

SHORT REPORTS

Variation in allele frequencies among caste groups of the Dhangars of Maharashtra, India: An analysis with Wright's F_{ST} statistic

RANAJIT CHAKRABORTY, ARAVINDA CHAKRAVARTI
and KAILASH C. MALHOTRA

*Center for Demographic and Population Genetics, University of Texas at Houston
and Anthropometric and Human Genetic Research Unit, Indian Statistical Institute,
Calcutta, India*

[Received 18 January 1977]

Summary. Variations in allele frequencies at 11 serological and biochemical loci among 21 endogamous Dhangar castes of Maharashtra, India, have been studied using Wright's F_{ST} statistic. Gene differentiation among these castes is only about 4 per cent at these polymorphic loci. The distributional form of F_{ST} is explicable by random genetic drift at this early stage of differentiation of these caste groups. The empirical relationship between the average allele frequencies and the ratio of within to total gene diversity of the caste groups is also studied.

1. Introduction

The Dhangar caste-cluster of Maharashtra, consisting of 23 endogamous castes, constitutes a significant component of the nomads in India. During the past seven years we have engaged in intensive multidisciplinary studies on these caste groups to characterize their ethnographic, cultural, social, and biological make-up. The ethnographic accounts and the archaeological remains indicate that the ancestors of the present three million Dhangars have migrated into this western state of India from the northwestern part of the country around 5000-10 000 years ago, and possibly represent more than one wave of migration into their present habitats (Malhotra, 1976; Malhotra, Chakraborty and Chakravarti, 1977). The marriage pattern and the social organization of the Dhangars also lead us to believe that apart from caste endogamy, a further degree of subdivision exists among each caste at the administrative level of the talukas, each taluka containing one to several villages. Recently we (Malhotra *et al.*, 1977) examined the genetic differentiation among these castes by using gene frequency data on several polymorphic serological and biochemical loci and observed that the average genetic distance as measured by Nei's (1972) measure is only 0.01 net codon difference per locus. The purpose of this note is to study the population structure of the Dhangar caste-cluster by using the same data employing F -statistics (Wright, 1943, 1951).

2. Materials

The present analysis is based on 11 serological and biochemical systems (A_1A_2BO , MN, Le(a), Le(b), P, Rh, Hp, Hb, LDH, Cp and AcP) pertaining to 21 of the 23 endogamous castes of the Dhangars. The gene frequencies for each system are

published in Das, Malhotra, Mukherjee and Das (1974), Mukherjee, Das and Malhotra (1977), Undevia, Malhotra and Dhodwalla (1973), and Malhotra, Das, Mukherjee, Undevia and Bhanu (1976).

3. Results and discussion

When a population consists of infinitely many subpopulations, each with effective size N , Wright (1943, 1951) obtains $F_{ST} = \sigma_x^2 / \bar{x}(1 - \bar{x})$, where \bar{x} and σ_x^2 are the mean and variance of frequencies of an allele among subpopulations. In practice, however, estimates of F_{ST} will vary from locus to locus, and in fact in a locus with more than two alleles, each allele generates a different estimate of F_{ST} . We may further note that any natural population in fact consists of only a finite number of subpopulations. So, F_{ST} as defined above is also subjected to random genetic drift and is expected to show considerable variation. The expected mean and variance of F_{ST} in such events have been worked out recently under the effects of genetic drift with and without an island model type of migration (Nei and Chakravarti, 1977; Nei, Chakravarti and Tateno, 1977).

Allele	Number of castes	Mean allele frequency (\bar{x})	Variance of allele freq. (σ_x^2) $\times 10^{-3}$	F_{ST}	
O	21	0.559	6.5	0.026	1.019
A ₁	21	0.168	2.4	0.017	1.014
A ₂	21	0.030	0.2	0.008	1.006
B	21	0.242	4.5	0.025	1.018
M	19	0.582	4.8	0.020	1.013
Le(a+)	17	0.130	2.0	0.018	1.011
Le(b+)	17	0.341	3.8	0.017	1.014
P	17	0.424	7.8	0.032	1.026
R ₁ (cDe)	15	0.504	21.5	0.086	1.023
R ₁ (cDE)	15	0.014	0.3	0.020	1.016
R ⁺ (Cde)	15	0.083	11.4	0.150	1.017
R ₂ (cDE)	15	0.087	1.8	0.023	1.052
R ₀ (cDE)	15	0.147	11.5	0.092	1.034
r(cde)	15	0.155	5.3	0.040	1.016
R ⁺ (cdE)	15	0.009	0.2	0.020	1.019
R ₁ (CdE)	15	0.001	0.02	0.016	1.014
Hp ¹	18	0.166	3.3	0.024	1.016
Hb ^{normal}	10	0.961	3.7	0.099	1.082
L.DH ^{normal}	10	0.914	5.5	0.070	1.036
Cp ^B	8	0.965	3.8	0.110	1.112
AcP ^A	9	0.163	1.0	0.007	1.006

Table 1. F_{ST} values for the different allele frequency distributions among caste groups of the Dhangars of Maharashtra.

We first compute the F_{ST} values for each allele of the eleven polymorphic loci from the gene frequencies of 21 Dhangar caste groups. However, for the systems with two alleles, F_{ST} values are the same for each allele and so, in table 1 we present the computations for only one of the two alleles for such loci. Furthermore, the gene frequency data are not available for all the subpopulations for each system. Instead

of restricting ourselves to only common subpopulations, we have taken all the available data to get as much information as possible. This gives 21 F_{ST} values corresponding to the alleles shown in table 1. From the F_{ST} values so obtained, we first calculate the mean and variance between them as

$$F_{ST} = \frac{1}{m} \sum_{i=1}^m F_{ST}(i),$$

and

$$s_{F_{ST}}^2 = \frac{1}{m} \sum_{i=1}^m [F_{ST}(i) - F_{ST}]^2$$

where m is the number of estimates of F_{ST} , and $F_{ST}(i)$ is the estimate of F_{ST} obtained from the i th allele. From table 1 we have $m=21$, $F_{ST}=0.0438$, $s_{F_{ST}}^2=0.0015$. Thus, the 21 Dhargar castes have an average F_{ST} of 0.044 ± 0.009 . Alternatively, the average F_{ST} may be computed as

$$F_{ST} = \frac{\sum_{i=1}^m s_x^2(i)}{\sum_{i=1}^m \bar{x}_i(1 - \bar{x}_i)} = 0.0399.$$

The mean value of F_{ST} for both the methods of calculations shows rather similar values, around 4 per cent. The fact that only highly polymorphic systems are considered for this analysis (average heterozygosity around 39–47 per cent) further confirms our assertion that the gene differentiation among the Dhangers is only at an early stage (Malhotra *et al.*, 1977).

Although not all of the F_{ST} values of table 1 are independent, the empirical distribution of F_{ST} values (figure 1) has an interesting implication when compared with

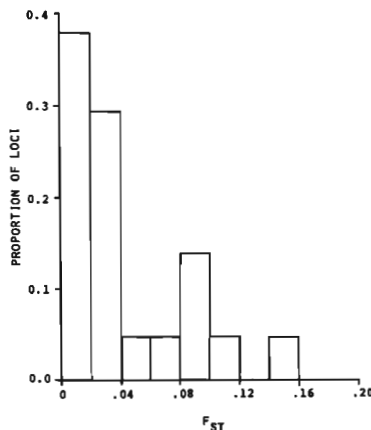


Figure 1. Frequency distribution of F_{ST} values among the Dhangers of Maharashtra.

the simulation results of Nei and Chakravarti (1977). The distribution as shown in figure 1 is *L*-shaped. The harmonic mean of the effective sizes of these populations is estimated as $N=4000$. Taking a tentative estimate of t as 250–500 generations (approximately 5000–10 000 years), we have $T=t/2N$ as 0.031–0.063. The distributional form of F_{ST} at such an early stage of differentiation is also seen to be similar (*L*-shaped) in simulations of Nei and Chakravarti (1977).

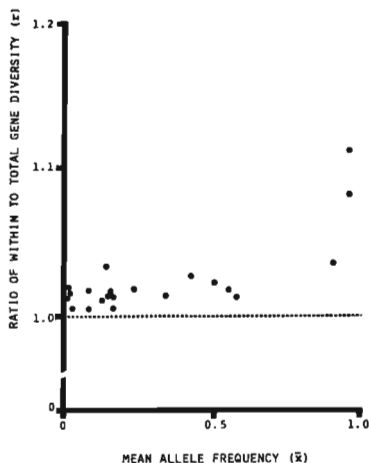


Figure 2. Empirical relationship between the gene frequency (\bar{x}) and the ratio of within-caste gene diversity to the total gene diversity (r) among the Dhangars of Maharashtra.

Recently Maruyama (personal communication) suggested another method of analysis which relates the ratio of total gene diversity to within subpopulation diversity with the average frequency of an allele under a mutation-drift balance. Mathematically, the average frequency of an allele, $\bar{x} = \sum N_i x_i / N$ and, the ratio of total gene diversity to within subpopulation diversity, $r = \bar{x}(1 - \bar{x}) / H(x)$, where $H(x) = \sum N_i x_i (1 - x_i) / N$, x_i being the frequency of the allele in the i th subpopulation; N_i , the effective size of the i th subpopulation and N , the total effective size of the entire population. Note that x and r can be computed for each allele from the data presented in table 1. Figure 2 represents the empirical relationship of x and r as obtained from the 21 allele frequencies of table 1. Two factors prevent us from showing the theoretic relationship between x and r . First, Maruyama considers a mutation-drift balance which cannot be achieved by examining only a set of highly polymorphic systems as in our case. Secondly, the subpopulations of Dhangars (the different caste groups) are not yet separated for a considerable period of time to come to a steady-state stage of differentiation. However, we note that with considerable inter-caste migration the expected relationship should have been represented essentially by a straight line (shown by a dotted line in figure 1) which is the

case for a panmictic population. The expected relationship is in fact a bell-shaped one symmetric about $x=0.5$. Acknowledging the limitations of our data, the figure at least indicates that inter-caste gene migration is not substantial. Direct estimate of inter-caste admixture rate is also seen to be rather small (of the order of 1 in 10 000 or more marriages per year).

One may be tempted to argue that the F_{ST} values computed as above may be used to detect the effect of selection pressures on genetic differentiations among Dhangers through Lewontin and Krakauer's (1973) k -statistic. In view of the historic relationship of the different Dhangar caste groups, and because of our lack of knowledge regarding the gene frequencies among the ancestors of the Dhangers we shall avoid such arguments since inferences reached from the k -statistic are shown to be rather inconclusive in some recent findings (see, e.g. Nei, 1975; Nei and Maruyama, 1975; Robertson, 1975; Nei and Chakravarti, 1977; Nei *et al.*, 1977). On the other hand, the present analysis reaffirms our assertion that gene differentiation among the Dhangers is at an early stage, and roughly is in conformity with expected differentiation by random genetic drift alone.

Acknowledgments

The field work on which the analysis is based was supported by a grant from the University Grants Commission, New Delhi, India. The data analysis was possible through U.S. public health service grants from the United States National Institutes of Health. We are grateful to Professors W. J. Schull and M. Nei for their comments and interest in this work.

References

- Das, S. R., Malhotra, K. C., Mukherjee, B. N., and Das, S. K. (1974). Genetic variants of lactate dehydrogenase in Dhangers of Maharashtra. *Japanese Journal of Human Genetics*, **18**, 305-309.
- Lewontin, R. C., and Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175-195.
- Malhotra, K. C. (1977). Ethnography of the Dhangers of Maharashtra (in preparation).
- Malhotra, K. C., Das, S. K., Mukherjee, B. N., Undevia, J. V., Bhanu, B. V., Yee, S., and Majumdar, P. P. (1976). Distribution of blood groups among 21 Dhangar castes of Maharashtra. *ISI Tech. Rep. Anthropol.* 10/76.
- Malhotra, K. C., Chakravarty, R., and Chakravarti, A. (1977). Gene differentiation among the Dhangar caste-cluster of Maharashtra, India. *Human Heredity* (in the press).
- Mukherjee, B. N., Das, S. K., and Malhotra, K. C. (1976). Serum protein and enzyme polymorphism in the Dhangers of Maharashtra. *Indian Journal of Medical Research*, **64**, 1240-1244.
- Nei, M. (1972). Genetic distance between populations. *American Naturalist*, **106**, 283-292.
- Nei, M. (1975). *Molecular Population Genetics and Evolution*. North Holland-American Elsevier, Amsterdam.
- Nei, M., and Chakravarti, A. (1977). Drift variances of F_{ST} and G_{ST} statistics obtained from a finite number of isolated populations. *Theoretical Population Biology* (in the press).
- Nei, M., Chakravarti, A., and Tatenno, Y. (1977). Mean and variance of F_{ST} in a finite number of incompletely isolated populations. *Theoretical Population Biology* (in the press).
- Nei, M., and Maruyama, T. (1975). Lewontin-Krakauer test for neutral genes. *Genetics*, **80**, 395.
- Robertson, A. (1975). Gene frequency distributions as a test of selective neutrality. *Genetics*, **81**, 775-785.
- Undevia, J. V., Malhotra, K. C., and Dhodwalla, F. (1973). G-6-PD deficiency and abnormal haemoglobin among the Dhangers of Maharashtra. *Proceedings of the 28th Annual Conference of Indian Physicians*, Vol. 41.
- Wright, S. (1943). Isolation by distance. *Genetics*, **28**, 114-138.
- Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, **15**, 323-354.

Address correspondence to: Dr. Ranajit Chakravarty, Center for Demographic and Population Genetics, P.O. Box. 20334, Houston, Texas 77025, U.S.A.

Zusammenfassung. Die Variationen der Allelfrequenzen von 11 serologischen und biochemischen Loci bei 21 endogamen Dhangar-Kasten in Maharashtra in Indien wurden mit Wrights F_{ST} -Statistik untersucht. Die Gendifferenzierung in den Kasten beträgt nur etwa 4 Prozent bei diesen polymorphen Loci. Die Verteilungsform von F_{ST} ist durch zufällige genetische Drift im Frühstadium der Differenzierung dieser Kastengruppen erklärbar. Der empirische Zusammenhang zwischen den durchschnittlichen Allelfrequenzen und dem Verhältnis von Binnen- zu Gesamt-Gendifferenzierung der Kastengruppen wird ebenfalls untersucht.

Résumé. La variation des fréquences alléliques à 11 locus sérologiques et biochimiques parmi 21 castes endogames Dhangar du Maharashtra (Inde) a été étudiée à l'aide de la statistique F_{ST} de Wright. La différenciation génique parmi ces castes est seulement d'environ 4 pour cent à ces locus polymorphiques. La forme de la distribution de F_{ST} peut s'expliquer par la dérive génétique aléatoire à ce stade précoce de différenciation de ces castes. La relation empirique entre les fréquences alléliques moyennes et le quotient de la diversité génique à l'intérieur des groupes à la diversité génique totale est aussi étudiée.