Age, Area, and Acheiropody

N. E. Morton and C. A. A. Barbosa

Population Genetics Laboratory, University of Hawaii at Manoa, 1980 East-West Road, Honolulu, HI 96822, USA

Summary. Analysis of the geographic distribution of acheiropody suggests spread from São Paulo along the São Francisco valley at a rate consistent with current parent-offspring distribution and 20 generations of gene flow.

Introduction

Many rare genes have a localized distribution which suggests that most copies in a restricted region are descended from a single founder. The number of descendants poses grave mathematical problems, being dependent on the unknown distribution of family size in successive generations. The dispersion of descendants is much more tractable, and gives indirect information about the time since the founder lived, his place of residence, and the rate of migration. Here we develop the theory and apply it to acheiropody, a characteristic Brazilian disease due to a unique recessive mutation. The principles are equally applicable to Tay-Sachs disease in Ashkenazi Jews and other spatial clusters.

Material and Methods

The data on 22 sibships were provided by Freire-Maia et al. (1975b), who sent letters to mayors, priests, and physicians throughout Brazil. All affected were born in six states, with not a single case from a large city. The number of acheiropody genes was estimated to be 18,000 in 1960 (Freire-Maia et al. 1975a).

In Table 1 the 22 sibships are referred to their birth places. The control population for each state is taken as the rural 1960 census (in thousands, to bring ascertained frequencies close to carrier frequencies), assigned to the capital if there were no sibships in the state, and distributed equally among birth places of sibships otherwise. This is the closest we could get to the ideal that data form an r x n table, where the rows corresponds to different degrees of relationship, and the columns are (for determinancy) the minimum number of locations to specify carriers and controls without loss of required detail. Geographic coordinates in km were computed on the Hayford spheroid following the method of Morton et al. (1969), with Pirapora as the origin. These coordinates are denoted X, Y. The coefficient of relationship to a carrier is R. Sample size at a given set of X, Y is Nxy.

Geographic and social barriers are hard to model and have no general interest. We therefore assume isotropic migration, which we believe to be realistic in this context, where migration is small relative to the area studied. Invoking the central limit theorem, we take the distribution of carrier descendants after several generations to be proportional to

\[ C_{xy} = e^{-((X-E)^2 + (Y-F)^2)/2C^2} \]

where E, F are the coordinates of the founder’s residence and C is the standard deviation of dispersion around that residence (Sawyer 1976). Ascertainment is assumed to be uniform.
Table 2. Tests of hypotheses. Assumed parameters in parentheses

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>$d \times 10^2$</th>
<th>$E$</th>
<th>$F$</th>
<th>$k \times 10^2$</th>
<th>$C$</th>
<th>$\chi^2 + \text{const.}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No clustering</td>
<td>0.595</td>
<td></td>
<td></td>
<td>(0)</td>
<td></td>
<td>370.81</td>
</tr>
<tr>
<td>$d = 0$</td>
<td>(0)</td>
<td>79</td>
<td>-446</td>
<td>1.246</td>
<td>823</td>
<td>357.84</td>
</tr>
<tr>
<td>$E, F \sim$ Jacobina</td>
<td>0.578</td>
<td>(475)</td>
<td>(677)</td>
<td>0.222</td>
<td>139</td>
<td>370.68</td>
</tr>
<tr>
<td>$E, F \sim$ mean</td>
<td>0.593</td>
<td>(-57)</td>
<td>(-251)</td>
<td>0.298</td>
<td>105</td>
<td>370.81</td>
</tr>
<tr>
<td>General</td>
<td>0.270</td>
<td>-293</td>
<td>-643</td>
<td>3.395</td>
<td>196</td>
<td>345.72</td>
</tr>
<tr>
<td>Standard error</td>
<td>±0.097</td>
<td>±91</td>
<td>±111</td>
<td>±1.282</td>
<td>±48</td>
<td></td>
</tr>
</tbody>
</table>

The frequency of ascertained carriers at coordinates $X, Y$ is

$$P_{XY} = d + kC_{XY}$$

where $d$ is the frequency of ascertained carriers not due to the founder, assumed to be constant within the region, and $k$ is the frequency at the center of dispersion of ascertained carriers who received the gene by descent from the founder. The maximum frequency of ascertained carriers is $d + k$. The frequency of apparent noncarriers at coordinates $X, Y$ is

$$P_{XY0} = 1 - P_{XY1} = 1 - d - kC_{XY}$$

Relatives of known carriers have fractional values of $R$, such as 1/2 for the random parent of a carrier. Their frequency is

$$P_{XYR} = RP_{XY1} + (1 - R)P_{XY0}$$

$$= R(d + kC_{XY}) + (1 - R)(1 - d - kC_{XY})$$

This theory will be used most effectively if birth places of ancestors of carriers are ascertained, but typically only the birth places of carriers (or, less informatively their children) are reported. Consanguineous marriages permit recognition of carrier ancestors, who have reduced dispersion around the founder. Extension of the theory should consider nonrandom ascertainment, constant for a given part of the region, and estimable from numbers of ascertainment per proband or probands per sibship. If different values of $R$ greater than 0 are treated simultaneously, the number of classified relatives of a particular degree per carrier must be considered.

If migration is homogeneous in time, the expected value of the variance is

$$C^2 = t \sigma^2$$

where $\sigma^2$ is the variance of parent-offspring distance in the plane, and $t$ is the number of generations back to the founder. Otherwise $\sigma^2$ is an average value over generations. For any known value of $t$, the estimate of the standard deviation of migration is

$$\sigma = C\sqrt{t} \pm SE(C)\sqrt{t}$$

where $SE(C)$ is the standard error of $C$.

Estimates of the parameters $C, E, F, d,$ and $k$ are obtained by a computer program CENDIS (Morton 1980), which uses a procedure for optimization of general nonlinear functions (Lalouel 1979). As the function to be minimized we took

$$f = -\Sigma N_{XYR} \ln P_{XYR}$$

In the theory of large samples $2f + \text{const}$ has a $\chi^2$ distribution when $P_{XYR}$ is true.

Discussion

Assuming that ascertainment is uniform (i.e., no greater in the state of São Paulo than in the Northeast), our results indicate that the center of origin is the Planalto Paulista, the cradle of the frontiersmen in the old Capitanía of São Vicente. Their founders were the mameluco community of Santo André da Borda do Campo, the union of João Ramalho and a handful of Portuguese adventurers with Tupi-Guarani women of the Guianá tribe (Ellis 1944; Morse 1965). A principal founder of Santo André was Tibiriçá, the chief of Inhapuambuçu, three of whose daughters had large progenies with Portuguese men. Their descendants intermarried with later immigrants and scoured Brazil for Indian slaves, precious stones and metals, and grazing land. To the north the São Francisco valley was an open thoroughfare. Bands of hundreds of men spent years in the backlands. Expeditions were sent against the Dutch and Indians of the Northeast. As late as 1694 a band of 400 Paulistas under Domingos Jorge Velho was roaming the backlands of Pernambuco and Alagoas, where they were recruited to conquer the "republic" of runaway Negro slaves known as Palmares. Domingos Jorge Velho, according to the Bishop of Pernambuco, was attended by several Indian concubines. It seems likely that the gene for acheiropody was introduced into the São Francisco valley by the energy of such men. Since congenital malformations were not noted in vital records, it is perhaps impossible now to infer whether the gene originated in a Tupi Indian, such as Tibiriçá, or a Portuguese such as João Ramalho. However, an African origin is unlikely, since Negro slaves first appeared in São Paulo in the late sixteenth century, and along the São Francisco were vastly outnumbered by...