Genetic Distances Between Certain New Guinea Populations Studied Under the International Biological Programme

P. B. Booth

Phil. Trans. R. Soc. Lond. B 1974 **268**, 257-267
doi: 10.1098/rstb.1974.0029

Phil. Trans. R. Soc. Lond. B. **268**, 257–267 (1974) [257]
Printed in Great Britain

Genetic distances between certain New Guinea populations studied under the International Biological Programme

By P. B. Booth
Immunohaematology Department, Christchurch Hospital, Christchurch, New Zealand
(lately Red Cross Blood Transfusion Service, Papua New Guinea)

The populations fell into two groups – highland and coastal. Comparisons were mainly between individual populations, classified linguistically, within each group. Calculated gene frequencies derived from the results on 12 systems (5 red cell antigen, 2 serum protein and 5 red cell enzyme), which showed useful variation, were used to estimate pair-wise genetic distances. Cluster analysis showed good correlation between linguistic and genetic information. The assumption that drift rather than selection had been the main cause of the observed variations between the populations studied seemed justified by comparisons of time scales produced by lexico-statistical methods with those derived from genetic distances. One particular measure of variation (f_θ) may be useful in providing some approximate and tentative estimate of the times of separation of related New Guinean population groups.

POPULATIONS STUDIED

The Australia/United Kingdom study of human adaptability in the tropical lowlands and central highlands of New Guinea selected Kaul village on Karkar Island, off Madang, as the base for its coastal investigations, and Lufa, not far from Goroka in the Eastern Highlands, as the highland base.

While blood sampling for genetic markers was concentrated on Karkar Island and around Lufa, examinations were also undertaken in adjoining areas as a basis for comparison. In order to comprehend the results of the genetic surveys, some knowledge of the populations found in each region is needed. New Guinea possesses what might be regarded as an excessive number of languages, there being 700–1000 (depending upon how one defines a language) for a total population of 2½ million. Workers from the Australian National University in Canberra and the Summer Institute of Linguistics, based in the Eastern Highlands of New Guinea, have succeeded in analysing and classifying a high proportion of these languages. In what follows, the classifications of Z'Graggen (1971) in respect of the languages of the Madang area, and of Wurm (1964) regarding the Highland languages, have been followed.

The primary division of New Guinean tongues is into Austronesian or non-Austronesian (Papuan). The former, numbering about 200 and almost entirely coastal, are the only ones having related languages outside New Guinea. They form the Melanesian subgroup of the Austronesian family that also includes Indonesian, Micronesian and Polynesian subgroups. However, the Melanesian languages are said not to be as closely related to each other as are the languages of the other subgroups, and show some anomalous features, possibly related to contact with non-Austronesian groups.

The situation with regard to the non-Austronesian languages is that some 20 phyla and 100 isolates have been established, which account for almost 400 of the 500 or so non-Austronesian languages. About 100 thus remain totally unknown. About 15 of the 20 phyla are related in one large 'macro-phylum' (McElhanon & Voorhoeve 1970).

Turning now first to the Madang/Karkar Island area (see figure 4), the southeastern half of the island is inhabited by speakers of the Takia language, while the northwestern half is populated by Waskia speakers. Takia is an Austronesian language, the only one included in the present surveys. It appears on the mainland as the language used in two coastal villages, and is closely related to that used in Bilbil village near Madang. The Bilbil people were dwellers on an off-shore island until the turn of this century, and were included in the surveys.

Waskia is a non-Austronesian language, a member of Z'Graggen's Adelbert Range phylum, one of the three phyla into which he classified the non-Austronesian languages of this region. Waskia appears on the mainland as the language of the inhabitants of one coastal village. On the New Guinea north coast opposite Karkar Island, the speakers of another Adelbert Range phylum language, Mugil, were investigated in three villages, while in the Madang hinterland, the non-Austronesian Madang phylum language speakers of the Gogol Valley were also examined in three villages. These latter people are unlike the coastal dwellers, and it has been suggested that they have possible Highland affiliations.

TABLE 1. EXTENT OF GENETIC TESTING

useful variation	*little or no variation*	
blood groups:		
ABO	C^w	Js^a
MNSs	Vw	Di^a
P	He	Wr^a
Rh	Lutheran	Radin
Gerbich	Kell	
serum proteins		
haptoglobin		
transferrin		
red cell enzymes		
acid phosphatase	adenylate kinase	
6-phosphogluconate dehydrogenase	phosphohexose isomerase	
phosphoglucomutase	lactate dehydrogenase	
malate dehydrogenase		
adenosine deaminase		
not all populations tested		
blood groups		
Lewis		
Duffy		
Kidd		
I^T		
saliva		
ABH secretion		
enzymes		
glucose-6-phosphate dehydrogenase		

Leaving the coast for the highlands, again a complex linguistic situation exists (see figures 5, 6). In the Eastern Highlands the languages in the main belong to the Eastern or East Central families. Yagaria, the language spoken around Lufa, belongs to the latter family, and is linguistically classed in a subfamily with Kamano and Keiagana. Other East Central subfamilies included in the present survey were those comprising Gahuku/Asaro/Bena Bena and that containing Fore. Of the Eastern family languages, Agarabe, Tairora and Auyana groups were studied. Each of these three languages is classed in a different subfamily.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

GENETIC TESTING AND DISTANCES

Table 1 sets out the systems investigated. They are blood groups, serum proteins and red cell enzymes. The full data are being published elsewhere. Abnormal haemoglobins were also sought for and found, but will not be dealt with here.

Population genetic surveys as extensive as those carried out under the International Biological Programme in New Guinea tend to produce a rather indigestible mass of raw information, and in order to analyse the differences between the various populations studied, a method which takes into account all the observed genetic variation is needed.

A number of measures of genetic distance between populations, which serve this purpose, have been described, all of which seem to give similar results. The one used here was presented to the Royal Society by Professor Cavalli-Sforza in 1965, and has been developed subsequently by him and his fellow workers (Cavalli-Sforza & Bodmer 1971).

Results from 12 systems (5 red cell antigen, 2 serum protein and 5 red cell enzyme), for which all populations were tested, showed useful variation, and the calculated gene frequencies were used to estimate genetic distances between various of the populations, which were classified by linguistic group. These distances were used in two ways. First, 'tree' diagrams were produced, and assessed to establish whether the distribution and clustering of the populations conformed to expectations based on other known relationships (for example, linguistic and geographical) between them. Secondly, a further measure of variation, the mean kinship coefficient (f_θ) (Cavalli-Sforza 1969) was calculated, and its possible correlation with time scales investigated. Comparisons were chiefly between individual populations within each main group, highland and coastal.

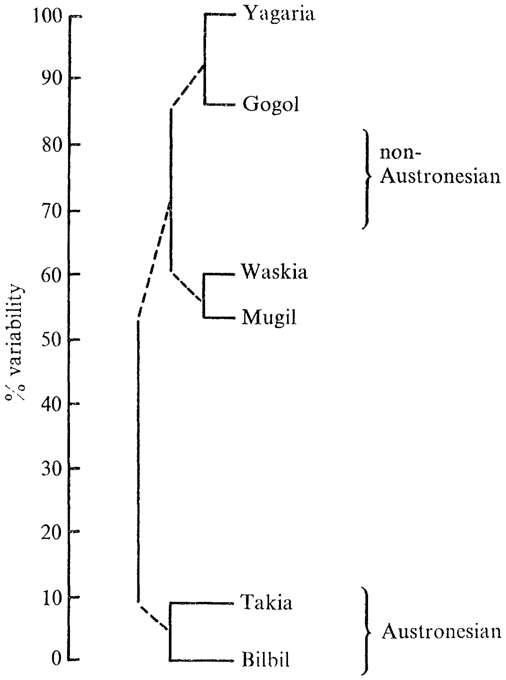

FIGURE 1. Clustering and relationships of linguistic groups in the Madang area, together with one Eastern Highland population (Yagaria). The sum of squares about the mean is 0.5672 (= 100% variability).

Cluster analysis

In the first instance, it was assumed that drift rather than selection would have been the main cause for differentiation between related populations, and cluster analysis was performed, using the squares of the genetic distances (the square being proportional to time for divergence due to drift) in an analysis of variance. This method finds the two most compact population clusters, and the process is repeated sequentially, to produce a 'tree' diagram.

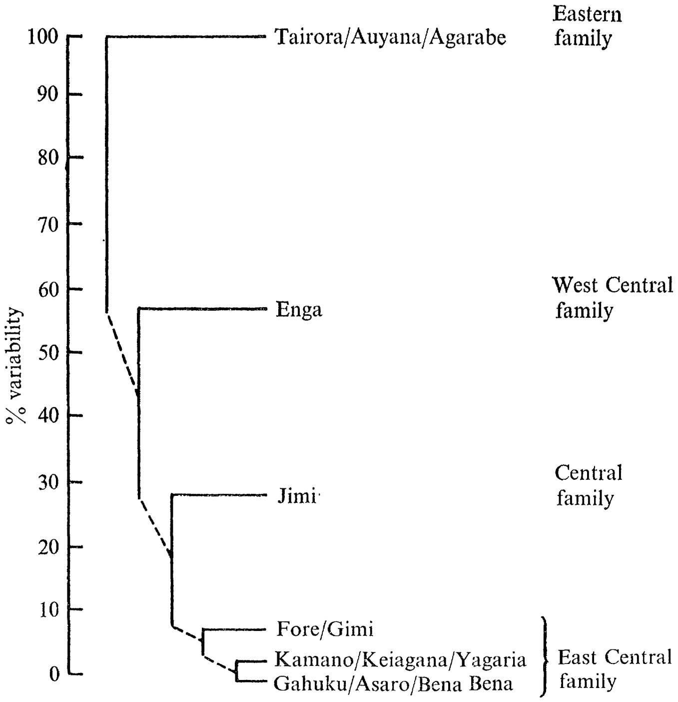

Figure 2. Clustering and relationships of representatives of Highland language families. Sum of squares = 0.2465.

Figure 1 depicts the best tree for the Karkar Island–Madang area. Because of the suggested Highland connexions of the Gogol Valley people, one eastern population, arbitrarily selected, was included in this analysis. The choice of Yagaria would seem to have been justified by the results. The first split removes the Austronesian speakers from the non-Austronesian, while the next separates the Adelbert Range phylum speakers (Waskia and Mugil) from the Yagaria/Gogol Valley pair. The genetic distance between the Austronesian and non-Austronesian speakers is marked, while of the pairs, Waskia and Mugil are closest. The supposed affinity of the Gogol Valley dwellers with Highland people is given support by this analysis. It must be remembered that trees such as this presuppose a common origin of the populations included, and that this is almost certainly not the case here. Nevertheless, the clustering demonstrates such genetic similarities as exist, and conforms to expectations based upon linguistic information.

On Highland populations, a number of analyses were performed. First, four main language families were considered, with results as shown in figure 2.

The figures for the West Central Enga were derived from the report of Sinnett *et al.* (1970), while those for the Central family Jimi are a side issue from the I.B.P. work and will be reported at a later date. The language families, or representatives therefore, are well separated genetically, with three subfamilies of the East Central family clustering close together. It is of interest that the earliest split seems to have separated the Eastern family from the rest, with subsequently the West Central, Central and East Central families splitting off in that order. Wurm (1966), in discussing the evidence for east to west migration in the Highlands, suggested that the main thrust of migration may have penetrated only as far as the present border between the East Central and Central families. This does not conflict with the genetic information, for it could be supposed that differentiation, probably already begun outside the Highlands, continued firstly in the Eastern Highlands, with subsequent migration from the central population into the Western Highlands across the Chimbu–Chuave Divide. Lastly, differentiation within the central region occurred.

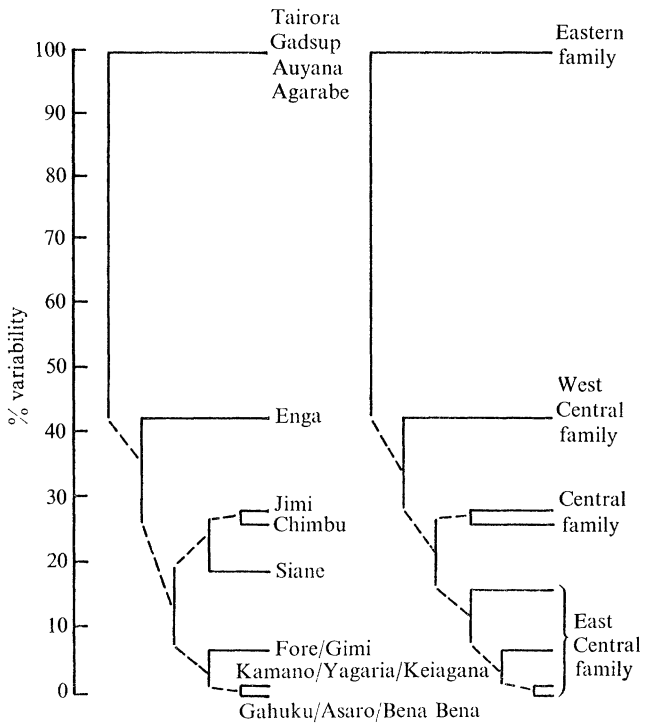

FIGURE 3. Clustering and relationships of Highland populations, including Chimbu, Siane and Gadsup, but from ABO, Rh and MNSs gene frequencies only (cf. figure 2). Two trees, of almost equal likelihood, are depicted. Sum of squares = 0.1122.

If only blood-group gene frequencies are considered, a further two populations can be included, namely Chimbu and Siane, with results as shown in figure 3. Here the two best trees are depicted. It is obvious that the general form is similar to that of figure 2, with Chimbu close to Jimi, both being members of the Central family. The position of Siane is anomalous and interesting, being open to two possibilities, of almost equal likelihood. The best cluster is formed with the Central family language, the second best with other languages of the East Central family, to which Siane belongs. Siane territory adjoins that of the Central family, however, so here geographical may outweigh or at least balance linguistic considerations, though the limitations of cluster analysis must be borne in mind.

TIME SCALES

Differentiation of populations is the outcome of past events, some remote, some more recent, but there are few dates in human evolution available to assist in fitting a time scale to observed genetic variation. In New Guinea several approximate dates, based upon language statistics, have been calculated, and correlation of these with the mean kinship coefficient (f_θ) has produced interesting results.

TABLE 2. SOME COMPARISONS OF NEW GUINEA LINGUISTIC GROUPS

comparison	t	f_θ	$-\ln(1-f_\theta)$	$2\sqrt{f_\theta}$
Eastern *v.* East Central Highland	3500	0.0350	0.0356	0.374
Guhu-Samane *v.* Binandere	4000	0.0407	0.0416	0.404
Atzera *v.* Takia	5000	0.0475	0.0487	0.436

t = time of separation in years, based on lexico-statistical evidence.
f_θ and its functions – see text.

Table 2 shows some comparisons of New Guinea populations. The times of separation shown are those derived from linguistic evidence (Wurm 1966, and personal communications from Summer Institute of Linguistics), and are of course only approximate. The logarithmic function of f_θ is proportional to elapsed time if the observed differences between the populations are entirely due to drift. If selection alone were the cause, then $\sqrt{f_\theta}$ would be appropriate. It can be seen, first, that at these magnitudes, the quantity f_θ may be roughly equated with the logarithmic function; secondly, that drift, rather than selection, best fits the time scale; and thirdly, that $10^5 f_\theta$ appears to give a good approximation to the lexico-statistical estimates. These New Guinea f_θ values are similar to those mentioned by Cavalli-Sforza (1969) relating to Australian aboriginal tribes, African main groups, and indeed isolated populations in the upper Parma Valley in northern Italy. Thus roughly the same order of microgeographic variation is found in these widely separated regions.

In the Madang/Karkar area (figure 4 and table 3) for groups closely related linguistically namely Waskia/Mugil and Takia/Bilbil, a time lapse of 2000–2500 years is indicated. For Waskia/Takia, who share Karkar Island, 3500 years is the figure, and something of this order separates the Gogol Valley people from the Eastern Highland Yagaria. Mugil is equidistant at 4000 years, from both Yagaria and Gogol Valley. Other comparisons show relationships, which if they really exist at all, are very remote.

It is not reasonable to suppose that the Austronesian Takia and the non-Austronesian Waskia shared a common ancestry, and it seems more likely that the present genetic distance between them represents a convergence, rather than a divergence. This hypothesis can be tested if it is assumed that the Austronesian speakers now inhabiting areas near Lae and near Madang had a common origin. The former, as represented by the Atzera people, show a fairly distant relationship to the Takia (table 2), but are considerably closer to the Bilbil people ($100 f_\theta$ about 1.5). From table 3 it can be seen that the Bilbil are much farther from the non-Austronesian groups than are the Takia. The distances for Atzera are slightly greater than for Bilbil.

Assuming, then, that Atzera are nearest to the Austronesian ancestral type, and that Waskia and Mugil, who are genetically close, may therefore not have diverged too greatly from the non-Austronesian ancestral type, it is possible to test whether the Takia may now represent

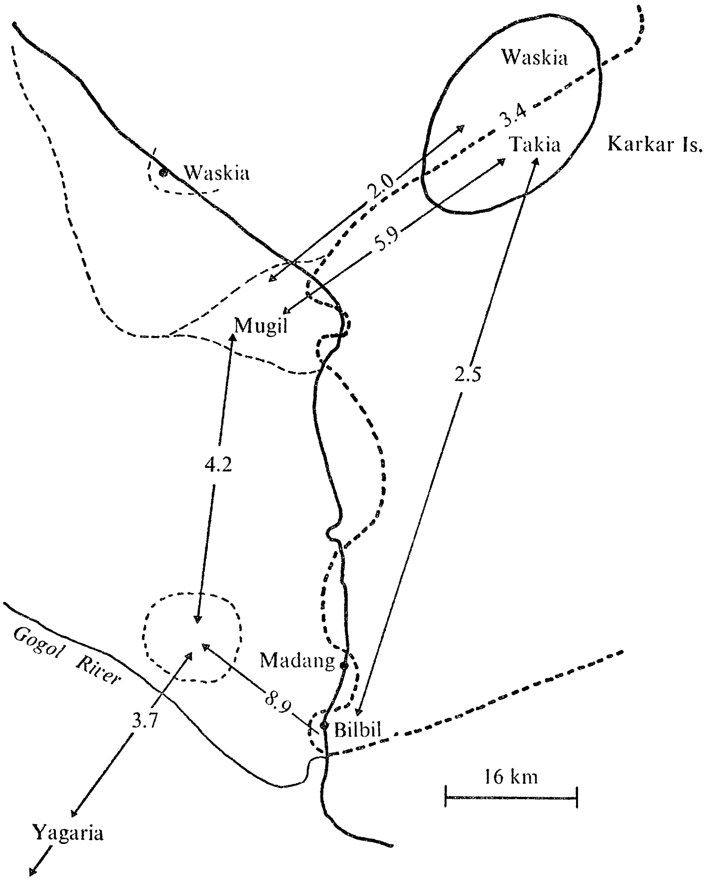

FIGURE 4. Relationships between groups in the Madang area. The numbers are $100f_\theta$, and may represent times of separation in millennia. The heavy dotted line separates the Austronesian (Takia and Bilbil) from the non-Austronesian language speakers. The Eastern Highland Yagaria population lies about 100 km (60 miles) to the south of the Gogol Valley group.

TABLE 3. $100f_\theta$ VALUES FOR MADANG AREA POPULATIONS PLUS YAGARIA (THE DISTANCES MAY REPRESENT SEPARATIONS IN MILLENNIA)

	Adelbert Range		Austronesian		non-Austronesian	
	Waskia	Mugil	Takia	Bilbil	Gogol	Yagaria
Waskia	—	*2.0*	*3.4*	6.6	5.6	5.6
Mugil	—	—	5.9	6.8	*4.2*	*4.2*
Takia	—	—	—	*2.5*	5.6	6.0
Bilbil	—	—	—	—	8.9	10.5
Gogol Valley	—	—	—	—	—	*3.7*
Yagaria	—	—	—	—	—	—

a mixture of these two putative ancestral groups. If this were so, the genetic distance between the ancestral groups should be equal to the sum of the distances between the presumed mixture and each ancestral group.

Using distances ($\sqrt{f_\theta}$) derived from blood groups only (ABO, Rh, MNSs, P and Gerbich – 15 alleles), and computing standard errors from the variation between loci, we have:

$$\begin{array}{ccccccc} \text{Atzera/Waskia} & & \text{Atzera/Takia} & & \text{Waskia/Takia} & & \text{sum} \\ 0.364 \pm 0.082 & = & 0.210 \pm 0.074 & + & 0.232 \pm 0.048 & = & 0.442 \pm 0.088 \end{array}$$

The sum does not differ significantly from the Atzera–Waskia distance ($t = 0.65$, $P > 0.5$). Alternatively:

$$\begin{array}{ccccccc} \text{Atzera/Mugil} & & \text{Atzera/Takia} & & \text{Mugil/Takia} & & \text{sum} \\ 0.393 \pm 0.093 & = & 0.210 \pm 0.074 & + & 0.253 \pm 0.048 & = & 0.463 \pm 0.088. \end{array}$$

Again the sum does not differ significantly from the Atzera–Mugil distance ($t = 0.55$, $P > 0.5$).

Thus the Takia might be genetically a mixture of Atzera with either Waskia or Mugil. Takia could represent a mixture of 45 % Mugil, 55 % Atzera, or of 48 % Waskia, 52 % Atzera. No other possibilities of admixture have so far been detected in the populations studied.

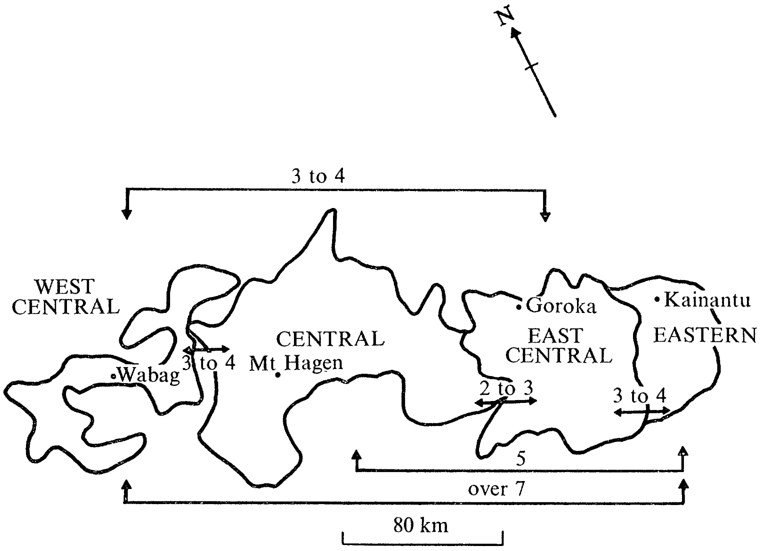

FIGURE 5. Relationships between main Highland language families. The numbers are $100f_{\theta}$, which may represent times of separation in millennia.

TABLE 4. $100f_{\theta}$ VALUES FOR HIGHLAND POPULATIONS

	FG	KYK	GAB	Jimi	Enga
Eastern family					
Tairora/Auyana/Agarabe	3.4	3.1	4.0	4.9	7.6
East Central family					
Fore/Gimi (FG)	—	0.5	0.9	2.8	3.5
Kamano/Yagaria/Keiagana (KYK)	—	—	0.4	1.9	3.1
Gahuku/Asaro/Bena Bena (GAB)	—	—	—	1.9	3.3
Central family					
Jimi	—	—	—	—	3.8
West Central family					
Enga	—	—	—	—	—

In the Highlands (figure 5 and table 4) the separation between adjacent language families, or representatives thereof, mainly fall with Wurm's estimate of 3500 ± 500 years, the exception being the East Central and Central families, where a time of 2000–3000 years is indicated by the kinship coefficient. Time of separation increases with geographical distance, the Eastern family appearing to be about 5000 years from the Central, and even more distant from the West Central family.

The closely related subfamilies of the East Central family (figure 6) seems to be 500 to 1000 years apart, and individual languages within these subfamilies show similar distances when compared. It can be seen that the northwestern groups are closer than those inhabiting more rugged territory to the south. For some languages, only ABO, Rh and MNSs results are available, and though, as has been seen (figures 2, 3), proportionality is retained, f_θ values based upon blood groups alone may be higher than those where serum protein and red cell enzyme results are available, simply because New Guinea populations in general display less variation in these polymorphisms than in red cell antigen systems.

Distances shown in figure 6 have all been computed from blood group gene frequencies only, and thus probably represent a time about 25 % shorter than indicated, giving separation times of 300–750 years for East Central family language groups.

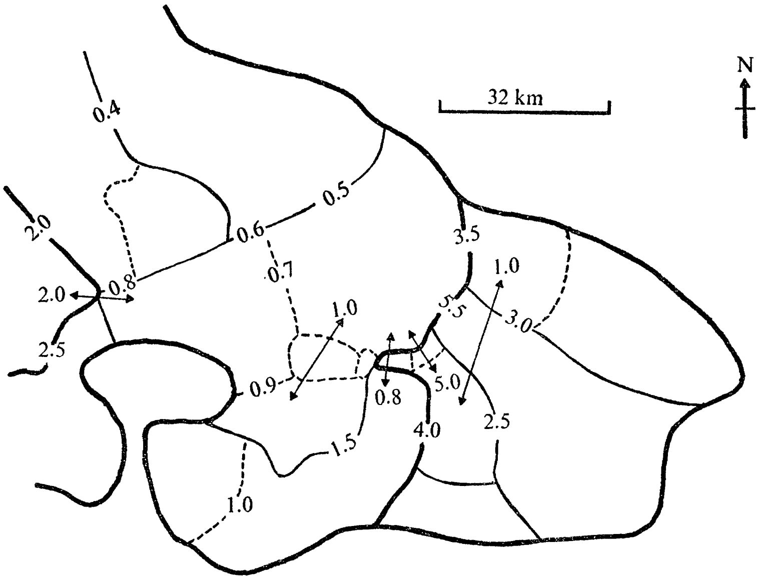

FIGURE 6. Relationships between Eastern Highlands linguistic groups. The numbers are 100f_θ, which may represent times of separation in millennia. Thick lines indicate boundaries of language families, thin lines boundaries of subfamilies, and dotted lines boundaries of individual languages. Calculations were based on ABO, Rh and MNSs systems only.

TABLE 5. 100f_θ VALUES FOR EASTERN FAMILY LANGUAGES, COMPARED TO TIMES IN MILLENNIA DERIVED FROM GLOTTO-CHRONOLOGICAL DATA (IN PARENTHESES)

	Agarabe	Tairora
Auyana	1.0 (1.65)	2.3 (2.55)
Agarabe	—	2.9 (2.40)

For Eastern family languages, some comparisons based on Wurm's glotto-chronological data (quoted by Watson, Zigas, Kooptzoff & Walsh 1961) are available, as shown in table 5. The genetic and linguistic data agree in placing Tairora at something like 2500 years from Auyana and Agarabe, and indicating a much closer relationship between the latter two groups.

It might appear from the figures that differentiation of many Eastern Highland language groups began 500–1000 years ago, but deviation from expectations based on f_θ values may be marked unless the populations have been of constant size and density, and have followed some regular pattern of migratory change. (Although Ward (1972), using these methods, successfully predicted genetic relationships between groups of Yanomama Indians over a range of conditions which departed to a varying degree from the requirements of the model.) The introduction of the sweet potato to the Highlands would have produced considerable alteration in the factors mentioned. The population explosion which is believed to have followed its introduction, and which may not have occurred until after the sixteenth century when the Spanish visited New Guinea, may well have resulted in major redistributions of populations, with alterations in linguistic territories. It seems not unlikely that acceleration of differentiation of individual languages may have become accentuated since that time.

So, while the larger f_θ values may provide useful information, the smaller ones relating to separations of individual language groups may be less satisfactory, even when computed from a comprehensive range of polymorphic systems, and probably represent a shorter time than the $10^5 f_\theta$ which seems to fit the separations occurring in the more distant past. If this were so, it would support the view that sweet potato introduction may be dated to 400–500 years ago.

It is clear that in the two very different regions of New Guinea dealt with here, the genetic and linguistic data give mutual support. Statistically, the correlation is high, if one compares f_θ values with the percentage of cognate forms shared by pairs of populations. For 24 such comparisons, the correlation coefficient was -0.58, making P about 0.001.

To establish, however tentatively, a time scale appropriate to the diversity observed seems a useful first step if we hope to learn more about the operation of drift and selection in the differentiation of primitive man. In the generally small population groups of New Guinea, chance effects, which may produce dramatic shifts in gene frequencies, must surely have outweighed the slower processes of adaptive selection when genetic data are viewed in the round.

However, the next step might be to investigate whether selection may be operating in some of the systems. This can be done in the absence of adequate information on migration or population sizes, by testing the homogeneity of f_θ values for different genes. As already noted, red cell antigen systems display greater variation than do the serum protein and red cell enzyme systems. In general, about 75% of the variability is accounted for by the red cell antigens, and thus the operation of some balancing force for the other loci must be considered. The preliminary analysis of the New Guinea I.B.P. data encourages continuance along these lines.

SUMMARY

First, cluster analysis based on calculated gene frequencies has shown good correlation between genetic and linguistic information. Secondly, the assumption that drift rather than selection has been the main cause of the observed variations between the populations studied seems justified by comparisons of time scales derived from genetic distances with those produced by lexico-statistical methods. Thirdly, the mean kinship coefficient may be useful in giving some approximate and tentative estimate of the times of separation of related New Guinea population groups.

REFERENCES (Booth)

Cavalli-Sforza, L. L. 1965 Population structure and human evolution. *Proc. R. Soc. Lond.* B **164**, 362–379.

Cavalli-Sforza, L. L. 1969 Human diversity. *Proc. XIIth Int. Congr. Genet.* **3**, 405–416.

Cavalli-Sforza, L. L. & Bodmer, W. F. 1971 *The genetics of human populations.* San Francisco: W. F. Freeman.

McElhanon, K. A. & Voorhoeve, C. L. 1970 The trans-New Guinea phylum: explorations in deep-level genetic relationships. *Pacific Linguistics* (B), no. 16. Canberra: Australian National University.

Sinnett, P., Blake, N. M., Kirk, R. L., Lai, L. Y. C. & Walsh, R. J. 1970 Blood, serum protein and enzyme groups among Enga-speaking people of the Western Highlands, New Guinea, with an estimate of genetic distance between clans. *Archaeol. and phys. Anthropol. in Oceania* **5**, 236–252.

Ward, R. H. 1972 The genetic structure of a tribal population, the Yanomama Indians. V. Comparisons of a series of genetic networks. *Ann. Hum. Genet.* **36**, 21–43.

Watson, J. B., Zigas, V., Kooptzoff, O. & Walsh, R. J. 1961 The blood groups of natives in Kainantu, New Guinea. *Hum. Biol.* **33**, 25–41.

Wurm, S. A. 1964 Australian New Guinea Highland languages and the distribution of their typological features. *Am. Anthropol.* **66**, 77–97.

Wurm, S. A. 1966 Language and agriculture in New Guinea. *New Scient.* **31**, 216–218.

Z'Graggen, J. A. 1971 Classifactory and typological studies in languages of the Madang District. *Pacific Linguistics* (C), no. 19. Canberra: Australian National University.