecha (Ding *et al.*, 2002; Wang *et al.*, 2004).

La secuenciación del VNTR de 48 pb del gen DRD4 no ha sido descrita con anterioridad en una población mestiza de origen mexicana. A partir del tamizaje de genotipos, y del análisis de secuencia encontramos variaciones nuevas dentro del motivo de repetición de este VNTR. Los haplotipos con frecuencias poblacionales altas como es el caso de la forma 1-2-3-4 y 1-2-6-5-2-5-4, coinciden con los datos publicados por Ding *et al.* (2002) y Grady *et al.* (2003).

El número de posiciones que segregan y la diversidad nucleotídica están dentro de los rangos de otros genes y se podría asegurar que al menos la diversidad del alelo de 7R es relativamente alta.

La variación de nucleótidos puede ser explicada mediante el modelo ISM, derivado de la Teoría de la Neutralidad (Kimura, 1983). Una forma de evaluar si existe neutralidad en los datos es con la prueba de Tajima (D), la cual es adecuada para secuencias cortas y se espera que haya correlación con el número de sitios polimórficos y el número promedio de diferencias nucleotídicas (Tajima, 1989).

Las variaciones nucleotídicas de los alelos 2R y 4R, en general, se comportan de forma neutral, y puede deberse a que están ubicados en un lugar funcional del gen en donde la tasa de mutación es baja, lo que afecta la densidad de los SNPs y los valores de la D de Tajima, y se muestra en los datos como bajo polimorfismo de apenas 3 sitios (en el caso de 4R). En cambio, las regiones con alta tasa de mutación promueven una baja frecuencia y crea una débil correlación entre los SNP cercanos. Por lo tanto en los parámetros de la mutación neutral de los alelos 2 y 4, se espera que el bloque de haplotipos sea más pequeño que el predispuesto por la baja tasa de mutación. Esta mutación rompería el tamaño del haplobloque, utilizado en la genética de la enfermedad. Es decir no son adecuados para buscar la relación con psicosis. El análisis de AMOVA y F_{ST} entre los subgrupos de controles y esquizofrénicos no mostró diferencias estadísticamente significativas lo que sugiere que la estructura genética entre las poblaciones con secuencias 4R y 7R de las subpoblaciones, es pequeña y se podría tomar como un resultado de un flujo genético contemporáneo. El grado de diferenciación entre poblaciones no sólo depende del aislamiento reproductivo debido a la enfermedad que muchas veces incapacita física y socialmente, sino también es influenciado por procesos evolutivos como la selección natural, la deriva génica, la tasa de mutación y el tiempo de aislamiento (Mettler y Gregg, 1982). A pesar de existir mayor variabilidad dentro de las poblaciones, el establecimiento de divergencias suele ser complejo, porque la estructura social y geográfica así como el tamaño de la población reproductiva pueden afectar el tiempo de diversificación genética de la población, manteniendo la homogeneidad entre poblaciones aisladas por largo tiempo (Rocha-Olivares y Sandoval-Castillo, 2003). Por tanto, la falta de heterogeneidad entre las poblaciones puede deberse a un aislamiento por baja frecuencia de apareamiento de los esquizofrénicos.

Un efecto de la selección puede observarse en el valor general de D (-2.21, $p < 0.01$) encontrado en las mutaciones del alelo 7R, el cual sugiere un exceso de variación rara tanto de alelos de baja como de alta frecuencia, consistente con el crecimiento poblacional o selección positiva (Thomas *et al.*, 2006). Esto podría explicarse por que la población de individuos con los alelos 7R está creciendo más que la población con alelos 2R y 4R. Este alelo tiene un efecto de adecuación que propiamente tiene una ventaja sobre los demás alelos (2R y 4R).

El efecto adaptativo del 7R podría estar influenciado por las migraciones pasadas

o del producto del mestizaje durante la colonización. Las mutaciones de estos alelos pueden formar algunas relaciones de los SNP cercanos manteniendo el haplobloque para usarse en la genética de la enfermedad relacionada con el gen DRD4 como la psicosis (en análisis de DL).

Otros estudios reportan un efecto de selección direccional o positiva relacionado con el alelo de 7R (Ding *et al.*, 2001). Estos autores analizan el origen de los alelos y encuentran que el alelo 7R es de reciente aparición, tal vez este efecto de selección influye sobre su frecuencia poblacional ya que aumenta a nivel mundial. También, este hallazgo puede estar correlacionado con los patrones de migración en el pasado. Chen *et al.* (1999) propone que las variaciones de este gen pueden estar correlacionadas con dichos patrones en épocas de la prehistoria, en donde las poblaciones migratorias (macro-migración) tienen altas frecuencias de los alelos de 7R o largos. Sujetos con dichos alelos se asocian a comportamientos como *novelty seeking*, hiperactividad y exploración, por lo que tuvo un valor adaptativo en las sociedades migratorias (Chen *et al.*, 1999). La distribución geográfica del alelo 7R con una alta proporción en América, el cual se puede relacionar con la característica de la personalidad del *novelty seeking*, o al déficit de atención e hiperactividad (ADHD), tal vez benefició a ciertos estilos de vida de la sociedad haciendo sujetos más activos y eficientes por tener un carácter de búsqueda de la novedad (Willians y Taylor, 2005).

En relación con el análisis de la Teoría de la Neutralidad en las variaciones de las secuencias de 7R, se separaron las subpoblaciones de sujetos diagnosticados con esquizofrenia y controles. En esta prueba, la D de Tajima pierde fuerza estadística pero se sigue observando una tendencia hacia un efecto de selección positiva que es más clara entre el subgrupo de controles (-1.48, $p= 0.058$) y entre los esquizofrénicos (-0.93, $p = 0.21$) hay mayor número de alelos de baja frecuencia. Aunque sería más evidente al aumentar el tamaño de la muestra.

DESEQUILIBRIO DE LIGAMIENTO Y CONSERVACIÓN

De los resultados obtenidos del análisis de la prueba de Tajima en donde el alelo 7R tiene un efecto de selección direccional, se puede deducir una dinámica poblacional en donde se espera que haya un desequilibrio de ligamiento (DL). De los tres tipos de secuencias analizadas se encontró fuerte DL entre los sitios parsimónicos de los alelos

7R, lo que confirma que hay sitios bajo presión selectiva. Así, nosotros encontramos en nuestra muestra de estudio que al menos 3 sitios variables en las secuencias de 7R, estaban en fuerte DL entre ellos, con una $D' = 1.0$ ($p < 0.0001$).

Además, este alelo mantiene valores de baja entropía sugiriendo un alto grado de conservación. Esta estrategia de usar la prueba de desequilibrio de ligamiento, así como el análisis de entropía logró identificar sitios útiles o marcadores que sugieren puede usarse en el mapeo de fenotipos relacionados con el alelo de 7R del gen DRD4.

En especial, recientemente el alelo de 7R ha sido asociado a problemas de comportamientos en niños como el apego desorganizado hacia su madre, que es el sistema innato manifestado en momentos de amenaza ante la ausencia de la figura de apego primaria (materna), que evolutivamente es importante para la sobrevivencia de los recién nacidos (Lakatos *et al.*, 2002), o al trastorno TADH, en el que se incrementan haplotipos raros (Grady *et al.*, 2005).

La dinámica de población se puede observar con eventos claros de mutaciones, que están sujetos a eventos de mestizaje, y a la entrada de otras poblaciones por la colonización, los cuales pudieron aportar mayor variación.

La estrategia de buscar con marcadores moleculares, sitios o regiones conservadas que posiblemente tienen un origen desde un ancestro común, permite valorar las implicaciones de riesgos epidemiológicos y de asociación a enfermedades (Rosenberg *et al.*, 2002).

Un estudio de continuidad iría destinado a aumentar la muestra para identificar más puntos de mutación y verificar regiones de desequilibrio, las cuales refieren a la dinámica de la región genómica de estudio antes de inferir conclusiones poblacionales.

Posteriores estudios podrían ir dirigidos a la identificación de otras mutaciones en este gen lo que incrementaría y ratificaría la trascendencia e importancia de los resultados y hallazgos encontrados en este trabajo, al examinar más genes con efecto moderado relacionado a los trastornos psiquiátricos asociados con el gen DRD4.

RECONSTRUCCIÓN FILOGENÉTICA

Un cladograma nos ilustra la probabilidad de que dos organismos o secuencias están cercanamente relacionados una de la otra, con respecto a una tercera. Así mismo puede ser usado para la formulación de nuevas hipótesis que nos arrojan una nueva “pista” en la interpretación de los datos moleculares empleados. La selección de los taxones, caracteres (secuencias), y relaciones genealógicas en el presente estudio,

generaron la obtención de resultados los cuales nos permitieron vislumbrar asociaciones entre grupos monofiléticos con un estado apomórfico compartido por dos taxones para constituir la sinapomorfía, rasgo distintivo de la monofilia y evento de especiación. (Cuando comparamos dos caracteres ya sea cualitativo o cuantitativo que forman una serie, llamamos plesiomórfico al que representa el estado ancestral y apomórfico al que deriva de él).

Con respecto al origen y paralelismo de las secuencias objeto de estudio y la secuencia de 7R de *Samieris boliviensis*, existe la posibilidad de que el carácter en un estado plesiomórfico (se infiere que se hallaba en el antecesor del grupo de estudio), se transforme (estado apomórfico) y luego esté presente en dos o más taxones (simplesiomorfía), para constituir la representación gráfica de los patrones de relación y longitud de ramas, así como las relaciones evolutivas entre las 22 taxas descritas en la figura 1 de los resultados.

En relación a otros estudios de cladística o sistemática filogenética molecular, con el objeto de reconstruir la historia evolutiva de los genes o especies, los resultados son controvertidos. En un análisis de 15 genes nucleares los Tarsiodeos son colocados como una rama separada entre los primates con alta significancia estadística (Eizirik *et al.*, 2004). Otro análisis sugiere que los Tarsios pueden representar un grupo monofilético con los Platiirinos (ceboidea), y catarrinos (Cercophitecoidea) (hipótesis Haplorrhine) (Chiu *et al.*, 1996).

Para el caso particular del presente estudio, el punto central a discutir es la presencia de una secuencia consenso funcional en los monos del nuevo mundo (platiirinos) con una divergencia de 26 millones de años respecto al humano, y la variación polimórfica en un segmento de ADN (VNTRS), y SNP con un papel fisiológico del receptor ligando.

MODELAJE MOLECULAR

El modelaje por homología y el análisis de mutación dentro del dominio de variación codificado por el exón 3 del receptor D4 muestra que la variación podría alterar la hidrofobicidad y/o el tamaño del aminoácido sustituido que probablemente tenga un efecto estructural creando un plegamiento diferente y tal vez proteínas inestables, como en el caso de las proteínas con dos repeticiones del VNTR del exón 3.

El análisis de secuencia de aminoácidos sobre la conservación de la unidad del VNTR del exón 3 mostró que la región estaba altamente conservada, a pesar de la

diversidad nucleotídica en la unidad del VNTR-exón 3, lo cual no produce ningún efecto en la secuencia de aminoácidos de la proteína, así como las propiedades fisicoquímicas de la secuencia original se mantienen conservadas. Sin embargo, el número de repeticiones en la secuencia de aminoácidos, como variaciones de longitud del asa citoplásmica larga (como alelos de 7R), puede afectar la acción farmacológica del receptor y en donde se pueden contrastar las diferentes propiedades bioquímicas de la proteína (Ashgari *et al.*, 1995). Por ejemplo, alelos largos de 7R y 10 R pueden tener la función de inhibir el acople de la adenil ciclasa 2 a 3 veces más reducido que las variantes de 2 y 4R (Jovanovic *et al.*, 1999) y en la sensibilidad de unión del antipsicótico clozapina (Ashgari *et al.*, 1995).

También, la variante de 7R parece disminuir la expresión del gen reportero luciferasa comparado a las formas 2R y 4R, lo que sugiere una expresión diferencial debido al número de repeticiones, lo cual podría explicar los diferentes comportamientos y trastornos (Schoots y Van Tol, 2003).

VI. CONCLUSIONES

Las variaciones de tres regiones del gen DRD4 están sometidas a procesos genómicos de creación, mantenimiento y eliminación de variabilidad, lo cual fue versátil y dinámico en las poblaciones de estudio. Los polimorfismos encontrados contribuyeron de forma modesta al conocimiento del contexto genómico, con respecto a la secuencia del gen DRD4 candidato a enfermedades psiquiátricas como la psicosis. La diversidad nucleotídica en la secuencia de los 3 alelos estudiados (2R, 4R y 7R), respondía a mutaciones puntuales y agrupó a nuevos haplotipos que sugieren pueden usarse como referencia en estudios de asociación de trastornos psiquiátricos asociados al gen DRD4.

La variación que agrupó estos eventos de mutación fue neutral en los alelos 2R y 4R y en las secuencias de los alelos de 7R presentan una selección positiva lo que sugiere que estos alelos pueden tener un papel en nuestra población de estudio. Los sitios nucleotídicos como el 301, 304, 309 y 313 de las secuencias de los alelos 7R muestran presión selectiva y están en desequilibrio de ligamiento estadísticamente significativo, las cuales son probablemente regiones o marcas de sitios susceptibles en el mapeo de trastornos psiquiátricos asociados con el gen DRD4 como la esquizofrenia o trastornos psicóticos.

El alineamiento y del análisis de entropía entre *Hylobates* y *Homo sapiens* mostró regiones conservadas a través de los taxones, fenómeno que permitió encontrar sitios que podrían ser auxiliares en la identificación de variaciones para tratar de entender el origen del fenotipo relacionado a la entidad clínica. Nuestros datos sugieren que los sitios con fuerte desequilibrio de ligamiento se ubican en dichas regiones conservadas, lo que confirma la identificación de marcadores SNP en los alelos de 7R.

Los estudios de asociación bajo la distribución de las frecuencias de genotipos y alelos, sugieren que el VNTR de 48pb y el SNP -521 pueden usarse como marcadores moleculares relacionados con los trastornos psicóticos. Utilizar otros polimorfismos para formar estructuras de haplotipos resultaría una herramienta útil para identificar enfermedades, aunque no se encontró una asociación con la distribución de frecuencia del VNTR de 12pb del exón 1, en donde se aceptó la hipótesis nula de que no hay diferencia entre las poblaciones de psicóticos y controles.

El cladograma de las secuencias de 7R de humanos muestra que los polimorfismos de sujetos esquizofrénicos y sujetos controles 19 y 39 (con variaciones

encontrados en la muestra de estudio) son de aparición reciente, en cambio las formas comunes (1-2-6-5-2-5-4) y control 70 (con formas reportadas en este estudio) tienen un origen más ancestral que se une al nodo interno de *Saimiri boliviensis*. El análisis cladista mediante el método Bayesiano logró detectar de forma eficiente sitios o haplotipos susceptibles a la enfermedad de esquizofrenia en la muestra estudiada. Resultado que no se observó oportunamente en el estudio de asociación que llevamos a cabo, lo que nos lleva a concluir que es necesario hacer análisis más detallados cuando se trata de pacientes diagnosticados con psicosis.

VII. REFERENCIAS

- Aguirre Samudio AJ, Nicolini H. El gen receptor a dopamina d4 y su asociación con los trastornos mentales. *Rev Inv Clin* 2005; 57:65-75.
- Apiquián R, Fresán A, Ulloa RE. Diagnóstico y tratamiento de los pacientes con primer episodio psicótico. *Salud Mental*. 1999, Número especial:79-87.
- Asghari V, Sanyal S, Buchwaldt S, Peterson A, Jovanovic V, Van Tol HHM. Modulation of intracellular cyclic AMP levels by different human dopamine D4 receptor variants. *J Neurochem* 1995; 65:1157-1165.
- Aitta-aho T. Neonatal viral brain infection and the development of prepulse inhibition: A neurodevelopmental model schizophrenia. *Pro gradu*. Department of Pharmacy, University of Helsinki, Sweden. 2003.
- Baeza Pertegaz I, Payá González B, Crespo Pacorro B. La hipótesis gabaérgica en la fisiopatología de la esquizofrenia. *Psiquiatría Biológica* 2000; 7:20-28.
- Bamshad M, Mummidi S, Gonzalez E, Ahuja SS, Dunn DM, et al. A strong signature of balancing selection in the 5' cis-regulatory region ccr5. *Proc Natl Acad Sci USA*, 2002; 99: 10539-10544.
- Bamshad M, Wooding SP. Signature of natural selection in the human genome. *Nat Rev Genet* 2003; 4: 99-111.
- Barnes N M, Sharp T. A review of central 5-HT receptors and their function; *Neuropharmacology* 1999; 38 1083–1152
- Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM Worldwide human relationships inferred from genome-wide patterns of variation. *Science*. 2008 Feb 22;319:1100-4.
- Berger MA, Defagot M, Villar MJ, Antonelli M. D4 dopamine and metabotropic glutamate receptors in cerebral cortex and striatum in rat brain. *Neurochem Res*, 2001; 26:345-352.
- Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, et al: Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* 2004; 74: 1111-1120.
- Bookman EB, Taylor RE, Adams-Campbell L, Kittles RA. DRD4 promoter SNPs and gender effects on extraversion in African Americans. *Mol Psychiatry* 2002; 7:786-789.

- Bouvier C, Bunzow J, Guan H-C, Unteutsch A, Civelli O, Grandy D, Van Tol HH. Functional characterization of the human dopamine D4.2 receptor using vaccinia virus as an expression system. *Eur J Pharmacol*, 1995; 290:11-17.
- Caraveo AJ: La prevalencia de los trastornos psiquiátricos en la población mexicana. Estado actual y perspectivas. *Salud Mental*, 1996; (Supl. Abril): 8-13.
- Chakravarti A. Population genetics – making sense out of sequence. *Nature Genetics* 1999; 21: 56-60.
- Catalano M, Nobile M, Novelli E, Nöthen M, Smeraldi E. Distribution of a novel mutation in the first exon of the human dopamine D4 receptor gene in psychotic patients. *Biol Psychiatry*, 1993; 34:459-464.
- Chang FM, Kidd JR, Livak K, Pakstis AJ, Kidd KK, The world-wide distribution of allele frequencies at the human dopamine D4 receptor locus. *Hum Gent* 1996; 98: 91- 101.
- Choy Wong AH, Buckle CE, Van Tol HHM. Polymorphisms in dopamine receptors: what do they tell us? *Eur J Pharmacol* 2000; 410 183-203.
- Chen CH, Chiu YL, Shaw CK, Tsai MT, Hwang AL, Hsiao KJ. Detection of Borna disease virus RNA from peripheral blood cells in schizophrenic patients and mental health workers. *Mol Psychiatry*, 1999;4:566-640.
- Chiu CH, Schneider H, Schneider MP, Sampaio I, Meireles C, Slightom JL, Gumucio DL, Goodman M. Reduction of two functional gamma-globin genes to one: an evolutionary trend in New World monkeys (infraorder Platyrrhini). *Proc Natl Acad Sci USA* 1996; 93:6510-6515.
- Cichon S, Nöthen M, Catalano M, Di Bella D, Maier W, Lichtermann D, Mingos J, Albus M, Borrmann M, Franzeck E, Stöber G, Weigelt B, Körner J, Rietschel M, Propping P. Identification of two novel polymorphisms and a rare deletion variant in the human dopamine D4 receptor gene. *Psychiatric Genet*. 1995; 5:97-103.
- Civelli O, Bunzow JR, Grandy DK, Zhou QY, Van Tol HH Molecular biology of the dopamine receptors. *Eur J Pharmacol*. 1991 14;207:277-86.
- Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res* 1998; 8 1229-31.
- Denoeud F, Vergnaud G, Benson G. Predicting human minisatellite polymorphism. *Genome Res*. 2003; 13:856-67.2003

- Dubrova, Y.E., Plumb, M., Brown, J., Fennelly, J., Bois, P., Goodhead, D., Jeffreys, A.J.. Stage specificity, dose response, and doubling dose for mouse minisatellite germ-line mutation induced by acute radiation. *Proc. Natl. Acad. Sci.* 1998; 95: 6251-6255.
- Dubrova, Y.E., Nesterov, V.N., Krouchinsky, N.G., Ostapenko, V.A., Vergnaud, G., Giraudeau, F., Buard, J., and Jeffreys, A.J.. Further evidence for elevated human minisatellite mutation rate in Belarus eight years after the Chernobyl accident. *Mut. Res.* 1997; 381: 267-278.
- Ebstein RP, Novick O, Umansky R, Priel B, Osher Y, Blaine D. Dopamine D4 receptor (D4DR) exon III polymorphism associated with the human personality trait of novelty seeking. *Nature Genet.* 1996; 12:78-80.
- Eizirik, E, Murphy WJ, Springer MS, O'Brien SJ. Molecular Phylogeny and Dating of Early Primates Divergentes. In *Anthropoid Origins* (eds. Kay, R and Ross, C) 2004 New Visions.
- Excoffier, L., Smouse, P.E., and Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.
- Fallin D, Schork N. Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphase diploid genotype data. *Am J Hum Genet* 2000, 67:947-959.
- Fatemi SH, Emamian ES, Sidwell RW, Kist DA, Stary JM, Earle JA, Thuras P. Human influenza viral infection in utero alters glial fibrillary acidic protein immunoreactivity in the developing brains of neonatal mice. *Mol Psychiatry*, 2002;7:633-640.
- Karlsson H, Bachmann S, Schroder J, McArthur J, Fuller Torrey E, Yolken R. Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia *PNAS*; 2001, 98:4634–4639.
- Fontdevila A. El mantenimiento de la variabilidad genética de las poblaciones. *Investigación y Ciencia* 1978, 20: 94-103.
- Fu Y-X, Li W-H: Statistical tests of neutrality of mutations. *Genetics* 1993; 133: 693-709.
- Gelernter J, Kennedy J, Van Tol HHM, Civelli O, Kidd KK. The D4 dopamine receptor (DRD4) maps to distal 11p close to HRAS. *Genomics* 1992; 13:208-210.
- Gray IC, Campbell DA, Spurr NK. Single nucleotide polymorphisms as tools in human genetics. *Hum Mol Genet* 9 (2000) 2403-8. Jorde LB. Linkage disequilibrium and the search for complex disease genes. *Genome Res.* 2000;10:1435–1444. doi: 10.1101/gr.144500.

- Goldstein DB, Weale ME. Population genomics: linkage disequilibrium holds the key. *Current Biology* 2001; 11: R576–R579.
- Gordon D, Finch S, Nothnagel M, Ott J. Power and sample size calculations for case-control genetic association tests when errors are present: Application to single nucleotide polymorphisms. *Hum Hered* 2002; .54:22-33.
- Gordon D, Levenstien MA, Finch S.J, Ott J. Errors and linkage disequilibrium interact multiplicatively when computing sample sizes for genetic case-control association studies. *Pacific Symposium on Biocomputing* 2003; 5: 490-501.
- Gottesman II, McGuffin P, Farmer AE. Clinical genetics as clues to the real genetics of schizophrenia. *Schizophrenia Bull* 1987; 13:23-47.
- Hamosh A, Scott Alan, Amberger J, Bocchini C, McKusick V. Online Mendelian inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 2005; 33: D514-D517.
- Hellmann I, Prüfer K, Ji H, Zody MC, Pääbo S, Ptak SE. Why do human diversity levels vary at a megabase scale? *Genome Res* 2005, 15:1222-31
- Hedrick P W. Genetic of population. 2^a Edidtion. Jones and Bartlett Publishers, Sudbury, Massachusetts. 1999. U.S.A., cap 9.
- Hughes AL, Yeager M. Natural selection at major histocompatibility complex loci of vertebrates. *Annu Rev Genet.* 1998; 32: 415-35.
- Huelsenbeck J P, Ronquist F. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 200117:754-755.
- Inoue-Murayama M, Takenaka O, Murayama Y. Origin and divergence of tandem repeats of primate D4 dopamine receptor genes. *Primates* 1998; 39: 217–224.
- International SNP Map Working Group: A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001, 409: 928-933.
- Ito H, Nara H, Inoue-Murayama M, Shimada M, Koshimura A, Ueda Y, Kitagawa H, Takeuchi, Mori Y et al. Allele frequency distribution of the canine dopamine receptor D4 gene exon III and I in 23 Breeds. *J Vet Med Sci* 2004; 7: 815-820.
- Jensen-Seaman MI, Deinart AS, Kidd KK. Modern African ape populations as genetic and demografpic models of the last common ancestor of humans, chimpanzees, and gorillas. *J Heredity* 2001, 92: 475-480.
- Jönsson E, Ivo R, Forslund K, Mattila-Evenden M, Rylander G, Cichon S, Propping P, Nöthen M, Asberg M, Sedvall G. No association between a promoter dopamine D4 receptor gene variant and schizophrenia. *Am J Med Genet* 2001;105:525-528.

- Jorde LB. Linkage disequilibrium and the search for complex disease genes. *Genome Res*, 2000; 10:1435-1444.
- Kalani MY, Vaidehi N, Hall S, Trabanino R, Freddolino p, Kalani M, Floriano W, Tak Kam V, Goddard III W. The predicted 3D structure of the human D2 dopamine receptor and the binding site and binding affinities for agonists and antagonists. *PNAS* 2004; 101: 3815-3820.
- Kamakura S, Iwaki A, Matsumoto M, Fukumaki Y. Cloning and characterization of the 5'-flanking region of the human dopamine D4 receptor gene. *Biochem Biophys Res Commun* 1997; 235:321-326.
- Kaplan NL, Hudson RR, Langley CH. The "Hitchhiking effect" revisited. *Genetics* 1989; 123: 887-899.
- Kato M, Sekine A, Ohnishi Y, Johnson TA, Tanaka T, Nakamura Y, Tsunoda T: Linkage disequilibrium of evolutionarily conserved regions in the human genome. *BMC Genomics* 2006, 7:326 doi:10.1186/1471-2164-7-326.
- Kelley J. Genética compleja: su influencia en la biología molecular y la enfermedad. *Rev Med Clin Condes* 2007; 18:306-312.
- Kennedy, G.C., German, M.S., Rutter, WJ. The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nat. Genet* 1995. 9: 293–298.
- Khoury MJ, Beaty TH. Applications of the Cases-Control method in genetic epidemiology. *Epidemiologic Reviews*. 1994; 16:134-150.
- Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences *J Mol Evol* 1980; 16: 111-120.
- Kruglyak L, Nickerson DA. Variation is the spice of life. *Nat Genet*. 2001; 27: 234-236.
- Laruelle, M., Abi-Dargham, A., Gil, R., Kegeles, L., Innis, R. Increased dopamine transmission in schizophrenia: Relationship to illness phases. *Biological Psychiatry* 1999; 46: 56-72.
- Lewontin RC. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*. 1964; 49:49–67.
- Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science*. 2008; 319:1100-4.
- Lichter JB, Barr CL, Kennedy J, Van Tol H, Kidd K, Livak K. A hypervariable segment in the human dopamine receptor D4 (DRD4) gene. *Hum Mol Genet* 1993; 2:767-773.

- Liebermann JA, Sheitman BB, Kinon BJ. Neurochemical sensitisation in the pathophysiology of schizophrenia: Deficits and dysfunction in neuronal regulation and plasticity. *Neuropsychopharm* 1997;17:205-229.
- Livak KJ, Rogers J, Lichter JB. Variability of dopamine D₄ receptor (*DRD4*) gene sequence within and among nonhuman primate species. *Proc Natl Acad Sci USA* 1995 92: 427–431.
- Livingston RJ, von Niederhausern A, Jegga AG, Crawford DC, Carlson CS, Rieder MJ, Gowrisankar S, et al. Pattern of sequence variation across 213 environmental response genes. *Genome Res* 2004; 14:1821-1831.
- López Mato A, Boulosa O, Marquez C, Vieitez A, Illa G. Tratamiento del primer episodio en esquizofrenia. *ALCMEON Rev Argentina de Clínica Neuropsiquiátrica* 2000, 4. www.alcmeon.com.ar .
- Luo YQ, Kukkonen GC, Wung XT, Neve KA, Roth GS. D2 Dopamine receptor stimulate mitogenic through pertuss toxin-sensitive G proteins and...
- Meador-Woodruff JH, Grandy DK, Van Tol HH, Damask SP, Little KY et al. Dopamine receptor gene expression in the human medial temporal lobe. *Neuropsychopharmacology* 1994, 10: 239-240.
- Mettler L, Gregg T G. Genética de las poblaciones y evolución. Prentice – may, E.U.A. Traducción UTHA S.A de C.V., 1982.
- Mitsuyasu H, Hirata N, Sakai Y, Shibata H, Takeda Y, Ninomiya H, Kawasaki H, Tashiro N, Fukumaki Y.. Association analysis of polymorphisms in the upstream region of the human dopamine D4 receptor gene (*DRD4*) with schizophrenia and personality traits. *J Hum Genet* 2001; 46:26-31.
- Mooney SD, Klein T E. The functional importance of disease-associated mutation. *BMC Bioinformatics* 2002; 3:24.
- Mogensen L, Christian Kinze C, Werge T, Berg Rasmussen H. Identification and Characterization of a Tandem Repeat in Exon III of the Dopamine Receptor D4 (*DRD4*) Gene in Cetaceans. *J Heredity* 2006; 3: 279-284.
- Mulcrone J, Kerwin RW. The regional pattern of D4 gene expression in human brain. *Neuroscience Letters*, 1997; 234: 147-150.
- Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in human. *Genetics* 2000; 156: 297-304.
- Nachman MW. Single nucleotide polymorphism and recombinationrate in human. *Trend Genet* 2001; 17: 481-485.

- Nei M. *Molecular Evolutionary Genetics*. Columbia University Press, New York, 1987, cap 3.
- Nordborg M, Tavares S, Linkage disequilibrium: what history has to tell us. *TRENDS Genet*, 2002; 18:83-90.
- Nothen MM, Cichon S, Hemmer S, Hebebrand J, Remschmidt H, Lehnkuhl G, Poustka F, Schmidt M, Catalano M. Human dopamine D4 receptor gene: frequent occurrence of a null allele and observation of homozygosity. *Hum Mol Genet*, 1994; 3:2207-2212.
- Oak J, Lavaine N, Van Tol HH. 2001. Dopamine D4 and D2L receptor stimulation of the mitogen-activated protein kinase pathway is dependent on trans-activation of the platelet-derived growth factor receptor. *Molecular Pharmacology* 60: 92-103.
- O'Dwod BF, Hnatowich M, Caron MG, Lefkowitz RJ, Bouvier M. Palmitoylation of the human β 2-adrenergic receptor: Mutation of Cys241 in the carboxy tail leads to an uncoupled nonpalmitoylated form of the receptor. *J Biol Chem* 1989; 264:7564-7569.
- Okuyama Y, Ishiguro H, Toru M, Arinami T: A genetic polymorphism in the promoter region of DRD4 associated with expression and schizophrenia. *Biochem Biophys Res Commun*, 1999; 258: 292-295.
- Okuyama Y, Ishiguro H, Nankai M, Shibuya H, Watanabe A, Arinami T. Identification of a polymorphisms in the promoter region of DRD4 associated with the human novelty seeking personality trait. *Mol Psychiatry*, 2000; 5:64-69.
- Oliva R: *Genoma humano*. Barcelona , Masson S.A., 1996; p. 53-55.
- O'malley KL, Harmon S, Tang L, Tood RD. the rat dopamine D4 receptor: sequence, gene structure and demonstration of expression in the cardiovascular system. *New Biologist* 1992, 4: 137-146.
- Ortega Soto H, Valencia Collazos M. Esquizofrenia estado actual y perspectivas. Publicaciones del Instituto Nacional de Psiquiatría Ramón de la Fuente, México. 2001; p: 636.
- Parkhomchuk D. Di-nucleotide Entropy as a Measure of Genomic Sequence Functionality *Quantitative Biology > Genomics* (q-bio.GN)2006, arXiv:q-bio/0611059 [pdf].
- Petronis A, Gottesman I, Kan P, Kennedy J, Basile V, Paterson A, Pependikyte V. Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? *Schizophrenia Bull*, 2003; 29:169-178.
- Petronis A. The genes for major psychosis: aberrant sequence or regulation?. *Neuropsychopharmacol*, 2000; 23:1-12.

- Prasad S, Semwal P, Deshpande S, Bhatia T, Nimgaonkar VL, Thelma B.K. Molecular genetics of schizophrenia: past, present and future. *J Biosciences* 2002; 24S1: 35-52.
- Przeworski M. The signature of positive selection at randomly chosen loci. *Genetics* 2002; 160:1179-89.
- Pritchard J.K, Przeworski M: Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* 2001; 69:1-14.
- Pungliya MS, Salisbury BA, Nandabalan K, Stephens JC. Genetic variability and evolution of two pharmacologically important classes of genes. *Pharmacogenomics* 2004, 5: 115-127.
- Savage SA, Stewart Bj, Eckert A, KilleyM, Liao JS, Chanock SJ. Genetic variation, nucleotide diversity, and linkage disequilibrium in seven telomere stability genes suggest that these genes may be under constraint. *Hum Mutat* 2005, 26: 343-350.
- Savage S, Gerstenblith M, Goldstein A, Mirabello L, Concetta Fargnoli M, et al. Nucleotide diversity and population differentiation of the Melanocortin I receptor gene, MC1R. *BMC Genetics* 2008, 9:31 doi:10.1186/1471-2156-9-31
- Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko, JV, Patterson NJ, McDonald GJ, et al: Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002; 419: 832–837.
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; 409: 928-33.
- Schetz J, Benjamin P, Sibley D. Non conserved residues in the second transmembrane-spanning domain of the D4 dopamine receptor are molecular determinants of D4 – Selective pharmacology. *Mol pharmacol* 2000; 57:144-152.
- Schneider S, Roessli D, Excoffier L. 2000. Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- Schneider JA, Pungliya MS, Choi JY, Jiang R, Sun XJ, Salisbury BA, Stephens JC. DNA variability of human genes. *Mech Ageing Dev* 2003, 124:17-25.
- Seeman P, Guan H, Van Tol HH. Dopamine D4 receptors elevated in schizophrenia. *Nature* 1993; 365: 441-445.
- Seeman P. Dopamine receptor sequences therapeutic levels of neuroleptics occupy D2 receptors, clozapine occupies D4. *Neuropsychopharm* 1992; 7:261-284.

- Seeman M, Chang FM, Deinard A, Quiñones A, Kidd K. Evolution of exon 1 of the dopamine D4 receptor (DRD4) gene in primates. *J Exp Zool*, 2000; 288:32-38.
- Sham PC, Curtis D. Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann Hum Genet* 1995; 59:97-105.
- Stefanis NC, Bresnick JN, Kerwin RW, Schofield WN, McAllister G. Elevation of D-4 dopamine receptor mRNA in postmortem schizophrenic brain. *Mol Brain Res*. 1998; 53:112-119.
- Strader CD, Candelore M, Hill WS, Sigal I S, Dixon R A. Identification of two serine residues 79 and 113 of the beta-adrenergic receptor have different roles in receptor function. *J Biol Chem*, 1989; 263: 13572-13578.
- Strader CD, Candelore M, Hill WS, Sigal I S, Dixon R A. Identification of two serine residues 79 and 113 of the beta-adrenergic receptor have different roles in receptor function. *J Biol Chem*, 1989; 263: 13572-13578.
- Sunyaev S, Kondrashov FA, Bork P, Ramensky V. Impact of selection, mutation rate and genetic drift on human genetic variation. *Hum Mol Genet*, 2003, 24: 3325-3330.
- Ren R, Mayer BJ, Cicchetti P, Baltimore D. Identification of a tern-amino acid praline rich SH3 binding site. *Science* 1993; 259: 1157.
- Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*. 1996; 273:1516–1517.
- Risch N. searching for genetics determinants in the news millennium. *Nature* 2000, 405: 847-856.
- Reich D, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, Richter D, Lander ES, Altshuler D Human genome sequence variation and the influence of gene history, mutation and recombination. *Nature Genetics* 2002, 32: 135-142.
- Rinetti G, Camarena B, Cruz C, Apiquian R, Fresan A, Paez F, Nicolini H. Dopamine D4 receptor (DRD4) gene polymorphism in the first psychotic episode. *Arch Med Res* 2001; 32: 35-38.
- Ronai Z, Barta C, Guttman A, Lakatos K, Gervai J, Staub M, Sasvari-Szekely M. Genotyping the -521C/T functional polymorphism in the promoter region of dopamine D4 receptor (DRD4) gene. *Electrophoresis* 2001; 22:1102-1105.
- Rosenberg NA, Pritchard J, Weber JL, Cann H, Kidd K, Zhivotovsky L, Feldman M. Genetic structure of human populations. *Science* 2002; 298: 2381-2385.
- Tajima, F: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989; 123: 585-595.

- Tarazi F, Baldesarini R. Brain dopamine D4 receptors: basic and clinical status. *Int J Neuropsychopharmacol*, 1999; 2:41-58.
- Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, et al: Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* 2001; 293:455-462.
- Thomas D, Trumbower H, Kern A, Rhead B, Kuhn R, et al: Variation resources at UC Santa Cruz. *Nucleic Acids Res* 2006; 00: D1-D5.
- Thompson EE, Kuttab-Boulos H, Witonski D, Yang L, et al: CYP3A variation and the evolution of SALT-sensitivity variants. *Am J Hum Genet* 2004; 75: 1059-1069.
- Turri, M.G., Cuin, K.A., and Porter, A.C. Characterization of a novel minisatellite that provides multiple splice donor sites in an interferon-induced transcript. *Nucleic Acids Res.* 1995; 23: 1854-1861
- Van Craenenbroeck K, Clark SD, Cox M, Oak JN, Liu F, Van Tol HH. 2005. Folding efficiency is rate-limiting in dopamine D4 receptor biogenesis. *J Biol Chem* 280: 19350-1937.
- Van Tol HHM, Bunzow JR, Guan H, Sunahara R, Seeman P, Niznik HB, Civelli O.. Cloning of the gene for a human dopamine D4 receptor with high affinity for the antipsychotic clozapine. *Nature*, 1991; 350: 610-614.
- Van Tol HHM. The dopamine D4 receptor. *NIDA Res Monogr* 1996; 161: 20-38.
- Van Tol HHM, Wu CM, Guan H-CH, Ohara K, Bunzow JR, Civelli O, Kennedy J, Seeman P, Niznik HB, Jovanovic V. Multiple dopamine D4 receptor variants in the human population. *Nature* 1992; 358:149-152.
- Vergnaud G, Denoeud F.. Minisatellites: Mutability and genome architecture. *Genome Res.* 2000 10: 899-907.
- Voight B, Kudaravalli S, Wen X, Pritchard J: A map of recent positive selection in the human genome. *Plos Biology* 2006; 4: 0446-0456.
- Wang E, Ding Y-C, Flodman P, Kidd JR, Kidd K, Grady DL, Ryder O, Spence MA, Swanson JM, Moyzis RK. The genetic architecture of selection at the human dopamine receptor D4 (DRD4) gene locus. *Am J Hum Genet* 2004; 74: 931-944
- Winterer G, Weinberger DR. Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends Neurosci* 2004; 27: 683-690.
- Williams J, Taylor E. The evolution of hyperactivity, impulsivity and cognitive diversity. *J R Interface.* 2005 p 1-15 (doi:10.1098/rsif.2005.0102)

- Yan Z, Bielawski JP: Statistical methods for detecting molecular adaptation. *Trends Ecol, Evol* 2000; 15: 496-503.
- Yu N, Jensen-Seaman M, Chemnick L, Ryder O, Li W-H. Nucleotide diversity in gorillas. *Genetics* 2004, 166: 1375-1383.
- Zhao H, Pfeiffer R, Gail M. Haplotype analysis in population genetics and association studies *Pharmacogenomics*. 2003, 2: 171-178.
- Zhang K, Zhaohui Q, Liu J, Chen T, Waterman M, Sun F. Haplotype block partitioning and tag SNP selection using genotype data and their applications to association studies. *Genome Res* 2004, 908-916.
- Zondervan KT, Cardon L. The complex interplay among factors that influence allelic association. *Nat Review* 2004; 2: 89-100.

APÉNDICE

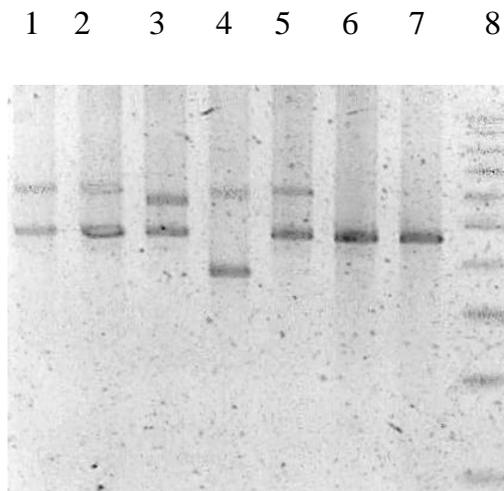


Figura 1. Gel de agarosa al 3%. Carril 1, 2 y 5 son heterocigotos 4/7, en el 3: 4/6, en el 4: 2/7. Los homocigotos 4/4, se muestran en los carriles 6 y 7. Y el marcador de peso molecular de 50pb está en el 8.

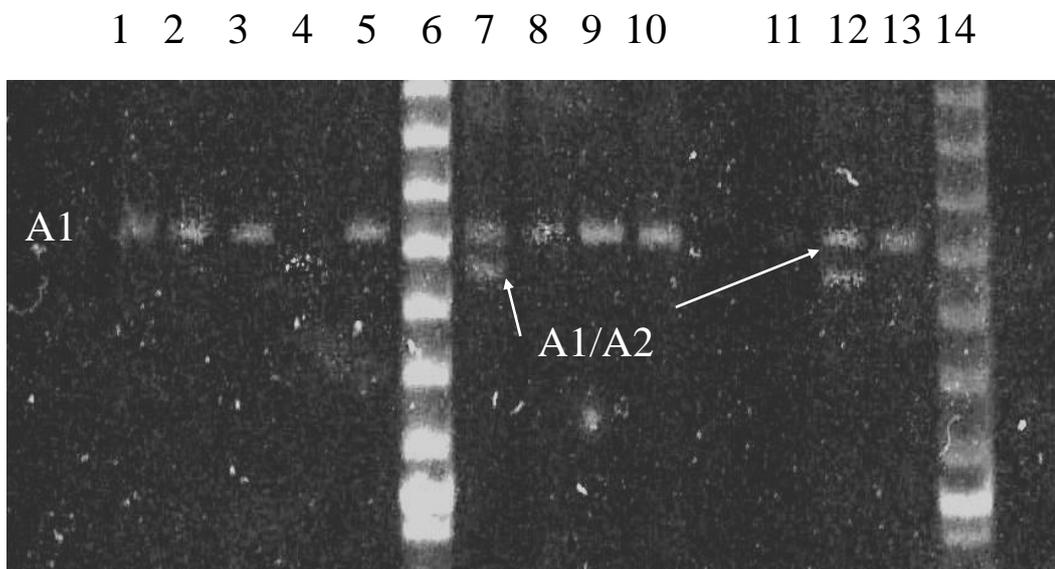


Figura 2. Gel de poliacrilamida al 6%. Carril 1-3,5 homocigotos A1, carril 7 y 12 heterocigotos A1/A2, 6 y 14 marcador de peso molecular de 20 pb.

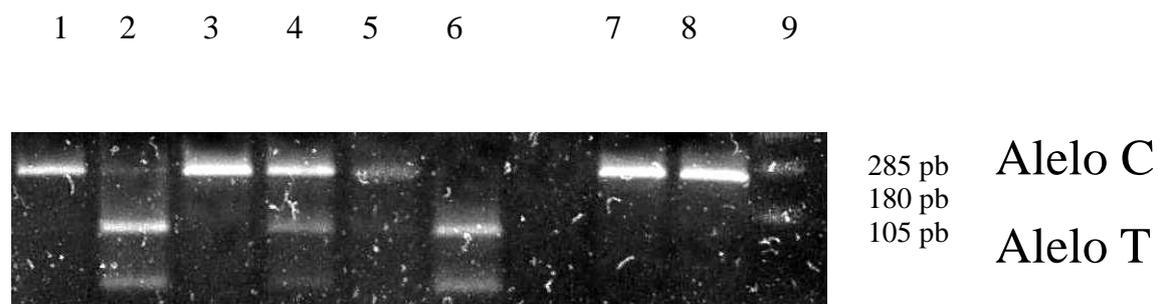


Figura 3. Gel de agarosa de 2.3%, carril 1-8 Muestras de pacientes, 9: marcador de peso molecular. Se observan los genotipos del polimorfismo SNP del sitio -521. Heterocigotos se muestran en el carril 4 y 5. Homocigotos TT están en el carril 2 y 6; y CC en los carriles 1, 3, 7, 8.

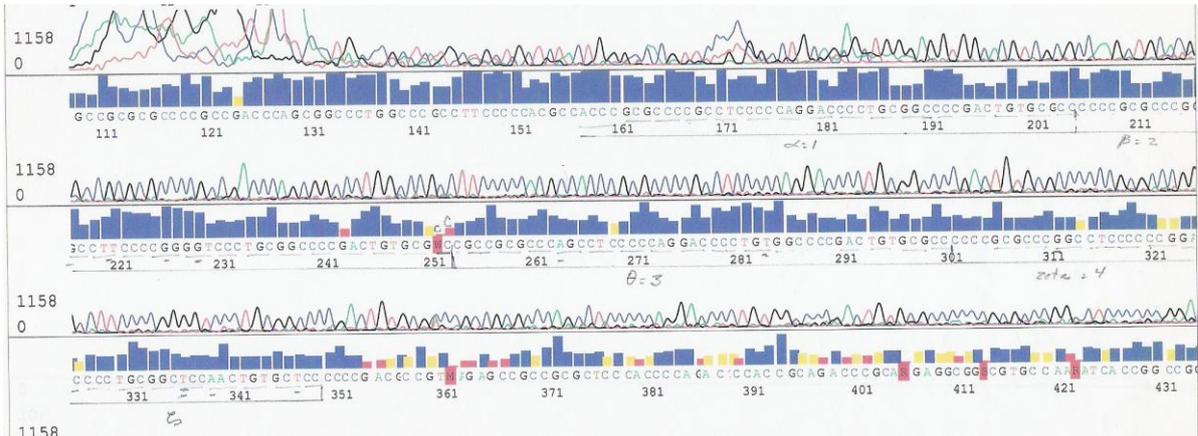


Figura 4. Secuencia de un cromosoma del alelo de 4R, se describe como el haplotipo 1-2-3-4. Las barras de colores enzima de cada nucleótido hacen referencia al análisis de calidad del algoritmo KB (*basecalling*). Las barras altas indican que los nucleótidos son estadísticamente aceptables.

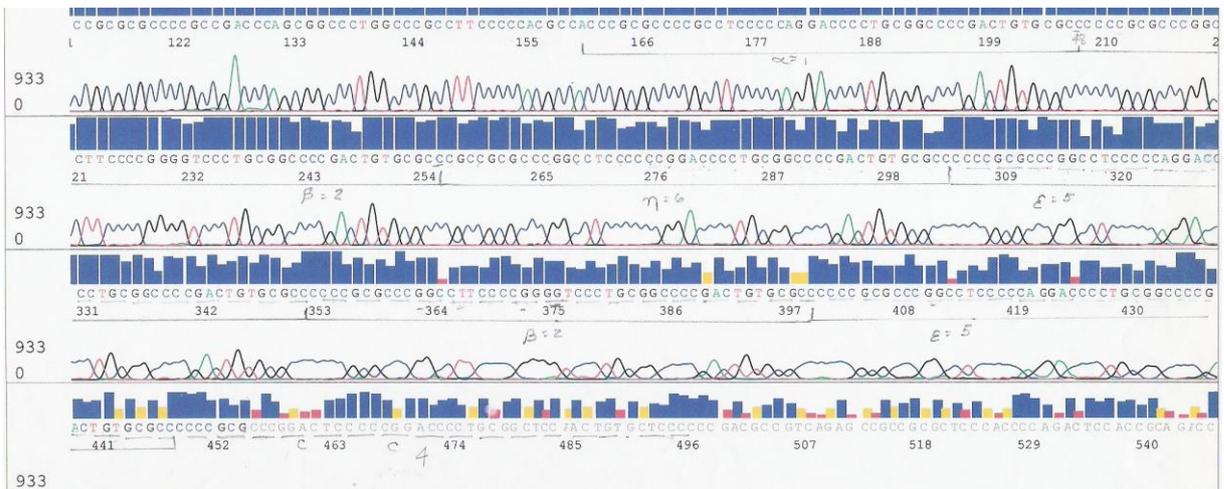


Figura 5. Secuencia de un cromosoma del alelo de 7R, se describe como el haplotipo 1-2-6-5-2-5-4. Las barras enzima de cada nucleótido hacen referencia al análisis de calidad del algoritmo KB (*basecalling*). Las barras altas indican que los nucleótidos son estadísticamente aceptables.

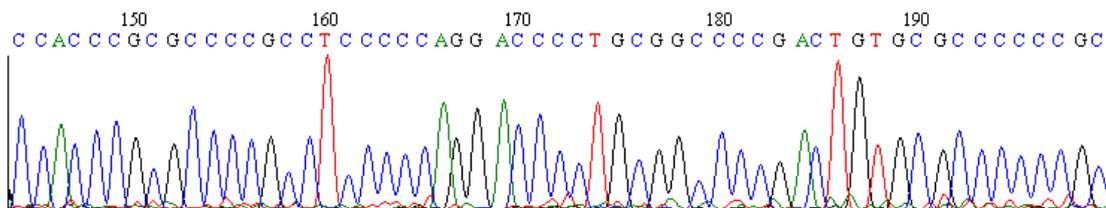


Figura 6. Secuencia que describe la primera unidad de repetida del VNTR (de 48pb) del exón 3 del gene DRD4, casi siempre empieza con A C C C, y es muy conservada entre los haplotipos.

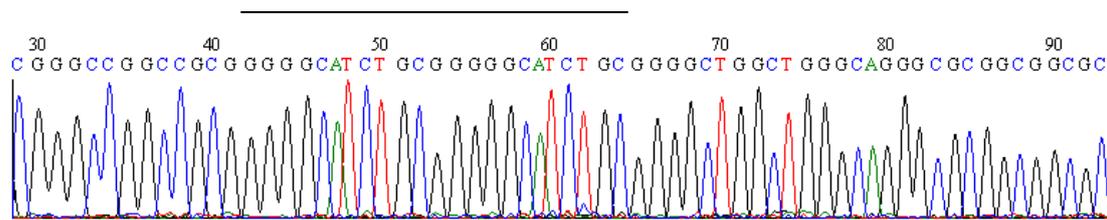


Figura 7. Electroferograma del VNTR de 12pb del exón 1 del gene DRD4. La secuencia se repite 2 veces y corresponde al alelo A1, el cual comienza en el nucleótido 42 y termina en el 65.

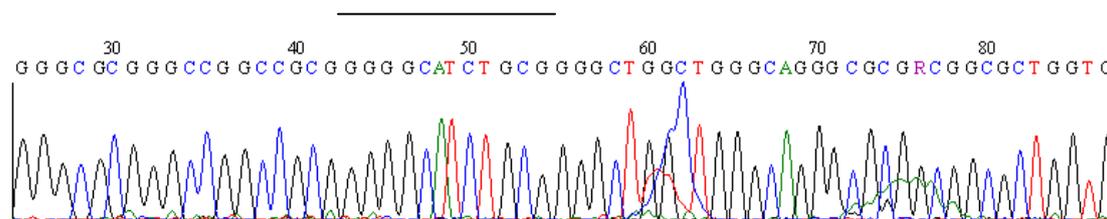


Figura 8. Electroferograma del VNTR de 12pb del exón 1 del gene DRD4. La secuencia solo esta en una unidad de repetidas y corresponde al alelo A2, el cual comienza en el nucleótido 42 y termina en el 54.

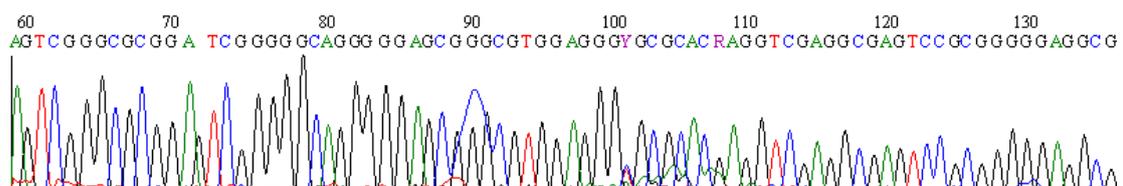


Figura 9. Electroferograma de la región 5' flanqueante del promotor del gene DRD4. La región fue clonada y secuenciada debido al alto contenido de GC, lo que resultaba complicado realizar una secuenciación directa con los primers reportados por Okuyama *et al.* (1999). El SNP -521 se ubica en el sitio 101, donde se observa un heterocigoto CT.

