

# DNA, proteins and human diversity

*Francisco M. Salzano*

## ABSTRACT

A review is made of studies at the protein and DNA levels performed by the Porto Alegre group in association with several others, inside and outside Brazil. Topics examined were: (a) the determination at the molecular level of mutations in the albumin and glucose-6-phosphate molecules, as well as their population distribution; (b) investigations involving "classical" markers; (c) nucleotide diversity of the mtDNA of four South American Indian tribes; and (d) the population prevalence of alleles and haplotypes from two nuclear DNA systems: D1S80 and beta-globin gene cluster. The relationship of these results with the question of the early peopling of the Americas was systematically considered.

## GENETICS AND THE PEOPLING OF THE AMERICAS

Probably all the possible hypotheses have already been made in relation to the peopling of the Americas. Most of the early studies have been reviewed by Pericot y Garcia (1962). In relation specifically to the contribution that genetics can give to the solution of this riddle, I have considered this question repeatedly in the past (Salzano, 1963, 1964, 1976, 1985; Salzano and Callegari-Jacques, 1988). If there are genetic markers that are specific to a given ethnic group and are geographically restricted, their presence in the putatively derived populations may furnish clear clues about origins. The problem is that such specific markers are very rare. We have, therefore, to rely on relative resemblances, with all its problems. Genetic similarity does not necessarily imply common origin; moreover, presently observed values could not have been characteristic of previous populations. The recent possibility of studying ancient DNA may remove this last restriction, but this approach has its own problems (Hermann and Hummel, 1994). The most recent attempts in the investigation of this question resorted to the study of mitochondrial DNA (mtDNA), with no generally agreed consensus, however (see Szathmary, 1993; Cann, 1994, and the ensuing discussion which appeared in the American Journal of Human Genetics). In what follows I will review some of the studies our group has made in recent years,

describing and evaluating different aspects of our species genetic variability, and relating them to the problem under consideration.

## DNA OR PROTEIN?

The first question we have to address is: What level of investigation should be used? Theoretically, if we are interested in genetics, the best approach would be the investigation of the genetic material (deoxyribonucleic acid or DNA) itself. We should, however, keep in mind that only a small fraction of the human genome is translated into protein. This raises the following **dilemma**: Which level of variability is really significant for the understanding of our evolutionary history? The answer to this question will depend on the kind of specific problem that is being considered (for instance, historical relationships or given selective pressures). Our research group is investing in the study of human variability at both the DNA and protein levels, and some examples will now be given.

## ALBUMINS

The albumin molecule presents a repeated pattern in helix form of nine loops joined by 17 disulfide bridges, with three homologous domains, each having 195 amino acid residues. Table I shows the population surveys we had performed up to 1989 in 34 South American populations, involving 14,650 individuals. As a result, 12 instances of

**Table I** - Studies in the albumin molecule performed by the Porto Alegre group in collaboration with other centers.**Surveys**

| Ethnic characterization               | Number of   |             |
|---------------------------------------|-------------|-------------|
|                                       | Populations | Individuals |
| Whites                                | 4           | 2110        |
| White/Black or White/<br>Indian/Black | 13          | 9615        |
| Indians                               | 17          | 2925        |

**Variants determined at the molecular level**

| Name  | Change        |
|---|---------------|
| Coari I, Porto Alegre I   | 358 GLU → LYS |
| Porto Alegre II, Manaus I, Fortaleza I<br>(Vancouver, Birmingham, Adana, Lambadi) | 501 GLU → LYS |
| Oriximiná I (Maku)  | 541 LYS → GLU |

Source: Arai *et al.*, unpublished

albumin variants had been detected. Some, however, could not be distinguished from each other by classical horizontal starch gel electrophoresis (Franco and Salzano, 1985). Studies at the molecular level performed by Arai *et al.* (1989) indicated that three types of mutations were involved (Table I). A recently detected variant, found in Fortaleza by Dr. Maria Z.P.J. Bezerra, was found to have the same molecular change as albumins Porto Alegre II and Manaus I (unpublished results of Dr. S.O. Brennan, Canterbury Health Laboratories, Christchurch, New Zealand) and is presently being submitted to binding studies with several substances by Dr. Ulrich Kragh-Hansen (University of Aarhus, Aarhus, Denmark).

While the first two mutations listed in Table I are probably of Caucasian origin, Oriximiná I (Maku) was found only among South American Indians or those of mixed Indian ancestry. Studies in groups that probably furnished the ancestors of the Amerindians may indicate whether the mutation occurred before or after the early colonization of the Americas.

## GLUCOSE-6-PHOSPHATE DEHYDROGENASE (G6PD)

G6PD is a dimeric protein, with each of the polypeptide chains having 495 amino acids. The population surveys we performed up to 1993 studying this enzyme are listed in Table II. They involved 28 South American populations and 7,794 individuals, and as a result we

detected nine variants in addition to the B, A, A<sup>-</sup>, and Mediterranean types. Four of them underwent DNA analysis, of which three were found to have been already reported elsewhere. One of them, however (G6PD São Borja) was observed for the first time in this locality of southern Brazil. It is interesting to note that the DNA of a male who consistently showed double electrophoretic bands, and that of a female with an also consistent slow variant, did not show revealing mutations, despite the fact that all the 13 exons have been examined for single-stranded conformation polymorphisms (SSCP) and the majority sequenced, for a total of 2,000 base pairs. Two other variants (provisionally named "Farroupilha" and "Lages") are currently under investigation.

The ethnic distribution of the variants is curious, since although the samples obtained among the Indians were larger than those from Whites or admixed persons, not a single variant was observed among these Indians. This locus therefore is at present uninformative about the question of Amerindian origins.

## "CLASSICAL" MARKERS

Our group has extensively studied South American Indians, using the "classical" techniques of immunotyping and electrophoresis. Most of the early studies, which began in 1958, were summarized in Salzano and Callegari-Jacques (1988). A recent multivariate analysis involving 29 Amazonian tribes and 17 systems (Callegari-Jacques *et al.*, 1994) disclosed a geographical dichotomy, the dividing line being the Amazon river. Two explanations can be presented for these findings: (a) the Amazon river would constitute a barrier to north-south gene flow; or (b) the differences would reflect past migrations that originated from distinct western regions at different latitudes.

Another study, also recently finished (Arnold *et al.*, 1995), refers to Factor XIII (transglutaminase), which acts in the final stage of the coagulation cascade. The protein occurs in tetrameric form (A<sub>2</sub>B<sub>2</sub>), the B unit being the carrier of Factor FXIII<sub>A</sub>. Using isoelectric focusing it is possible to distinguish four common alleles, responsible for different forms of the protein. The interesting finding is that the FXIII<sub>A</sub>\*1A allele, practically absent among the Macushi and Içana river Indians, is relatively common in Japanese (27%). Since the Asiatic origin of Amerindians is well recognized, this is unexpected.

## MITOCHONDRIAL DNA (mtDNA)

Many studies have recently been conducted in two hypervariable segments of the control region (D-loop) of this cytoplasmic organelle. The investigations we have performed in the first of these segments, based on individuals from four South American tribes, are summarized in Table III. Despite some variation (the Xavante showing consistently

**Table II** - Studies in the glucose-6-phosphate dehydrogenase (G6PD) molecule performed by the Porto Alegre group in collaboration with other centers.

| Ethnic characterization           | Number of   |             |         |
|-----------------------------------|-------------|-------------|---------|
|                                   | Populations | Individuals |         |
|                                   |             | Males       | Females |
| Whites                            | 1           | 1408        | 938     |
| White/Black or White/Indian/Black | 5           | 825         | 884     |
| Indians                           | 23          | 1776        | 1963    |

**Variants determined at the molecular level**

| Characteristics   | G6PD Variants |                   |  |              |
|-------------------|---------------|-------------------|--|--------------|
|                   | Mediterranean | Seattle           | Anaheim                                  | São Borja    |
| <b>Probands</b>   |               |                   |  |              |
| Sex               | Male          | Male              | Male                                     | Female       |
| Age               | 24            | 33                | 3  | 18           |
| Ethnic group      | White         | White             | White                                    | White        |
| Origin            | Russia        | Germany or Poland | Italy                                    | Portugal     |
| Clinical symptoms | Asymptomatic  | Asymptomatic      | Neonatal jaundice,<br>3 hospitalizations | Asymptomatic |
| <b>Molecule</b>   |               |                   |  |              |
| Specific change   | C → T, nt563  | G → C, nt844      | G → A, nt1178                            | G → A, nt337 |

Source: Weimer *et al.* (1993).**Table III** - Mitochondrial DNA sequence diversity considering the first 360 nucleotides of the control region in four South American tribes studied by the Porto Alegre group in collaboration with other centers.

| Characteristic                     | Tribes      |             |             |             |
|------------------------------------|-------------|-------------|-------------|-------------|
|                                    | Gavião      | Zoró        | Xavante     | Mapuche     |
| No. indiv. studied                 | 27          | 30          | 25          | 39          |
| No. different lineages             | 7           | 9           | 4           | 13          |
| 9 bp deletion                      | +           | +           | +           | +           |
| Gene diversity <sup>a</sup> ± S.E. | 0.87 ± 0.00 | 0.78 ± 0.14 | 0.68 ± 0.03 | 0.91 ± 0.00 |
| Nucleotide diversity ± S.E.        |             |             |             |             |
| $\pi^b$                            | 1.12 ± 0.32 | 1.12 ± 0.29 | 0.84 ± 0.28 | 1.53 ± 0.37 |
| E(s) <sup>c</sup>                  | 4.15 ± 1.04 | 4.30 ± 1.04 | 2.65 ± 0.84 | 4.97 ± 1.08 |

<sup>a</sup>Sequence heterozygosity.<sup>b</sup>Average of the differences between sequences ( $\times 10^2$ ).<sup>c</sup>Number of polymorphic (segregating) sites per nucleotide site.Source: Ginther *et al.* (1993); Ward *et al.* (in press).

lower values than the other groups), the indices of nucleotide diversity are not much different. The analyses of the Gavião, Zoró and Xavante are now being extended to the second segment. In addition, information about two other tribes (Surui, Wai Wai) are already at hand, and we have DNA isolated from five other tribes for additional investigations.

Global analyses including these and all the sequences available in data banks indicate that: (a) there should be much more than just four Amerindian founding lineage clusters, as has been proposed, for instance, by Schurr *et al.* (1990); and (b) entrance of the earliest colonizers in the continent occurred more than 15,000 years before present.

## NUCLEAR DNA

Hypervariable DNA polymorphisms are advantageous for the study of population relationships, because as the name indicates they have a large number of alleles, and consequently high heterozygosity per locus. One of such loci, commonly called the Variable Number of Tandem Repeat

(VNTR) locus, is D1S80, which owes its variation to the repetition of a 16-bp core sequence. Our studies of this locus in Porto Alegre and in two Brazilian Indian tribes are shown in Table IV. We found a striking reduction in the number of alleles present in these two tribes (5, 6) in relation to those found in Porto Alegre Whites and Blacks (18). Not surprisingly, the alleles most common in Amerindians are not the same as those found among non-Indians. We already have data available for three other Indian groups (Gavião, Surui and Wai Wai), and eventually it will be possible to establish specific tribal patterns for this locus.

Another nuclear DNA system of interest is the beta-globin gene cluster, detectable by the restriction fragment length polymorphism technique. Using a set of endonucleases it is possible to detect multiple haplotypes, and the results we obtained in five Brazilian Indian tribes are shown in Table V. The Xavante depart markedly from the four other groups; they have the highest number of haplotypes (5), and four of them have relatively high frequencies.

When this haplotypic diversity is quantified using the Gini-Simpson Index, and the average of the five

**Table IV** - D1S80 allele distribution observed in Porto Alegre and in two Brazilian Indian tribes.

| Characteristic                         | Populations         |                     |              |                 |
|--|---------------------|---------------------|--------------|-----------------|
|  | Porto Alegre Whites | Porto Alegre Blacks | Zoró Indians | Xavante Indians |
| No. chromosomes studied                | 150                 | 106                 | 50           | 50              |
| No. of alleles found                   | 18                  | 18                  | 5            | 6               |
| Alleles present in one population only | 16, 38, > 41        | 19, 20, 33          | none         | none            |
| Most common allele                     | 24                  | 24                  | 30           | 18              |
| Frequency (%)                          | 35                  | 29                  | 56           | 44              |

Source: Heidrich *et al.* (1995).

**Table V** - Beta-globin gene haplotype distribution observed in five Brazilian Indian tribes.

| Characteristic                          | Tribes  |       |        |      |         |
|---|---------|-------|--------|------|---------|
|   | Wai Wai | Surui | Gavião | Zoró | Xavante |
| Haplotypes <sup>a</sup>                 |         |       |        |      |         |
| 1 (----)                                | 0       | 0     | 2      | 0    | 0       |
| 2 (+----)                               | 60      | 82    | 88     | 93   | 60      |
| 3 (----+)                               | 0       | 0     | 2      | 0    | 0       |
| 5 (-+++)                                | 15      | 5     | 0      | 0    | 15      |
| 6 (-++++)                               | 18      | 11    | 3      | 7    | 18      |
| 7 (-+++)                                | 5       | 0     | 5      | 0    | 5       |
| 11 (---++)                              | 0       | 2     | 0      | 0    | 0       |
| 16 (-+---)                              | 2       | 0     | 0      | 0    | 2       |
| Gini-Simpson Index (x 10 <sup>2</sup> ) | 22      | 31    | 22     | 12   | 58      |

<sup>a</sup>Values in percentages.

Source: Bevilaqua *et al.* (in press).

populations compared with those obtained among Europeans and Asiatics, a marked decrease in variability is observed in the Amerindians as compared to Europeans (56%), but much less compared with Asiatics (8%). This suggests the absence of an important bottleneck effect in the early colonization of South America.

## CONCLUSIONS

1. There is more genetic variability in ours and other species than would have dreamt the science of some years ago.

2. When the sites of the mutations identified in our studies are compared with those in which other mutations have been found by other investigators, it is clear that their distribution is not random. There are *limitations* in the variability that can be generated, and variation at one level is not necessarily correlated with that which occurs at another level (as is exemplified by the Xavante low variability in the mtDNA, but high diversity in the beta-globin gene cluster).

3. For the understanding of problems like the peopling of the Americas, or those related to the maintenance and change dynamics of our genetic heritage, *interdisciplinary studies* are necessary. Our species is too complex, and research in just one area of knowledge is insufficient to explain the enigmas of its variation.

## ACKNOWLEDGMENTS

The studies reviewed herein have been performed in collaboration with many colleagues, inside and outside Brazil. Proper credit to them is given in the references cited; but I would like to take this opportunity to thank the members of our specific research group and of our Department in general for marvelous working conditions. Presently our studies are financed by Financiadora de Estudos e Projetos (FINEP) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

## RESUMO

Realizou-se uma revisão de estudos desenvolvidos tanto a nível de proteína como de DNA pelo grupo de Porto Alegre, em associação com diversos outros, brasileiros e do exterior. Os tópicos examinados foram: (a) a determinação a nível molecular de mutações nas moléculas da albumina e da glicose-6-fosfato desidrogenase, assim como a sua distribuição populacional; (b) investigações envolvendo marcadores clássicos; (c) a diversidade nucleotídica do DNA mitocondrial de quatro tribos indígenas sul-americanas; e (d) a prevalência populacional de alelos e haplótipos de dois sistemas de DNA nuclear: D1S80 e agrupamento do gene da beta-globina. A relação desses resultados com a questão do povoamento pré-histórico das Américas foi considerada de maneira sistemática.

## REFERENCES

- Arai, K., Huss, K., Madison, J., Putnam, F.W., Salzano, F.M., Franco, M.H.L.P., Santos, S.E.B. and Freitas, M.J.M.** (1989). Amino acid substitutions in albumin variants found in Brasil. *Proc. Nat. Acad. Sci., USA* 86: 1821-1825.
- Arnold, G., Kömpf, J. and Salzano, F.M.** Coagulation Factor XIIIa subtypes in Brazilian Indians. *Hum. Hered.* (in press).
- Bevilaqua, L.R.M., Mattevi, V.S., Ewald, G.M., Salzano, F.M., Coimbra, C.E.A., Jr., Santos, R.V. and Hutz, M.H.** Beta-globin gene cluster haplotype distribution in five Brazilian Indian tribes. *Am. J. Phys. Anthropol.* (in press).
- Callegari-Jacques, S.M., Salzano, F.M., Weimer, T.A., Hutz, M.H., Black, F.L., Santos, S.E.B., Guerreiro, J.F., Mestriner, M.A. and Pandey, J.P.** (1994). Further blood genetic studies on Amazonian diversity - data from four Indian groups. *Ann. Hum. Biol.* 21: 465-481.
- Cann, R.L.** (1994). mtDNA and native Americans: A southern perspective. *Am. J. Hum. Genet.* 55: 7-11.
- Franco, M.H.L.P. and Salzano, F.M.** (1985). A comparative study of albumin variants found in Brazil. *Hum. Hered.* 35:34-38.
- Ginther, C., Corach, D., Penacino, G.A., Rey, J.A., Carnese, F.R., Hutz, M.H., Anderson, A., Just, J., Salzano, F.M. and King, M.-C.** (1993). Genetic variation among the Mapuche Indians from the Patagonia region of Argentina: Mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. In: *DNA Fingerprinting: State of the Science* (Pena, S.D.J., Chakraborty, R., Epplen, J.T. and Jeffreys, A.J., eds.). Birkhäuser Verlag, Basel, pp. 211-219.
- Heidrich, E.M., Hutz, M.H., Salzano, F.M., Coimbra, C.E.A., Jr. and Santos, R.V.** (1995). D1S80 locus variability in three Brazilian ethnic groups. *Hum. Biol.* 67: 311-319.
- Hermann, B. and Hummel, S.** (1994). *Ancient DNA*. Springer Verlag, Berlin, pp. 263.
- Pericot y Garcia, L.** (1962). *América Indígena*, Tomo I. Salvat, Barcelona, pp. 1182.
- Salzano, F.M.** (1963). A origem do homem americano. *Rev. Antropol.*, São Paulo 11: 1-8.
- Salzano, F.M.** (1964). Genetics of South American Indians and the origin of American Man. *Annals, II Encontros Intelectuais de São Paulo, Instituto de Pré-História, Universidade de São Paulo, São Paulo*, pp. 371-379.
- Salzano, F.M.** (1976). Cultura, estrutura populacional e variabilidade genética em índios sul-americanos. *Interciência* 1: 155-158.
- Salzano, F.M.** (1985). The peopling of the Americas as viewed from South America. In: *Out of Asia: Peopling the Americas and the Pacific* (Kirk, R. and Szathmary, E., eds.). Journal of Pacific History Special Publication, Canberra, pp. 19-29.
- Salzano, F.M. and Callegari-Jacques, S.M.** (1988). *South American Indians. A Case Study in Evolution*. Clarendon Press, Oxford, pp. 259.
- Schurr, T.G., Ballinger, S.W., Gan, Y.Y., Hodge, J.A., Merriwether, D.A., Lawrence, D.N., Knowler, W.C., Weiss, K.M. and Wallace, D.C.** (1990). Amerindian mitochondrial DNAs have rare Asian mutations at high

frequencies, suggesting they derived from four primary maternal lineages. *Am. J. Hum. Genet.* 46: 613-623.

**Szathmary, E.J.E.** (1993). mtDNA and the peopling of the Americas. *Am. J. Hum. Genet.* 53: 793-799.

**Ward, R.H., Salzano, F.M., Bonatto, S.L., Hutz, M.H., Coimbra, C.E.A., Jr. and Santos, R.V.** Mitochondrial DNA polymorphism in three Brazilian Indian tribes. *Am. J. Hum. Biol.* (in press).

**Weimer, T.A., Salzano, F.M., Westwood, B. and Beutler, E.** (1993). Molecular characterization of glucose-6-phosphate dehydrogenase variants from Brazil. *Hum. Biol.* 65: 41-47.

**(Received October 9, 1995)**