

The Evolution of Surnames: An Analysis of Their Distribution and Extinction*

N. YASUDA,[†] L. L. CAVALLI-SFORZA, M. SKOLNICK, AND A. MORONI[‡]

Department of Genetics, Stanford University

Received July 30, 1973

INTRODUCTION

Theoretical developments around surnames have occurred along two main lines. One of them is the relation between identity of surnames of spouses and their genetic relationship. It was noted about a 100 years ago by G. Darwin (cited by Morton, 1969) that identity of surnames of spouses can help to ascertain if they are relatives. For instance, first cousins have identical surnames when both are children of brothers. On some simplifying assumptions, this is expected to happen in one-fourth of all cousins. Crow and Mange (1965) have recently extended this thinking by noting that the frequency of identical surnames among spouses who are relatives is, again under the same assumptions, equal to four times their inbreeding coefficient. Moreover, a comparison of the *isonymy*, i.e., equality of surnames, of spouses with that expected if mating were entirely random could supply information on the genetic structure of a population.

This consideration has raised considerable interest in the analysis of surnames, usually a much easier task than all other methods of study of population structure, ranging from full pedigrees to gene and/or genotype frequency distributions. No attempt to review the literature will be made here. A few relevant conclusions may be cited, however.

A comparison of observed and expected (random) isonymy in Japan, for instance (Yasuda and Furusho, 1971), has shown very clearly the preference for matings between relatives, which is a characteristic, though disappearing, social custom in that country. In one case, it has been shown, on the basis of historical data, that the distribution of surnames has reached equilibrium, as judged from the trend of random isonymy in time, but also showed some changes with time, which may be easily interpreted. On the same occasion (Cavalli-Sforza and Bodmer, 1971), an analysis of possible sources of error in carrying over conclu-

* Research supported by USAEC Grant No. (04-3)326 #34 and USPHS 10452-09.

[†] Research supported by WHO research training grant. Address: National Institute of Radiological Sciences, Angawa, Japan.

[‡] Address: Istituto di Zoologia, Università, Parma, Italy.

sions from surnames to genes has been given. Of these sources of error, Morton and Hussels (1970) have given importance to the fact that many surnames were, at the start, identical (polyphyletism).

Another line of research has dealt with the extinction of surnames. Again, interest in this problem arose in the last century and gave rise to early theoretical developments like the Galton-Watson branching process. More recently, Lotka (1931, 1939) and Steffensen (cited by Lotka, 1939) have given attention to this problem. Application to real data of extinction of surnames, however, has been scarce.

This paper is concerned with the application of modern theory to data on surnames that were collected in a survey of a population in the Parma Valley, Northern Italy. (For other aspects of this research, see L. Cavalli-Sforza, 1969.) In particular, our attention was directed to the fact that surnames can be considered as alleles of one locus, transmitted only along the male line. Since individual surnames usually can be considered to have no special advantages or disadvantages (with rare exceptions), it was anticipated that they might be considered as not being directly affected by natural selection, i.e., to be selectively neutral. The development of a complete theory of the behavior of neutral mutations in finite populations of constant size by Karlin and McGregor (1967) prompted us to test the applicability of this theory to surnames. This is useful also for understanding the possibility of extending conclusions obtained from surnames to population genetic structure.

2. THE MODEL OF NEUTRAL MUTATIONS APPLIED TO SURNAMES

In many cultures, surnames are transmitted regularly from the father to all of his children (patrilineally). In other cultures, rules of transmission may differ; some of the exceptions can be accommodated in the theory. Here we will consider only patrilineal transmission.

For the purposes of the model, one can study the male population and consider it formally as made of haploid organisms reproducing asexually.¹ In practice, surnames of females may be useful, and in some cases, as in the example that we shall see, it may be convenient to use surnames of females rather than those of males, but this is only for estimation purposes in a special situation. As only males are considered in the model, the effective population size will be roughly half as large as that of the same population considered bisexually.

We will largely restrict our attention to the case when population size is constant over time, as assumed in the model by Karlin and McGregor (1967)

¹ In such a system, surnames behave formally like alleles of a gene located on Y chromosomes.

for "the number of mutant forms maintained in a population". Their symbolism will also be employed. Karlin (1969) has also given a treatment of the same model with changing population size, but it will not be used in this paper.

The model assumes a population of N individuals, only males in our case, in which m_i is the number of individuals carrying the i th surname (in our case; the i th allele in the general model). There exist r different surnames. Composition of the population in a given instant can be represented by the vector:

$$(m_1, m_2, \dots, m_i, \dots, m_r),$$

where $\sum^r m_i = N$, and some of the m_i values (but not all) may be zero if the surname is not represented.

Karlin and McGregor suggest two models, and we will use their Model II in its simpler version.

At every unit of time, there is a transition which may lead to a birth of a new individual, in which case another individual disappears to keep N constant. In this transition, with probability $1 - p_1 - p_2$, there is no change (the population vector remains unaltered). With probability p_1 , one individual with the same surname as his parent is formed, and therefore one of the m_i values increases by one, while another one decreases by one (the choices being random). With probability p_2 , the new individual formed acquires a new surname (any of the $r - 1$ residual ones). The probabilities p_1 and p_2 are defined in terms of a parameter β , so that $r\beta = \nu$ is the overall mutation rate per unit of time (i.e. the probability of change of surname). The chance of reproduction and mutation is equal for all individuals, and therefore this has been called the model of (selectively) neutral mutations.

In a more elaborate version of their Model II, Karlin and McGregor have considered that the number of new individuals formed at every time unit has a specified probability distribution. If this is Poisson, results are identical (at least under some conditions) with those of the simpler version; otherwise, a correction of the results is necessary.

In the application to surnames, we have made use of a series of data collected in the upper Parma Valley that will be further characterized later. In this, as in practically all other cases, surnames are very numerous. In particular, about 3000 different surnames were recorded in a period of over 300 years in an area where the population size rarely exceeded 15,000 individuals. It thus seems useful to use formulas deduced from the limiting case in which r , the total number of surnames, tends to infinity. We will consider some possible consequences and limitations of this assumption in Section 6.

The problem of the distribution of progeny size will be considered in more detail in Section 4, but we may anticipate that it seems to give rise to no major problem.

Finally, the model can be used, extending the parameter ν , to estimate mutation plus migration, rather than mutation alone.

With these qualifications, we will give here the formulas which have been employed from Karlin and McGregor's theory of neutral mutations.

The theory allows us to predict the distribution of $N^*(k)$, the number of surnames, which are represented by k individuals, i.e., for which $m_i = k$. The expected value of $N^*(k)$ is, for $r \rightarrow \infty$,

$$E[N^*(k)] = \frac{1}{k} \frac{N\nu}{1-\nu} \frac{\left(\frac{N}{1-\nu} - (k+1) \right)}{\frac{N-k}{\left(\frac{N}{1-\nu} - 1 \right)}}. \quad (2.1)$$

The mean of this distribution is equal to the expected number of surnames in the population, n_s . Thus,

$$E(N^*) = \frac{N\nu}{1-\nu} \left\{ \frac{\Gamma' \left(\frac{N}{1-\nu} \right)}{\Gamma \left(\frac{N}{1-\nu} \right)} - \frac{\Gamma' \left(\frac{N\nu}{1-\nu} \right)}{\Gamma \left(\frac{N\nu}{1-\nu} \right)} \right\}. \quad (2.2)$$

The second moment about the origin of (2.1) is

$$I = 1/(N\nu + 1 - \nu) \quad (2.3)$$

and is equal to the expected value of "random isonymy", i.e., the probability that the surnames of two individuals taken at random are identical. Its reciprocal can be called, in analogy to the "effective number of alleles" of Kimura and Crow (1964), n_e , the effective number of surnames, and is equal to the number of surnames that would obtain if all surnames present had equal frequency, i.e., $m_k = m = 1/n_e$. If the m_i 's differ, as is usually the case, $n_s > n_e$. Thus,

$$n_e = 1/I = N\nu + 1 - \nu. \quad (2.4)$$

The distribution is completely specified by N and ν . If, for instance, the value of N is known, one can obtain from empirical data the value of ν either from (A) n_s , the total number of surnames in the sample, and use of formula (2.2), or from (B) n_e or I , using formulas (2.3) or (2.4). The validity of the distribution can then be tested. The estimation of both N and ν from a sample distribution of surnames on the basis of (2.1) may also be attempted but is not easy.

The value n_s is easily obtained from the data, being the total number of surnames, but there is no explicit solution of (2.1), and we have therefore

prepared Table I, giving n_s as a function of a set of N and ν values. This table can then be used to estimate ν , given N and n_s , or N , given ν and n_s , or n_s , given N and ν .

Also, the values n_e and its reciprocal I are easily obtained from the data, I being estimated as

$$I = \sum x_i^2; \quad n_e = 1/I,$$

when x_i is the relative frequency of the i th surname, and the sum is extended to all surnames present ($\sum x_i = 1$). This is only slightly more laborious than simply counting the total number of surnames, n_s , and then formula (2.4) provides an explicit solution for either N or ν , given n_e . This method, (B), may seem to be simpler than method A, but it is likely that a statistic based on the second moment rather than the first has a lower efficiency. The analysis by Ewens (1971) of the maximum likelihood estimation of the number of alleles from a sample of population would seem to confirm, though indirectly, that method (A) is more efficient. Ewens' estimation, however, considers the case in which a sample of the population is analyzed, while here we will often be able to consider the whole population. Should a partial sample be used, then n_s should be estimated from the sample on the basis of Ewens' method, but his tables cannot be easily extended to cover the case of very large n_s as is necessary for surnames. For all cases not covered in Table I, formulas (2.3) or (2.4) can be used. We have partial evidence to date that method (B) is less efficient than method (A), but we have not investigated the matter thoroughly.

The precision of fit of the theoretical distribution to empirical data may provide some insight to the applicability of this theory to surnames. As we shall show in the next section, however, one can look for further support by estimating the fundamental parameters independently and comparing them with those obtained by the fit of the distribution.

3. DISTRIBUTION OF SURNAMES: APPLICATION IN THE UPPER PARMA VALLEY.

Observed Distributions

Our data represent a preliminary sample of a larger set that will be more fully examined and published elsewhere. They come from marriage books available in 40 parishes of the Upper Parma Valley, Northern Italy. In some of the parishes, the books start earlier than the year 1600, and in others, at various later times. We have investigated the distribution by parish and have used only female surnames because, by custom, females marry in the parish of their birth. Female surnames thus give a more accurate picture of the male surnames of the immediately preceding generation in the same parish than a mixture of male

TABLE I

Expected Number of Surnames n_s for Values of N (Effective Population Size) and ν (Mutation and Immigration) According to Formula 3.8 of S. Karlin and J. McGregor, 1967

| ν/N | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 |
|---------|-----|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 0.01 | 1.1 | 1.3 | 1.5 | 1.7 | 1.9 | 2.1 | 2.3 | 2.5 | 2.7 | 3.0 | 3.2 | 3.4 | 3.6 | 3.9 | 4.1 | 4.3 | 4.5 | 4.8 | 5.0 | 5.2 |
| 0.02 | 1.2 | 1.5 | 1.9 | 2.2 | 2.6 | 3.0 | 3.4 | 3.8 | 4.2 | 4.6 | 5.0 | 5.3 | 5.7 | 6.1 | 6.5 | 6.9 | 7.3 | 7.7 | 8.1 | 8.5 |
| 0.03 | 1.3 | 1.8 | 2.3 | 2.8 | 3.3 | 3.8 | 4.4 | 4.9 | 5.4 | 6.0 | 6.5 | 7.0 | 7.6 | 8.1 | 8.7 | 9.2 | 9.7 | 10.3 | 10.8 | 11.4 |
| 0.04 | 1.4 | 2.0 | 2.6 | 3.3 | 3.9 | 4.6 | 5.2 | 5.9 | 6.6 | 7.2 | 7.9 | 8.6 | 9.2 | 9.9 | 10.6 | 11.2 | 11.9 | 12.6 | 13.2 | 13.9 |
| 0.05 | 1.5 | 2.2 | 2.9 | 3.7 | 4.5 | 5.3 | 6.0 | 6.8 | 7.6 | 8.4 | 9.2 | 10.0 | 10.7 | 11.5 | 12.3 | 13.1 | 13.9 | 14.7 | 15.5 | 16.3 |
| 0.06 | 1.5 | 2.4 | 3.2 | 4.1 | 5.0 | 5.9 | 6.8 | 7.7 | 8.6 | 9.5 | 10.4 | 11.3 | 12.2 | 13.1 | 14.0 | 14.9 | 15.7 | 16.6 | 17.5 | 18.4 |
| 0.07 | 1.6 | 2.6 | 3.5 | 4.5 | 5.5 | 6.5 | 7.5 | 8.5 | 9.5 | 10.5 | 11.5 | 12.5 | 13.5 | 14.5 | 15.5 | 16.5 | 17.5 | 18.5 | 19.5 | 20.5 |
| 0.08 | 1.7 | 2.7 | 3.8 | 4.9 | 6.0 | 7.1 | 8.2 | 9.3 | 10.4 | 11.5 | 12.6 | 13.7 | 14.8 | 15.8 | 16.9 | 18.0 | 19.1 | 20.2 | 21.3 | 22.4 |
| 0.09 | 1.8 | 2.9 | 4.1 | 5.3 | 6.4 | 7.6 | 8.8 | 10.0 | 11.2 | 12.4 | 13.6 | 14.8 | 15.9 | 17.1 | 18.3 | 19.5 | 20.7 | 21.9 | 23.1 | 24.3 |
| 0.10 | 1.9 | 3.1 | 4.3 | 5.6 | 6.9 | 8.1 | 9.4 | 10.7 | 12.0 | 13.3 | 14.5 | 15.8 | 17.1 | 18.4 | 19.6 | 20.9 | 22.2 | 23.5 | 24.8 | 26.0 |
| 0.11 | 1.9 | 3.2 | 4.6 | 5.9 | 7.3 | 8.7 | 10.0 | 11.4 | 12.7 | 14.1 | 15.5 | 16.8 | 18.2 | 19.6 | 20.9 | 22.3 | 23.6 | 25.0 | 26.4 | 27.7 |
| 0.12 | 2.0 | 3.4 | 4.8 | 6.3 | 7.7 | 9.1 | 10.6 | 12.0 | 13.5 | 14.9 | 16.4 | 17.8 | 19.2 | 20.7 | 22.1 | 23.6 | 25.0 | 26.5 | 27.9 | 29.4 |
| 0.13 | 2.1 | 3.5 | 5.0 | 6.6 | 8.1 | 9.6 | 11.1 | 12.6 | 14.2 | 15.7 | 17.2 | 18.7 | 20.3 | 21.8 | 23.3 | 24.8 | 26.4 | 27.9 | 29.4 | 30.9 |
| 0.14 | 2.1 | 3.7 | 5.3 | 6.9 | 8.5 | 10.0 | 11.6 | 13.2 | 14.8 | 16.4 | 18.0 | 19.6 | 21.2 | 22.8 | 24.4 | 26.0 | 27.6 | 29.2 | 30.8 | 32.4 |
| 0.15 | 2.2 | 3.8 | 5.5 | 7.1 | 8.8 | 10.5 | 12.2 | 13.8 | 15.5 | 17.2 | 18.8 | 20.5 | 22.2 | 23.9 | 25.5 | 27.2 | 28.9 | 30.6 | 32.2 | 33.9 |
| 0.16 | 2.2 | 4.0 | 5.7 | 7.4 | 9.2 | 10.9 | 12.6 | 14.4 | 16.1 | 17.9 | 19.6 | 21.4 | 23.1 | 24.9 | 26.6 | 28.4 | 30.1 | 31.8 | 33.6 | 35.3 |
| 0.17 | 2.3 | 4.1 | 5.9 | 7.7 | 9.5 | 11.3 | 13.1 | 14.9 | 16.8 | 18.6 | 20.4 | 22.2 | 24.0 | 25.8 | 27.6 | 29.5 | 31.3 | 33.1 | 34.9 | 36.7 |
| 0.18 | 2.4 | 4.2 | 6.1 | 8.0 | 9.8 | 11.7 | 13.6 | 15.5 | 17.4 | 19.2 | 21.1 | 23.0 | 24.9 | 26.8 | 28.6 | 30.3 | 32.4 | 34.3 | 36.2 | 38.1 |
| 0.19 | 2.4 | 4.3 | 6.3 | 8.2 | 10.2 | 12.1 | 14.0 | 16.0 | 17.9 | 19.9 | 21.8 | 23.8 | 25.7 | 27.7 | 29.6 | 31.6 | 33.5 | 35.5 | 37.4 | 39.4 |
| 0.20 | 2.5 | 4.5 | 6.5 | 8.5 | 10.5 | 12.5 | 14.5 | 16.5 | 18.5 | 20.5 | 22.5 | 24.5 | 26.6 | 28.6 | 30.6 | 32.6 | 34.6 | 36.6 | 38.6 | 40.6 |

| | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 0.21 | 2.5 | 4.6 | 6.6 | 8.7 | 10.8 | 12.9 | 14.9 | 17.0 | 19.1 | 21.1 | 23.2 | 25.3 | 27.4 | 29.4 | 31.5 | 33.6 | 35.7 | 37.7 | 39.8 | 41.9 |
| 0.22 | 2.6 | 4.7 | 6.8 | 8.9 | 11.1 | 13.2 | 15.3 | 17.5 | 19.6 | 21.7 | 23.9 | 26.0 | 28.2 | 30.3 | 32.4 | 34.6 | 36.7 | 38.8 | 41.0 | 43.1 |
| 0.23 | 2.6 | 4.8 | 7.0 | 9.2 | 11.4 | 13.6 | 15.8 | 18.0 | 20.1 | 22.3 | 24.5 | 26.7 | 28.9 | 31.1 | 33.3 | 35.5 | 37.7 | 39.9 | 42.1 | 44.3 |
| 0.24 | 2.7 | 4.9 | 7.2 | 9.4 | 11.7 | 13.9 | 16.2 | 18.4 | 20.7 | 22.9 | 25.2 | 27.4 | 29.7 | 31.9 | 34.2 | 36.4 | 38.7 | 40.9 | 43.2 | 45.4 |
| 0.25 | 2.7 | 5.0 | 7.3 | 9.6 | 11.9 | 14.2 | 16.6 | 18.9 | 21.2 | 23.5 | 25.8 | 28.1 | 30.4 | 32.7 | 35.0 | 37.3 | 39.7 | 42.0 | 44.3 | 46.6 |
| 0.26 | 2.8 | 5.1 | 7.5 | 9.8 | 12.2 | 14.6 | 16.9 | 19.3 | 21.7 | 24.0 | 26.4 | 28.8 | 31.1 | 33.5 | 35.9 | 38.2 | 40.6 | 43.0 | 45.3 | 47.7 |
| 0.27 | 2.8 | 5.2 | 7.6 | 10.1 | 12.5 | 14.9 | 17.3 | 19.7 | 22.2 | 24.6 | 27.0 | 29.4 | 31.8 | 34.3 | 36.7 | 39.1 | 41.5 | 44.0 | 46.4 | 48.8 |
| 0.28 | 2.9 | 5.3 | 7.8 | 10.3 | 12.7 | 15.2 | 17.7 | 20.2 | 22.6 | 25.1 | 27.6 | 30.1 | 32.5 | 35.0 | 37.5 | 40.0 | 42.4 | 44.9 | 47.4 | 49.9 |
| 0.29 | 2.9 | 5.4 | 8.0 | 10.5 | 13.0 | 15.5 | 18.1 | 20.6 | 23.1 | 25.6 | 28.2 | 30.7 | 33.2 | 35.8 | 38.3 | 40.8 | 43.3 | 45.9 | 48.4 | 50.9 |
| 0.30 | 3.0 | 5.5 | 8.1 | 10.7 | 13.3 | 15.8 | 18.4 | 21.0 | 23.6 | 26.2 | 28.7 | 31.3 | 33.9 | 36.5 | 39.1 | 41.6 | 44.2 | 46.8 | 49.4 | 52.0 |
| 0.31 | 3.0 | 5.6 | 8.2 | 10.9 | 13.5 | 16.1 | 18.8 | 21.4 | 24.0 | 26.7 | 29.3 | 31.9 | 34.5 | 37.2 | 39.8 | 42.4 | 45.1 | 47.7 | 50.3 | 53.0 |
| 0.32 | 3.1 | 5.7 | 8.4 | 11.1 | 13.8 | 16.4 | 19.1 | 21.8 | 24.5 | 27.2 | 29.8 | 32.5 | 35.2 | 37.9 | 40.6 | 43.2 | 45.9 | 48.6 | 51.3 | 54.0 |
| 0.33 | 3.1 | 5.8 | 8.5 | 11.3 | 14.0 | 16.7 | 19.5 | 22.2 | 24.9 | 27.6 | 30.4 | 33.1 | 35.8 | 38.6 | 41.3 | 44.0 | 46.8 | 49.5 | 52.2 | 54.9 |
| 0.34 | 3.1 | 5.9 | 8.7 | 11.5 | 14.2 | 17.0 | 19.8 | 22.6 | 25.3 | 28.1 | 30.9 | 33.7 | 36.5 | 39.2 | 42.0 | 44.8 | 47.6 | 50.3 | 53.1 | 55.9 |
| 0.35 | 3.2 | 6.0 | 8.8 | 11.6 | 14.5 | 17.3 | 20.1 | 22.9 | 25.8 | 28.6 | 31.4 | 34.2 | 37.1 | 39.9 | 42.7 | 45.5 | 48.4 | 51.2 | 54.0 | 56.9 |
| 0.36 | 3.2 | 6.1 | 8.9 | 11.8 | 14.7 | 17.6 | 20.4 | 23.3 | 26.2 | 29.1 | 31.9 | 34.8 | 37.7 | 40.5 | 43.4 | 46.3 | 49.2 | 52.0 | 54.9 | 57.8 |
| 0.37 | 3.3 | 6.2 | 9.1 | 12.0 | 14.9 | 17.8 | 20.8 | 23.7 | 26.6 | 29.5 | 32.4 | 35.4 | 38.3 | 41.2 | 44.1 | 47.0 | 50.0 | 52.9 | 55.8 | 58.7 |
| 0.38 | 3.3 | 6.3 | 9.2 | 12.2 | 15.1 | 18.1 | 21.1 | 24.0 | 27.0 | 30.0 | 32.9 | 35.9 | 38.9 | 41.8 | 44.8 | 47.8 | 50.7 | 53.7 | 56.6 | 59.6 |
| 0.39 | 3.3 | 6.3 | 9.3 | 12.4 | 15.4 | 18.4 | 21.4 | 24.4 | 27.4 | 30.4 | 33.4 | 36.4 | 39.4 | 42.4 | 45.5 | 48.5 | 51.5 | 54.5 | 57.5 | 60.5 |
| 0.40 | 3.4 | 6.4 | 9.5 | 12.5 | 15.6 | 18.6 | 21.7 | 24.7 | 27.8 | 30.8 | 33.9 | 37.0 | 40.0 | 43.1 | 46.1 | 49.2 | 52.2 | 55.3 | 58.3 | 61.4 |
| 0.41 | 3.4 | 6.5 | 9.6 | 12.7 | 15.8 | 18.9 | 22.0 | 25.1 | 28.2 | 31.3 | 34.4 | 37.5 | 40.6 | 43.7 | 46.8 | 49.9 | 53.0 | 56.1 | 59.2 | 62.3 |
| 0.42 | 3.4 | 6.6 | 9.7 | 12.9 | 16.0 | 19.1 | 22.3 | 25.4 | 28.6 | 31.7 | 34.8 | 38.0 | 41.1 | 44.3 | 47.4 | 50.5 | 53.7 | 56.8 | 60.0 | 63.1 |

and female surnames or male surnames alone. This would not necessarily be true of sources of surnames other than marriage books, e.g., birth (baptismal) books, censuses, etc.

Marriage books have the advantage that they allow an almost complete ascertainment of the population. Moreover, "effective population size" (the number of individuals reproducing per generation) can be estimated directly, even though approximately, in a fairly simple way, as long as the marriage books are a complete and faithful representation of events, as we believe they are in the present situation. From the definition just given, the number of marriages that occur, on average, over a time span equivalent to a generation is equivalent to N , the number of reproducing males. Accordingly, we have subdivided the total period into 30-year intervals, 30 years being approximately the duration of one generation. In fact, considering that male generations are involved and that this is a little higher than 30 years in the area (see Cavalli-Sforza, Barrai, and Edwards, 1964), a slightly longer period should have been chosen, but the error is likely to be trivial. The total number of marriages in the 30-year period was then taken as the estimate of N for that period for the parish. There is some variation of N for a given parish over time; on average, there is a decline during the 17th and 18th centuries and an increase thereafter, but short-term changes are small and were considered unlikely to affect the results of application of the theory, which assumes N constant over time. We will call "parish-period" the observations made of female surnames of all marriages that took place in one parish during a 30-year period, periods beginning in 1600 (or, in some parishes, earlier) and ending in 1960.

For a comparison of the actual distribution of the number of surnames per parish-period, all parish-periods in which N was equal to 30 ± 2 (i.e., between 28 and 32) were pooled and were altogether 27. Table II gives the observed data. The frequencies summed over the 27 parish-periods are given in the second column and averaged per parish-period in the third. The average number of surnames per parish-period is the sum of the second column divided by 27 (or the sum of the third column), which is 12.55. This is an estimate of n_s for the data accumulated over the 27 parish-periods. From it and Table I, given $N = 30$, one can estimate $\nu = .202$. Using formula (2.1), $N = 30$, $\nu = .202$, the values $E[N^*(k)]$ were obtained and multiplied times 27 to make them comparable to the expected frequencies in the second column. Values were pooled, as indicated by brackets in the second column, for the purpose of computing χ^2 values. The computation of expectations was repeated, using the observed I value ($I = .1277$), from which $\nu = .236$ was obtained (last column of Table II). The I value expected for $\nu = .202$ previously used is $I_{\text{exp}} = .1458$. The fit is satisfactory only with the computations based on the ν value obtained from n_s and the use of formula (2.1).

Similar computations were carried out using 11 parish-periods in which the

TABLE II^a

| Observed $N^*(k)$ | | | Expected $N^*(k)$ | |
|--------------------------------|-----|-------------------|-------------------------------|------------------------------|
| k | Sum | Per parish-period | Method A (based on n_s) | Method B (based on I) |
| 1 | 164 | 6.07 | 168.1 | 178.64 |
| 2 | 80 | 2.96 | 68.46 | 69.43 |
| 3 | 43 | 1.59 | 36.94 | 35.77 |
| 4 | 12 | 0.44 | 22.27 | 20.54 |
| 5 | 14 | 0.52 | 14.21 | 12.47 |
| 6 | 11 | 0.41 | 9.37 | 7.81 |
| 7 | 3 | 0.11 | 6.30 | 4.98 |
| 8 | 2 | 0.07 | 4.29 | 3.20 |
| 9 | 3 | 0.11 | 4.94 | 3.41 |
| 10 | 2 | 0.07 | | |
| 11 | 1 | 0.04 | 4.12 | 2.75 |
| 12 | 2 | 0.07 | | |
| 13 | 1 | 0.04 | | |
| 16 | 1 | 0.04 | | |
| Sum | 339 | 12.55 (approx.) | | |
| $n_s = 339/27 = 12.55$ | | | 0.202 | 0.236 ν |
| $I = 0.1277; n_e = 1/I = 7.83$ | | | 0.1458 | 0.1277 I |
| | | | 11.17 | 20.31 $\chi^2[8 \text{ df}]$ |

^a Comparison of the observed and expected distribution of surnames. Frequencies are obtained summing 27 periods of 30 years each, during which marriages were recorded in various parishes, with only female surnames considered. Parish-periods used had a total number of marriages varying from 28 to 32 ($N = 30$). ν was estimated by the two methods described in the text. k is the number of females with given surname; $N^*(k)$ is the number of surnames carried by k females, summed over 27 parish-periods in the second column, and averaged per parish-period in the third.

total number of marriages was between 59 and 61 ($N = 60$). The fit is slightly better with the estimate of ν based on n_s , but is satisfactory also with that obtained on the basis of I (see Table III).

The fit thus obtained is very encouraging. The estimation of N is straightforward; although other methods might be devised, and the choice of the period may have to be changed slightly to make it closer to the true generation time, there is little doubt that it cannot be too far off. The estimate of ν , however, may

TABLE III^a

| Observed $N^*(k)$ | | | Expected $N^*(k)$ | |
|--------------------------|-----|-------------------|-------------------------------|-----------------------------|
| k | Sum | Per parish-period | Method A (based on n_s) | Method B (based on I) |
| 1 | 94 | 10.36 | 104.43 | 97.75 |
| 2 | 60 | 5.45 | 44.58 | 43.05 |
| 3 | 23 | 2.09 | 25.32 | 25.24 |
| 4 | 20 | 1.82 | 16.13 | 16.61 |
| 5 | 11 | 1.00 | 10.94 | 11.63 |
| 6 | 4 | 0.86 | 7.70 | 8.47 |
| 7 | 2 | 0.18 | 5.56 | 6.33 |
| 8 | 4 | 0.36 | 4.09 | 4.81 |
| 9 | 6 | 0.55 | | |
| 10 | 1 | 0.09 | 5.33 | 6.60 |
| 11 | 4 | 0.36 | | |
| 14 | 1 | 0.09 | 7.92 | 12.51 |
| 19 | 1 | 0.09 | | |
| 36 | 1 | 0.09 | | |
| Sum | 232 | 21.09 (approx.) | | |
| $n_s = 232/11 = 21.09$ | | | 21.09 | 18.73 ν |
| $I = 0.1151; n_e = 8.69$ | | | 0.0980 | 0.1151 I |
| | | | 12.20 | 13.21 $\chi^2[8]$ |

^a Eleven parish-periods with total number of marriages close to 60. Computations as in Table II.

leave some doubt until it can be confirmed by independent sources. It is therefore important to consider, at this stage, the exact meaning of ν and the validity of our identification of ν with migration.

Surnames are subject to very little change, so that there is little true mutation. Altogether, the following cases should be considered as contributing to ν :

- (1) Illegitimate birth, mother known. The child usually takes the surname of the mother unless legitimized later by the natural father.
- (2) Both parents unknown. The infant receives a new surname, often indicating illegitimacy.

(3) True change of surname: very rare in this population in the period covered by the analysis. Minor changes due to misspellings have regularly been corrected at the time of coding of the data.

We do not have good estimates of these events, but they should all be low in this population, certainly much smaller than the immigration rate. Moreover, children in categories 1 and 2 may have a lesser change of marrying because of their lower social status as illegitimates, and thus contribute less than proportionately to ν . Thus ν , which should be the sum of mutation and immigration, will be made almost entirely from immigrants, but the immigration that counts has to be defined: it is formed by the males that take residence in the parish, originating elsewhere, and have progeny in it.

An estimate of immigration that comes very close to it was obtained in a fraction of these villages on the population that was living there in 1957–58. The fraction of children born in one parish, whose parents were also born in the same parish, were estimated to be $.821 \pm .019$ for father offspring, $.773 \pm .022$ for mother offspring. The difference between these two frequencies shows that in this area, migration is mostly patrilocal, though perhaps less so in recent times than in the past. The estimate of immigration in which we are interested is therefore $1 - .821 = .179 \pm .019$.

As these data were obtained on the population living at a fairly recent date, the appropriate time period was chosen (1900–1960) in the parish records available. Number of marriages, N , and the number of surnames, n_s , were available for all of the 40 parishes for this period. The estimate of ν obtained was $.168 \pm .019$, averaging the estimates over all parishes. The empirical standard error is appended. The agreement between this and the observed immigrations seems quite satisfactory, assuming that true mutation, as defined before, is low. This quantity would inflate the estimate of ν above that due to immigration alone. In spite of the absence of direct estimates, it seems that the true mutation is at most of the order of 2 or 3%. Therefore, the agreement between the male immigration rate observed and that estimated from surname distribution seems excellent.

4. THE DISTRIBUTION OF PROGENY SIZE

The theory of Karlin and McGregor can be applied to surname data if the assumption of progeny size distribution is correct. In the theory, a Poisson process of one birth accompanied by one death is assumed; this model corresponds fairly well with one in which property is passed patrilineally from father to son, as is the case in this highly rural population (see discussion). Here, the right of primogeniture (the whole estate is passed to the first-born male child) has been in operation for most of the period considered until recently. However, there

are exceptions to the rule: parents may have no male progeny, for instance. A distribution, perhaps with an unusually small variance, may then be expected to operate.

The distribution desired is that of the male progeny that takes residence in the parish and continues the surname. There are no such distribution data available for this (or, to our knowledge, any other) region.

One distribution studied in this area may, however, supply the basis for estimation. This is the distribution of all children born in the same parish to fathers born in the parish (Skolnick *et al.*, 1971). This distribution was fitted, with only moderate success, with both the geometric distribution and the negative binomial. The fit required truncation at zero, as the frequency at zero was inflated by several factors that do not necessarily interest us. In other cases (see, for example, Cavalli-Sforza and Bodmer, 1971), truncation at zero was not necessary. The distributions analyzed in Skolnick *et al.* may have other sources of bias which will not be further considered here.

The distribution of all children ever born is not directly the one desired. For one thing, only children of male sex should be counted for our purposes. We found the following way of reconstructing on its basis the desired probability distribution:

- (1) If the population is stationary, the mean number of males per male parent who take residence and bring up a family in the parish must be $m = 1$. If the population is increasing or decreasing at constant rate r , then $m = e^{rT}$, where T is the generation time and r the net reproduction rate. If population counts are available at times t_1 and t_2 , being N_1 and N_2 , and $(t_1 - t_2)/T = t$, then

$$m = (N_1/N_2)^{1/t}. \quad (4.1)$$

- (2) Assuming that the distribution of total progeny size is a negative binomial, the probability of n children being born in a family is

$$p_n = \binom{\alpha + n - 1}{n} p^\alpha q^n, \quad (4.2)$$

where α , p ($q = 1 - p$) have been estimated to be, from the 18th and 19th century data in this area, $p = .37$, $\alpha = 2.2$ (18th century); $p = .33$, $\alpha = 1.65$ (19th century).

- (3) If none of the progeny is male, the surname will be extinct, and for a family of n , the chance of extinction will be $(1/2)^n$ or more exactly $(p_f)^n$, if p_f is the frequency of a female birth. Given that the progeny distribution is as above, the probability of extinction for lack of male progeny will be:

$$P_0 = \sum_{n=0}^{\infty} p_n (p_f)^n = p/(1 - qp_f). \quad (4.3)$$

With the numerical values of p and α given above, and $p_f = 1/2$, one can estimate $P_0 = .258$ for the 18th century and .314 for the 19th century. Averaging, P_0 is .286.

(4) The distribution of male children continuing the surname in the parish has been computed as in (1). and chance of extinction computed as in (3), or related ways. Assume that the distribution of male children continuing the surname in the parish is shaped as a geometric dudded at zero (truncated at zero) so that

$$P_n = bc^{n-1} \quad \text{for } n = 1, 2, \dots, \quad (4.4)$$

and also

$$p_0 = (1 - b - c)/(1 - c) \quad \text{for } n = 0. \quad (4.5)$$

This distribution (see, e.g., Karlin, 1969) has special interest for the theory to be used in the next section. It would be valid for the number of males, with changed coefficients, if the distribution of the number of children irrespective of sex were also geometric (see Lotka, 1939). Even if this is not entirely correct, it is not too far from truth. Given m and $P_0 = p_0$, computed as in (4.1) and (4.3), b and c are

$$\begin{aligned} b &= (1 - P_0)^2/m, \\ c &= 1 - (1 - P_0)/m. \end{aligned} \quad (4.6)$$

The mean and variance of the distribution are

$$\begin{aligned} m &= b/(1 - c)^2, \\ \text{var} &= m(1 - b)/(1 - c)^2. \end{aligned} \quad (4.7)$$

Taking $P_0 = .286$ from (3), and $m = 1$ (a stationary population),

$$b = .510, \quad c = .286.$$

This distribution can be employed for computing probabilities of extinction of surnames, to be discussed in the next section. It may be noted here that the variance of the empirical distribution (from 4.7) is .961 and, therefore, very close to the mean 1. Although the distribution is not Poisson, given the assumptions made, the ratio of variance to mean happens to be very close to a Poisson, justifying further the application of the theory mentioned in Section 2.

5. PROBABILITY OF EXTINCTION AFTER SEVERAL GENERATIONS

The basic formulas that we shall use are the following (see Karlin, 1969), which all apply if the dudded geometric can represent the distribution of male progeny remaining in the parish, as hypothesized in Section 4.

The probability of extinction at the n th generation of a surname represented by one individual at generation zero is

$$\phi_n = nc/(1 + (n-1)c) \quad \text{for } n = 1, 2, \dots, \quad (5.1)$$

if $m = 1$, and therefore $c = P_0$, the probability of extinction in one generation; if $m \neq 1$,

$$\phi_n = 1 - m^n \left(\frac{1 - P_0/c}{m^n - P_0/c} \right) \quad (5.2)$$

when $c = 1 - (1 - P_0)/m$.

If a surname is represented by more than one individual at generation zero, say, by $x_0 > 1$ individuals, and all families reproduce independently with the same distribution of male progeny per generation, then the probability of extinction is

$$(\phi_n)^{x_0}. \quad (5.3)$$

If the distribution of surnames at generation zero is given by (2.1), the probability of extinction of a surname, independently of its frequency in generation zero is

$$P_E = \sum_{k=1}^{\infty} E(N^*(k)) \phi_n^k, \quad (5.4)$$

when $E(N^*(k))$ is given in (2.1) and ϕ_n is given above. Given that N and ν at generation zero are known, it is thus possible to compute the average extinction probability of a surname. For $m = 1$, and calling $A = N/(1 - \nu)$,

$$P_E = \nu A \sum_{k=1}^{\infty} \frac{(nc)^k}{(1 - c + nc)^k} \frac{1}{k} \frac{(A - k - 1)! N!}{(N - k)! (A - 1)!}. \quad (5.5)$$

For purposes of testing the validity of application of these formulas, it is better to consider surnames individually rather than on average basis, as the data provide information on k and the testing is thus more precise.

The longest time which can be tested in the Parma material refers to ten generations. The first period we formed starts at a variable date and ends in 1630. Data are available for the second period (years 1630–1660) for seven parishes, there being data also for these parishes for the first period (before 1630) and for all subsequent ones, including the last (1930–1960). Existence of data for

the first period assures us that the records start earlier than 1630 and are complete for the second period, which will be taken as generation zero. The number of females in the marriages of this period carrying a given surname was counted; it was then tested if this surname was still represented among the females of the marriages that took place in the same parish ten generations later (1930-1960). Table IV indicates the number of surnames still represented and extinct after ten generations.

TABLE IV

Surnames Still Present, and Those Extinct, After Ten Generations
as a Function of the Number of People Carrying These Surnames at the Start.

Results are Accumulated for the Seven Parishes.

Expected Extension Probabilities are Explained in the Text.

| No. of individuals with given surname at the start | After ten generations, No. of surnames that are: | | | Probability of Extinction | | |
|--|---|---------|-------|---------------------------|--------------|--------------|
| | Present | Extinct | Total | Observed | Expected (1) | Expected (2) |
| | | | | | $m = 1$ | $m = 1.049$ |
| 1 | 9 | 36 | 45 | 0.8182 | 0.8002 | 0.7673 |
| 2 | 4 | 7 | 11 | 0.6364 | 0.6404 | 0.5887 |
| 3 | 6 | 5 | 11 | 0.4546 | 0.5124 | 0.4517 |
| 4 | 4 | 4 | 8 | 0.5000 | 0.4101 | 0.3466 |
| 5 | 6 | 6 | 12 | 0.5000 | 0.3281 | 0.2659 |
| 6 | 3 | 2 | 5 | 0.1667 | 0.2626 | 0.2040 |
| 7 | 6 | 1 | 7 | | 0.2101 | 0.1565 |
| 8 | 1 | 0 | 1 | | 0.1681 | 0.1201 |
| 9 | 1 | 0 | 1 | | 0.1346 | 0.0922 |
| 10 | 2 | 0 | 2 | | 0.1077 | 0.0707 |
| 11 | 1 | 0 | 1 | | 0.0862 | 0.0543 |
| — | | | | | | |
| 19 | 1 | 0 | 1 | | 0.0144 | 0.0065 |

The probability of extinction, as could be expected, decreases approximately geometrically with the number of representations of the surname at the start. The formulas given before allow us to compute expectations. The first computation given in Table IV uses the assumption that the population is stationary ($m = 1$) and the extinction probability per generation, $P_0 = .286$, taken here for the estimate of c , as explained in Section 4 and formula (5.1). The agreement is very good.

The second computation takes account of the fact that from beginning to end there has been a net modest increase in population size. Taking the number of marriages in the 30-year period as an estimate proportional to census size, the average increase per parish per generation was computed from (4.1) to be such that $m = 1.049$. Then, given as before $P_0 = .286$, $c = .319$ (from 4.6). The use of formula (5.2) generates the expectation given in the last column of Table IV.

The fit of the formula assuming $m = 0$ seems to be slightly preferable, in spite of the fact that there was a net increase in population size. It should be noted, however, that, as already mentioned, the population size first decreased somewhat, and then increased rapidly in the last few generations. The distribution obtained in Section 4, and used to estimate the probabilities of extinction, was obtained quite independently from these data, although on somewhat simple assumptions, but the agreement shown in Table IV certainly helps in reinforcing its validity. Alternatively, one could have estimated the probability $P_0 = c$ of extinction of a surname per generation from the data in Table IV and formula (5.1), but the result for P_0 would have changed only very slightly from the value obtained in Section 4.

6. DISCUSSION AND CONCLUSIONS

Surname evolution has similarities with that of gene evolution. The model we have used compares them to genes of haploid organisms reproducing asexually, as if only males existed. For reasons of social custom, we have limited the investigations to surnames of females in marriage books, as females, marrying usually where they are born, indicate more accurately than males the local surnames when using marriage data.

Instead of using marriage data, surnames might be sampled from birth or death data, or from censuses or population samples. Marriage data were found easy to handle by subdividing them into periods corresponding approximately to a generation time.

The agreement between theory and data has been tested in two ways: the distribution of the number of surnames carried by a given number of individuals and the agreement between the immigration rates per generation obtained from direct census and from the relation of number of surnames to population size N . Both comparisons show that the theory of neutral mutations by Karlin and McGregor can be satisfactorily applied to surnames.

The following general conclusions seem worth being commented and expanded upon. We will consider separately the fit of theories to surname data and the use of surnames to understand genetic population structure.

With regard to the first aspect, it would seem that either the number of sur-

names n_s or the random isonymy index I can be employed to obtain the value of immigration ν given for N (or vice versa). As already discussed, the use of n_s is probably more satisfactory. The agreement of immigration values observed directly, say, ν' , from male immigration, e.g., father-offspring data, and that inferred from the surname distribution (ν) implies:

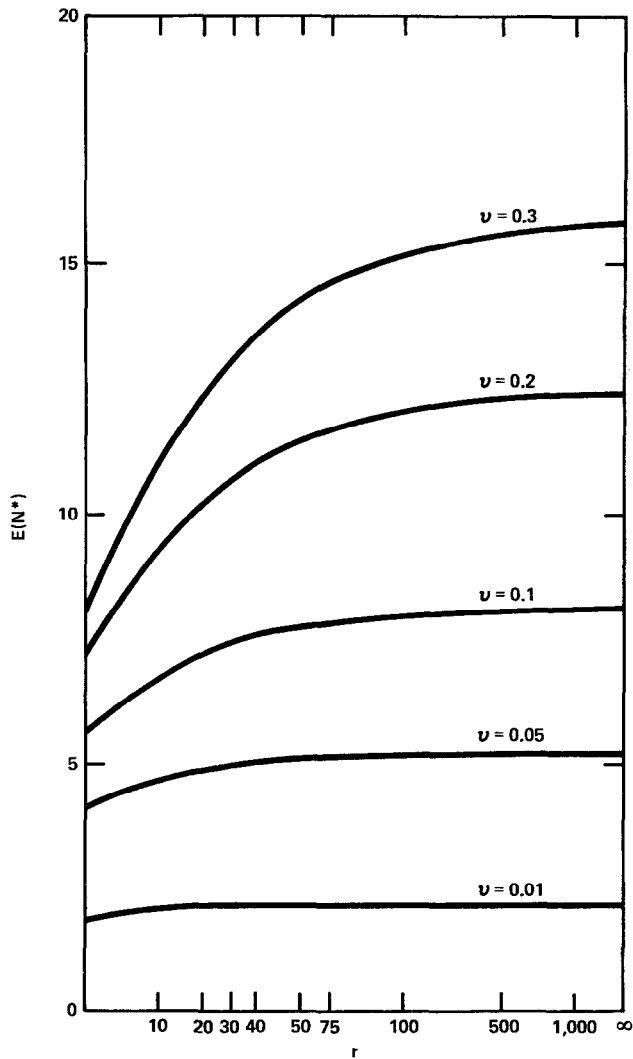


FIG. 1. Finite total number of surnames r and expected number of surnames in a population with $N = 30$ and various migration values ν according to the Karlin-McGregor theory.

- (1) That the population is stationary.
- (2) That the rate of mutation is negligible (or else it should be equal to the difference between ν and ν' ; in our case, this difference is not significantly different from zero, and the mutation rate is known to be low. "Mutation" in the present case is defined in Section 3.
- (3) That the number of surnames is very high. The theory assumes an infinite number of surnames. The effect on n_s of a finite number of surnames has been computed for the case $N = 30$ on the basis of formulas of Karlin and McGregor (see Fig. 1) and found to be negligible for the total number of surnames truly observed, which is in the thousands. It may be more important for larger populations than those considered here and may then have to be taken into account.

These hypotheses seem to be satisfied in this case on the basis of independent evidence.

(4) One difference of the real situation from the model of Karlin and McGregor should be noted, the importance of which is hard to assess as of now. The model assumes a Poisson process of births (automatically accompanied by deaths). This leads to an exponential distribution of generation times, while the real one has a mode different from zero and is more similar to a gamma variate. Thus, the real situation is somewhat intermediate between the Karlin-McGregor model and a model with fixed generation times (nonoverlapping generations). There is no theoretical treatment of the latter, but a simulation with fixed generation times and Poisson distribution of progeny indicates that n_s and n_e are lowered, the latter being close to $1/(1 + 2N\nu)$ (Bodmer and Cavalli-Sforza 1972).

(5) The effect of a Poisson (or other) distribution of progeny size by itself is not entirely known. Karlin and McGregor give bounds that can be computed for n_s , but they are wide. For a Poisson distribution, when N is large and $N\nu$ finite, however, they show an approximate result which is identical to that obtained in the simpler version, which ignores progeny size distribution. The joint effect of deviation from the last two hypotheses may perhaps be estimated to lead to a decrease in observed vs expected n_e and n_s by a factor smaller than two, on the basis of the simulation results just cited. Such a decrease is not observed.

A corollary which can be drawn regards the problem of "polyphyletism" of surnames (as already mentioned, the fact that many surnames may have had multiple origins). Surnames are very often patronymics, that is, derived from the father's first name. If a given name is frequent (or was frequent at the time surnames originated, which was around the 13th-15th centuries in Europe), the surnames derived from it will have multiple origins. Morton and Hussels (1970)

have analyzed a population in Switzerland and have concluded that their material was heavily affected by polyphyletism. Their evidence is indirect. It would seem that in our material, polyphyletism is present as surnames connected with common jobs (such as Ferrari, analogous to Smith) are among those most frequent. The amount of polyphyletism might be defined as the ratio of the total number of ancestors who acquired surnames divided by the total number of different surnames, thus measuring the average multiplicity of origin of surnames. In our case, however, it would appear that the amount of polyphyletism must be small. If it were, say, of the order of 2, i.e., each surname had a duplicate origin, the number of surnames n_s would be halved, and the amount of migration ν computed from N and n_s would also be approximately halved. It would then be definitely smaller than the one observed (ν'). As this does not happen, the amount of polyphyletism cannot be high.

There are hints on why polyphyletism may be especially low in the present material. Over 60% (and probably more, up to perhaps 90%) of surnames are clearly patronymic, being simple modifications of a first name, usually an archaic one. Many of the surnames found, however, clearly descend from the same first name (e.g. Berti, Bertini, Bertoli, Bertorelli, Bertolini, etc.) but with the variation imposed by the obviously wide use of diminutives derived from first names. In the case of one name, over 30 different surnames obviously derived from it by way of diminutives were counted. This probably created sufficient variation that the effects of duplication of first names were largely avoided.

The problem of extending conclusions from surnames to the genetic population structure, as in the attempts of inferring inbreeding coefficients from isonymy, will be discussed more fully elsewhere. It may be enough to conclude here that the main difficulty is that all the demographic parameters involved refer only to males: (1) the generation time in surname analysis is that of males which is usually higher than that of females; (2) the population effective size is that of males, which is approximately half that of the total population; (3) the distribution of progeny is that of male progeny per male parent *who marry*, and this may be deeply affected by social custom such as the right of primogeniture; (4) the immigration that counts is that of males, which may be very different from that of females in patrilocal or matrilocal situations. For genetic population structure, instead, parameters from both females and males should be considered and given equal weight.

Some of these sources of difference, and perhaps all of them, may be corrected, given adequate knowledge of the demographic quantities involved. If this is lacking, however, definite biases can be expected if one tries to estimate inbreeding coefficients or other genetic parameters from the study of surnames.

SUMMARY

Surname distribution and evolution can be analyzed by a theory put forward for neutral mutation by Karlin and McGregor. This theory allows us to predict the distribution of surnames, and to predict immigration on the basis of knowledge of population size, and of the total number of different surnames or, in lieu of the latter, a quantity known as random isonymy.

Another approach allows us to predict probabilities of extinction after any number of generations and various sets of conditions. The predictions of these theories have been tested successfully on a number of empirical data from the Upper Parma Valley.

REFERENCES

- BODMER, W. F., AND CAVALLI-SFORZA, L. L. 1971. Variation in fitness and molecular evolution. Vol. 5: "Darwinian, Neo Darwinian and Non Darwinian Evolution," Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, pp. 155-175.
- CAVALLI-SFORZA, L. L. 1969. Genetic drift in an Italian population. *Sci. Amer.* **221**(2), 30-37.
- CAVALLI-SFORZA, L. L., BARRAI, I., AND EDWARDS, A. W. F. 1964. Analysis of human evolution under random genetic drift. *Cold Spring Harbor Symp. Quant. Biol.* **24**, 9-20.
- CAVALLI-SFORZA, L. L., AND BODMER, W. F. 1971. "The Genetics of Human Populations," W. H. Freeman Publishing Co., San Francisco.
- CROW, J. F., AND MANGE, A. P. 1965. Measurement of inbreeding from the frequency of marriages between persons of the same surname. *Eugen. Quart.* **12**, 199-203.
- EWENS, W. J. 1972. The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3**(1), 82-112.
- KARLIN, S., 1969. "Stochastic Processes," Academic Press, New York.
- KARLIN, S., AND MCGREGOR, J. 1967. The number of mutant forms maintained in a population. *Proc. Fifth Berkeley Symp. Math. Stat. Prob.* **4**, 415-438.
- LOTKA, A. J. 1931. The extinction of families. *J. Wash. Acad. Sci.* **21**, 377.
- LOTKA, A. J. 1939. "Theorie Analytique des Associations Biologiques," deuxième partie, Hermann and Co., Paris.
- MORTON, N. E. 1969. Human Population Structure. *Ann. Rev. Genet.* **3**, 53-74.
- MORTON, N. E., AND HUSSELS, I. 1970. Demography of inbreeding in Switzerland. *Hum. Biol.* **42**(1), 65-78.
- SKOLNICK, M., MORONI, A., CANNINGS, C., AND CAVALLI-SFORZA, L. L. 1971. The reconstruction of genealogies from parish books. In "Mathematics in the Archeological and Historical Sciences" (F. R. Hodson, Ed.), Edinburgh University Press, Edinburgh, pp. 335-346.
- YASUDA, N., AND FURUSHO, T. 1971. Random and nonrandom breeding revealed from isonymy study. I. Small cities of Japan. *Hum. Genet.* **23**(3), 303-316.