Chapter from the book *Genetic Diversity in Microorganisms*
Downloaded from: http://www.intechopen.com/books/genetic-diversity-in-microorganisms

Interested in publishing with InTechOpen?
Contact us at book.department@intechopen.com
1. Introduction

It all began back in the 1990’s when the need to compare the mitochondrial DNA for matrilineages arose. It was then when the search for polymorphic markers of the Y-specific region started, in order to test whether the histories of female and male lineages were the same.

The human Y chromosome has an intermediate mutational rate between the one from autosomes and the X chromosome. Its mode of inheritance is exclusively patrilinear and its lower effective number (1/4th of the autosomal and 1/3rd regarding the X chromosome) makes it highly susceptible to genetic drift (Jobling and Smith, 2003).

Given its specific characteristics, the Y chromosomes have shown some degree of continental differentiation, providing a way of distinguishing among European, American, African, and Asian lineages. A tetranucleotidic microsatellite (DYS19, Roewer et al. 1992), an Alu insert (YAP+ or M1, Hammer et al. 1994), a single nucleotide polymorphism (SNP) (M2) associated to YAP+ (Seielstad et al. 1994), and variants of the alphoid system (Santos et al. 1995) were the first polymorphic systems to be studied. Later on, the fruitful search for polymorphisms led Underhill and collaborators to publish the first phylogeny compiling 166 SNPs geographically correlated (Underhill et al. 2000). Since then the nomenclature was normalized so that different research groups could compare their work (Y Chromosome Consortium, 2002), and an improved phylogeny was published in 2008 by Karafet and collaborators. The accepted phylogeny spans in a tree of 20 major clades that represent haplogroups (Figure 2), where the accumulation of polymorphisms along the lineages determine their diversification and the configuration of sub-branches.
Haplogroups are defined by SNPs, which have a mutational rate of $13.5 \times 10^{-10}$ (Anagnostopoulou et al. 1999), so that the probability of homoplasy results extremely low. Thus, the number of SNPs determines the phylogenetic status of the analyzed chromosome.

For the American continent, early studies that characterized the alphoid system and the DYS19 microsatellite showed that a large amount of Native American Y chromosomes carried the combination II-A for those loci (Pena et al. 1995, Santos et al. 1996a,b). In 1996, Underhill and collaborators described a mutation at the DYS199 locus (now named M3), that resembled the haplotype II-A described by Pena et al. (1995) in the sense that it was widely distributed in Native Americans but absent in populations of other continental origins. In a previous work of our team we correlated the three systems and described the frequency of six polymorphic sites, among which there was a new SNP, currently named SRY2627 (Bianchi et al. 1997), specific for European populations (Hurles et al. 1999).

Most males of Native American ancestry in South America carry a Y-chromosome lineage named sub-haplogroup Q1a3a (Karafet et al. 2008), which is characterized by the aforementioned SNP M3, and the derived states for M242 and M346. It has been found in all Native American populations from Alaska to the Magellan Strait, with an average frequency of 60% (Bianchi et al. 1998, Bortolini et al. 2003).

Present evidence makes researchers suppose that Q1a3a diverged from its predecessor during or shortly before the crossing of the Bering Strait (Lell et al. 2002). There is no doubt that Siberia was the main source of colonization of America, the last continent to be populated. Multidisciplinary evidence supports the hypothesis that strong bottlenecks happened during this process. Yet, when it happened and how many migratory waves occurred is still debated.

All these previous studies showed that the peopling of the Americas bore a strong founder effect regarding the paternal lineages, as a result of which most Y chromosomes described belong to a single lineage. It was only when Y-specific microsatellites began to be studied that its intrinsic diversity could be described through defining haplotypes. One of them was identified as ancestral to the diversification of Y chromosomes bearing M3. Several authors estimated the time of divergence of these lineages through maximum parsimony methods, obtaining different results: 22,770 years (13,500-58,700) (Bianchi et al. 1998); 11,456 (9,423-13,797) (Ruiz-Linares et al. 1999); 7,570 (SE 681) (Bortolini et al. 2003). Temporal differences among those three reports depended mostly on the adjustments of the estimation of mutational rates and the generation time considered. While the first two considered 27 years, Bortolini et al. used a shorter generation time of 25 years.

In 2003 Seielstad and collaborators described a new mutation at locus M242 that proved to be immediately ancestral to M3, and that now defines haplogroup Q. Haplogroup Q shows a wide distribution in the Old World, with an average frequency of 19% in Siberia that drops to 5% in Central Asia (Seielstad et al. 2003, Karafet et al. 2002). Lineages with membership in Q but lacking M3 have also been described in Native Americans. Formally designated as paragroup Q* or Q(xM3), these lineages attain high frequencies of up to 47% in some North American populations (Zegura et al. 2004, Bolnick et al. 2006), yet they are poorly represented in most Central and South American natives, with frequencies rarely surpassing 6% (Bortolini et al. 2003, Bailliet et al. 2009, Bisso-Machado et al. 2010).
Although 17 polymorphisms have been described so far inside haplogroup Q, most Native American Q lineages seem to be phylogenetically close to Q1a3a, sharing three of these SNPs (Karafet et al. 2008, Bailliet et al. 2009). It has been recently demonstrated that Q(xM3) lineages present in Native Americans, with the sole exception of some Eskimos, share the presence of M346 mutation and thus belong to clade Q1a3* (Bailliet et al. 2009, Bisso-Machado et al. 2011). Deviant from this pattern are two Q1a* lineages recently reported for one ancient Paleo-Eskimo (~4000 years BP) and one extant Eskimo (Rasmussen et al. 2008, Bisso-Machado et al. 2011).

Even though the chromosomes bearing M3 are the most frequent clade in Native populations, many research groups have tried to find out whether other haplogroups entered the American continent from Asia. In 1999, Bergen and collaborators described one SNP in the RPS4Y gene that defines what is now known as haplogroup C, present in populations of Asia, Australo-Melanesia and also the Americas, where they are can be found in populations belonging to the three major linguistic stocks, namely Eskimo-Aleuts, Na-Dene and Amerindians (Bergen et al. 1999, Karafet et al. 1999, Capelli et al. 2001, Hammer et al. 2001, Bosch et al. 2003, Malhi et al. 2008). C lineages with one further mutation at M217 belong to the C3b clade, which spread through central and eastern Asia and America. While Native North American C3 lineages share the P39 mutation that defines the monophyletic C3b group, the few C lineages known from South American populations belong to its presumed sister group C3(xP39). Although not systematically searched for in Central and South America, C lineages have so far been found in three Waorani and one Kichwa individuals from Ecuador, and two Wayúú from Venezuela, out of a total of 767 individuals from 48 populations (Lell et al. 2002, Bortolini et al. 2003, Zegura et al. 2004, Mazieres et al. 2008, Geppert et al. 2010).

1.1 Going narrower than haplogroups

When researchers need to work on a finer scale, they define haplotypes. Haplotypes are built from microsatellites or Short Tandem Repeats (STR), which are sequences of 2-6 nucleotides repeated in tandem. They have an average mutation rate of $3.35 \times 10^{-3}$, which translates into a possible differentiation at the father-son transmission if a high number of STRs is studied (Ballantyne et al. 2010). When STRs are analyzed within the lineages of a haplogroup, there are mathematical methods to estimate the time of divergence considering the mutational rate for each STR (Ballantyne et al. 2010).

So far the mutation rate has been estimated through the changes observed in certified father-son transmissions, and 186 STR mutational rates have been published (Ballantyne et al. 2010), so it is possible to chose the most adequate STRs depending on the problem to solve, and to use commercial kits for amplifying jointly 12 and 17 STRs which are resolved by capillary electrophoresis.

Lineages can be built by combining haplogroups and haplotypes. They represent monophyletic lineages that share a common ancestor and among which the diversification occurred by mutation acting at the STR level (Jobling et al. 2004).

There is relevant information regarding samples analyzed for C and Q haplogroups from many locations with a vast geographic distribution, which allows us to analyze the peculiarities of the geographic distribution of male lineages and their diversity in America. In the present work we analyze the distribution of major Y-chromosome haplogroups in 279
individuals from ten Native American populations from Chile, Paraguay, and Argentina, and 638 individuals from 12 cosmopolitan, urban populations of Argentina.

2. Materials and methods

We analyzed 917 samples of randomly selected male voluntary donors from 10 self-acknowledged Native American populations: Ayoreo (9) and Lengua (25) from Paraguay; Huilliche (26) and Pehuenche (18) from Chile; Wichí (118), Toba (5), Chorote (9), Mocoví (27), Mapuche (26), and Tehuelche (20) from Argentina. Twelve urban populations from Argentina were also included in the sample: Susques (19), Cochinoca (6), Rinconada (6), San Salvador de Jujuy (85), Catamarca (106), Salta (77), Tucumán (16), La Rioja (87), Mendoza (75), Tartagal (105), Aguaray (12), and Azampay (44) (Fig. 1)

Fig. 1. Map of South American sampling localities
We amplified by PCR different fragments carrying the markers of interest and then identified the polymorphism through enzymatic digestion, since restriction enzymes recognize specific sequences and cut the DNA that bears that sequence. In case there is a mutation on the sequence, the enzyme does not cut the DNA because the recognition site is different. This way, the alteration in one base is transformed into a difference in size, and it is for this reason that they are called Restriction Fragment Length Polymorphisms (RFLP).

<table>
<thead>
<tr>
<th>SNP / indel</th>
<th>Reference SNP ID</th>
<th>Forward primer</th>
<th>Reverse primer (5’-3’)</th>
<th>PCR Size (pb)</th>
<th>Enzymes</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>YAP</td>
<td></td>
<td>cagggggaagataaagaata</td>
<td>actgctaaaagggcatgatgta</td>
<td>150-450</td>
<td></td>
<td>Hammer, 1994</td>
</tr>
<tr>
<td>M3</td>
<td>rs3894</td>
<td>taatcagtctctcccagca</td>
<td>taggtcagctctctcattaat</td>
<td>201</td>
<td>MfeI</td>
<td>Present paper</td>
</tr>
<tr>
<td>M9</td>
<td>rs3900</td>
<td>ccagcattaaacacttcagg</td>
<td>aaacctaaactttgcacaagc</td>
<td>340</td>
<td>HinfI</td>
<td>Underhill et al. 1997</td>
</tr>
<tr>
<td>P27</td>
<td>rs9786896</td>
<td>ccctcaggaacttgtggaactaattg</td>
<td>aggcacacattttgtgcaaac</td>
<td>288</td>
<td>BanI</td>
<td>Present paper</td>
</tr>
<tr>
<td>M89</td>
<td>rs2032652</td>
<td>acagaaaggtcgtctcgattctccaa</td>
<td>ccaactcagccaaagttgagacat</td>
<td>87</td>
<td>NlaII</td>
<td>Present paper</td>
</tr>
<tr>
<td>M168</td>
<td>rs2032595</td>
<td>agtttgaggtagaatctggttagttctctcagtcgttc</td>
<td>aatctcataggctctgtgctgct</td>
<td>473</td>
<td>HinfI</td>
<td>Present paper</td>
</tr>
<tr>
<td>M207</td>
<td>rs2032658</td>
<td>ggggcaaagttaactcagacg</td>
<td>ttttagctgtgctgtgctgtgct</td>
<td>164</td>
<td>DraI</td>
<td>Present paper</td>
</tr>
<tr>
<td>M242</td>
<td>rs8179021</td>
<td>tcagatggcagagttgtagttgtaaagtcag</td>
<td>aaaacagcagtaggacgagtcagtcg</td>
<td>151</td>
<td>Hpy188III</td>
<td>Present paper</td>
</tr>
<tr>
<td>M346</td>
<td></td>
<td>agatgggaaagggcagccaaagcro</td>
<td>tctgtcactggctgtcgg</td>
<td>419</td>
<td>-</td>
<td>Sengupta et al. 2006</td>
</tr>
</tbody>
</table>

Table 1. SNP and Indel. -1: Sequencing

We used 9 markers to define membership in 9 clades (Fig. 2): YAP, M168, M89, M9, P27, M207, M242, M346, and M3 (Table 1, Fig 2). Primers were designed in order to reduce the fragment length for optimal amplification with degraded DNA and, where necessary, creating a mismatch in order to generate a recognition site for RFLP (present work)(Table 1, Fig.2). The basic set of seven STRs: DYS19, DYS389 I and II, DYS390, DYS391, DYS392, DYS393 was employed to establish haplotypes (Kayser et al. 1997, Pascali et al. 1999).

In the case of the lineages belonging to the Q1a3* paragroup, we built a Median Joining Network (Bandelt et al. 1999). Median Joining Networks allow researchers to estimate phylogenies at the intra-species level, without choosing between trees and employing non-recombining population data. The most popular software to compute them is NETWORK (http://www.fluxus-engineering.com/sharenet.htm), which allows the assignment of a definite weight to each character, so that less likely events could be given a higher weight (considered as more “decisive” since they are rare) than events with a high probability (since they are likely to have occurred many times). It is essential to employ a good criterion to weigh each character, with differential importance for each mutation depending on their rate, since the final network depends on it. We used a formula designed by our team (Muzzio et al. 2010), which establishes a precise mathematical scale that concurs with the probability of change in each marker.
Fig. 2. Phylogenetic tree according to Karafet et al. [2008]. Solid lines indicate haplogroups, which can be typed by the Y-SNP RFLP assays. Markers that have been typed are indicated upon the lines. Dotted lines indicate haplogroups, which are not included in the study.
The Differentiation Index Fst (Excoffier et al. 1992) was applied together with the Arlequin software (Schneider, Roessli, and Excoffier 2000). For the MNS analysis we used the NTSYS 2.11S (Exeter Software) from the Da Distance (Nei 1972). MNS analysis represents the information in space, where each axis represents a component that involves the variability found. In our case, the use of two axes or dimensions was enough and simplified the interpretation.

3. Results

3.1 Autochthonous lineages

Haplogroup Q1a3a attained frequencies between 46 and 96% for our Native American populations, with values equal to or higher than 80% among Wichi, Toba, Chorote, and Pehuenche (Table 2). This clade accounted for only 7-17% of the lineages in Mendoza, La Rioja, Catamarca, Tucumán, Azampay, and Aguaray but gathered together one to two thirds of patrilinages in Tartagal, SS Jujuy, Salta, Cochinoa, and Rinconada. The small urban sample of Susques stands out as having a share of 95% for Q1a3a lineages.

<table>
<thead>
<tr>
<th>Populations</th>
<th>Province</th>
<th>N</th>
<th>AB</th>
<th>DE</th>
<th>F (xK)</th>
<th>K (xQ,R)</th>
<th>Q1a3*</th>
<th>Q1a3a</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayoreo</td>
<td>Paraguay</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0.111</td>
<td>0.222</td>
<td>0.556</td>
<td>0.111</td>
<td></td>
</tr>
<tr>
<td>Lengua</td>
<td>Paraguay</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0.040</td>
<td>0.040</td>
<td>0.280</td>
<td>0.680</td>
<td>0</td>
</tr>
<tr>
<td>Wichi</td>
<td>Salta</td>
<td>25</td>
<td>0</td>
<td>0.040</td>
<td>0</td>
<td>0</td>
<td>0.960</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Wichi</td>
<td>Formosa(LY)</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0.133</td>
<td>0.033</td>
<td>0</td>
<td>0.800</td>
<td>0.033</td>
</tr>
<tr>
<td>Wichi</td>
<td>Formosa(IJ)</td>
<td>59</td>
<td>0</td>
<td>0</td>
<td>0.102</td>
<td>0.017</td>
<td>0</td>
<td>0.831</td>
<td>0.051</td>
</tr>
<tr>
<td>Chorote</td>
<td>Salta</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0.111</td>
<td>0</td>
<td>0</td>
<td>0.889</td>
<td>0</td>
</tr>
<tr>
<td>Mocoví</td>
<td>Santa Fé</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0.074</td>
<td>0.037</td>
<td>0.074</td>
<td>0.481</td>
<td>0.333</td>
</tr>
<tr>
<td>Mapuche</td>
<td>Chubut</td>
<td>26</td>
<td>0</td>
<td>0.077</td>
<td>0.115</td>
<td>0</td>
<td>0.038</td>
<td>0.538</td>
<td>0.231</td>
</tr>
<tr>
<td>Tehuelche</td>
<td>Neuquén</td>
<td>20</td>
<td>0</td>
<td>0.150</td>
<td>0.095</td>
<td>0</td>
<td>0</td>
<td>0.650</td>
<td>0.150</td>
</tr>
<tr>
<td>Huilliche</td>
<td>Chile</td>
<td>26</td>
<td>0</td>
<td>0.056</td>
<td>0.308</td>
<td>0.038</td>
<td>0.038</td>
<td>0.462</td>
<td>0.154</td>
</tr>
<tr>
<td>Pehuenche</td>
<td>Chile</td>
<td>18</td>
<td>0</td>
<td>0.056</td>
<td>0.056</td>
<td>0</td>
<td>0</td>
<td>0.833</td>
<td>0.056</td>
</tr>
<tr>
<td>Susques</td>
<td>Jujuy</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0.053</td>
<td>0</td>
<td>0</td>
<td>0.947</td>
<td>0</td>
</tr>
<tr>
<td>Cochinoa</td>
<td>Jujuy</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0.167</td>
<td>0</td>
<td>0</td>
<td>0.667</td>
<td>0.167</td>
</tr>
<tr>
<td>Rinconada</td>
<td>Jujuy</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0.167</td>
<td>0</td>
<td>0</td>
<td>0.667</td>
<td>0.167</td>
</tr>
<tr>
<td>SS Jujuy</td>
<td>Jujuy</td>
<td>85</td>
<td>0</td>
<td>0.012</td>
<td>0.094</td>
<td>0.071</td>
<td>0</td>
<td>0.459</td>
<td>0.365</td>
</tr>
<tr>
<td>Catamarca</td>
<td>Catamarca</td>
<td>106</td>
<td>0</td>
<td>0.094</td>
<td>0.208</td>
<td>0.009</td>
<td>0</td>
<td>0.104</td>
<td>0.585</td>
</tr>
<tr>
<td>Azampay</td>
<td>Catamarca</td>
<td>44</td>
<td>0</td>
<td>0.023</td>
<td>0.068</td>
<td>0</td>
<td>0</td>
<td>0.159</td>
<td>0.750</td>
</tr>
<tr>
<td>Salta</td>
<td>Salta</td>
<td>77</td>
<td>0</td>
<td>0.078</td>
<td>0.078</td>
<td>0.013</td>
<td>0.013</td>
<td>0.494</td>
<td>0.325</td>
</tr>
<tr>
<td>Tartagal</td>
<td>Salta</td>
<td>105</td>
<td>0</td>
<td>0.010</td>
<td>0.105</td>
<td>0.152</td>
<td>0</td>
<td>0.362</td>
<td>0.371</td>
</tr>
<tr>
<td>Aguaray</td>
<td>Salta</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.167</td>
<td>0.833</td>
</tr>
<tr>
<td>Tucumán</td>
<td>Tucumán</td>
<td>16</td>
<td>0</td>
<td>0.063</td>
<td>0.125</td>
<td>0.250</td>
<td>0</td>
<td>0.125</td>
<td>0.438</td>
</tr>
<tr>
<td>La Rioja</td>
<td>La Rioja</td>
<td>87</td>
<td>0</td>
<td>0.149</td>
<td>0.241</td>
<td>0.057</td>
<td>0</td>
<td>0.103</td>
<td>0.448</td>
</tr>
<tr>
<td>Mendoza</td>
<td>Mendoza</td>
<td>75</td>
<td>0</td>
<td>0.147</td>
<td>0.293</td>
<td>0.040</td>
<td>0</td>
<td>0.067</td>
<td>0.453</td>
</tr>
</tbody>
</table>

Table 2. Haplogroup frequency in populations studied. LY: Laguna Yema, IJ: Ingeniero Juarez.
A total of 374 lineages were assigned to Q1a3a, and complete STR profiles were obtained for 137 of them, resulting in 97 different haplotypes. Fixation Index (Fst) for these lineages was 0.112, and the mean gene diversity was 0.501. We observed a great allele frequency differentiation of Q1a3a haplotypes. In the MDS plot (Fig. 3), the first axis separates populations by their geographic location: Northwestern Andean populations as Rinconada, Cochinoca, and Humahuaca on the upper left side, while the Northeastern Gran Chaco populations, as Wichi, Toba, Chorote, Ayoreo, and Lengua, occupied the upper right side.

Only 13 individuals were assigned to paragroup Q1a3*, all of them derived from indigenous populations except for one donor from Salta. All cases were singletons, except for Paraguayan Ayoreo and Lengua, populations for which high frequencies of 22-30% were found. Although reduced in total number, they showed a considerably high allele frequency differentiation with a mean gene diversity of 0.478.

Fig. 3. Two-dimensional scaling of genetic distances (Nei, 1972) based on seven Y-chromosome STRs haplotypes belonging to Q1a3a haplogroup from 17 populations from Argentina and Paraguay. SSJ: San Salvador de Jujuy. STRESS1 = 0.11876.

Network analysis of the Q1a3* haplotypes showed three Lengua at the central position, while the only haplotypes that differed in one or two allelic changes from these were from Lengua or Ayoreo populations. The other haplotypes diverged in more allele changes, while three median vectors (indicating absent haplotypes in the sample) were interposed between the central and derived haplotypes. This is concordant with the hypothesis of severe drift acting over these less frequent haplotypes. Lengua 2 and Lengua 3 carried 2 identical haplotypes each (Fig.4).
3.2 Allochthonous haplogroups

The most frequent allochtonous haplogroups were R and F(xK). High frequencies of 32-83% for R were found in all the urban populations except for the three highland samples of Rinconada, Susques, and Cochinoca. Conversely, R never surpassed the maximum of 34% in the Native Americans ones, and was even absent in four of them. Paragroup F(xK) showed overall frequencies lower than 30% in the urban populations, and values under 15% in the Native ones, with the remarkable exception of 31% in Chilean Huilliches. Surprisingly, frequencies for F(xK) were equal or higher than those for R in Amerindian populations of Formosa: Wichi, Chorote, Huilliche, Tehuelche, and Pehuenche (Table 1).

Haplogroup DE had a maximum frequency of around 15% in Mendoza and La Rioja, and paragroup K(xQ,R) did not exceed values of ~7% when only reasonably sized samples (i.e. N $\geq$ 10) were considered. We only found two Y chromosomes that belonged to the AB haplogroups.

3.3 Genetic differentiation among total haplogroups

Population variation was 17.23% and the differentiation coefficient observed among populations was quite high (Fst= 0.17).

Figure 5 represents the two-dimensional Da distance matrix (Nei, 1972). The stress value was 0.0297, which suggests good adjustment. A highlight in the figure represents the first
axis (R1), which explains 100% variation and separates all Native American populations from the urban ones, with the exception of those from Jujuy: SS Jujuy, Rinconada, Cochinoca, and Susques. The high frequencies of Q1a3a, Q1a3*, and K(xQ,R) determine that portion of the plot. The three populations of Wichi plus Susques, Chorote, Pehuenche and Rinconada and Cochinoca (these last two indistinguishable) represent one group, while the other one comprises Mapuche, Tehuelche, Huilliche, Mocovi, and SS Jujuy samples. The samples from Lengua and Ayoreo were situated far in relation to the aforementioned, probably because of their highest proportion of Q1a3*.

The other half of the plot is influenced by the foreign haplogroups AB, DE, F(xK), and R, whose frequencies determine the position of the urban and semi-urban populations positions, excluding those from Jujuy. Salta, Aguaray and Azampay constitute one group; La Rioja and Mendoza form another group, and finally Tartagal, Catamarca, and Tucumán do not configure or take part in any group (Fig. 5).

Fig. 5. Two-dimensional scaling of Nei’s (1972) distance matrix; ● Populations. □ Haplogroups; Stress=0.02970

4. Conclusion

We summarize below a few important facts about the autochthonous and allochthonous haplogroups that were studied in this research work:

4.1 Autochthonous haplogroups to America

Q1a3a (bearer of the derived state for M242, M346, and M3) is the most frequent and widely distributed clade in the Americas (Underhill et al. 1996; Bianchi et al. 1998; Bortolini et al. 2003; Bisso Machado et al. 2010, Geppert et al. 2010; Toscanini et al. 2011) and is considered
autochthonous to that Continent. In Q1a3a, 5 mutations have been identified: a) M19 T-A which defines Q1a3a1 (Underhill et al. 1996) has been found in 22 of 33 Ticuna and 2 of 19 Wayuu (Bortolini et al. 2003), and in 2 Toba from Argentina (Toscanini et al. 2011); b) M194 T-C which defines Q1a3a2 (Underhill et al. 2001) was described in one Maya (Shen et al. 2000); c) M199 is an insertion defining Q1a3a3 (Underhill et al. 2001), which was found in 1 Suruí (Shen et al. 2000); d) SA01 C-T which defines the new sublineage Q1a3a4, which has been identified in the Andean populations of South America (Jota et al. 2011). However, none of these variants have been found in the series of samples analyzed in the present work.

Members of Q1a3* are the sister group to major sub-haplogroup Q1a3a (Seielstad et al. 2003, Bortolini et al. 2003, Bailliet et al. 2009). This clade has a very low frequency and patchy distribution in South America (Bortolini et al. 2003, Bailliet et al. 2009) but attains higher frequencies in North America (Bolnick et al. 2006, Malhi et al. 2008). Out of the 17 polymorphisms described for Q by Karafet et al. (2008), all Q* haplogroups so far analyzed for America present the derived states for M242, P36.2, MEH2, and M346 (Karafet et al. 2008; Bailliet et al. 2009; Bisso-Machado et al. 2010; Geppert et al. 2011; Toscanini et al. 2011), except for one ancient Paleo-Eskimo and one extant Eskimo individuals assignable to Q1a* (Rasmussen et al. 2008, Bisso-Machado et al. 2011).

It should be highlighted that Q1a3* was found in the 9.2% of 885 males from 16 ethnic groups of Siberia and East Asia. The age for this subhaplogroup was estimated in South Siberia at about 4.5±1.5 thousand years ago (Ka), while the divergence time between clade Q1a3* and American-specific haplogroup Q1a3a was equal to 13.8±3.9 Ka, pointing to a relatively recent entry date to America (Malyarchuk et al. 2011).

Haplogroup C is present in Asia in variable frequencies. In Mongolia, C3-M217 and C3a-M48) are the benchmark haplogroups, with frequencies of 13% and 46% respectively (Chen et al. 2011), while among the Kazakhs those are present in 9% and 57% of the cases (Nasidze et al. 2005).

In North America the subhaplogroups C3* and C3b were described with frequencies averaging 5.8% (Bergen et al. 1999; Zegura et al. 2004; Bolnick et al. 2003). Only C3* was found in South America, among 2 Wayuu described by Zegura et al. (2004) and 3 Waorani and 1 Kichwa from the Ecuador Amazon region (Geppert et al. 2010).

4.2 Allochthonous haplogroups

A-B are almost exclusive to sub-Saharan Africa. While clade A chromosomes occurs with high frequencies of 30-66% in Southern and Eastern African populations and are also present at lower values in North and Central Africa (Hassan et al. 2008, Cruciani et al. 2002), haplogroup B chromosomes are specially frequent in Central and Western Africa (Hammer et al. 2001; Underhill et al. 2001; Jobling and Tyler-Smith, 2003).

Our YAP+ chromosomes are most probably members of clade E, widely distributed in Africa and West Eurasia.

Under paragroup F(xK) we have probably detected an assortment of lineages of both European and Middle Eastern/North African origin belonging to haplogroups G, H, I and J,
whose presence in Native and cosmopolitan populations of Argentina has already been reported (Corach et al. 2010, Blanco-Verea et al. 2010).

R is the most frequent haplogroup in Europe (Jobling and Tyler-Smith 2003), and is also the most common haplogroup in Argentinean urban populations (Ramallo et al. 2009a, Corach et al. 2010).

Contemporary self-acknowledged Native American populations keep bearing an important number of paternal Native lineages. Although this also happens in admixed urban contexts, the populations from Jujuy differ from the rest because of their high Native American contribution. Such component comes from the ethnographic and historical characteristics of Jujuy, since it was one of the most highly populated regions during pre-Columbian times and offered a strong resistance to the Spanish colonization (Hernández, 1992; Pucci, 1998). On the other hand, there is evidence of a lower proportion of admixture probably due to altitude, which may have acted as a barrier or dissuasive effect for the inhabitants of a European origin (Dipierri et al. 1997, 1998, 2000).

Our results also show a connection among Mapuche, Huiliche, and Tehuelche, which is possible to be interpreted within a historical context: Mapuche from Argentina and Huiliche from Chile have the same origin, and contact between them is quite well documented (Martínez Sarasola, 1992).

In reference to the foreign haplogroups, it is the first time that AB is described for Argentina; so far the African presence had only been found through YAP+ chromosomes (Bravi et al. 2000n).

The third most frequent haplogroup in the Argentine populations studied thus far in our laboratory is F(xK), while K(xQ,R) is a minor haplogroup among South American samples and involves subhaplogroups of Asian origin (Su et al. 2000; Hammer et al. 2001; Su et al. 1999; Underhill et al. 2001).

The high frequency of R can be explained by the strong European migration that took place during the late XIX and early decades of the XX centuries, specifically with the arrival of Italian and Spaniard migrants. Something similar was observed for other American countries such as Brazil (Bortolini et al. 2003), Mexico (Rangel-Villalobos et al. 2008), and the United States (Zegura et al. 2004; Bolnick et al. 2006). Even the R1b subhaplogroup was described in 11% urban samples from the city of La Plata, Argentina (Bianchi et al. 2007).

4.3 Concluding remarks

Native American male lineages found in self-acknowledged Native American populations can also be found in urban contexts, although at lower frequencies. Likewise, even among self-acknowledged Native American populations foreign haplogroups are present, depending on the recent history of human migrations.

The distribution of autochthonous lineages is the result of a complex admixture process that occurred in many Latin American populations. We are currently employing those Native traces to explain other historical events such as the peopling of the Americas, by describing the possible bottlenecks and founder effects that the lineage distribution shows.
5. Acknowledgment

We thank all DNA donors for making this work possible. We are very grateful to Néstor O. Bianchi and Susana Salceda for their help during different stages of our work. Grant sponsors: CONICET, CICPBA, ANPCyT, UNJu, Antorchas Foundation of Argentina. G Bailliet and CM Bravi are members to the “Consejo Nacional de Investigaciones Científicas y Técnicas de la República Argentina” (CONICET).

6. References


relationship of native populations in South America. Forensic Science International: Genetics, Vol.5, pp.100–104


Genetic Diversity in Microorganisms presents chapters revealing the magnitude of genetic diversity of microorganisms living in different environmental conditions. The complexity and diversity of microbial populations is by far the highest among all living organisms. The diversity of microbial communities and their ecologic roles are being explored in soil, water, on plants and in animals, and in extreme environments such as the arctic deep-sea vents or high saline lakes. The increasing availability of PCR-based molecular markers allows the detailed analyses and evaluation of genetic diversity in microorganisms. The purpose of the book is to provide a glimpse into the dynamic process of genetic diversity of microorganisms by presenting the thoughts of scientists who are engaged in the generation of new ideas and techniques employed for the assessment of genetic diversity, often from very different perspectives. The book should prove useful to students, researchers, and experts in the area of microbial phylogeny, genetic diversity, and molecular biology.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: