Análisis de Doce Marcadores Polimórficos en el ADN de una Muestra Poblacional del Valle Central Costarricense

Eugenio Rojas, Jorge Lobo y Pedro León

Objetivo: Desarrollo de bases de datos de frecuencias alélicas en una muestra de la población costarricense del Valle Central.

Métodos: Muestras de sangre periférica de más de 40 individuos del Valle Central de Costa Rica fueron utilizadas para aislar ADN y analizar 10 marcadores polimórficos de repeticiones dinucleotídicas y 2 marcadores de tipo minisatélite por medio de la reacción en cadena de la polimerasa. Los alelos se identificaron usando como referencia ADN de las familias del Centro de Estudios de Polimorfismos Humanos (CEPH). La identificación de los alelos se hizo marcando uno de los dos iniciadores con 32P en la punta 5’, y separando los productos de la amplificación por electroforesis en geles de secuenciación y por autoradiografía.

Resultados y conclusión: El análisis de las frecuencias alélicas en esta muestra indica que la población se encuentra en equilibrio de Hardy-Weinberg y no se presenta evidencia de desequilibrio gamético entre diferentes marcadores. En este trabajo se compara la muestra costarricense con otras poblaciones estudiadas con los mismos marcadores genéticos, y se encuentra similitud entre ellas en cuanto a las frecuencias alélicas. Es notable que la muestra costarricense presenta, con 4 de 10 marcadores, los niveles más bajos de heterocigosidad, seguidos por la muestra de Cerdeña. En contraste, las muestras de origen africano presentan los más altos índices de heterocigosidad, con más alelos presentes.

Descriptores: micro y minisatélites del ADN, PCR de polimorfismos humanos, identificación humana

Introducción

Los estudios del genoma humano han revelado la existencia de secuencias repetitivas en tandem en muchos sitios a lo largo de todos los cromosomas, que varían en el número de veces que la unidad de repetición está presente. Cuando la unidad repetitiva del marcador genético es corta, de 2 a 5 nucleótidos, se les denomina microsatélites, o "short tandem repeats" (STRs) mientras que si la secuencia repetitiva es de 9 a 64 nucleótidos se les denomina minisatélites. También se ha acuñado el término 'variable number tandem repeat' (VNTR) para designar a los minisatélites, aunque esto se presta a confusión porque todos estos polimorfismos, finalmente, representan alguna forma de repetición en tandem. En el genoma nuclear se han descrito miles de microsatélites multifalácicos con un alto grado de heterocigosidad (>70%) en todas las poblaciones estudiadas hasta la fecha. Ha sido precisamente con base en marcadores dinucleotídicos que el consorcio francés GENETHON elaboró los primeros mapas completos del genoma humano, con una resolución menor de 5 cM. Esto permite, en principio, el

Abreviaturas: CEPH, Centro de Estudios de Polimorfismos Humanos; STR, repeticiones pequeñas en tandem; VNTR, repeticiones en tandem en número variable; cM, centímetros; PCR, reacción en cadena de la polimerasa; CA, citosina-adenina.

1. Perkin-Elmer, Santiago, Chile.
2. Escuela de Biología, Universidad de Costa Rica.

Correspondencia: Pedro León, CIBCM, Universidad de Costa Rica.
ma-peo de cualquier gen causante de enfermedades hereditarias simples o complejas.\textsuperscript{14,12}

Los microsatélites también han resultado de mucha utilidad para resolver casos de paternidad dudosos y de identidad.\textsuperscript{1,10-22} En contraste, los marcadores clásicos (de tipo inmunológico, grupos sanguíneos, proteínas séricas, HLA, etc.) son muy pocos y presentan una baja heterocigosidad. Por lo tanto, la tendencia en el ámbito mundial es la de utilizar repeticiones tri- y tetranucleotídicas, muchas de las cuales tienen alelos pequeños, fáciles de resolver aún en geles de agarosa. Además, se ha visto que estas son más estables durante el proceso de amplificación y tienen una tasa de mutación más baja durante la replicación que las repeticiones dinucleotídicas.\textsuperscript{24}

Las tecnologías moleculares basadas en la reacción en cadena de la polimerasa (PCR) han facilitado el acceso de muchos laboratorios a participar en el mapeo genético y la identificación forense, incluso en países en vías de desarrollo.\textsuperscript{23-29} La PCR permite analizar cantidades muy pequeñas de ADN, por lo que aún minúsculas muestras de material biológico aportan evidencias de identidad.\textsuperscript{27,29} Además, como ya se indicó, los alelos de los microsatélites son de tamaño pequeño, por lo que es posible analizar ADNs degradados, en contraste con los minisatélites con alelos más grandes. El uso de las técnicas de hibridación de Southern tiene la desventaja de que requiere gran cantidad de ADN no degradado para el análisis.\textsuperscript{25,31,29}

Este trabajo representa el primer estudio genético hecho con micro y minisatélites en una muestra de la población costarricense, usando ADN nuclear para fines forenses. Previosamente, Barrantes y colaboradores han estudiado muestras de amerindios costarricenses con marcadores clásicos\textsuperscript{30} y con microsatélites multialélicos en el cromosoma Y y en el gen de la miotonia quinasa.\textsuperscript{31-32} Los datos poblacionales que se ofrecen en este trabajo fueron utilizados para establecer las probabilidades de paternidad en 10 casos analizados como parte de una tesis de maestría.\textsuperscript{33} En este artículo se presentan los resultados del análisis de estos polimorfismos genéticos en la muestra de la Meseta Central y se comparan con las muestras de Cerdeña, Egipto y África al sur del Sahara estudiadas por Di Rienzo \textit{et al.}\textsuperscript{34} con estos mismos marcadores.

**Materiales y Métodos**

Muestra poblacional: La muestra proviene de 10 parejas no relacionadas entre sí, referidas para pruebas de paternidad por el Laboratorio del Organismo de Investigaciones Judiciales de la Corte Suprema de Justicia, al CIBCM en la Universidad de Costa Rica. Otras 26 personas no relacionadas con 4 apellidos comunes en el Valle Central (E. Fournier, comunicación personal) donaron voluntariamente, previo consentimiento informado, una muestra de 20 ml de sangre venosa en tubos con ácido cítico-dextrosa (Becton Dickinson Vacutainer Systems, Franklin Lakes NJ 07417-1885). Se utilizaron métodos previamente descritos para el aislamiento de núcleos, digestión con Proteína K y extracción con solventes orgánicos antes de precipitar el ADN con etanol.\textsuperscript{34,35} Las muestras de ADN se mantuvieron concentradas y en congelación a -20°C.

**Tipo genético con marcadores de ADN:** Los métodos de amplificación y análisis con estos marcadores genéticos han sido descritos.\textsuperscript{36}\textsuperscript{37} Los marcadores genéticos usados fueron aportados por el Dr. Nelson Freimer y analizados parcialmente en su laboratorio en la Universidad de California en San Francisco. La mayoría de estos pertenecen a la serie de James Weber en Marshfield, USA con el dinucleotido citosina-adenina (CA) repetido de 15 a 30 veces.\textsuperscript{8} Las condiciones para la amplificación de estos marcadores están detalladas en el trabajo de Di Rienzo \textit{et al.}\textsuperscript{34} y en Rojas.\textsuperscript{33} Para el análisis de los STRs se marcó uno de los iniciadores en la punta 5' con \textsuperscript{33}P por la reacción de la T4-kinasa, según el protocolo descrito por Freimer \textit{et al.}\textsuperscript{36} y los productos de amplificación se separaron en geles de secuenciación de poliacrilamida al 6%. Los dos minisatélites, APO B y D1S80, se analizaron en geles de agarosa al 1% teñidos con bromuro de etidio. Los genotipos fueron registrados por dos personas, independientemente, para mayor consistencia.

**Análisis estadístico:** Las frecuencias alélicas, el grado de heterocigosidad, y el análisis de Hardy-Weinberg se realizaron con el programa BIOSYS-1.\textsuperscript{37} También se analizaron las frecuencias genotípicas observadas y esperadas usando una prueba de Chi-cuadrado para muestras pequeñas y se determinó el índice de fijación, que cuantifica el exceso o deficiencia relativa de heterocigotos en la población (F=1 cuando solo se observan heterocigotos); el Chi-cuadrado evalúa estadísticamente la significancia de la observación.\textsuperscript{38}

**Resultados**

El Cuadro 1 muestra la lista de marcadores amplificados, su locus genético, el número de cromosomas analizados, el nivel de heterocigosidad detectado y otras características de esta muestra del Valle Central. También se presentan los valores del Chi-cuadrado y los valores del índice de fijación, que son muy bajos para casi todos los marcadores, lo cual es consistente con la hipótesis de que la población está próxima a un estado de equilibrio. En algunos casos este índice toma un pequeño valor negativo, lo cual sugiere que existe una deficiencia de homocigotos y un exceso de heterocigotos para ese marcador. Estas desviaciones son muy leves y no significativas, como tampoco lo son los valores de Chi-cuadrado del cálculo de Hardy-Weinberg. El Cuadro 2 presenta las frecuencias alélicas observadas en la muestra. El tamaño del alelo incluye las regiones únicas donde se hibridan los iniciadores y otras secuencias adyacentes, que deben de restarse para establecer el número de repeticiones, según explica la leyenda del Cuadro 2. El análisis de los datos también revela la ausencia de ligamiento entre los marcadores utilizados (detalles en 33).

El análisis de los resultados refleja una distribución alélica comparable con las muestras de Cerdeña, Egipto y África al sur
Cuadro 1

Nombre del locus genético, el marcador específico en paréntesis, la localización cromosómica, heterocigósidad (H), número de alelos detectados, rango de tamaño de los alelos, número de cromosomas en esta muestra (N), el valor de Chi-cuadrado corregido para el tamaño de la muestra (λ2*) y el índice de fijación (F) para cada marcador en esta muestra poblacional del Valle Central de Costa Rica.

<table>
<thead>
<tr>
<th>Locus genético</th>
<th>Cromosoma</th>
<th>H</th>
<th># alelos</th>
<th>Rango tamaño</th>
<th>N</th>
<th>λ2*</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>APO B</td>
<td>2p23-24</td>
<td>0.78</td>
<td>9</td>
<td>480-905</td>
<td>96</td>
<td>0.934</td>
<td>-0.040</td>
</tr>
<tr>
<td>DIS80(MCT118)</td>
<td>1p35-36</td>
<td>0.85</td>
<td>16</td>
<td>433-689</td>
<td>90</td>
<td>0.567</td>
<td>0.092</td>
</tr>
<tr>
<td>APOA2 (Mfd3)</td>
<td>1q21-23</td>
<td>0.78</td>
<td>8</td>
<td>131-147</td>
<td>90</td>
<td>0.030</td>
<td>-0.067</td>
</tr>
<tr>
<td>D18S35(MFD32)</td>
<td>18q21</td>
<td>0.64</td>
<td>8</td>
<td>100-124</td>
<td>86</td>
<td>0.028</td>
<td>-0.042</td>
</tr>
<tr>
<td>D22S156(MFD33)</td>
<td>22q11-12.12</td>
<td>0.64</td>
<td>6</td>
<td>98-108</td>
<td>94</td>
<td>0.043</td>
<td>0.117</td>
</tr>
<tr>
<td>D4S175(MFD38)</td>
<td>4q25-34</td>
<td>0.82</td>
<td>11</td>
<td>110-132</td>
<td>88</td>
<td>1.958</td>
<td>0.017</td>
</tr>
<tr>
<td>D4S174(MFD59)</td>
<td>4q11-13</td>
<td>0.85</td>
<td>10</td>
<td>177-197</td>
<td>74</td>
<td>0.836</td>
<td>0.093</td>
</tr>
<tr>
<td>D12S59(MFD75)</td>
<td>12q</td>
<td>0.86</td>
<td>13</td>
<td>162-192</td>
<td>90</td>
<td>0.014</td>
<td>-0.018</td>
</tr>
<tr>
<td>D21S156(MFD55)</td>
<td>21q22.3</td>
<td>0.82</td>
<td>11</td>
<td>77-105</td>
<td>90</td>
<td>0.057</td>
<td>-0.040</td>
</tr>
<tr>
<td>D8S184(MFD104)</td>
<td>8q13-22.1</td>
<td>0.75</td>
<td>12</td>
<td>162-201</td>
<td>88</td>
<td>0.441</td>
<td>0.016</td>
</tr>
<tr>
<td>D18S178(M139)</td>
<td>19q13.2</td>
<td>0.70</td>
<td>12</td>
<td>153-186</td>
<td>90</td>
<td>0.028</td>
<td>0.022</td>
</tr>
<tr>
<td>D4S109(MFD142)</td>
<td>4p21-q24</td>
<td>0.76</td>
<td>7</td>
<td>95-107</td>
<td>86</td>
<td>1.440</td>
<td>-0.090</td>
</tr>
</tbody>
</table>

Del Sahara,\textsuperscript{24} que son de tamaño similar. No se presentan diferencias extremas en la frecuencia de los alelos más comunes entre las poblaciones. Por ejemplo, para los marcadores D18S35, D19S178 y D22S156 podemos observar que el alelo más común es el mismo en todas las poblaciones, mientras que con el marcador D4S175 el alelo predominante es el mismo en tres de las cuatro poblaciones comparadas. Algunos de estos marcadores presentan perfiles multimodales, como se observa claramente con APOA2, D19S35 y D19S178 en las cuatro poblaciones. Aunque la muestra del Valle Central no se diferencia notablemente de las otras poblaciones, en particular con marcadores muy conservados como APOA2, D22S156 y D19S178, el marcador D4S175 presenta un alelo muy común en la población costarricense (alelo 112 con frecuencia de 0.30) que en las otras poblaciones está presente en menos del 0.10 de la población. Dado el tamaño pequeño de las muestras es imposible hacer averiguaciones muy certeras sin estudios adicionales.

Discusión

La comparación de marcadores en la muestra de la Meseta Central de Costa Rica con las poblaciones estudiadas por Di Rienzo et al.\textsuperscript{24} confirma la similitud genética de las poblaciones humanas. La muestra de la Meseta, sin embargo, presenta, con 4 de los 10 marcadores, la tasa de heterocigósidad más baja\textsuperscript{24} (Cuadro 3), seguida por la muestra de Cerdeña con 3 de 10 marcadores. Esto podría deberse a efectos fundadores de una población que se expande, con algunos alelos sobrerepresentados y otros ausentes,\textsuperscript{11,24,38,45} propiciados por condiciones de “insularidad”.

La distribución multimodal en estos STRs ha sido relacionada con modelos mutacionales de dos eventos, que introducen mutaciones y variabilidad al acervo genético de la población.\textsuperscript{34,41} Por un lado, las mutaciones sencillas de adicción o deleción de una unidad de repetición generan una distribución bimodal alrededor de una media. Por otro, eventos mutacionales menos comunes, que ocurren posiblemente en la meiosis, involucran cambios de muchas unidades que alteran la distribución bimodal de estas repeticiones. La persistencia de perfiles similares de frecuencias alélicas con algunos marcadores en todas las poblaciones humanas, sugiere que este patrón se estableció en un antepasado común, antes de la separación de subpoblaciones y del surgimiento de diferentes etnias.

La disponibilidad de bases de datos sobre las frecuencias alélicas en la población costarricense es útil para estimar las probabilidades de identidad en procesos judiciales. De hecho, estos datos fueron utilizados para resolver los primeros 10 casos de paternidad dudosa solicitados por el Poder Judicial.\textsuperscript{32} La probabilidad de que alguien posea un alelo dado depende de la frecuencia de ese alelo en la población. La probabilidad de poseer una serie de alelos de diferentes marcadores no ligados, se establece como el producto de la frecuencia alélica de cada miembro de la serie. Con esta estadística se pueden calcular índices de verosimilitud para la probabilidad de paternidad o de identidad forense.\textsuperscript{32,38} Es evidente que se requieren estudios de otras subpoblaciones en el territorio nacional, para llegar a tener perfiles genéticos representativos de todo el país.

Cuando no se dispone de bases de datos de las poblaciones locales, se usan datos de otras poblaciones humanas, o se hacen aproximaciones\textsuperscript{25,30} que utilizan valores de frecuencia muy altos, restando poder incriminatorio a cada marcador. Esto favorece al imputado, siguiendo el principio legal in dubio pro reo, y aumenta el costo de la prueba, pues para lograr el mismo grado de
### Cuadro 2

Frecuencias alélicas de 10 micro y 2 minisatelites en el ADN nuclear. Excepto por el marcador D1S80, el tamaño de los alelos (en negrita) es en pares de bases. El tamaño de las secuencias, a ambos lados de la repetición, en pares de bases: APOA2, 104; D16S55, 79; D22S156, 50; D4S174, 136; D4S175, 71; D4S193, 62; D8S184, 115; D12S59, 100; D19S178, 124. Para D1S80 se utilizó la nomenclatura validada por Cosso y Reynolds 

<table>
<thead>
<tr>
<th>Locus</th>
<th>Frecuencia alélica debajo de los alelos (negrita)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APOA2</td>
<td>131 133 135 137 139 141 143 145 147</td>
</tr>
<tr>
<td></td>
<td>0.23 0.19 0.00 0.31 0.09 0.04 0.21 0.02 0.01</td>
</tr>
<tr>
<td>D4S174</td>
<td>177 179 181 183 185 187 189 191 193 197</td>
</tr>
<tr>
<td></td>
<td>0.28 0.13 0.09 0.06 0.12 0.08 0.08 0.06 0.05 0.01</td>
</tr>
<tr>
<td>D4S175</td>
<td>110 112 114 116 118 122 124 126 128 132</td>
</tr>
<tr>
<td></td>
<td>0.02 0.30 0.03 0.01 0.03 0.01 0.14 0.13 0.06 0.05 0.18</td>
</tr>
<tr>
<td>D4S193</td>
<td>95 97 99 101 103 107 107</td>
</tr>
<tr>
<td></td>
<td>0.01 0.03 0.04 0.17 0.22 0.37 0.14</td>
</tr>
<tr>
<td>D8S164</td>
<td>163 165 167 169 171 173 187 193 195 197 199 201</td>
</tr>
<tr>
<td></td>
<td>0.01 0.15 0.05 0.02 0.43 0.08 0.01 0.14 0.05 0.01 0.01 0.01</td>
</tr>
<tr>
<td>D12S59</td>
<td>162 166 168 170 172 174 176 178 180 182 186 188 192</td>
</tr>
<tr>
<td></td>
<td>0.02 0.10 0.22 0.03 0.10 0.04 0.20 0.13 0.06 0.02 0.02 0.02 0.01</td>
</tr>
<tr>
<td>D18S35</td>
<td>100 104 106 108 110 120 122 124</td>
</tr>
<tr>
<td></td>
<td>0.01 0.55 0.06 0.05 0.05 0.04 0.04 0.18 0.01</td>
</tr>
<tr>
<td>D19S178</td>
<td>153 155 157 169 171 173 175 177 179 181 183 185</td>
</tr>
<tr>
<td></td>
<td>0.01 0.52 0.02 0.02 0.01 0.04 0.11 0.04 0.04 0.04 0.03 0.02</td>
</tr>
<tr>
<td>D21S156</td>
<td>77 85 89 91 93 95 97 99 101 103 105</td>
</tr>
<tr>
<td></td>
<td>0.34 0.03 0.06 0.12 0.03 0.04 0.08 0.07 0.12 0.07 0.02 0.01</td>
</tr>
<tr>
<td>D22S156</td>
<td>98 100 102 104 106 108</td>
</tr>
<tr>
<td></td>
<td>0.01 0.01 0.20 0.52 0.20 0.05</td>
</tr>
<tr>
<td>APO B</td>
<td>480 550 603 641 689 722 782 825 866 905</td>
</tr>
<tr>
<td></td>
<td>0.01 0.12 0.16 0.39 0.03 0.04 0.04 0.11 0.01</td>
</tr>
<tr>
<td>D1S80</td>
<td>18 20 21 22 23 24 25 26 27 28 29 30</td>
</tr>
<tr>
<td></td>
<td>0.30 0.03 0.01 0.01 0.01 0.16 0.12 0.03 0.03 0.07 0.07 0.05</td>
</tr>
<tr>
<td></td>
<td>0.07 0.01 0.01 0.02</td>
</tr>
</tbody>
</table>

### Cuadro 3

Valores de heterocigosidad de las cuatro poblaciones estudiadas con diez marcadores genéticos**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costa Rica</td>
<td>0.64</td>
<td>0.85</td>
<td>0.75</td>
<td>0.70</td>
<td>0.76</td>
<td>0.85</td>
<td>0.78</td>
<td>0.64</td>
<td>0.62</td>
<td>0.86</td>
</tr>
<tr>
<td>Cerdeña</td>
<td>0.79</td>
<td>0.87</td>
<td>0.77</td>
<td>0.76</td>
<td>0.78</td>
<td>0.79</td>
<td>0.67</td>
<td>0.70</td>
<td>0.76</td>
<td>0.85</td>
</tr>
<tr>
<td>Egipto</td>
<td>0.77</td>
<td>0.88</td>
<td>0.79</td>
<td>0.77</td>
<td>0.75</td>
<td>0.88</td>
<td>0.73</td>
<td>0.70</td>
<td>0.87</td>
<td>0.84</td>
</tr>
<tr>
<td>África*</td>
<td>0.74</td>
<td>0.80</td>
<td>0.79</td>
<td>0.62</td>
<td>0.69</td>
<td>0.87</td>
<td>0.84</td>
<td>0.66</td>
<td>0.83</td>
<td>0.85</td>
</tr>
</tbody>
</table>

*África al sur del Sahara
**Se indican en negrita el menor valor de cada serie


AMC, setiembre 1999, vol 41 (3)
confiabilidad estadística es necesario aumentar el número de marcadores (por ejemplo, de 6 a 10).

**Agradecimientos**

Nuestro agradecimiento a muchos compañeros en el CIBCM que hicieron diferentes aportes: Sandra Silva, Reynaldo Pereira, Henriette Ravenst, Heidy Villalobos y Amy C. Peterson. También agradecemos el apoyo del Dr. Nelson Freimer del laboratorio de Neurogenética de la Universidad de California en San Francisco. Además de la Vice Rectoría de Investigación, este trabajo recibió el apoyo del CONICIT (801-90567) y de la IAEA (COS 013) para la pasantía de E. R. en el laboratorio del Dr. Freimer.

**Abstract**

**Aim:** To establish databases of allele frequencies in a Costa Rican Central Valley population sample.

**Methods:** Peripheral blood samples from more than 40 individuals were used to isolate DNA and analyze each sample with 10 dinucleotide repeat genetic markers and with 2 minisatellite repeats, using the polymerase chain reaction. Alleles were identified by comparison with DNA from CEPH family members. Genotypes were determined by labelling one of the two PCR primers with $^{32}P$ before amplification, electrophoresis in sequencing gels and autoradiography.

**Results and Conclusions:** Analysis of this data set indicates that this sample is in Hardy-Weinberg equilibrium and shows no evidence of linkage disequilibrium between markers. These data are compared with results from other human populations analyzed with the same markers, finding similarities in allele frequencies among them. Notably, the Costa Rican sample presents the lowest heterozygosity value, with 4 of the 10 dinucleotide markers tested, followed by a Cerdenial sample. In contrast, the two African samples presented the highest heterozygosity indexes with a larger number of alleles.

**Referencias**


5. Weber JL. Informativeness of human (dC- dA) and (dG- dT) polymorphisms. Genomics 1990; 7: 524-530.


