An Analysis of the ABO Gene Frequency Cline in Japan : A Migration Model *

Kenichi AOKI** and Keiichi OMOTO

*Department of Anthropology, Faculty of Science,
The University of Tokyo, Tokyo*

Abstract The well known cline in ABO blood group gene frequencies in Japan was analysed based on the working hypothesis of a transient cline formed by the admixture of two heterogeneous populations by diffusion. Mathematically, the model corresponds to solving the diffusion equation with a step function as the initial condition. Using the A allele frequency data, the position of the initial discontinuity in the gene frequency is estimated by simulation as having been in the vicinity of present day Gifu City, and again assuming 20 km per generation of 25 years for a standard deviation of migratory distance the time as approximately 2,000 years since the commencement of admixture. This and other results of simulation were discussed in the light of possible population movements induced by rice crop farming during the Yayoi era in western Japan.

INTRODUCTION

It has been noted that the distribution of the ABO blood group gene frequencies in the Japanese Islands is non-random and that a feeble but well-defined cline in the frequency of the A allele in particular exists along its length (TANAKA 1959, FURUHATA *et al* 1966). NEI and IMAIZUMI (1966) showed that this regional variation was compatible with random differentiation, but were inclined to the view that the cline was caused by regional differences in selection, suggesting that tem-

perature differences might be the environmental factor involved. CAVALLI-SFORZA and BODMER (1971) have estimated the order of magnitude of the selection coefficient which would be involved in maintaining such a cline, using the equilibrium models of HALDANE (1948) and of FISHER (1950), and are of the opinion that perhaps it is too small in this case for consistency.

A cline (or geographical gradient) in the gene frequencies is a striking example of genetic variation. An analysis of the mechanism(s) involved in the formation of an actually observed cline may provide insight into the more general problem of the maintenance of genetic variability in natural populations. Deterministically speaking, a cline can be of two types: it may be an equilibrium cline such as

*) This study was aided by the grant of Ministry of Education, Science and Culture.

**) Present address: Department of Genetics, The University of Wisconsin, Madison, Wisconsin, USA

Fig.1. The gene frequency of allele O (top), allele A (middle), allele B (bottom) of the ABO blood group system plotted against the distance measured along the cubic regression curve fitted to Japan. The point on the curve corresponding to Kagoshima is taken as the origin. Distances are measured in units of approximately 111km.

is maintained by the balanced forces of selection and diffusion (HALDANE 1948, FISHER 1950, KIMURA 1960), or it may be a transient cline. Transient clines may be formed by many diverse causes.

In this paper, we wish to analyse the possibility that the cline observed in the ABO blood group system is a transient cline due to the admixture of two heterogeneous populations by diffusion (i.e., short range migration that is random in direction). The work was done in the period 1974-1976 and constituted a substantial part of the first author's M. Sc. thesis. The publication of *JIBP Synthesis: Human Adaptability Vol. 2, Anthropological and Genetic Studies on the Japanese* in 1975 provided the opportunity for this study. The new data on ABO gene frequencies recently collected by FUJITA, TANIMURA and TANAKA (1978) could not be included in this study, since it was published after the completion of this study.

MATERIALS

Data both published and unpublished on the gene frequencies of various genetic systems in Japan have been brought together in a compact volume, *JIBP Synthesis: Human Adaptability Vol. 2, Anthropological and Genetic Studies on the Japanese*, edited by S. WATANABE, S. KONDO, and E. MATSUNAGA (University of Tokyo Press, Tokyo, 1975). The data on the ABO blood group system, the majority of which were sampled before 1959 (see TANAKA 1959), are especially numerous.

We restricted our attention to data

sampled from the larger communities of the three major islands, i.e., Kyushu, Shikoku, and Honshu. Hokkaido, the colonization of which was a relatively recent event, differs both historically and genetically from the rest of Japan, and therefore requires to be treated separately. Also, although gene frequency data from the smaller islands, particularly Okinawa, are available, they were not used, for the present analysis is intended to be strictly deterministic. The sample size ranged from about 100 to more than 10,000. Data showing significant deviation from the Hardy-Weinberg proportions were disregarded. This left us with usable data from 42 localities.

The clines exhibited by each allele of the ABO system are plotted against the distance in Figure 1.

METHODS

1. Migration model

Along a linear habitat, we assume an initial discontinuity in the geographical distribution of the gene frequencies represented by the step function

$$u(x) = \begin{cases} p_1 & \text{for } x \leq c \\ p_2 & \text{for } x > c \end{cases} \dots\dots\dots (1)$$

where x measures the distance and c is the coordinate of the discontinuity. After a finite time interval a transient cline is induced by the action of diffusion alone. (See Figure 2 below.)

Mathematically, this corresponds to solving the diffusion equation $\frac{\partial u}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 u}{\partial x^2}$ with equation (1) as the initial condition. The solution representing the gene frequency $u(x, t)$ as a function of the dis-

Fig. 2. Schematic representation of the migration model. See text for explanation.

tance x and the time t (the curve in Fig. 2) can be expressed as

$$u(x, t) = p_1 + \Delta p \Phi\left(\frac{x-c}{\sigma\sqrt{t}}\right) \dots\dots\dots (2)$$

where $\Delta p = p_2 - p_1$, σ is the standard deviation of the migration distance per generation, and $\Phi(x) = 1/\sqrt{2\pi} \int_{-\infty}^x e^{-y^2/2} dy$

2. The geographical description

The shape of Japan (unless otherwise specified, Japan refers to the three major islands of Kyushu, Shikoku, and Honshu) is essentially linear with a maximum breadth about one-eighth of its length. In this paper, we analyse clines along the length of Japan using a one-dimensional model. The assumption is made that conditions are uniform in the transverse direction, which is not unreasonable if the ratio of width to σ (the standard deviation of migration distance per generation) is not excessively large.

Outlined below is the method used to reduce Japan to a linear habitat. The 41 prefectural capitals of Honshu and Kyu-

shu were taken as the representative points, from both a geographical and demographic viewpoint. To mitigate the distortion of distance that occurs with a change of latitude, a simple transformation of the longitude ξ and the latitude η coordinates of these 41 localities was effected

$$y = \eta - \bar{\eta} \\ x = (\xi - \bar{\xi} \cos \eta / \cos \bar{\eta}) \cos \bar{\eta} \dots\dots\dots (3)$$

The bar over the variables stands for taking the mean of 41 values. The origin of the new coordinate system becomes (136.27E, 35.55N). Incidentally, one unit is approximately equal to 111km.

A polynomial regression curve was then fitted to the 41 transformed coordinates. In fitting a cubic regression curve, the coefficient of the highest order (i.e., cubic) term was significant at the 5 percent level (a one-sided t test). However, the fourth and fifth degree polynomials did not yield significant highest order coefficients, and it would have been meaningless to pursue

Fig. 3. The cubic regression curve fitted to a map of Japan. See text for details. The formula is given below. Circles indicate prefectural capitals; the full circles in particular indicate centres of sampling for the ABO blood groups. $Y = 0.0236X^3 + 0.0753X^2 + 0.2941X - 0.3678$

higher order regression. The cubic regression curve obtained by this method is superimposed on a map of Japan in Figure 3.

In the actual analysis of cline, the position of a locality for which gene frequency data are available was specified by its longitude and latitude coordinates. These coordinates were transformed by formula (3) and then projected onto the nearest point on the cubic regression curve. Distances were then measured along the curve by numerical integration using Simpson's one-third rule. The point on the curve corresponding to Kagoshima was taken as the origin.

In order to treat Japan as a linear habitat, NEI and IMAIZUMI (1966) used the shortest railroad distances measured from Aomori as the parameter designating position. The correlation between the

two types of distance is very high (.988). An advantage of the distance measured along the cubic regression curve is that it is more nearly equal to geodesic distance, permitting a more or less direct comparison with the work of AMMERMAN and CAVALLI-SFORZA (1971a). (See footnote 4 following DISCUSSION.)

RESULTS

The migration model was fitted to the clines exhibited by each allele of the ABO system by interpreting equation (2) as a linear regression equation of the gene frequency u on the transformed distance Φ (see Figure 4). The two parameters in the argument of the function were assumed to take the values $c = 8.03$ and $\sigma\sqrt{t} = 1.6$; the justification of these values will be discussed later. In each case, the regression was significant; at the 1% level for A and O, and at the 5% level for B (a two-sided t test).

DISCUSSION

Equation (2) of the migration model mentioned above was fitted to the data on the ABO system by choosing the four parameters p_1 , Δp , c , and $\sigma\sqrt{t}$ so as to minimize the residual variance. This was accomplished by a simple simulation with c and $\sigma\sqrt{t}$ as the running variables. The parameters c and $\sigma\sqrt{t}$ were run with intervals of .1 and .05 respectively. The estimates of the four parameters in the case of allele A are

Fig.4. Fit of the migration model. Abscissa: the distance transformed by the cumulative standardized normal distribution function with mean 8.03 and standard deviation 1.6. Ordinate: gene frequency of allele O (top), allele A (middle), allele B (bottom).

$$\begin{aligned}
 p_1 &= .288 \pm .002 \\
 4p &= -.031 \pm .004 \quad \dots\dots\dots(4) \\
 c &= 8.03 \\
 \sigma\sqrt{t} &= 1.6
 \end{aligned}$$

the residual variance being $.903 \times 10^{-4}$. The behaviour of the residual variance in the neighbourhood of the minimum is given in Table 1. There is, however, no assurance that other minima do not exist. If the migration model is strictly applicable, we should expect similar estimates of c and $\sigma\sqrt{t}$ when the remaining alleles are considered. For allele O, we have $c=8.43$ and $\sigma\sqrt{t}=1.8$; for allele B, we have $c=7.63$ and $\sigma\sqrt{t}=1.2$. However, the standard errors of c and $\sigma\sqrt{t}$ cannot be obtained by this method so that it is difficult to decide whether the estimates are sufficiently alike or not. The allele A with the most prominent cline would seem to provide the most reliable estimates of c and $\sigma\sqrt{t}$; if we accept these estimates and evaluate the regression coefficient for each allele, they are significant as was previously mentioned.

Table 1 The residual variance ($\times 10^4$).

c $\sigma\sqrt{t}$	7.73	7.83	7.93	8.03	8.13	8.23	8.33
1.45	.945	.926	.915	.913	.955	.943	.940
1.50	.940	.920	.909	.907	.949	.938	.936
1.55	.937	.917	.906	.904	.946	.936	.935
1.60	.937	.917	.905	.903	.946	.936	.936
1.65	.939	.919	.907	.905	.948	.939	.938
1.70	.944	.924	.912	.909	.952	.944	.944
1.75	.951	.931	.919	.916	.959	.951	.951

Figure 4 suggests that the plots of the frequency of allele A are consistently below the regression line in the interval $.2 < \Phi < .5$ and consistently above it for

$.5 < \Phi < .8$. This is attributable to the fact that the actual cline is more linear than is predicted by the model. The discrepancy can be reduced if we assume a less abrupt form for the initial gene frequency distribution.

It is of some interest to consider if the migration model is compatible with the archaeological and historical evidence. If we assume $\sigma=20$ km, per generation time of 25 years (see footnote 4), the fourth equation of (4) gives $t=1971$ years since the commencement of admixture. Moreover, the position of the initial discontinuity in the gene frequency is estimated to have been in the vicinity of present day Gifu City ($c=8.03$) (see footnote 1). Although these estimates are not to be taken very seriously, they would seem to suggest that the cline observed in the ABO system reflects an admixture of heterogeneous populations which began during the Yayoi Era.

The Yayoi Era is believed to have been a period of transition. An important feature of this era was the introduction of rice crop farming by an immigrant body of farmers. There is no consensus among the archaeologists as to the number that actually arrived; however, the remarkably rapid expansion of rice crop farming (within a matter of several decades) from its origin in northern Kyushu as far as the Kansai area suggests that their numbers may have been quite substantial (SAHARA 1975). In effect, this point of view assumes that quite a number of experienced farmers actually settled throughout western Japan at much the same time

(see footnote 2). The migration model above would seem to be applicable in this situation if the alien population of farmers possessed a gene frequency differing from that of the indigenous inhabitants. The difference in the gene frequencies need not have been tremendously large, if the alien population of farmers initially enjoyed a reproductive advantage due to their relatively advanced technology (see footnote 3).

Of course, the initial condition postulated above as part of the model (a step in the gene frequencies) is an over-simplification; we might, for instance, expect a quasi-linear cline to have formed temporarily as the immigrant farmers quickly moved from northern Kyushu toward the Kansai area in a search for suitable land. (It was seen that such a hypothesis would improve the fit.)

There is thus reason to believe that events occurring during the Yayoi Era may have been responsible for initiating the formation of the presently observed cline. Let us restrict our attention to the Yayoi Era and consider a perhaps more realistic variation of the migration model. This new model is a simulation of the spread of early farming by two means: the actual geographical displacement of farmers, and/or the transmission by learning of the knowledge and skills necessary for farming (see footnote 4). It is a simplification (and we hope an improvement) of the model constructed by SGARAMELLA-ZONTA and CAVALLI-SFORZA (1973) for the detection of cline caused by the spread of early farming in Europe. It

assumes that the initial distribution of the alien farmers was restricted to the northern Kyushu area, and that the propagation of the rice crop farming technology was mainly due to the diffusion of a cultural stimulus. A population growth occurs in the farming communities which in conjunction with a modest amount of local migration causes a diffusion of the alien genes in an easterly direction. The simulation is conducted to see whether the differential population growth of the alien farmers was sufficient to induce a cline.

We represent the relevant area of the Japanese Islands (northern Kyushu to northernmost Honshu) by a one-dimensional stepping stone model (KIMURA 1960) with a finite array of colonies twenty in number: the colonies are numbered in order from the left. We assume that each colony initially consists of 700 members, which gives a total population size of 14,000 (see footnote 5). Of these, the first colony is an immigrant population of farmers while the remainder are inhabited by the indigenous hunter-gatherers. The gene frequency of a particular allele (which we take to correspond with the A allele of the ABO system) is assigned the value .30 in the immigrant population and .25 in the 19 indigenous colonies. We assume that any farming group (the size of this group need not be a whole number) grows logistically with an intrinsic rate of increase of .04 per year, and eventually attains an equilibrium population size of $700 \times 200 = 140,000$. (This value of increase permits a doubling about

every 18 years (see footnote 6). The equilibrium population size corresponds to a final density of about 1000 persons per 100 square kilometers (see footnote 5). We also assume that any hunting-gathering group experiences no growth. The migration rate between adjacent colonies is taken as .003 per year. (This value corresponds to the standard deviation of migration distance $\sigma=20$ km. per generation of 25 years.) Population growth and migration are taken to occur in that order. We define the acculturation factor f as the minimum fraction of farmers in a colony necessary for the conversion of all its members to farming. This factor determines the rate of cultural diffusion: smaller values correspond to more rapid diffusion. For the first 10 colonies we take $f=.01$, corresponding to the view that the initial propagation was extremely rapid; for the remaining 10 colonies we take $f=.1$. The position of the front is defined as the number of the colony in which all the members have most recently been converted to farming.

The results of the simulation are given in Figure 5 for year number 0, number 50, number 300, and number 2200. It can be seen that the successive positions of the front are in good agreement with the archaeological evidence. Moreover, the simulation demonstrates that by the year 2200 (assumed to correspond with the present) the initial genetic difference has been sharply reduced. As we are making deterministic calculations over a finite time period, however, a small cline remains extending from the first colony to

the eighth with its steepest gradient between the second to the fifth. We think this simulation suggests that if the conditions postulated in this model are true, the alien farmers would not have left an appreciable effect on the gene pool of the Japanese: their reproductive advantage would have lasted for too short a time.

ACKNOWLEDGEMENTS

We are deeply indebted to Dr. N. YASUDA of the National Institute of Radiological Sciences for helpful criticism and the generous giving of his valuable time to stimulating discussions. We also thank Mr. T. UEKI and Mr. K. IMAMURA for stimulating discussions on the archaeological aspects of the problem; needless to say the authors are totally responsible for any factual errors.

FOOTNOTES

- 1) A deviation of one interval in c corresponds to 11.1km.; a decrease of one interval in $\sigma/\sqrt{t}$ gives $t=1850$ years, whereas a corresponding increase gives $t=2096$ years.
- 2) An alternative view to be developed in the following pages is that the presence of an incipient farming among some indigenous populations of western Japan facilitated the rapid transition to a rice crop farming economy (IZUMI 1968). According to this view, the immigrant body need not have been large.
- 3) That the introduction of rice crop farming permitted a rapid population growth is indicated by an increase in the number and distribution of sites from the end of the early phase onwards. The population growth was resolved by the budding off of new settlements rather than by contained growth within each village; it was only in the late

phase that large settlements began to appear (KONDO 1966).

- 4) AMMERMAN and CAVALLI-SFORZA (1971a, 1971b) demonstrated that the spread of early farming in Europe could have occurred by the actual displacements of the farmers rather than by the diffusion of a cultural stimulus. Their major argument is that the radiocarbon datings of the earliest arrival time in each locale show a remarkably high linear correlation with the geodesic distance measured from a hypothetical diffusion centre in the middle-east. This is to be expected from the theoretical calculations of SKELLAM (1951) and KENDALL (1948) who showed that when short-range migration random in direction occurs concurrently with an exponential population increase, a geographical expansion sets in proceeding at a constant radial rate proportional to the geometric mean of the growth rate and the variance of the migratory distance. AMMERMAN and CAVALLI-SFORZA's estimate of the rate of advance is approximately 1 km. per year, from which the standard deviation of the migration distance per generation time of 25 years is estimated as 17.9 km. (We took 20 km. as a conveniently round number.)

In neolithic Japan, the Yayoi culture is believed to have spread from northern Kyushu to northern Honshu, a distance of nearly 1500 kilometres in about 300 years. Moreover, it is widely held that in the first half century, at most, it went almost halfway. The situation is apparently different from that of Europe. Furthermore, we probably cannot use the carbon 14 datings (even if they existed in sufficient numbers for a statistical analysis) for the standard error attached to these measurements is rather large (at least a hundred years) relative to the total propagation time.

- 5) BRAIDWOOD and REED (1957) have hazarded estimates of population densities at various stages of economy. The more spe-

cialized food-collecting economy of late glacial and early post-glacial times is estimated to have sustained a mean density of about 5 persons per 100 square kilometres. This value is subject to much variance with an economy based on salmon fishing in particular permitting a concentration 20 times as high. The primary village-farming economy has been estimated as sustaining a population density of 1,000 persons per 100 square kilometres, and the primary urban community stage (e. g., Sumer) as twice this value.

In the absence of estimates for the Japanese Islands, we take 5 persons per 100 square kilometres as the population density of the terminal Jomon, and 1,000 persons per 100 square kilometres as that of the Yayoi Era after the prominent population expansion permitted by rice crop farming. These densities correspond to absolute numbers of 14,000 and 2,800,000 respectively.

There is evidence that these values are not altogether irrelevant. SUDA (1952) advanced an estimate of population density for the Heian Era (circa 900 A.D.) which is chronologically well after the establishment of a central government; his estimate of the total population size based on an official record of rice yield for each province is about 5,470,000 persons for the three major islands of Kyushu, Shikoku, and Honshu, which gives an average density of just under 2,000 persons per 100 square kilometres. Of course, there is much variability between the provinces, and cultural and political centres exhibit extremely high densities, but as an overall estimate, it is in remarkably good agreement with BRAIDWOOD and REED's (1957) estimate for the primary urban community level of development.

- 6) BIRDSELL (1957) has observed that the growth curve of a specialized hunting-gathering or horticultural community which has newly occupied "empty space" is essentially logistic in form with an intrinsic

rate of increase which permits a doubling in approximately 20 years. In this respect, the backmarshes created by the recession of the sea could have provided an unoccupied ecological niche for the early farmers.

REFERENCES

- AMMERMAN, A. J. and CAVALLI-SFORZA, L. L., 1971a: Measuring the rate of spread of early farming in Europe. *Man* 6: 674-688.
- AMMERMAN, A. J. and CAVALLI-SFORZA, L. L., 1971b: A population model for the diffusion of early farming in Europe. *Proc. Res. Sem. Arch. Rel. Subj. Sheffield*, Dec. 1971.
- BIRDSELL, J. B., 1957: Some population problems involving Pleistocene man. *Cold Spring Harbor Symp. Quant. Biol.* 22: 47-69.
- BRAIDWOOD, R. J. and REED, C. A., 1957: The achievement and early consequences of food-production: a consideration of the archaeological and natural-historical evidence. *Cold Spring Harbor Symp. Quant. Biol.* 22: 19-31.
- CAVALLI-SFORZA, L. L. and BODMER, W. F., 1971: *The Genetics of Human Populations*. W. H. Freeman and Co., San Francisco.
- FISHER, R. A., 1950: Gene frequencies in a cline determined by selection and diffusion. *Biometrics* 6: 353-361.
- FUJITA, T., TANIMURA, M. and TANAKA, K., 1978: The distribution of the ABO blood groups in Japan. *Jap. J. Hum. Genet.* 23: 63-109.
- FURUHATA, T., ISEKI, S., and KITAHAMA, M., 1966: Japanese viewed from the standpoint of blood groups. *The Eleventh Pacific Congress Symp. No.* 45.
- HALDANE, J. B. S., 1948: The theory of a cline. *J. Genet.* 48: 277-284.
- IZUMI, S., 1968: in *symp.: The Origins of Japanese Farming Culture*. (ed. IZUMI, S.) Kadogawa Shoten, Tokyo. (in Japanese)
- KENDALL, D. G., 1948: A form of wave propagation associated with the equation of heat conduction. *Proc. Cambr. Phil. Soc.* 44: 591.
- KIMURA, M., 1960: *Outline of Population Genetics*. Baifukan, Tokyo. (in Japanese)
- KONDO, Y., 1966: The development of the Yayoi culture and the change in social relations. *The Archaeology of Japan III Yayoi Era*. (ed. WAJIMA, S.) Kawade Syoten, Tokyo. (in Japanese)
- NEL, M. and IMAIZUMI, Y., 1966: Genetic structure of human populations I. Local differentiation of blood group gene frequencies in Japan. *Heredity* 21: 9-35.
- SAHARA, M., 1975: The commencement of agriculture and the formation of a class society. in *Japanese History 1, Primitive and Classical Eras 1*. Iwanami Shoten, Tokyo. (in Japanese)
- SGARAMELLA-ZONTA, L. and CAVALLI-SFORZA, L. L., 1973: A method for the detection of a demic cline. in *Genetic Structure of Populations, Population Genetics Monographs, Vol. III* (ed. MORTON, N. E.) Univ. Press of Hawaii.
- SKELLAM, J. G., 1951: Random dispersal in theoretical populations. *Biometrika* 38: 196-218.
- SUDA, A., 1952: The transition of population densities and the movement of cultural centres in our country. in *The Japanese Race*. ed. The Anthropological Soc. of Nippon. Iwanami Shoten, Tokyo. (in Japanese)
- TANAKA, T., 1959: A study on the Japanese from the standpoint of the blood groups. *Hanzaigakuzasshi* 25 (1) Supplement: 37-67 (in Japanese with English summary)
- WATANABE, S., KONDO, S., and MATSUNAGA, E., (editors) 1975: *JIBP Synthesis: Human Adaptability Vol. 2. Anthropological and Genetic Studies on the Japanese*. University of Tokyo Press, Tokyo.

(Received August 10, 1979)

日本における ABO 式血液型遺伝子頻度の地理勾配の分析：移住モデル

青木健一・尾本恵市

東京大学理学部人類学教室

日本人の ABO 式血液型遺伝子頻度につきよく知られている地理勾配が、遺伝子頻度を異にし東西に対立する 2 集団の漸次混合により生じた過渡的勾配であるとの作業仮説を検討した。資料は本州、四国、九州の 42 箇所の遺伝子頻度データを用いた。勾配形成のモデルは遺伝子頻度の不連続な変化を初期条件とする拡散方程式の解として導かれる。A 遺伝子頻度にもとづく模擬試験により、初期の 2 集団の境界は現在の岐阜市附近、混合開始の時期は約 2,000 年前と推定された。さらに他の模擬試験の結果も加え、西日本の弥生時代に農耕により引き起された移住があったとの観点より検討を行なった。