

R-Matrix Analysis and Patterns of Gene Flow in India

B. MOHAN REDDY¹

Abstract An appraisal on the usefulness of **R**-matrix analysis and the model of Harpending and Ward (1982) to study the population structure of Indian populations is made in the light of prevalent marriage patterns, caste structure, and cultural specificity of the region. With the help of available data on the migration histories and geographical backgrounds of marine fishermen on the east coast of India, and through the historical analysis of marriage patterns, it is demonstrated that the routine inferences based on the regression plots of average heterozygosity versus genetic distance from the centroid may not generally be apt for the Indian situation. Increased heterozygosity among migrant fishermen seems to have resulted from mating patterns within this community rather than from external gene flow. In either case, the genetic implications are supposedly identical. Nevertheless, when making inferences regarding local population structure it is important to have detailed knowledge of mating patterns and the cultural context of the region under study.

Local genetic structure, which is the primary concern of anthropological geneticists, is almost exclusively driven by migration and drift. Harpending and Ward (1982) developed a model that helps in understanding the effect of these forces on the local population structure. According to this model, the expected heterozygosity within a population (H_i) is a function of the total heterozygosity (H_r) of the entire region and of the distance of the population from the regional mean centroid of the allele frequencies (r_{ii}). If gene flow from outside the region varies in amount from population to population, this linear relationship no longer holds. Very isolated populations will be less heterozygous than the linear prediction, while populations receiving more gene flow from outside will be more heterozygous, suggesting greater external gene flow. The theory thus indicates that we might gain useful insights by examining the outliers in the regression plot of mean heterozygosity versus r_{ii} . Harpending and Ward (1982) applied this model to a number of regional groups, and the results in general endorsed the usefulness of this approach by way of concurring with the historical patterns of gene flow.

¹ Anthropology and Human Genetics Unit, Indian Statistical Institute, Calcutta 700 035, India.

Human Biology, February 2001, v. 73, no. 1, pp. 57-69.

Relethford and Blangero (1990) extended this model to quantitative variables and demonstrated its usefulness by applying it to the dermatoglyphic traits among nine Nepali villages and to the anthropometric measurements of seven isolated populations in western Ireland. The results clearly portrayed the historically known patterns of gene flow among these regional groups. Encouraged by these findings, several other workers employed this model to study the structure of human populations from different regions (Devor et al. 1984; Relethford 1985; Jantz and Meadows 1995; Puppala and Crawford 1996; and others), using different sets of variables—genetic markers as well as the quantitative variables.

India, with its unique population structure, offers innumerable situations to which this model may be applied for understanding dynamics within the substructured local populations. We have recently applied this model to a cluster of marine fishermen groups, both migrants and their parental counterparts, and to a dozen groups of northwestern India, representing the entire gamut of socioeconomic heterogeneity in the region (Reddy and Chopra 1999; Reddy et al. 1999). We were probably the first to have applied this model to different sets of data (anthropometry, dermatoglyphics, and genetic markers) from the same set of populations. The fit of the model to these data sets was consistently good; hence the results reiterated, more emphatically, the usefulness of this approach in unravelling the local population structure. The regression plots of heterozygosity versus r_{ii} suggested the possibility of external gene flow into certain populations, since they lie above the *expected* regression line as outliers. However, in retrospect, a careful consideration of the mating patterns and historical backgrounds of these populations seems to suggest that the increased heterozygosity might have resulted from factors other than external gene flow.

The present paper attempts to illustrate such factors in the case of marine fishermen of the East Coast of India, with the help of certain population structural measures and the demographic data available for these populations.

Materials and Methods

The Harpending-Ward Model. The basic steps involved in the application of the Harpending and Ward (1982) model are given below.

Given n loci with two alleles at each locus, the expected heterozygosity of population i [$E(H_i)$] is a function of the heterozygosity of the total region (H_t) and the genetic distance between population i and the regional gene frequency centroid (r_{ii}). That is,

$$E(H_i) = H_t(1 - r_{ii}). \quad (1)$$

This equation provides an expected linear regression line of heterozygosity

on genetic distance from the centroid with intercept H_i and slope $-H_i$. For two alleles at each locus the heterozygosity of the total region is computed under the assumption of complete panmixia as

$$H_t = \sum 2p_k q_k / n, \quad (2)$$

where p_k and $q_k (= 1 - p_k)$ are the weighted mean allele frequencies for locus k and summation is over all n loci. The mean allele frequencies are computed as

$$p_k = \sum w_i p_{ik}, \quad (3)$$

$$q_k = 1 - p_k, \quad (4)$$

where w_i is the ratio of the census size of population i to the total census size over all groups, p_{ik} is the frequency of one allele at locus k in population i , and summation is over all groups.

The genetic distance of a population to the regional centroid is computed as the diagonal of the R matrix of scaled variances and covariances about the regional mean allele frequencies (Harpending and Jenkins 1973; Workman et al. 1973). For each allele the elements of the R matrix for populations i and j are computed as

$$r_{ij} = (p_i - p)(p_j - p)/p(1 - p). \quad (5)$$

The overall R matrix is then averaged over all alleles. The R matrix provides an estimate of genetic kinship relative to the contemporary region. That is, it measures deviations from the contemporary mean allele frequencies.

The observed heterozygosity of population i is computed as

$$H_i = \sum 2p_{ik} q_{ik} / n, \quad (6)$$

where summation is over all n loci. Under the assumption that all populations experience the same amount of gene flow from the same source (a homogeneous "outside world"), the expected heterozygosity and observed heterozygosity for population i will be the same. If either the rate or source of external gene flow is different among populations, then the expected relationship will not hold for all populations. In particular, Harpending and Ward show that populations having greater than average external gene flow will have observed heterozygosities greater than expected. Comparison of expected heterozygosity with observed heterozygosity allows assessment of which populations, if any, have experienced greater than average external gene flow [$H_i > E(H_i)$] or less than average external gene flow [$H_i < E(H_i)$].

Table 1. Sample Sizes of Different Data Sets for Different Populations, Migrant and Parental Fishermen of the East Coast

<i>Population</i>	<i>Type</i>	<i>Anthropometry^a</i>	<i>Dermatoglyphs^b</i>	<i>Genetic markers^c</i>
Jalari (JPP)	Parental	141	142	93
Vadabalija-V (VVP)	Parental	147	145	83
Vadabalija-P (VPP)	Parental	165	155	102
Palle (AP)	Parental ^d	151	158	120
Vadabalija of Penticotta (VP)	Migrant	208	160	121
Vadabalija of Vadapeta (VV)	Migrant	200	102	101
Jalari of Puri (JP)	Migrant	65	132	77

a. Anthropometric measurements included were stature, sitting height, head length, head breadth, nasal height, nasal breadth, biacromial breadth, bicristal breadth, and chest girth.

b. Dermatoglyphic variables are larger ridge counts on each of the 10 fingers, total ulnar count, total radial count, total number of triradii on fingers, main line index and a-b ridge count, separately for right and left hands.

c. Genetic markers include *A1A2BO*, *MN*, *RhD*, *ADA*, *AK*, *PGM*, *PGD*, *ACP*, *HP*, and *GC*.

d. Migrant counterpart for this population was not studied at Puri.

Data. In our earlier study of the biological affinities between migrant and parental populations of marine fishermen of the East Coast (Reddy and Chopra 1999), we applied *R*-matrix analysis to three sets of data: genetic markers (Harpending and Ward 1982), dermatoglyphs, and anthropometric measurements (Relethford and Blangero 1990). The same data form the base for the current study, although we present only the regression plots of heterozygosity/phenotypic variance versus the distance from the centroid for the three sets of data.

For details on data, population backgrounds, and sampling procedures, readers may consult our earlier paper. However, for the sake of convenience, the names of populations, their abbreviations and population-wise sample sizes for different sets of data are furnished in Table 1. Information on migration histories, geographical backgrounds, marriage distance, and village endogamy on the 1409 couples from the three migrant groups of fishermen from Puri supplement these data.

Subjects. We studied a total of seven fishermen groups. Of those, three were migrant groups living in Puri on the Orissa Coast for four to five generations. These migrant groups were drawn to Puri from about 100 villages distributed along the 400-mile-long stretch of the Orissa and Andhra coasts, south of Puri. The parental counterparts were sampled from 42 of those 100 villages. For comparative analysis we have included a related group of fishermen living adjacent to or cohabiting with one of the parental groups included in the study. All these groups were endogamous. There is no inter-

marriage between the migrant groups, except for a small and insignificant ($< 1\%$) frequency of exchange between the VP and VV, the two reproductive isolates of the same caste, Vadabaliya.

Results and Discussion

A brief review of the findings of our earlier paper (Reddy and Chopra 1999) may help put the problem of the present paper in proper perspective. Both the univariate and multivariate analyses suggested significant migration effect, diversifying migrants from the parental populations. The effect of migration had been particularly remarkable in the case of anthropometric measurements. The plot of group centroids based on discriminant analysis depicted a clear segregation of migrants from the parental populations. Due to relatively large effective population sizes and the short history of these populations at Puri, we could safely dismiss the role of genetic drift in the observed differentiation. Two possible alternative explanations for the observed differentiation are left: founder effect and external gene flow into migrants. Besides marked differences in the occupational patterns among the groups, there are strong indications that the founders, with reference to body dimensions, were a select group of fishermen, not migrants of random influx. Certain rare variants observed among migrants, not in the parental populations, also strengthened the possible role of founder effect in the divergence of migrants from their parental groups. On the other hand, the regression of mean phenotypic variance or heterozygosity on the distance from the gene frequency centroid of the groups suggested, as per the model, a possibility of external gene flow into the migrants as a factor in the observed differentiation.

Regression of Mean Heterozygosity versus Distance from the Centroid (R_{ii}). For ease of reference and to illustrate the present case, regression plots for the three sets of data that were given in our recent paper are reproduced (Figures 1–3). It is apparent from these figures that it is migrant groups that present themselves above the *expected* regression line, indicating greater degree of heterozygosity, and parental populations below the line, indicating greater isolation, hence less heterozygosity. This pattern is most evident in the case of genetic markers. The logical conclusion, as per the model, is that there was significant amount of external gene flow into the migrant groups. However, under the Indian social system, where caste endogamy is strictly followed, marriage outside the group is prohibited and in fact not observed among these migrant populations (Reddy 1984). Therefore, the unauthorized gene flow, even if it had occurred from outside the fishermen groups, cannot be so significant as to cause such a systematic pattern. Hence, while considering the results of regression analysis, it is necessary to give due focus to the nature of population formation and the mating patterns of the groups under

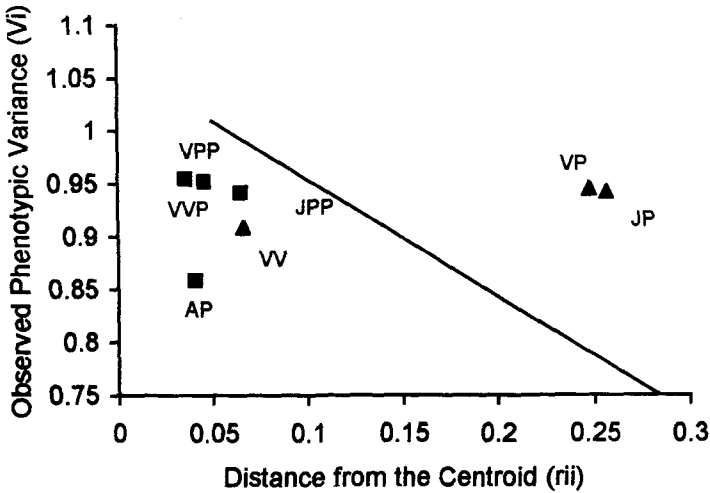


Figure 1. Plot of observed mean phenotypic variance versus anthropometric distance from the centroid of the seven fishermen groups, and the *expected* regression line. The triangles represent migrant populations and the squares the parental populations.

study. There are indications that the nature of distribution of the parental groups, the population formation of the migrant groups at Puri, and the mating interactions within these groups might have been in an intricate way responsible for this increase in heterozygosity, rather than the gene flow from outside the fishermen. Nevertheless, such subtle differences in the population structure generally escape attention of investigators, unless they have a detailed understanding of the population history and mating patterns. In subsequent sections of the paper, I shall focus on certain aspects of the population structure of the migrant groups of fishermen that are relevant in this context.

History of Migration and Population Formation at Puri. All marine fishermen at Puri are migrants, and the present populations are a result of several waves of migration at different times. They come from different stretches on the East Coast of India, south of Puri. The Vadabalija of Vadapeta (VV) and Jalary (JP) migrated a century ago from about 50 coastal villages distributed in Ganjam district of Orissa and in Srikakulam district of Andhra Pradesh. The Vadabalija of Penticotta (VP) migrated some 50 years ago from 48 coastal villages from three districts in Andhra Pradesh, namely, Visakapatnam, East Godavari, and West Godavari. The distribution ranges of the parental areas of the VP on the one hand and of the VV and JP on the other do not overlap, being separated by about 100 miles. While the latter two groups had already developed into self-contained endogamous units, almost

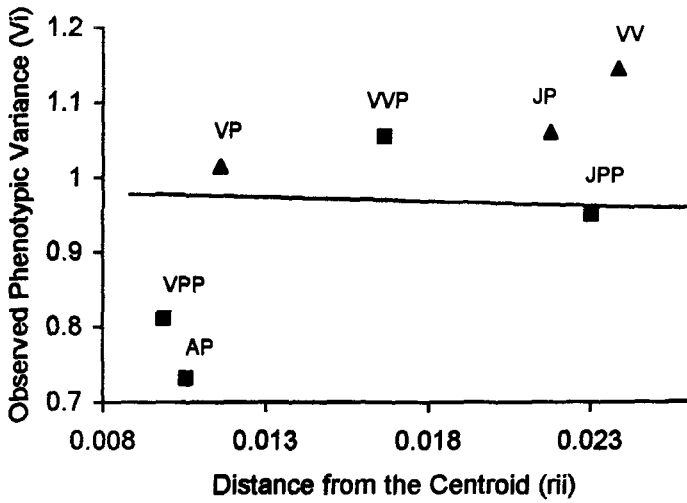


Figure 2. Plot of observed mean phenotypic variance versus dermatoglyphic distance from the centroid of the seven fisherman groups, and the *expected* regression line. The triangles and squares represent migrant and parental populations.

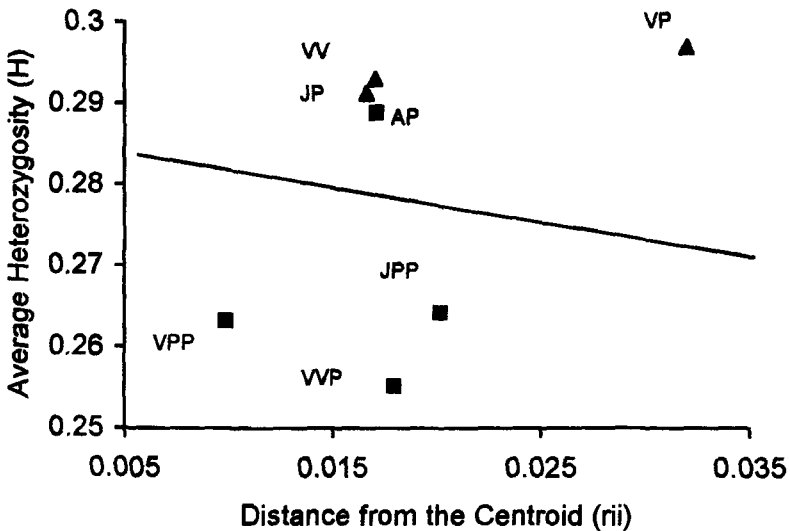


Figure 3. Plot of observed mean heterozygosity versus the genetic distance from the gene frequency centroid of the seven fisherman groups, and the *expected* regression line. The triangles and squares represent migrant and parental populations.

Table 2. Trends in Mean Marriage Distance (MMD) among the Migrant Groups

Generation ^a	VP			VV			JP		
	<i>n</i>	Mean	SD	<i>n</i>	Mean	SD	<i>n</i>	Mean	SD
I (64+)	47	4.6	8.9	46	36.9	39.7	11	19.6	31.9
II (43-63)	229	7.5	21.3	188	32.2	41.5	59	16.3	29.7
III (22-42)	445	9.3	24.8	260	26.7	41.7	100	10.1	25.2
IV (<22)	15	7.9	17.2	8	10.6	28.1	1	70.0	—

a. Based on the age of husbands (age given in parentheses).

Table 3. Trends in Village Endogamy among the Migrant Population of Puri

Generation	VP		VV		JP	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
I	47	53.2	46	28.3	11	72.7
II	229	40.2	188	30.9	59	57.6
III	445	33.7	260	45.4	100	74.0
IV	15	60.0	8	87.5	1	—

totally cut off from the parental villages, the VP still have contacts with their parental villages, not only through occasional marriage but also through the seasonal influx to Puri of fishermen from their parental villages. Every year, during the winter season, fishermen from the parental areas of the VP in Andhra Pradesh come to Puri with their boats and nets on fishing expeditions and go back after the prawn-harvesting season.

The population sizes of the VP, VV, and JP are 8000, 4000, and 800, respectively. Whereas most of the VV and JP fishermen, especially the younger generations, were born in Puri, only a small number of VP, even in the youngest generation of married couples, were born in the new habitat at Puri. While the majority of marriages among the VV and JP are contracted within the new settlements, an increasing proportion of such marriages over generations have been contracted by the VP in the new settlement.

Trend in Mean Marriage Distance and Village Endogamy. Tables 2 and 3 describe the trends in mean marriage distance (MMD) and village endogamy (VE) over the generations. Two distinct trends that are qualitatively different can be seen from these tables:

1. In the case of the VP, the MMD is small but shows an increase from the oldest to the youngest generation. Village endogamy is high but shows a decline from the oldest to the youngest generation.

2. In the case of the VV and JP, this trend is the opposite: mean marriage distance is relatively much larger and shows a declining trend, whereas village endogamy increases progressively from the older to the younger generations.

The qualitative difference in the trends can be explained given the history of migration, settlement patterns, and geographical backgrounds of these groups. For example, the VP, except for the youngest generation of couples who should have been born mostly in the present habitat, will be composed of couples most of whom should have been born in the parental villages and then accompanied their families/spouses to Puri. Therefore, most in the oldest age group would have married before or soon after they had migrated to Puri. Since it is known that endogamy rates are high in these populations, and the marriage contacts are usually restricted to small distances, the small mean marriage distance observed in the oldest generation can be explained. However, subsequently, with increasing settlement in the new habitat and with improving knowledge of and interaction with other families from distant villages, marriages within the new habitat might have gradually increased, giving rise to an increase in MMD over generations.

Families from about 50 villages situated along the 400 miles of coast have assembled in a single settlement at Puri. In a hypothetical model Reddy (1983) explained how social, psychological, and occupational needs would have promoted marital interactions in the new habitat. According to this model, marriage interactions in the new habitat were initially expected to have been guided by two qualitatively different sets of factors, positive and negative. The following factors were identified:

Positive Factors

- *Increasing knowledge.* Through occupational and sociocultural interactions in the unusually large new habitat, families came into increasing contact with other families.
- *Occupational assistance.* Fishing as an occupation relies on the collective efforts of, at least, a couple of families to handle boats and nets while at sea. Marrying within the new settlement would help form homogeneous units for fishing expeditions that usually last for a couple of days.
- *Increased choice of mates.* The large size of the settlement at Puri provides enormous choice of mates.
- *Physical proximity.* As a result of physical proximity at the new habitat, unmarried migrants can avoid the expense and time required for travel to their original places in search of mates.
- *Sociopsychological cohesion.* Group solidarity developed primarily to guard group interests against those of the other two earlier migrant

groups. The VV and JP had initially attacked the VP a couple of times, destroying their boats and nets in a bid to drive the new group away from Puri.

Negative Factors

- *Tradition.* Age-old ties tend to bind them with acquaintances from their original villages, ties that are reinforced by their occasional visits to parental villages.
- *Consanguinity.* The coastal fishermen of Andhra Pradesh are known to practice consanguineous marriages, involving uncle and niece and first cousins, with highest frequency in India (Sanghvi 1966). This would at least initially tempt them to opt for mates from the parental villages.
- *Instability of resources.* The initial uncertainty of resources at the new habitat would have discouraged marriages with the spouses from relatively unknown villages at the new settlement.

The model predicts that the extent of marital interactions within the new settlement is a direct result of the relative intensity of positive and negative factors operating in the migrant situation. However, with time, the negative factors would cease to operate or become weaker, while the positive factors would become overwhelmingly important in encouraging marriage alliances within the new settlement.

Decrease in mean marriage distance and increase in village endogamy among the more established migrant populations, the VV and JP, can be similarly explained. Both groups have been settled in their present habitats for four to five generations. Naturally, therefore, a certain proportion of even the first-generation couples among the VV and JP must have been born in the Puri settlements. Since a marriage between a spouse from the Puri settlements and one from the parental village meant a minimum distance of about 100 km, the MMD is generally larger among the VV when compared to the VP. The small JP population is known to have been completely settled much before the migration of the VV. Without much further influx from the parental places, the JD also show a relatively smaller MMD and much higher rates of village endogamy among them. The increase in the rate of VE reflects the increasing proportion of marriages between the spouses born in Puri settlements. Among the VV and JP, the MMD decreases from older to younger generations, since the marriage distance of couples born in the same village is considered to be zero.

The Genetic Implications of the Marriage Patterns. The migrant populations of Puri were drawn from a large number of villages scattered over an extended coastal stretch. Given that the population distribution along the

coast, especially of these traditional fishermen, is one dimensional, that marriage contacts are usually restricted to a small distance with high rates of village endogamy ($\sim 50\%$), and that close consanguineous marriages with a high degree of inbreeding (average $F = 0.03$) are practised among them, there is a possibility that each of these parental populations contains within it a number of smaller demes, reproductively isolated from each other (isolation by distance). Members of migrant Puri populations are drawn from many of those demes. Since mating interactions within the migrant populations are mostly restricted to Puri settlements, there had been a possibility of intermixing with the geographically and therefore probably genetically more heterogeneous individuals (from different demes) at Puri. Heterozygosity of the migrant populations most probably may have been enhanced when compared to that of the parental counterparts. Samples of these latter come from different demes within each of the parental stocks, hence the samples for them are likely to be mere aggregates of proportionately more homozygous individuals. Such small demes are supposed to be more inbred and homozygous when compared to larger populations. This effect is expected to be more prominent in the VP, who are drawn from a much longer coastal stretch compared to the area from where the VV and JP originate. The observed outliers in the regression plots, above the theoretical regression line, may be seen to reflect this aspect of population structure rather than the external gene flow. This observation is pertinent, because in our detailed genealogical and demographic investigations of the majority of Puri fishermen households we could not find any instance of a fishermen group exchanging mates with a nonfishermen group or, for that matter, even with another caste of fishermen. It is true not only of the current generation but also of the previous generations, as far as could be traced. It may be pertinent to note here that Eller (1999), on the basis of a large number of STR loci, studied population substructure in the three continental populations and observed that the greater heterozygosity of the Africans when compared to populations from the other continents is most probably due to their greater effective population sizes, rather than to greater gene flow into them.

The effect of heterozygosity by either means would be identical genetically but, nevertheless, the social ramifications/implications of the inference that it is due to the external gene flow will be generally inconsistent and inappropriate in the Indian context. Therefore, it is necessary to give due focus to the cultural context of the populations while drawing inferences on such issues.

In passing, I would like to refer to another aspect in the application of *R*-matrix analysis for inferring patterns of gene flow. To the best of my understanding, Harpending and Ward (1982) developed the model keeping in mind the local populations at the level of small regions or the linguistic groups with relatively recent history of common ancestry. In fact, the case studies to which they have applied this model all conform to this implicit assumption.

However, of late, we see that this approach is extended, in my opinion, inappropriately to macro situations, wherein inferences on the peopling of a continent or of the human origins are based on relative positioning of populations around the *expected* regression line. For example, it is not appropriate to use a heterogeneous mixture of populations—some African, some Asian, and some European, along with certain caste populations of India—in the regression plot of heterozygosity versus r_{ii} , and conclude that the Indian population with certain linguistic/ethnic affinity having been placed above the regression line, along with Africans, represents first settlers of the continent!

Acknowledgment I thank John H. Relethford for useful discussion on the issue, and for comments on the initial version of this manuscript.

Received 7 August 1999; revision received 11 May 2000.

Literature Cited

- Devor, E.J., M.H. Crawford, and V. Bach-Enciso. 1984. Genetic population structure of the Black Caribs and Creoles. In *Current Developments in Anthropological Genetics. Vol. 3: Black Caribs: A Case Study in Biocultural Adaptation*, M.H. Crawford, ed. New York, NY: Plenum Press, 365–380.
- Eller, E. 1999. Population substructure and isolation by distance in three continental regions. *Am. J. Phys. Anthropol.* 108:147–159.
- Harpending, H.C., and T. Jenkins. 1973. Genetic distances among Southern African Populations. In *Methods and Theories of Anthropological Genetics*, M.H. Crawford and P.L. Workman, eds. Albuquerque, NM: University of New Mexico Press, 177–199.
- Harpending, H.C., and R. Ward. 1982. Chemical systematics and human evolution. In *Biochemical Aspects of Evolutionary Biology*, M. Nitecki, ed. Chicago, IL: University of Chicago Press, 213–256.
- Jantz, R.L., and L. Meadows. 1995. Population structure of the Algonquian speakers. *Hum. Biol.* 67:375–386.
- Puppala, S., and M.H. Crawford. 1996. Genetic structure and phylogeny of populations of India: Andhra Pradesh. *Homo* 47:73–84.
- Reddy, B.M. 1983. Temporal trends in marriage distance and village endogamy among the migrant groups of fishermen of coastal Orissa. *Coll. Anthropol.* 7:125–135.
- Reddy, B.M. 1984. Demographic structure of the migrant groups of fishermen of Puri coast, India. *J. Biosoc. Sci.* 16:385–398.
- Reddy, B.M., and V.P. Chopra. 1999. Biological affinities between the migrant and parental populations of fishermen on the East Coast of India. *Hum Biol.* 71:803–822.
- Reddy, B.M., V.P. Chopra, B. Karmakar et al. 2000. Quantitative dermatoglyphics and population structure in Northwest India. *Am. J. Hum. Biol.* 12(3) (in press).
- Relethford, J.H. 1985. Isolation by distance, linguistic similarity, and the genetic structure on Bougainville Island. *Am. J. Phys. Anthropol.* 66:317–326.

- Relethford, J.H., and J. Blangero. 1990. Detection of differential gene flow from patterns of quantitative variation. *Hum. Biol.* 62:5-25.
- Workman, P.L., H.C. Harpending, J.M. Lalouel et al. 1973. Population studies on southwestern Indian tribes. VI. Papago population structure: A comparison of genetic and migration analyses. In *Genetic Structure of Populations*, N.E. Morton, ed. Honolulu, HI: University Press of Hawaii, 166-194.