Genetic variability in Amerindian populations of Northern Argentina

Cristina B. Dejean1, Brigitte Crouau-Roy2, Alicia S. Goicoechea1-3, Sergio A. Avena1-3 and Francisco R. Carnese1

1Universidad de Buenos Aires, Facultad de Filosofía y Letras, Instituto de Ciencias Antropológicas, Sección Antropología Biológica, Buenos Aires, Argentina.
2Université Paul Sabatier, Évolution et Diversité Biologique, Toulouse, France.
3Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina.

Abstract

The allelic variability of four dinucleotide microsatellites located in the HLA region (MOGc, D6S265, MIB, and TNFa) was analyzed in 67 individuals representing three Amerindian populations of the Argentine Gran Chaco: Toba, Wichi and Chorote. Genomic DNA was prepared from peripheral blood and DNA was extracted using the standard phenol-chloroform procedure. Alleles were identified by PCR, using an end-labelled reverse oligonucleotide primer (fluorescent 6 - Fam labeling). Despite the low number of samples studied, a high level of gene diversity was observed in each population and for each locus. Moreover, the mean number of alleles was 7.7, 5.3, 10.0, and 7.0 at loci MOGc, D6S265, MIB and TNFa, respectively. Differentiation tests between pairs of populations showed a clear differentiation between the Wichi and the other two groups. However, the proportion of the total genetic variability that is due to differences among populations, estimated by the Gst' index, was relatively low (6%). Almost all the genetic variation occurred at the intra-population level (96%). The high intra-populational genetic variation suggests the existence of an intensive gene flow among the Gran Chaco tribes. Historical information seems to confirm this result.

Key words: Aboriginal Argentinians, microsatellites, intra- and inter-populational variation.

Received: July 3, 2003; Accepted: May 5, 2004.

Introduction

Several studies have been carried out among Argentine Amerindians, involving investigations on blood groups, enzymes, serum proteins (Carnese et al., 1993, 1996, 2002; Carnese, 1995; Goicoechea et al., 1996) and DNA polymorphism including mtDNA, Y chromosome markers and nuclear DNA (Ginther et al., 1993; Baillet et al., 1994; Pena et al., 1995; Bianchi et al., 1997; Hutz et al., 1997; and Kaufman et al., 1998), and these data have been used to analyze genetic diversity and relationships between aboriginal Argentine populations and to compare them to other southern Amerindian populations.

The Argentine groups mentioned above exhibit the typical A, B, C, and D mitochondrial DNA haplogroups described for Amerindians and a small percentage of the so-called E haplogroup (Bravi et al., 1995). Recently, data on 23 genetic markers, such as blood groups, enzymes and serum proteins, obtained from the Toba, Wichi and Chorote populations inhabiting the North Argentine Gran Chaco Region, were compared with the Mapuche and Tehuelche native Argentine groups (Goicoechea et al., 2001). In this report, similar heterozygosis values were observed in the five groups. However, the phylogenetic tree divided the five tribes into two different groups (Chaco and Patagonia).

In previous investigations on aboriginal populations of the Argentine Chaco, antigens of the HLA system have been used. Pirosky et al. (1983) employed serological methods to study HLA-DR antigens in a Toba group and detected differences when they compared the data to those of a European control group. Vullo et al. (1992) analyzed, at the molecular level, the polymorphic variants of several HLA-A, -B, -C, -DR, -DQ, and -DP loci in a Wichi group. They found some haplotypes shared with other South American Indians, whereas others were present only in the population studied. A comparative study performed by Cerna et al. (1993) among several groups of the Argentine Gran Chaco Region (Toba, Wichi, Chorote and Pilagá) showed that the allele frequencies for HLA class II loci were similar in the Argentine tribes, but different from
those of the Brazilian Xavante, another group studied by
the same authors.

Despite the high number of studies performed in the
Gran Chaco populations, no data are available on micro-
satellite markers located in the HLA region, on the short
arm of chromosome 6 (see Foissac et al., 1997). Due to its
fundamental importance in the immune response, this re-
gion is under the action of highly selective forces and link-
age disequilibrium (Begovich et al., 1992).

In the present study, we have analyzed four
dinucleotide microsatellite polymorphisms located in the
HLA region (MOGc, D6S265, MIB, and TNFa, see Figure
1) in more than 60 individuals representing three Amerin-
dian populations: Toba, Wichi and Chorote. The variability
of these markers was analyzed in order to estimate levels of
 genetic structure and gene diversity, to shed light on the mi-
gation events that occurred in this geographical region and
to make hypotheses on their history.

Material and Methods

Populations

The groups investigated in this study were previously
described by Goicoechea et al. (2001). Chorote, Wichi and
Toba populations live near to the village of Santa Victoria
Este, Department of Rivadavia, Salta Province, located in
the Gran Chaco Region, in northern Argentina. These com-
munities surround Santa Victoria (62°42.5′ W; 22°16′ S)
like satellites. The distances separating this village from
each of the settlements vary from 3 to 70 km (Figure 2).

Both Wichi and Chorote individuals belong to the
Mataco branch of the same linguistic group, the Mataco-
Guaiicuru stock (Loukotka, 1968; Greenberg, 1987) living
in the Argentine Chaco, although each of them constitutes a
different ethnic group. The Chorote are thought to have ar-
ived from the Amazon basin, their subsistence is based on
hunting, gathering, fishing and a rudimentary agriculture
(Goicoechea et al., 2001). In 1984, there were 915 indi-
viduals, representing 5% of the aboriginal population of the
Salta Province. Our samples were obtained mainly in the
villages of La Merced and La Gracia. The Wichi, as they
call themselves, have a similar economy of subsistence and
are much more numerous, their population amounting to
about 20,000 in 1986 (Goicoechea et al., 2001). Samples
were obtained in the villages of La Paz, La Vertiente, Pozo
de Tigre and San Luis.

The third population studied, the Toba, speaks an-
other language that is also classified within the Guaiicuru
branch of the Mataco-Guaicuru stock (Loukotka, 1968;
Greenberg, 1987). They adopted the horse in the 18th cen-
tury, and have intensively interacted with all other Indian
populations of the region (for further details, see
Goicoechea et al., 2001). Many of them have migrated to
the peripheral zones of large cities like Salta, Rosario and
Buenos Aires. About 50,000 Toba individuals are distrib-
uted throughout the country (Goicoechea et al., 1996). The
Toba samples were obtained in Monte Carmelo. Five mem-
bers of the three tribes were sampled at La Curvita.

DNA typing and sequencing

Genomic DNA was prepared from peripheral blood. Samples
from unrelated individuals were obtained in 1995, 
refrigerated, sent to Buenos Aires and frozen at -20 °C.
DNA was extracted by the standard phenol-chloroform
procedure, in a final volume of 3 mL. An aliquot of the ex-
tracted DNA from each of 48 Wichi, 8 Toba and 11 Chorote
individuals was genotyped in Toulouse (France) for micro-
satellite mapping within the major histocompatibility com-
plex on chromosome 6: MOG C [(CA) n repeats], MIB
[(CA)n], D6S265 [(CA)n] and TNF a [(GT)n] (see review
in Foissac et al., 1997).

Alleles were amplified by PCR, using an end-labelled
reverse oligonucleotide primer (fluorescent 6-Fam labeling).
The PCR amplification was performed in a 25 µL re-
action volume, containing 1X buffer, 1.5 mM MgCl2,
0.2 mM dNTP, 1 pM of each denaturation primer, 0.25 U
Promega Taq polymerase, and 400 ng of DNA. After a

Figure 1 - Schematic representation of the major histocompatibility gene complex and positions of microsatellites. Dejean et al.
3-min step at 95 °C, 10 amplification cycles were carried out (94 °C for 30 s, 1 °C for 30 s, 72 °C for 30 s), followed by another 20 cycles (90 °C for 30 s, 2 °C for 30 s, 72 °C for 30 s), and a 5-min elongation step at 72 °C. In some cases, the use of two annealing temperatures (t1 and t2) was found to be more efficient (see Table 1). Microsatellite genotyping was resolved in a 6% denaturing polyacrylamide gel in an automatic 373A sequencer (Perkin Elmer, Applied Biosystems) with an internal size standard (Genescan-2500 TAMRA). Table 1 shows the repeat motif and primer sequences of each micro-satellite.

For the MIB locus, the PCR products were purified using Qiaquick spin columns (Qiagen) before ligation into a pGEM T easy vector (Promega) and cloning into Epicurian Coli XL1 Blue (sub-cloning grade competent cells, Stratagene). Plasmid DNA was then extracted and purified using Wizard Minipreps (Promega). Nucleotide sequencing was carried out by the dideoxytermination method using a dye primer (Perkin Elmer) in an automatic sequencer. Consensus sequences for each individual were obtained by aligning the forward and reverse sequences obtained.

Statistical analysis

The Hardy-Weinberg equilibrium was determined and the population structure was studied by estimating the allelic and genotypic distributions, using the GENEPOP 3.2 package (Raymond and Rousset, 1995a). The gene differentiation was tested by Fisher’s exact test (Raymond and Rousset, 1995b). To quantify the degree of genetic variation within and between populations, Nei’s gene diversity measures were used (Nei, 1973, 1986). These analyses were performed using the DISPAN-program (Ota, 1993).

Results

The distribution of the allele frequencies found at the four HLA microsatellite loci analyzed in the 3 studied populations is presented in Table 2. A large number of alleles (13, 9, 17, and 11) was observed in the 3 populations (67 individuals); despite the low number of samples genotyped, the mean number of alleles was 7.7, 5.3, 10.0, and 7.0 at loci MOGc, D6S265, MIB and TNFa, respectively. Three out of the 4 microsatellites analyzed displayed a multimodal distribution of allele frequencies in almost all the tested samples. Sometimes gaps in the allelic distribution (i.e. differences of more than one repeat) were observed. The gene diversity values for each of the four STRs in the three populations are presented in Table 3. Sorted by locus, the highest average value corresponds to TNFa and the lowest to D6S265. From a population perspective, the gene diversity estimate, averaged over all loci, was high, ranging between 0.781 ± 0.035 (Wichi population) and 0.861 ± 0.083 (Toba population). Microsatellite MIB in the three populations and MOGc in the Wichi and Chorote groups showed a significant departure from panmixia, due to a heterozygote deficiency, while for TNFa this departure is due to an increase in heterozygosis in the Toba and Chorote groups.

Performing the differentiation test of every population pair provided complementary information (Table 4). The Wichi and Toba showed highly significant differences in the MOGc locus and significant differences in the MIB locus; Wichi and Chorote also exhibited statistically significant differences in the MOGc locus and highly significant differences in the MIB locus, whereas Toba and Chorote were demonstrated to be different only regarding the MOGc locus. Loci TNFa and D6S265 showed no statistically significant differences when comparing the three populations.

When the analysis was made for paired populations, but including the data on the four loci, the P-value obtained for each population pair demonstrated that the Wichi group was clearly different from Toba and Chorote. The Toba and Chorote populations showed a similar distribution for all four loci (Table 4). Overall genetic differentiation (Gst’) indicated that 6% of the genetic variance was due to the genetic differences among the three populations. TNFa exhibited the lowest values of inter-populational variation (Gst’ = 1.6%) and MOGc the highest (Gst’ = 7.8%). Loci MOGc and MIB contributed much more than D6S265 and TNFa to the total heterozygosis found (Table 5).

Table 1 - Characteristics of the microsatellites, primers employed and PCR temperatures.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Repeat motif</th>
<th>Primer sequences (5’→3’)</th>
<th>Annealing temperatures</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOGc</td>
<td>(CA)n</td>
<td>(1)GAA ATG TGA GAA TAA AGG AGA</td>
<td>t1: 57 °C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2)GAT AAA GGG GAA CTA CTA CA</td>
<td>t2: 58 °C</td>
</tr>
<tr>
<td>MIB</td>
<td>(CA)n</td>
<td>(1)AAG AAG AAA TCC TGC CTC AG</td>
<td>t1: 56 °C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2)TAT CAG TCC AGA TTC TTC AG</td>
<td>t2: 60 °C</td>
</tr>
<tr>
<td>TNFa</td>
<td>(GT)n</td>
<td>(1)GCC TCT AGA TTT CAT CCA</td>
<td>t: 58 °C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2)GCA CTC CAG CCT AGG CCA CA</td>
<td></td>
</tr>
<tr>
<td>D6S265</td>
<td>(CA)n</td>
<td>(1)ACG TTC GTA CCC ATT AAC CT</td>
<td>t: 59 °C</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2)ATC GAG GTA AAC AGC AGA AA</td>
<td></td>
</tr>
</tbody>
</table>

$\text{t}_1$ = annealing temperature of the first phase of PCR.
$\text{t}_2$ = annealing temperature of the second phase of PCR.
Discussion

Most of the available information about the microsatellites studied here comes from researches on European populations. In Spanish Basque groups (SB) (Peña et al., 2002), for example, it was detected that the frequencies of alleles D6S265*131 (18%) and *133 (27%) were relatively high, but lower than the ones observed here for the Chaco populations (CHP), in which the means were 39% and 40% for alleles *131 and *133, respectively. Regarding the TNFa locus, both groups register similar and relatively high frequencies of the allele TNFa *97 (CHP: X = 16%;
In turn, alleles TNFa*107 (CHP: X = 30%; SB = 6%) and TNFa*115 (CHP: X = 32%; SB = 5%) were present, though with different frequencies, in the Chaco groups and in a population sample from Finland (Karelau et al., 2000), while, on the other hand, these populations shared none of the alleles of loci MOGc, D6265 and MIB.

In European populations, the most frequent allele within the locus TNFa seems to be TNFa*99 (20% - 32%), therefore considered a genetic marker characteristic of these populations (Roth et al., 1994). Yet, its frequency is also relatively high in the Wichi group of the Argentine Chaco (11%), which, in our opinion, cannot be explained by the genetic mixture with European non-Amerindians alone. It is evident that further studies should be performed to respond to these queries, as well as to establish continental and intercontinental comparisons.

On the other hand, all the data shown in the present study seem to indicate the existence of genetic differences among the populations included, in spite of their geographical proximity and history of previous contacts. Moreover, we had expected to find a low intra-population variability due to the small sample sizes, a high level of endogamy (as observed by Cabutti and Palatnik, 1975) and the existence of a moderated contact with non-Indian people (the interethnic admixture estimated was low varying from 2% in Chorote to 4% in Toba; see Goicoechea et al., 2001). Nevertheless, a surprisingly high intra-population genetic diversity was detected for the four loci, despite the low number of genotyped individuals. In this sense, the intra-population diversity (Hs = 0.76) represented 96% of the total genetic variation (0.79), whereas the inter-population diversity (Dm = 0.05) and the percentage of the gene differentiation coefficient (Gst' = 6%) were relatively low. This value was similar to those observed at the nuclear level among the Surui, Zoró and Gavião groups inhabiting the same geographical region (Northwest of Brazil) and belonging to the same linguistic stock, the Tupi-Mondé (Hutz et al., 1999), but lower than the values obtained in South Amerindian populations considering VNTR/STR (Gst' = 19%; see Bortolini et al., 1998).

Investigations on microsatellite markers are rare in Argentine aboriginal populations. Sala et al. (1999), studying 12 autosomal microsatellites (differently from the present study), demonstrated a high variability in the Wichi Indians, the mean number of alleles being five.

In turn, the markers examined in the present study are described for the first time in the Chaco groups, and there are no data available to compare them with other groups of Argentine Indians. However, there are other studies of the Chaco populations with different hierarchic levels of analysis. Recently, thanks to the determination of 21 genetic protein systems in the same samples identified in the present study, a high intra-populational variability (Ht = 0.227; Hs = 0.205) and a relatively low genetic differentiation coefficient (Gst' = 2.6%) (Carnese et al., 2002) were also observed. Similar results were obtained by Demarchi et al. (2001) by means of mtDNA polymorphism analysis in Toba, Wichi, Chorote and Pilaga groups from the same geographic region. These authors obtained a low value of inter-group genetic differentiation (Gst' = 6%). We believe, in agreement with them, that the high intra-group genetic variation suggests the existence of an intensive gene flow among the Gran Chaco tribes.

Historical information seems to confirm this result. People living in the Chaco area maintain intense cultural and commercial relations among themselves and with other groups outside the region, such as Andean, Amazonian and Mesopotamian tribes (Metraux, 1946; Martinez Sarasola, 1992). The Toba constitute the largest groups, having incorporated the horse after the arrival of the Spaniards. They were better warriors than other aboriginal groups (Martinez Sarasola, 1992). In the Toba invasions, females and children of other communities were taken captive to join their group, thus establishing the gene flow from the attacked tribes into their own gene pool. The Toba groups studied here are confined to a fixed settlement near the other two groups. This contact surely implies a certain degree of gene flow, but no demographic data are available to quantify it. Nevertheless, we believe that the existence of inter-ethnic couples is possible, not only due to their geographic proximity, but also to their close linguistic relatedness (see Goicoechea et al., 2001).

In conclusion, we suggest that gene flow could be responsible for the variability found, but further studies with the same markers on other Amerindian populations would be useful to find out how these markers are distributed in America and to elucidate the genetic relationships among these populations.

**Acknowledgements**

This work was supported by grants from UBACyT, CONICET, SECyT-ECOS and CNRS.

Table 5 - Gene diversity analysis.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Ht</th>
<th>Hs</th>
<th>Dm</th>
<th>Gst</th>
<th>Gst'</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOGc</td>
<td>0.860</td>
<td>0.814</td>
<td>0.069</td>
<td>0.054</td>
<td>0.078</td>
</tr>
<tr>
<td>D6S265</td>
<td>0.645</td>
<td>0.614</td>
<td>0.046</td>
<td>0.048</td>
<td>0.069</td>
</tr>
<tr>
<td>MIB</td>
<td>0.890</td>
<td>0.848</td>
<td>0.063</td>
<td>0.048</td>
<td>0.069</td>
</tr>
<tr>
<td>TNFa</td>
<td>0.759</td>
<td>0.751</td>
<td>0.012</td>
<td>0.009</td>
<td>0.016</td>
</tr>
<tr>
<td>All loci</td>
<td>0.789</td>
<td>0.757</td>
<td>0.048</td>
<td>0.040</td>
<td>0.059</td>
</tr>
</tbody>
</table>

Ht: total genetic variation.
Hs: intra-populational diversity.
Dm: inter-populational diversity.
Gst/Gst': gene differentiation coefficient.
The subjects of the investigation were informed about the aims of the study and gave their consent, which is gratefully acknowledged.

We would also like to thank Raymonde Jambou for her technical assistance.

References


Loukotka C (1968) Classification of South American Indian Languages. Latin American Center, University of California, Los Angeles.


Ota T (1993) Dispan: Genetic Distance and Phylogenetic Analysis. University Park, Institute of Molecular and Evolutionary Genetics, The Pennsylvania State University, University Park, PA, USA.


Estudio de la frecuencia del HLA-DR. Medicina (B. Aires) 43:281-284.


Associate Editor: Francisco Mauro Salzano