FIFTY YEARS OF HUMAN GENETICS:

A FESTSCHRIFT AND *LIBER AMICORUM* TO CELEBRATE THE LIFE AND WORK OF GEORGE ROBERT FRASER

Edited by

Oliver Mayo and Carolyn Leach

Wakefield Press

Brazilians and the variable mosaic genome paradigm

Sergio D. J. Pena [1]

Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais,
Belo Horizonte, MG 31270–901, Brazil

Xenophanes speaks thus:
And no man knows distinctly anything,
And no man ever will.
Diogenes Laertius 1925

Populations and individuals

There is considerable medical interest in studying human genomic variation and its influence on health and individual drug responses. For that we especially need means to characterize and study the geographical distribution of human genetic diversity. The conventional approach to this question is to divide humanity into populations that can be defined on the basis of race, geography, culture, religion, physical appearance or whatever other criterion that is convenient. It is becoming increasingly clear that such division of humanity into populations may not constitute the most appropriate approach to deal with human variation. Treating people, for instance, of European ancestry and African ancestry, as separate categories for genetic studies tends to contribute to the public perception that the primary difference between these ways of defining populations is biological (Foster and Sharp 2004). This view confounds several issues and obscures the important fact that Europeans are genealogically related to Africans, having evolved as an offshoot of the latter.

Human evolutionary history is remarkably short and the worldwide geographical distribution of genetic traits is basically due to dispersal, with ensuing mutation, selection and genetic drift. In essence, the genetic diversity observable in Europe, Asia, Oceania and the Americas is a merely a subset of the variation found in Africa (Yu *et al.* 2002). As pointed out by Paabo (2003), from a genomic perspective we are all Africans, either living in Africa or in quite recent exile outside of Africa.

The human genome is composed of hundreds of thousands of genomic blocks of high linkage disequilibrium (Gabriel *et al.* 2002a) each one with its own pattern of variation and genealogy. This forms the basis of a 'Variable Mosaic Genome' (VMG) paradigm, according to which, rather than conceptualizing humans as belonging to populations, ethnicities or races, we consider the genome of any particular individual as a mosaic of variable haplotypes (Paabo 2003). This shifts completely the focus from populations to individuals. As we hope to demonstrate below, the VMG paradigm is especially useful when dealing with genome variation in highly admixed populations, such as Brazilians.

Of course, one cannot envisage the population view and the individual/genealogical view as exclusively antagonistic. Rather, they could be seen as complementary in the description of

human genome variation. An analogy with quantum mechanics could be used to illustrate this point: the complementarity principle states that light and electrons have a wave-particle duality. It is impossible to observe both the wave and particle aspects simultaneously. Together, however, they present a fuller description than either of the two taken alone. Most importantly, depending on the experimental arrangement, the light and electrons sometimes behave wavelike and sometimes particle-like, i.e., the model adopted predetermines the kind of properties that will be experimentally observed. Likewise, our scientific interpretations of genetic findings on human genome variation depend on whether we adopt a population or an individual/genealogical model. For instance, under the VMG paradigm certain ideas, such as that of human races, become absolutely meaningless. Also, we cannot overlook the fact that, as pointed out by Avise (2000), strong connections exist between demography and phylogeny. The historical demographies of populations are of profound relevance to phylogeographical patterns over microevolutionary time scales by virtue of their impact on the structure of haplotype genealogies.

The people of Brazil

Brazilians form one of the most heterogeneous populations in the world, the result of five centuries of interethnic crosses between peoples from three continents: Europeans, represented mainly by the Portuguese, Africans and Amerindians. When the Portuguese arrived in 1500, there were roughly 2.5 million indigenous people living in the area of what is now Brazil (Salzano and Freire-Maia 1970). The Portuguese-Amerindian admixture started soon after the arrival of the first colonizers. Mating between European men and indigenous women became commonplace and later (after 1755) was even encouraged as a strategy for population growth and colonial occupation of the country (Mörner 1967). The Amerindian tribes underwent a drastic demographic decline due to conflicts with the European colonizers and diseases to which they were not adapted (Salzano and Freire-Maia 1970). Today there are *circa* 720,000 Amerindians in Brazil, living on land set aside for them by the federal government. From the middle of the 16th century, Africans were brought to Brazil to work on sugarcane farms and, later, in the gold and diamond mines and on coffee plantations. Historical records suggest that between 1551 and 1850 (when the slave trade was abolished), 3.5 million Africans arrived in Brazil (Salzano and Freire-Maia 1970, Curtin 1969). As to the European immigration, it is estimated that *circa* 500,000 Portuguese arrived in the country between 1500 and 1808 (Salzano and Freire-Maia 1970). From then on, after the Brazilian ports were legally opened to all friendly nations, Brazil received increasing numbers of immigrants from several parts of the world. Portugal remained by far the most important source of migrants, followed by Italy, Spain, and Germany. In the 20th century, Asian immigration took place, mainly from Japan, as well as from Lebanon and Syria. According to Callegari-Jacques and Salzano (1999), 58% of the immigrants who arrived in Brazil between 1500 and 1972 were Europeans, 40% were Africans, and 2% were Asians.

In this sense, Brazil might be seen as representing a 'meeting point' for the three major historical geographical components of humanity [Africans, Asians (represented by their Native American descendants) and Europeans]. Of course, one might argue that the same 'meeting point' metaphor could be applied to the USA, as it also has the same three genealogical roots. However, as we will show, the extent of admixture between the three components has been significantly larger in Brazil.

Genetic variation in Brazilians

In the past few years we have been using several different molecular tools to try to characterize the ancestry of Brazilians and the formation of the Brazilian people. We will describe briefly these studies, from which we could unravel evidence of genetic admixture in levels much higher than had previously been suspected. We will start with studies based on uniparental markers because they can be useful to bring forth the concept of haplotype genealogies.

Uniparental genetic markers in Brazilians

There are several types of genetic markers at DNA level, and they can be classified according to their molecular nature and genomic localization. Autosomal markers are excellent individuality markers because they are diploid and subject to recombination. They can also be useful as ancestry-informative markers (AIMs) as long as the allele frequency difference between two ancestral populations is large (Shriver *et al.* 1997, Parra *et al.* 1998). On the other hand, uniparental maternal (mitochondrial DNA – mtDNA) and paternal (non-recombinant regions of the Y chromosome – NRY) polymorphisms are excellent lineage markers because they are haploid and do not recombine. As such, blocks of genes (haplotypes) transmitted to the next generations remain unaltered in the matrilineages and patrilineages until a mutation supervenes. The mutations that occurred and reached high frequencies after the dispersion of modern man from Africa can be specific to certain regions of the globe and can serve as geographical markers. The mitochondrial DNA and the NRY provide complementary information that can be traced back several generations into the past and allow reconstituting the history of a people through migrations of women and men respectively. It is however relevant to remember that a lineage marker such as mtDNA and NRY provides information about, respectively, a single female or male ancestor of an individual and thus constitutes a small proportion of the genetic constitution of a person.

We examined DNA polymorphisms in the non-recombining portion of the Y chromosome to investigate the contribution of distinct patrilineages to the present-day white Brazilian population. Twelve unique-event polymorphisms were typed in 200 unrelated males from four geographical regions of Brazil and in 93 Portuguese males (Carvalho-Silva *et al.* 2001). In our Brazilian sample, the vast majority of Y chromosomes proved to be of European origin. Only 2% of the Y-chromosome lineages were from sub-Saharan Africa (haplogroup E3a*), and none was Amerindian (haplogroup Q3*). Indeed, there were no significant differences when the haplogroup frequencies in Brazil and Portugal were compared by means of an exact test of population differentiation. Likewise, there was no population differentiation among the four geographical regions of Brazil. Nevertheless, by typing with fast evolving NRY markers we later could uncover a higher within population haplotype diversity in Brazil than in Portugal, explicable by the input of diverse European Y chromosomes (Carvalho-Silva *et al.* 2005).

We also studied a slightly larger sample, with the same basic constitution, for mtDNA. This revealed a very different 'reality' from that yielded by the NRY analysis. Considering Brazil as a whole, 33%, 28% and 39% of matrilineages were of Amerindian, African and European origin, respectively (Alves-Silva *et al.* 2000). As expected, the frequency of different regions reflected their genealogical histories: most matrilineal lineages in the Amazonian region had Amerindian origin, while African ancestrality was preponderant in the Northeast (44%) and the European haplogroups in the South (66%).

In summary, these phylogeographical studies with white Brazilians revealed that the vast

majority of patrilineages have European origin, while most matrilineages (>60%) were Amerindian or African. Together, these results configure a picture of strong directional mating between European males and Amerindian and African females, which agrees with the known history of the peopling of Brazil since 1500. Although having little to do with pharmacogenetics *per se*, these studies boldly reveal that the genomes of most white Brazilians are mosaic, having mtDNA and NRY with different phylogeographical origins.

Haplotype blocks and the 'Variable Mosaic Genome' paradigm

We have seen above the genetic properties of non-reticulate uniparental lineage markers. The same kinds of genealogical principles apply in theory to nuclear genes, whose multi-generation transmission routes involve both genders (Avise 2000). In terms of formal theory the major difference from uniparental markers is a four-fold adjustment required to account for the larger effective population size of autosomal alleles. This leads to corresponding four-fold longer coalescent times. A second, less important, difference is that in autosomes, besides mutations, lineages can change because of intragenic (or intrahaplotype, see below) recombination events.

In the past few years it has become evident that much of the human genome is composed of haplotypic blocks ('hapblocks') where polymorphic markers (especially single nucleotide polymorphisms – SNPs) are strongly associated over distances as large as 170 Kb (reviewed in Paabo 2003, Tishkoff and Verrelli 2003, Wall and Pritchard 2003). The discussion of the origin of these haplotype blocks is beyond the objective of this review. Suffice to say that probably the length of haplotype blocks is influenced by both demographic factors (which is certainly responsible for most of the variation of block sizes among populations) and genomic factors, especially the existence of recombination hot-spots (Zhang *et al.* 2003, Greenwood *et al.* 2004). The existence of such hapblocks has high significance for the feasibility of mapping disease genes by marker association studies, since each block can be defined by typing only 4–5 SNPs. Thus, the number of SNPs needed to achieve fine genomic screening might be reduced from millions to a few hundred thousand (Tishkoff and Verrelli 2003).

We can then envisage the human genome as composed of hundreds of thousands of small genomic blocks of high linkage disequilibrium (like the mtDNA or Y chromosome), each one with its own pattern of variation and genealogical origin. This forms the basis of the VMG paradigm, i.e., that rather than thinking about populations, ethnicities or races, we consider the genome of any particular individual as a mosaic of variable haplotypes (Paabo 2003). As we already saw this shifts completely the focus from populations to individuals.

Biparental genetic markers, morphological characteristics and ancestry in Brazilians

Using a panel of genetic polymorphisms that display large differences in allelic frequencies (>0.40) between Europeans and Africans, Parra *et al.* (1998) showed that, at a population level, it was possible to estimate with great precision the degree of African and European admixture among Americans. We decided to ascertain whether this same panel of markers would be capable to estimating, on an individual level, the degree of African ancestry in Brazilians. For that, we selected ten of the AIMs used in the American study (Parra *et al.* 2003).

With the purpose of verifying the individual discrimination power of this set of ten AIMs we genotyped a small sample of Portuguese individuals from the Northern part of the country and a sample of individuals from the island of São Tomé, located in the Gulf of Guinea, on the west

coast of Africa. These population sources were chosen because they are geographically related to the European and African population groups that participated in the peopling of Brazil. From the genotype information we calculated for each individual an African Ancestry Index (AAI) that is the logarithm of the sum over all alleles of the ratio of the likelihood of a given genotype being from African origin to the likelihood of it being of European origin. There was no overlap between the AAI values obtained for the two groups and 21 logs separated the two medians (9.71 and -11.73, respectively). A complete individual discrimination between the European and African genomes was obtained with an existence of 7.6 logs between the lower African value (AAI5 = 2.86) and the highest European score (AAI5 = 24.86). It was thus clear that the 10-allele set of Parra *et al.* (2003) was highly efficient and provided reliable individual discrimination between European and African genomes.

Our Brazilian sample was composed of 173 individuals from a Southeastern rural community, clinically classified according to their Colour (white, black, or intermediate) with a multivariate evaluation based on skin pigmentation in the medial part of the arm, hair colour and texture, and the shape of the nose and lips. When we compared the AAI values for these individuals, we observed that the groups had much wider ranges than those of Europeans and Africans and that there was very significant overlap between them. However, the comparison of whites versus blacks with the Mann-Whitney U test still showed a modest significant value ($z = 2.62$; $P < 0.01$). On the other hand, comparisons of whites versus Portuguese and blacks versus São Tomé islanders yielded extremely high significance (respectively, $z = 5.08$, $P < 0.0001$ and $z = 25.24$, $P < 0.0001$). Thus, the differences in AAI values of the group of Brazilian blacks compared with Brazilian whites are very distinct and several orders of magnitude smaller than the differences observed between Africans and Europeans. Other studies showed that these data could be reliably extrapolated to the rest of Brazil (Parra *et al.* 2003).

If we consider some peculiarities of Brazilian history and social structure, we can construct a model to explain why Colour should indeed be a poor predictor of African ancestry at an individual level. Most Africans have black skin, genetically determined by a very small number of genes that were evolutionarily selected. Thus, if we have a social race identification system based primarily on phenotype, such as occurs in Brazil, we classify individuals on the basis of the presence of certain alleles at a small number of genes that have impact on the physical appearance, while ignoring all of the rest of the genome. Assortative mating based on Colour, which has been shown by demographic studies to occur in Brazil, will produce strong associations among the individual components of Colour. Indeed, we detected the presence of such positive associations at highly significant levels in a Southeastern Brazilian population (Parra *et al.* 2003). On the other hand, we expect that any initial admixture association between Colour and the AIMs will inevitably decay over time because of genetic admixture. It is easy to see how this combination of social forces could produce a population with distinct Colour groups and yet with similar levels of African ancestry.

In essence, our data indicate that, in Brazil as a whole, Colour is a weak individual predictor of African ancestry. Based on the data generated by the above study, we used the *Structure* software (Pritchard *et al.* 2000) to estimate the number of people in Brazil with African ancestry. In this specific case, *Structure* estimated the proportion of European ancestry in each individual studied. The results, which can be seen in Table 1, are remarkable. More than 75% of self-declared White Brazilians from the North, Northeast or Southeast of the country, present

European ancestry below 90%. Even in Southern Brazil, with its history of strong European immigration, this value was of the order of 49%. For comparison, we calculated, from the data of Shriver *et al.* (2003), the values for European Americans: only 29% present European ancestry below 90%. On the other hand, only 73% of Brazilian Blacks have a predominantly African ancestry, i.e. a proportion of European ancestry below 50%, as compared with 91% in the U.S. (Table 1). Obviously these estimates have been achieved by extrapolating from experimental results with small samples and thus have very wide confidence limits. However, they illustrate cogently the high amount of genetic admixture that has taken place in the formation of the Brazilian people.

Table 1. Proportion of African ancestry in several groups in Brazil and in the United States. (Original data from Shriver *et al.* 2003 and Pena and Bortolini 2004)

	Whites (%)					**Blacks (%)**	
European	**Brazil**				**U.S.**	**Brazil**	**U.S.**
Ancestry	South	North	Northeast	Southeast			
>90%	51	24	20	11	71	3	1
>80%	71	43	36	21	92	7	2
>70%	84	53	46	36	99	13	3
>60%	90	73	64	53	100	20	4
>50%	96	82	74	74	100	27	9
>40%	100	92	82	81	100	37	17
>30%	100	98	94	89	100	60	29
>20%	100	98	97	96	100	80	48
>10%	100	100	100	98	100	87	72

Brazilians constitute a trihybrid population with European, African, and Amerindian roots. Thus, we had to ascertain where the Amerindians would be positioned in reference to the AAI scale. For that, we tested DNA samples from 10 individuals from three Amazonian tribes (Karitiana, Surui, and Ticuna) and observed that they fell in the same range as the Europeans. This finding is not unexpected, because several of the population-specific alleles in Parra's set had diverging frequencies in Europe and Africa because of specific selective factors operating in the African environment (Parra *et al.* 1998). Amerindians, who are well known to have a distant Asian ancestry, did not share such environments and would thus, be expected to have frequencies similar to Europeans. To estimate the Amerindian contribution to the Brazilian population we thus needed new polymorphic markers that would be sensitive to the three ancestralities. We then turned to a set of 40 insertion-deletion (indel) polymorphisms that proved to be exquisitely discriminating in that regard (Bastos-Rodrigues *et al.* 2006). Our results showed that White Brazilians have widely varying degrees of European, African and Amerindian ancestry. Even more variable were the results with Black Brazilians sampled in the city of São Paulo (Bastos-Rodrigues *et al.* 2006).

It is obvious then that, regardless of their skin colour, the overwhelming majority of Brazilians have a significant degree of African ancestry. Likewise it could be easily demonstrated that, regardless of their skin colour, the overwhelming majority of Brazilians have a significant degree

of European ancestry. Finally, although we have not calculated the exact numbers, we can safely predict that regardless of their skin colour, a sizeable proportion of Brazilians have a significant degree of Amerindian ancestry!

It thus makes no sense talking about 'populations' of 'White Brazilians' or 'Black Brazilians' because of the poor correlation between colour and ancestry. Also, it does not make sense talking about African-Brazilians or European-Brazilians because most Brazilians will have significant proportions of African and of European (and of Amerindian) ancestry. Thus, the only possible basis to deal with genetic variation in Brazilians is on a person-by-person basis, according with the VMG paradigm, which allows any individual to have different ancestries in different genomic segments.

Acknowledgements

Research in the authors' laboratory was sponsored by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) of Brazil.

Notes

1. This paper is dedicated to George Fraser on the occasion of his 75th birthday, with admiration and gratitute for his important contributions to human genetics.