

Relevance of Surnames in Population Structure : A Study Among The Vadde, A Fishing Community of Kolleru Lake in Andhra Pradesh, India

B.M. Reddy and K.C. Malhotra

Anthropometry and Human Genetics Unit, Indian Statistical Institute, Calcutta 700 035, West Bengal, India

KEY WORDS Surnames. Preferential Mating. Population Subdivision. Genetic Heterogeneity.

ABSTRACT With the help of demographic data on 2263 couples from the Vadde fisherfolk of Kolleru Lake in Andhra Pradesh, India, an attempt has been made to demonstrate the relevance of surnames in the study of population structure. Preferential mating between surnames is used to ascertain the extent of departures from random mating, or the extent of breeding isolation within a population. Besides the number of surnames, a measure called surname diversity index is defined and its relationship with other population structural measures like village endogamy, N_e , α , and MMD, and also with fertility and mortality explored. A certain degree of preferential mating is present between the surnames in Vadde indicating significant departure from random mating ($\chi^2 = 265.8$, df. 26, $P < 0.0001$). A linear relationship between the surname diversity index and village endogamy suggests that the greater diversity of surnames provide better scope for choosing mates within the village. A significant relationship of this diversity measure with child mortality is also apparent in the data. The genetic implications of these findings are discussed.

INTRODUCTION

The genetic structure of a community is determined by its composition of genes and the ways in which these genes are combined in genotypes. Evolutionary forces such as selection, mutation, drift and gene flow determine the former, while the latter is determined by mating patterns. A number of sociocultural preferences are known to regulate the mating patterns within any population. The most important among them, especially in Indian situation, is the regulation of exogamy which applies to sections in to which a population is divided. Known as surnames, *gotras* etc., these exogamous units are transmitted through male line. Folk stories among many of these communities make it abundantly clear that the sur-

names or *gotras* have a mythical origin and this structure might have been developed as a response to increasing inbreeding and incest. It is generally believed that members of any surname or gotra have a common origin.

In the process of mate choice, there are several criteria that are followed—social, geographical, economic and biological. The most important of all these is the surnames of the spouses which should be different and are ascertained foremost. There are, however, some studies dealing with isonymous marriages, especially in some western communities (Crow, 1960; Crow and Mange, 1965; Devor et al., 1983; Raspe and Lasker, 1980) and Indian Muslims (Kashyap, 1980) and Bhatias (Bhalla and Bhatia, 1976). The inbreeding coefficient has been estimated based on the data from such marriages. Surnames have also been used to assess the coefficient of biological relationship between different populations (Lasker, 1980). The application of gene diversity analysis to surname diversity data has also been explored (Bhatia and Wilson, 1981).

While it is known that marriage between spouses of the same surname is invariably disallowed, the frequency of mating between different surnames itself is another interesting aspect, for the selective mating between surnames may lead to population heterogeneity and breeding isolation within a population. A possible interrelationship between the nature of distribution of these surnames within any particular village and/or in a population over space may have certain implications to the other population structural measures like mean

marriage distance (MMD), village endogamy (VE) and consanguinity. Village endogamy for example is expected to show a relationship with the diversity of surnames in a village, for the diversity of surnames may reflect the scope for mate choice within a village. This is particularly true in the southern Indian situation, where village endogamy is generally preferred. Similarly, the distance at which one chooses his mate may also depend, at least in part, on the nature of distribution of surnames over space. Thus the mating between surnames and the spatial distribution of surnames of a particular population poses an interesting problem for population geneticists, especially in the Indian situation. There is hardly any in-depth study which deals with the marital interaction between the surnames except for a frequency of matings between clans reported by Rao (1983) among the Gonds of Central India. The present paper therefore studies these aspects among the Vadde, a fishing community of Kolleru Lake in Andhra Pradesh.

Kolleru Lake is an extensive shallow depression formed by a gradual rise in the level of the country on either side by alluvial deposits from Krishna and Godavari rivers. The lake is situated bordering both Krishna and Godavari districts of Andhra Pradesh, and its area stretches up to about 285 sq. miles, during the rainy season. There are about 60 fishing villages in and around this lake. Most of those who live in this area are fishermen and are known as Vadde. They are estimated to number about 30000 individuals (Reddy, 1983). Fishermen, in general, are considered lower in social hierarchy, and Vadde caste is no exception to that. However, they claim that their ancestors were rulers and therefore, they belonged to a *varna* called Agnikula Kshatriya who were warriors. Surprisingly, this is the case with many other fishing groups with whom we had the opportunity to interact with.

The Vadde, an endogamous caste, is stratified into a number of exogamous units called *intiperus* (surnames) which act like

clans/gotras and guide marital interactions. Typical of many other fishing groups, Vaddes also claim to have a single gotra for the entire caste.

The Vaddes of Andhra Pradesh according to ethnographic accounts is a backward community who are by profession stone breakers. The existence of a fishing group called Vadde is not known. Recently, There have been some efforts by Census of India to enumerate this community. Knowledgeable elders from these villages say that they have migrated from coastal areas of Orissa. No apparent indications were available through their language or culture bearing influence of the area from which they have migrated. However, the Oriyas, in general, are referred to as Vaddes in coastal Andhra Pradesh, and therefore, one may surmise that the name given to them was because they were migrants from Orissa.

MATERIALS AND METHODS

Information on caste affiliation birth place, surname, consanguinity, marriage distance and reproductive histories etc. of 2263 couples of the Vadde fishermen of Kolleru Lake, Andhra Pradesh (AP), were collected as part of the project, "Mating Patterns and Reproductive Behaviour of the Vadde, a Fishing Community of Kolleru Lake, A.P.", between February and June, 1982, by interview method. The list of villages along with population sizes and the number of couples studied is given in table 1.

Among Hindus, the wife adopts husband's surname after marriage; however, wife's maiden surname was also recorded in each case. The distances between the birth places of spouses were estimated in miles through travel routes. The local terms for different kinds of relationships were found useful in ascertaining the degree of relationship. Taken casually, this method is deceptive. For example, they generally refer cousin sister (1st cousin) as well as a distant one as sister. During the present investigation many such cases were encountered. Repeated cross questioning proved extremely helpful in ascertaining the exact relationship

Table 1: List of studied villages along with population sizes and the number of couples studied

S. No.	Village	Size of the fishing population	No. of couple studied
1.	Gudaka Lanka	1510	234
2.	Prathikuda Lanka	1430	251
3.	Paidi Chintapadu	450	134
4.	Kornati Lanka	140	40
5.	Sringarayathota	1600	313
6.	Gummala Padu	320	102
7.	Pandiri Palle Gudem	1580	301
8.	Kolleti Kota	80	20
9.	Pedda Kottada	320	89
10.	Chinna Kottada	200	59
11.	Penchakallamarri	560	155
12.	Vadlakuda Thippa	400	130
13.	Daichinta Padu	770	133
14.	Kovvada Lanka	850	146
15.	Penumaka Lanka	700	156
Total		10910	2263

between the couple. Pregnancy histories were collected, interviewing women.

RESULTS

Number and Predominance of Surnames

Table 2 gives number of surnames and their relative strengths in each of the villages and in the total Vadde population. From this table, the following observations can be made.

A total of 25 surnames are encountered which may roughly correspond with the total number of surnames into which the Vadde population is divided. As many as 9 surnames of these 25 have been adapted from the non-Vadde neighbours, in the process of admitting male spouses into the caste. The number of surnames vary between six in Chinna Kottada to 16 in Gudaka Lanka. Of the total number of spouses, 92% belong to the six of the 25 surnames namely Gantasala, Bale, Saidu, Mungara, Jaimangala and Mori in the order of

decreasing predominance. The first two surnames, however, constitute as much as 50% of the total population. On the basis of the predominance of a particular surname, villages 1, 4, 9, 11, 13 and 15 can be assigned to Gantasala, 5, 6, 7 and 8 to Bale, 2 and 10 to Saidu, and 12 and 3 Mungara.

Mating Assortment Between Different Surnames

The frequency of marriages between different surnames can be deciphered from table 3. Of the most frequent six surnames, Gantasala and Bale had marital interaction with the 19 of the remaining 24 surnames, encountered in the present study; the exceptions are some of the non-Vadde surnames who have been recently admitted into the Vadde. The extent of interaction of a particular surname appears to have a direct relationship with its numerical strength; the larger its number it showed interaction with greater number of surnames. It is apparent that over 95% of the marital interactions are confined to the most prevalent 9 surnames which understandably also constitute about 95% of the total population. Another observation worth mentioning is that the Vadde with a particular surname can marry into any other surname than his/her own. This is evident from the fact that the members of Gantasala and Bale had marital interaction with most other surnames.

It is expected that the selective mating between the surnames can lead to subdivisions in the population with an overall increase in the genetic variance, but reduced genetic variance within subunits of the larger population. Therefore, it has been further examined whether the frequency of mating between different surnames conform to random expectations. This is examined using only 9 of the 25 surnames which are most frequent, and constitute over 95% of the total population. The results are presented in table 4. The frequency of mating between Bale and Gantasala, Bale

Table 2: Number and predominance of surnames in each of the villages and in the total population

S. No.	Surname	Village Number															Total	%	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
1.	Bairava Setty*						2											2	0.00
2.	Bale	50	81	5	8	156	37	184	15	9	21	2	6	36	19	31	660	22.90	
3.	Chintadi*					1												1	0.04
4.	Chintapalle*	1																1	0.04
5.	Erra*									5								5	0.20
6.	Gantasala	69	57	23	18	101	21	121	11	54	13	87	28	77	46	55	781	27.00	
7.	Govinda	9	6	1	3													19	0.70
8.	Guradasi	1			4			10			1		3		1			20	0.70
9.	Jaimangala	40	3	14	4	18		34	5	16		66	10	11	63	24	308	10.70	
10.	Jalluri	4	2		4	17		1				4						32	1.10
11.	Karanam		3			1		18										22	0.80
12.	Korlabandi*	1																1	0.04
13.	Malle									5				3				8	0.30
14.	Mandala														2			2	0.07
15.	Mori	48	12	43	2	12	9	5	3			8		16	32	24	214	7.40	
16.	Mungara	51	17	46	3	19	1	18	2		1	27	61	30	13	27	316	10.90	
17.	Nabigari	1				8	15	6	2				11			1	44	1.50	
18.	Nepala	5		2		22	1	9				2	7			1	49	1.70	
19.	Pantula*								1									1	0.04
20.	Paravalla*	2																2	0.07
21.	Pinnam*	4																4	0.10
22.	Pulluri*	2																2	0.07
23.	Saidu	37	105	32	8	43	35	14	1	10	35	11	31	9	3	9	383	13.30	
24.	Sirbiri					7	1											8	0.30
25.	Vanamala		1				1			2								4	0.10
Total		325	287	166	54	405	123	420	40	96	76	207	157	182	179	172	2889		

*surnames of non-Vadde origin

and Saidu, Bale and Mori, and Bale and Mungara etc. are considerably greater than expected, while between some other surnames the observed frequencies are considerably lower. The goodness of fit between observed and expected mating frequencies is poor and the χ^2 value is rather high ($\chi^2 = 265.83$, $df.26$, $P < 0.001$), indicating significant departure from random mating between the surnames.

The Surname Diversity Index (D) Defined

In order to quantify the diversity existing in respect of surnames in a village, the measure that can be used is the number of different surnames. This measure, however, does not

take into account the relative frequencies with which different surnames occur. Thus, if there are two villages, each with N couples and in each of which there are k- surnames, then with respect to surnames both these villages will be equally diverse irrespective of the frequencies with which these k surnames occur. This is not the best way of measuring diversity. Analogous with the average Heterozygosity measure (Nei, 1975) used in population genetics, we have defined the Surname Diversity Index as

$$D = 1 - \sum_{i=1}^k p_i^2$$

Table 4: Mating assortment between the 9 most frequent surnames among the Vadde which constitute over 95% of the total matings

S. No.	Pair of Surnames	Observed No.	Expected No.
1. Bale	X Gantasala	381	295.15
	X Jaimangala	132	123.50
	X Mori	62	102.26
	X Mungara	80	127.77
	X Saidu	233	159.35
	X Nabigari	8	14.38
	X Nepala	24	21.24
	X Jalluri	3	15.17
2. Gantasala	X Jaimangala	121	137.56
	X Mori	150	113.90
	X Mungara	180	142.32
	X Saidu	144	177.49
	X Nabigari	11	16.02
	X Nepala	18	23.66
3. Jaimangala	X Jalluri	23	16.90
	X Mori	66	47.66
	X Mungara	82	59.55
	X Saidu	17	74.27
	X Nabigari	9	6.70
	X Nepala	2	9.90
4. Mori	X Jalluri	1	7.07
	X Mungara	14	49.31
	X Saidu	57	61.49
	X Nabigari	1	5.55
	X Nepala	5	8.20
5. Mungara	X Jalluri	1	5.85
	X Saidu	70	76.83
	X Nabigari	7	6.94
	X Nepala	9	10.24
6. Saidu	X Jalluri	3	7.31
	X Nabigari	10	8.65
	X Nepala	8	12.77
7. Nabigari	X Jalluri	16	9.12
	X Nepala	3	1.15
8. Nepala	X Jalluri	1	0.82
	X Jalluri	5	1.22

$\chi^2 = 265.83$; d.f. 26; $p < 0.001$

regression analysis. Considering the slope of the scatter, and the nature of theoretical relationship between the two measures we set up the regression equation of the type

$$Y = \beta_x$$

which becomes

$$\%VE = \beta_D$$

The estimated β value is 30.59. This value is inserted into the above equation and the expected values of VE are calculated for each of the observed ones. ANOVA yields a F-ratio of 2.08 with 1 and 10 degrees of freedom which is statistically nonsignificant at 5% level (Table 6). The variation explained by regression is statistically not sufficient for the linearity assumption to hold good in this case.

The r-value is also very high between the D-value and the infant mortality rates. In order to see how far the heterogeneity of surnames within a village explains the infant mortality rate, regression analysis was carried out. The equation set up was

$$Y = \alpha + \beta_x$$

which becomes

$$\% \text{ offspring mortality} = \alpha + \beta_D$$

The constants, α and β , are 14.80 and 53.15, respectively. The F-value obtained by ANOVA (5.67 with 1 and 12 df) is significant at 5% level. Therefore, it implies that the surname diversity index is useful in predicting offspring mortality.

DISCUSSION

The results of the foregoing analysis bring out the following observations that need further discussion. Although a total of about 25 surnames are encountered, the predominant two surnames, the Gantasala and Bale, constitute 50% of the total population. Of the remaining, while four surnames constitute 42% the remaining 19 constitute only 8% of the total. Out of these 19, 9 surnames have non-Vadde origin and constitute only 1% of the couples. What does such a predominance indicate? Is it due to differential perpetuation of different surnames? It is hard to comprehend how such a wide difference can be accounted for by this process. There could be two possible explanations for this—firstly, since in the present study the entire range of distribution of the Vaddes could not be covered, it is likely that the surnames with small numerical strengths occur more frequently in unsurveyed villages.

Table 5: Effective population size, no. of surnames, D-value, MMD, VE, percentage of consanguineous marriages, inbreeding coefficient (α), mean no. of live births and % offspring mortality for each of the villages of Vadde

S. No.	Village	Effective population	No. of surnames	D-value	MMD	%VE	% Consanguineous	α	Mean no. of offspring live births	% mortality
1.	Gudaka Lanka	253	16	0.8553	5.64	39.3	31.04	0.019	5.96	16.3
2.	Prathikuda Lanka	228	10	0.7411	7.20	19.9	39.05	0.022	5.81	22.6
3.	Paidi Chintapadu	78	8	0.7916	6.60	23.3	31.34	0.020	7.28	29.3
4.	Komati Lanka	35	9	0.8210	4.90	35.0	37.50	0.021	7.19	31.9
5.	Sringaraya Thota	283	12	0.7677	9.13	24.9	27.33	0.017	5.66	20.5
6.	Gummalapadu	77	10	0.7786	9.56	14.7	30.69	0.018	5.07	21.6
7.	P.P. Gudem	262	11	0.7124	6.00	43.2	34.24	0.020	5.85	20.1
8.	Pedda Kottada	85	6	0.6330	14.76	6.7	31.46	0.017	5.82	21.9
9.	Chinna Kottada	40	6	0.6776	10.00	18.6	27.12	0.014	6.85	16.7
10.	Penchakallamarri	120	8	0.6998	7.27	37.4	32.45	0.019	4.84	16.5
11.	Vadlakuda Thippa	63	8	0.7655	8.78	24.6	24.22	0.013	7.00	24.6
12.	Dai Chintapadu	149	7	0.7406	5.02	41.4	39.69	0.025	6.55	24.1
13.	Kovvada Lanka	189	8	0.7612	5.75	33.6	28.08	0.013	5.45	24.9
14.	Penumaka Lanka	131	8	0.7989	9.95	12.2	7.14	0.024	6.77	23.2

Table 6: The values of r showing the relationship of number of surnames and D-value with Ne, MMD, VE, % consanguineous marriages, α , live births, % mortality and It

S. No.	Pair of variables	$r \pm SE$
1.	Ne : Number of surnames	0.71±0.13*
2.	Ne : D-value	0.15±0.26
3.	No. of surnames : D-value	0.58±0.18*
4.	D : MMD	-0.54±0.19*
5.	D : % VE	0.68±0.16*
6.	No. of surnames : % VE	0.41±0.22*
7.	D : % consang	0.17±0.27
8.	D : α	0.27±0.25
9.	D : live births	0.27±0.25
10.	D : Offspring mortality	0.63±0.15*
11.	D : It	0.31±0.24

*significant at $p < 0.05$

Table 7: Analysis of variance for linear regression of VE on Surname Diversity Index, (D)

Source of variation	SS	df	MS	F
Due to regression	163.67	1	163.67	2.08
Residual	788.91	10	78.89	
Total	952.58	11		

But such a possibility is unlikely because the mating contacts are spread throughout the entire range of the population distribution and assuming that the exchange between any two villages is balanced, such a possibility can be safely ruled out. It needs to be, however, empirically verified.

In explaining the alternative possibility, historical background seems important. The Vaddees are said to be migrants from Orissa. It is likely that although the migration occurred in the remote past, it was selective in respect of surnames or that the migration was limited to only those six predominant surnames and the less frequent ones had their origin from the surrounding local populations due to the admixture as noted earlier.

Table 8: Analysis of variance for linear regression of % offspring mortality on Surname Diversity Index

Source of variation	SS	df	MS	F
Due to regression	126.90	1	126.90	5.67
Residual	268.82	12	22.40	
Total	359.72	13		

Similarly, the results concerning mating frequency between different surnames shows that the mating assortment is highly selective, since significantly larger number of matings are observed compared to the expected between some surnames like Gantasala, Bale, Saidu and Mungara. This situation, as mentioned earlier, is expected to lead to population subdivision with an overall increase in genetic variance, but with reduced genetic variance within subunits.

Several investigators have earlier suggested that there exists a relationship between population and village size on the one hand and village endogamy and/or α on the other. This was thought true because bigger the village greater the opportunity for mate choice. But none seems to have considered the number and variability of surnames in any particular village to explain the variation observed with respect to village endogamy and α . The greater diversity of surnames represented by larger D will give greater scope for mate choice within the village. As has been found, the village size may become important in the frequency of VE only through its relationship with number and strength of surnames. It is logical, therefore, to assume that the surname diversity index would be a more meaningful measure to explain the village endogamy. Although the linear trend is apparent, except for a couple of outliers, the predictive value of D over VE is statistically not sufficient in the small sample of 12 villages.

This index will be zero when there is only one surname in a village, and one, at least effectively, when there are a large number of surnames with roughly equal strengths. While this index is useful in understanding marriage alliances within any particular villages, there are many other factors that underlie mate choices and perhaps, therefore, a linear relationship is not obtained with %VE.

This situation may also lead to an interesting hypothesis concerning the distribution of surnames in the villages in the diverse cultural background of the Indian subcontinent. In areas where village endogamy is preferred, the villages are expected to have at least two and often several surnames (belonging to a single

caste). Following the above logic it is expected that the northern Indian villages where VE is discouraged usually a single surname would be found. In the absence of empirical data, this remains to be an interesting hypothesis to be investigated in future.

The predictive relationship obtained between Surnames Diversity Index and rate of offspring mortality remains a curious finding. The kind of empirical relationship obtained between them is, however, difficult to interpret biologically. The only plausible explanation seems that the greater D-value reflects greater scope for village endogamy and α and resulting in greater homozygosity which might cause greater mortality. This at best a conjecture which needs to be further verified with genetic data.

REFERENCES

- Bhalla, V. and Bhatia, K.: Isonymy in a Bhatia Leut. *Ann. Hum. Genet.*, 39: 479-500 (1976).
- Bhatia, K. and Wilson, S.R.: The application of gene diversity analysis to surname diversity data. *J. Theor. Biol.*, 88: 121-133 (1981).
- Crow, J.F.: Estimation of inbreeding from isonymy. *Hum. Biol.*, 52: 1-14 (1960).
- Crow, J.F. and Mange, A.P.: Measurement of inbreeding from the frequency of marriages between persons of the same surnames. *Eug. Quart.*, 12: 199-203 (1965).
- Devor, E.J.: Matrix method for the analysis of isonymous and nononymous surname pairs. *Hum. Biol.*, 55: 277-288 (1983).
- Kashyap, L.K.: Trends of isonymy and inbreeding among the Ahmadiyyas of Kashmir. *J. Biosoc. Sci.*, 12: 219-225 (1980).
- Lasker, G.W.: Surnames in the study of human biology. *Amer. Anthropol.*, 82: 525-538 (1980).
- Nei, M.: Analysis of gene diversity in subdivided populations. *Proc. Nat. Acad. Sci., USA*, 70: 3321-3323 (1973).
- Rao, V.R.: Estimation of certain genetical demographic parameters for the Naik Gonds. *Hum. Sci.*, 32: 275-283 (1983).
- Raspe, P. and Lasker, G.W.: Biological relationships by isonymy among the Scilly Isles and of Scilly Isles with the rest of England and Wales. *Ann. Hum. Biol.*, 7: 401-410 (1980).
- Reddy, B.M.: *Mating Patterns and Reproductive Behaviour Among the Vadde, a Fishing Community of Kolleru Lake, Andhra, Pradesh, India*. IST Tech. Report No. 3 (1983).