

# Chapter 10. A Study of Genetic Distance and the Austronesian/Non-Austronesian Dichotomy

**Kuldeep Bhatia , Simon Easteal and Robert L. Kirk**

*Data on genetic distance and unique allele distributions are presented for a number of Austronesian and non-Austronesian (Papuan) linguistic populations in the western Pacific. These data confirm separate origins for both of these major populations, but also suggest the existence of much subsequent gene flow between them. Genetic links between Australia and New Guinea are probably very remote in time.*

## **Introduction**

In 1965, Giles, Ogan and Steinberg claimed a clear-cut discrimination based on tests for the Gm system between Austronesian (An) and non-Austronesian (NAn) speakers in the Markham River Valley of Papua New Guinea. Because of a failure later to find a similar discrimination between An and NAn speakers on Bougainville, there has been critical and sometimes heated debate on (a) the usefulness of genetics for studying An and NAn origins, and (b) the validity of the model which suggests that An and NAn-speakers have different biological origins.

These competing views have been highlighted from differing perspectives by John Terrell in his *Prehistory of the Pacific Islands* (1986), Jonathan Friedlaender in his concluding chapter of *The Solomon Islands Project* (1987), and by Sue Serjeantson and Adrian Hill in *The colonization of the Pacific* (1989). The last conclude (Serjeantson and Hill 1989:287):

... the extreme view taken by Terrell (1986) and White *et al.* (1988), that Polynesians evolved within Melanesia from a population resident there for at least 30,000 years, is untenable in the light of the genetic evidence.

In the present volume Serjeantson and Gao provide further evidence for this position, based on information derived from analysis of HLA genes. The present discussion complements the HLA analysis by reviewing evidence collected over the past 20 years for a large number of blood-genetic traits and subjecting the data to newer multivariate analytical techniques.

## The Nature of the Evidence

We are not concerned here with discrimination using anthroposcopic or anthropometric characters. In passing, however, it should be noted that such studies, particularly those for teeth and fingerprints, can be very informative. Our own surveys have been restricted to traits detectable in samples of blood under simple genetic control by loci on many different chromosomes. The enzyme and other protein systems used, and their distributions in Pacific populations, have been reviewed recently by Kirk (1989), and in more detail for Papua New Guinea by Kirk (1992).

The data can be analysed in two ways. Some genetic differences are unique to certain populations and their patterns of distribution suggest common ancestry. In addition, variations in frequency of genetic factors can be subjected to multivariate analysis to give "genetic" distances between populations. These distances can be used to construct evolutionary trees by a number of methods including cluster and maximum likelihood analysis, or by principal component analysis, to obtain the distribution patterns of populations.

## Unique Allele Distributions

Kirk (1992), reviewing previous studies of the distribution of unique alleles in the western Pacific, recognized three patterns relevant to understanding the relationships between linguistic and genetic differentiation. The first of these patterns, the "Australoid", is associated with the transferrin allele Tf<sup>D1</sup> and the GC<sup>1A1</sup> allele of the vitamin D-binding protein system. The second, or "Proto-Papuan", is characterized by alleles such as PGM<sup>1</sup><sub>3</sub>, PGM<sup>2</sup><sub>9</sub>, PGM<sup>2</sup><sub>10</sub>, PGK<sup>4</sup> and MDH<sup>3</sup>. None of these alleles is found in Australia, suggesting that they were brought to, or originated in, New Guinea after the separation of New Guinea and Australia at the end of the Pleistocene, 8-10,000 years ago. These "Proto-Papuan" alleles all have relatively high frequencies in the Papua New Guinea Highlands and in parts of Irian Jaya, with lower frequencies in New Guinea coastal areas and even lower frequencies in the Solomons, Banks Islands and Polynesian Outliers.

The third pattern is "Austronesian". Alleles in this group are not found in Australia and rarely in the Papua New Guinea Highlands. They have their highest frequencies in the Solomons, Polynesian Outliers, Banks Islands, some coastal areas in the north and east of Papua New Guinea, the western Carolines and Fiji. These alleles include PGM<sup>1</sup><sub>7</sub>, PGK<sup>2</sup>, probably HB<sup>Tongariki</sup>, Albumin<sup>NG</sup>, GPT<sup>3</sup> and GPT<sup>6</sup>.

## Genetic Distance Studies

Previous studies, reviewed by Kirk (1986, 1989, 1992), have shown discrimination between *Waskia* (NAn) and *Takia* (An) on Karkar Island. However, for 17 other

populations in the north coastal regions of Papua New Guinea the An speakers are not clearly differentiated from NAn speakers. In these cases more detailed analysis shows that geographic location is more important than linguistic division (Serjeantson *et al.* 1983).

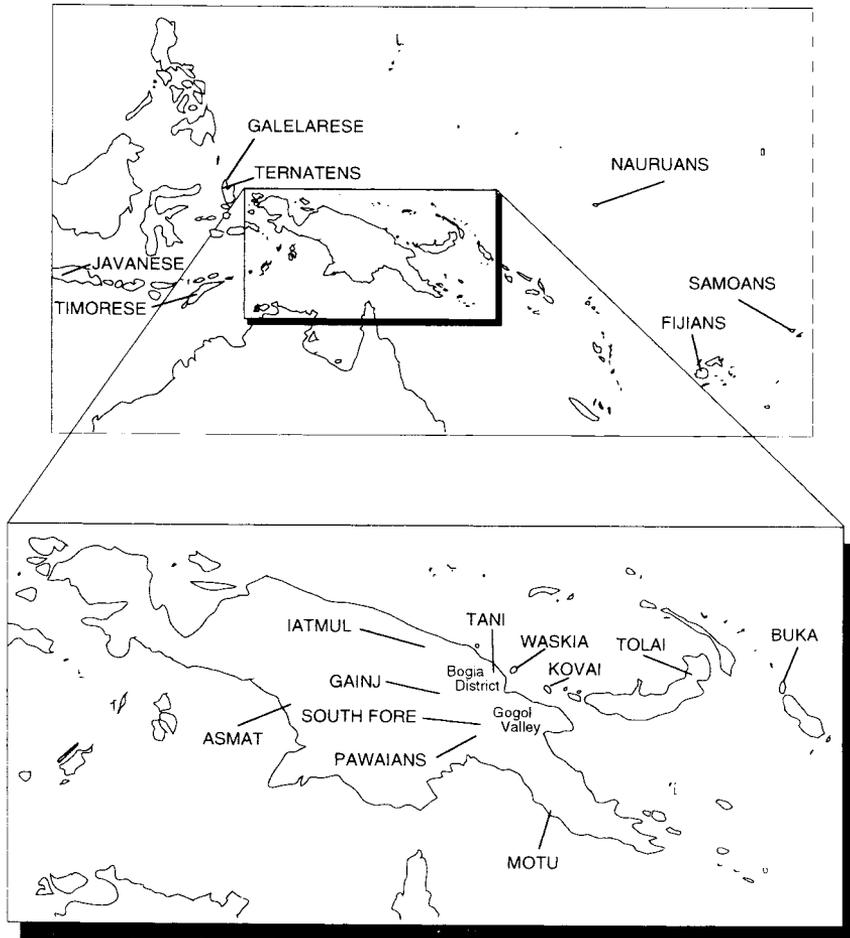
Nevertheless, consideration of populations over a wider geographic area, including many from the Highlands of Papua New Guinea and others from coastal areas and other parts of the western Pacific, show that language is an important discriminant, with the exception of the Mailu in southeast Papua. The exceptional position of the Mailu is due probably to the incorporation of An genetic components from neighbouring populations into a group which continues to speak a NAn language (see Kirk 1992 for further details).

In a detailed comparison of genetic distances between An-speaking Indonesian and other western Pacific populations, Sofro (1982) has shown that the Indonesian populations, including Ternatens and Galelarese from Halmahera whose languages are NAn, cluster with the An-speaking populations of New Guinea and elsewhere in the Pacific but are distinctive from the NAn-speaking populations both in Papua New Guinea and Irian Jaya.

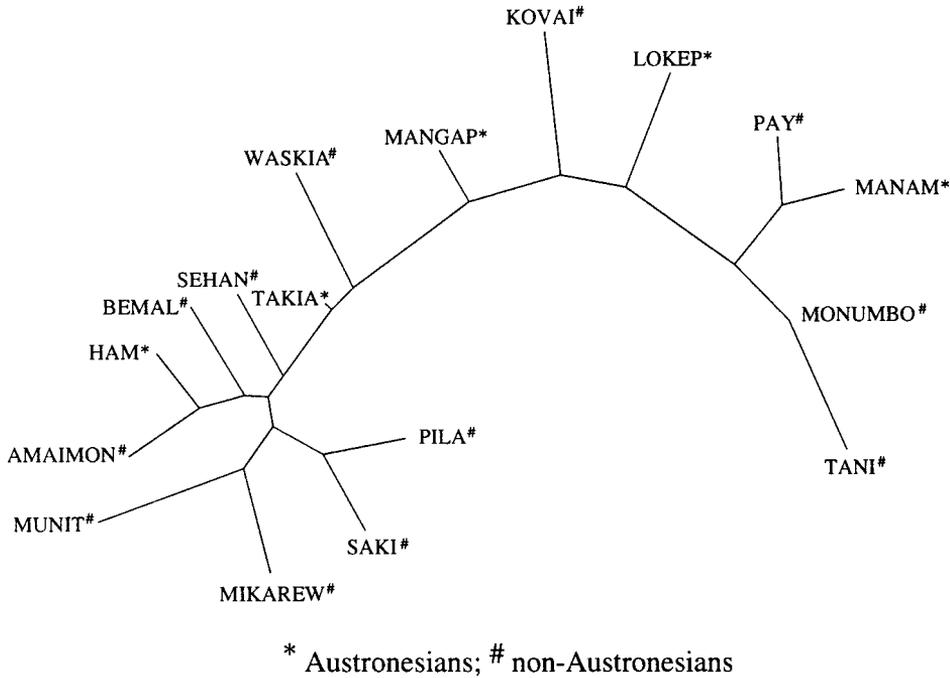
To examine further the question of linguistic and genetic relationships in the Pacific area we have used more recently developed statistical procedures to re-analyse some of our previous data, and have included some populations for which new genetic marker information is now available. Multi-locus allele frequency data were used to estimate the phylogeny of two population groups, using a partial maximum-likelihood method (Felsenstein 1981).

This method has been shown by Kim and Burgman (1988) to be more accurate than the more commonly used unweighted pair-group arithmetic average clustering (UPGAA) method of estimating phylogeny from allele frequency data, particularly when a relatively small number of loci are analysed and where rates of evolution may vary among populations. The maximum-likelihood approach results in an estimate of the evolutionary history of a group of populations in the form of a maximum-likelihood network (or tree) connecting them. The reliability of the estimate can be tested by comparing the "likelihood" of the maximum-likelihood network with that of other networks connecting the same populations through different patterns of branching.

The first group of populations analysed consisted of the same 17 populations referred to above, investigated by Serjeantson *et al.* (1983) and located in the Bogia District and Gogol Valley in northern coastal Papua New Guinea, and on the adjacent Manam, Karkar and Siassi Islands. These populations include both An and NAn speakers. The second group consists of An and NAn speakers from various localities on New Guinea and from throughout Indonesia, Island Melanesia, Micronesia and Polynesia (Map 1).



**Map 1. Localities of sampled populations. The linguistic groups sampled from the Bogia District are Saiki, Pila, Tani, Pay, Monumbo, Mikarew, and Manam; those sampled from the Gogol Valley are Munit, Sehan, Ham, Amaimon, and Bemal.**



**Figure 1. Maximum likelihood network connecting Austronesian and non-Austronesian-speaking populations from the Bogia District and Gogol Valley on the north coast of Papua New Guinea and from adjacent islands. Branch lengths are drawn in proportion to genetic distance.**

## **Austronesian and Non-Austronesian Populations on the North Coast of Papua New Guinea**

Our maximum-likelihood analysis of the data of Serjeantson *et al.* (1983) (Figure 1) confirms their conclusions that no clear genetic distinction exists between the An and NAn speakers residing along the north coast of Papua New Guinea, and that population affinities are based more on geographical proximity than on linguistic similarity. There are some differences between the branching patterns of our network and their dendrogram, but it remains the case that, for the most part, populations that are geographically close to each other are relatively similar genetically, irrespective of their linguistic affiliations. Thus, for example, the An Ham are more closely related to other NAn groups in the Gogol Valley than they are to the other An populations. Similarly, the An Manam resemble their NAn neighbours in the Bogia District more than they do other An populations. Two groups on Karkar Island (An Takia and NAn Waskia) are also close to each other in the network, as are the three groups from the Siassi Islands (An Mangap and Lokep and NAn Kovai).

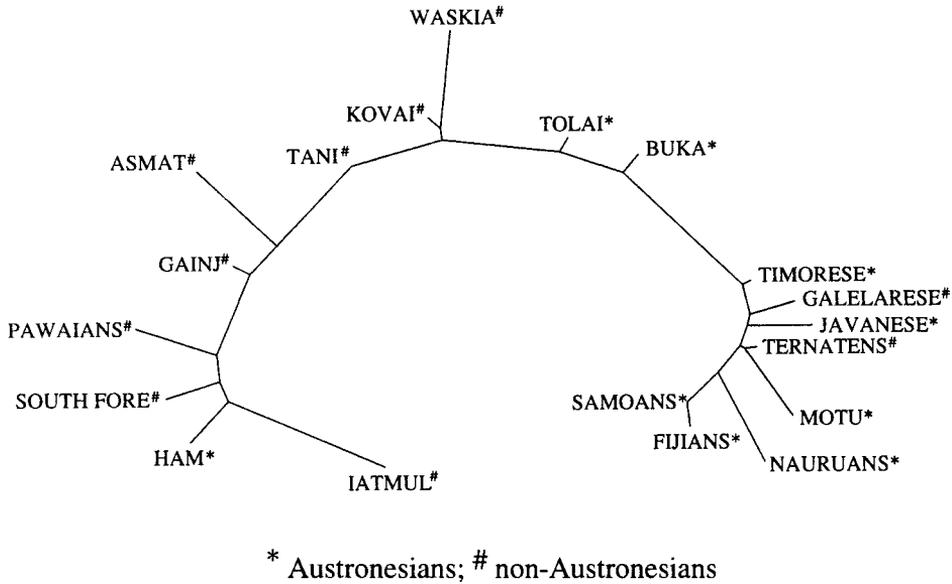
The situation with respect to the genetic affinities of NAn populations belonging to different phyla is less clear-cut, although, once again, geographical proximity appears to have some influence. Thus, in the Bogia District, Pay and Tani (Trans New Guinea Phylum, Adelbert Range Superstock) are more closely related to Monumbo (Torricelli Phylum) than they are to other Adelbert Range Superstock groups in the Gogol Valley (Amaimon) and on Karkar Island (Waskia). Other Adelbert Range groups in the Bogia District (Saki and Pila) are, however, also relatively distantly related to Pay and Tani.

### **Non-Austronesian Diversity and its Contribution to Austronesian Heterogeneity in Melanesia**

The lexicostatistical study of Dyen (1965) revealed significantly more diversity in the An languages spoken in Melanesia than among those spoken further west in Indonesia and Malaysia. It is now recognized that this diversity reflects heavy borrowing from the NAn languages which were significantly diversified at the time of Austronesian settlement.

Most of the An borrowing appears to have been from the smaller NAn phyla, with little influence from the two major NAn phyla, Trans New Guinea and Sepik-Ramu. The geographical distributions of these two phyla only overlap with the Austronesian speakers on mainland New Guinea, and that also marginally. Besides, the speakers of languages belonging to these phyla have only recently expanded into their present areas of distribution. The Highlands migration of the Trans New Guinea Phylum languages is considered to have begun around 5,000 to 2,000 years ago. The occupation, west to east, of the coastal areas of Sepik and Madang provinces by Sepik-Ramu speakers, who are essentially a riverine people, is much more recent a phenomenon. Investigation of the An/NAn dichotomy in Melanesia therefore must take into consideration this diversity of NAn languages and the extent to which it has influenced the An substratum.

To evaluate the relationships among An and NAn speakers on a wider scale we selected for analysis representatives of three different NAn phyla, namely, the North Halmahera Stock of West Papuan Phylum (Ternatens and Galelrese), the Iatmul of the Sepik-Ramu Phylum, and speakers of languages in the Trans New Guinea Phylum. In view of the extensive diversification of Trans New Guinea Phylum languages, we selected one population each from five different regions of New Guinea: the north coast (Pila), the northern Highlands-fringe (Gainj), the Highlands proper (South Fore), the southern Highlands-fringe (Pawaia) and the south coast (Asmat). In addition, two populations were added from the islands off the coast of New Guinea (Waskia and Kovai). We also included widely distributed An speakers (Figure 1).



**Figure 2. Maximum likelihood network connecting representative Austronesian and non-Austronesian-speaking populations from Indonesia, Melanesia, Micronesia and Polynesia. Branch lengths are drawn in proportion to genetic distance.**

The linguistic diversity among NAn speakers is reflected by their genetic diversity (Figure 2). Differences in the branching order between Figures 1 and 2 for those populations represented in both are due to slight differences in the data used in the two analyses.

The Ternatens and Galelarese from Indonesia (both belong to the West Papuan Phylum) are closely related to the majority of An speakers. They are quite distinct from the other NAn groups, which are themselves loosely clustered with no hierarchical structure to their relationships. It appears that the Trans New Guinea Phylum speakers failed to homogenize the genetic diversity underlying the linguistic substructure already in place at the time of their arrival.

Among the An populations, Ham clusters well within the NAn populations indicating that they have acquired an Austronesian language from outside. With the exception of Tolai and Buka (see below), the remaining An populations are relatively tightly clustered. This indicates close genetic affinities, despite the populations being widely distributed geographically. The pattern is consistent with these populations having spread rapidly and recently to occupy their present location.

It is now accepted that the Bismarck Archipelago was home to the progenitors of Lapita cultures, although an opposing view suggests that these cultures arrived fully formed into the region. Supporters of both these views, however, agree

that the islands of the Archipelago were central to the spread of Lapita people further east into the Pacific. Unfortunately, the genetic data on the contemporary Bismarck populations are scanty and the above argument is unlikely to be resolved without some information from New Ireland and New Britain. Genetic data on populations surrounding the Bismarck Archipelago are available, and it may be argued that the Bismarck populations are unlikely to be very different from these neighbouring populations. We argue, on the one hand, that the populations settled on both sides of St George's Channel, between New Britain and New Ireland, would more likely to have been influenced by the Lapita movement than those surrounding the Vitiaz Strait, between New Britain and New Guinea, if the colonizers of Polynesia originated in Southeast Asia and largely bypassed Melanesia. On the other hand, if the Lapita populations *did* develop entirely in the Bismarck Archipelago then one might expect greater genetic homogeneity among populations in the region of the Bismarck Sea.

To test these two opposing hypotheses we have included in the analyses populations bordering St George's Channel and the Vitiaz Strait. The Tolai occupy the western end of St George's Channel, whereas the Buka are located further east. For the Vitiaz Strait we have selected two NAn-speaking populations, the Kovai from Umboi Island and Waskia from Karkar Island.

The analysis reveals that the four Papua New Guinean island populations do not join either the group composed of Polynesians, Micronesians and Indonesians, or the remaining NAn populations (Figure 2) occupying an intermediate position in the network between the two. Waskia and Kovai share a common branch in Figure 2. Tolai and Buka branch quite separately and distinctly, but are closer to the remaining An populations. There is thus no homogeneity among the populations surrounding the Bismarck Sea and greater affinity between the main group of An populations and those adjacent to St George's Channel consistent with the hypothesis of a movement of Lapita culture through the region.

## **Linguistic Links Between Sepik-Ramu and Earlier Australian Languages**

The diversity of NAn languages raises the issue of their possible links with the languages which are now found in Australia. Wurm (1983) has suggested that linguistic traces of an early Australian (or Australoid) population, mixed with later arriving NAn speakers, can be seen in the languages of the Sepik-Ramu Phylum. According to him, Laycock (1973) has pointed out the general resemblance between the phonology of the languages of Ndu Family in the Middle Sepik Stock of the Sepik-Ramu Phylum and the general phonological features of the Australian languages. In addition, the occurrences in the Sepik-Ramu region of Australian cultural elements such as spearthrowers, bullroarers, flat surface and bark painting, and the resemblances of slit-gong melodies to didgeridoo melodies, are all considered to indicate possession of

common cultural traits. The connection between the speakers of Sepik-Ramu Phylum languages and Aboriginal Australians has been explained by a southward migration route passing through the Purari River area, possibly because there are similarities between the Sepik art styles and those of the Purari (Spieser 1937), where bullroarers are also found (Williams 1936).

However, further analysis of our data provides no indication of a connection between the Sepik-Ramu populations and Aboriginal Australians via the Purari River area. First, there is no close affinity between the Iatmul (Sepik-Ramu) and the Pawaians (Purari River) (Figure 2). Second, we repeated the analysis of the populations shown in Figure 2 with the inclusion of a population from central Australia (Waljbiri). The resulting network (not illustrated here) shows the Waljbiri are very distantly related to all other populations. The branch leading to the Waljbiri is nearly eight times as long as the next longest terminal branch on the network (leading to Iatmul), and the position at which it connects to the rest of the network could not reliably be determined. It would seem that if there are any genetic affinities underlying the cultural and linguistic similarities between Sepik-Ramu and Australian Aboriginal populations, as discussed by Wurm (1983), these are extremely remote.

## Conclusions

Our extensive data support the model that An speakers had a different biological origin from the NAn speakers in the western Pacific. However, the differences are not clear-cut in all cases and suggest that in many populations other factors, including intermarriage to various extents, have occurred to blur the edges of the linguistic boundaries.

## References

- Dyen, I.  
1965 A lexicostatistical classification of the Austronesian languages. *International Journal of American Linguistics*, Memoir 19.
- Felsenstein, J.  
1981 Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution* 35:1229-1242.
- Friedlaender, J. (ed.)  
1987 *The Solomon Islands Project*. Oxford: Oxford University Press.
- Giles E., E. Ogan and A.G. Steinberg  
1965 The gamma globulin factors (Gm and Inv) in New Guinea: anthropological significance. *Science* 150:1158-1160.
- Kim, J. and M.A. Burgman

1988 Accuracy of phylogenetic-estimation methods under unequal evolutionary rates. *Evolution* 42:596-602.

Kirk R.L.

1986 Human genetic diversity in south-east Asia and the western Pacific. In D.F. Roberts and G.F. de Stefano (eds) *Genetic diversity and its maintenance in tropical populations*, pp.111-134. Cambridge: Cambridge University Press.

1989 Population genetic studies in the Pacific: red cell antigen, serum protein and enzyme systems. In A.V.S. Hill and S.W. Serjeantson (eds) *The colonization of the Pacific: a genetic trail*, pp.60-119. Oxford: Clarendon Press.

1992 Population origins in Papua New Guinea. In R.D. Attenborough and M.P. Alpers (eds) *Human biology in Papua New Guinea: the small cosmos*, pp.172-197. Oxford: Clarendon Press.

Laycock, D.C.

1973 *Sepik languages — checklist and preliminary classification*. Pacific Linguistics Series B No. 25. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Terrell, J.

1986 *Prehistory in the Pacific Islands*. Cambridge: Cambridge University Press.

Serjeantson, S.W. and A.V.S. Hill

1989 The colonization of the Pacific: the genetic evidence. In A.V.S. Hill and S.W. Serjeantson (eds) *The colonization of the Pacific: a genetic trail*, pp.286-294. Oxford: Clarendon Press.

Serjeantson, S.W., R.L. Kirk and P.B. Booth

1983 Linguistic and genetic differentiation in New Guinea. *Journal of Human Evolution* 12:77-92.

Sofro, A.S.M.

1982 Population genetic studies in Indonesia. PhD thesis, The Australian National University, Canberra.

Spieser, F.

1937 Eine initiationszeremonie in Kambrango am Sepik, Neuguinea. *Ethnologischer Anzeiger* 4:153-157.

White, J.P., J. Allen and J. Specht

1988 Peopling the Pacific: the Lapita Homeland Project. *Australian Natural History* 22:410-416.

Williams, F.E.

1936 *Bullroarers in the Papuan Gulf* (Territory of Papua, Anthropological Report 12). Port Moresby: Government Printer.

Wurm, S.A.

1983 Linguistic prehistory in the New Guinea area. *Journal of Human Evolution* 12:25-35.