

Surnames in Western Europe: A comparison of the subcontinental populations through isonymy

Chiara Scapoli^{a,*}, Elisabetta Mamolini^a, Alberto Carrieri^a,
Alvaro Rodriguez-Larralde^b, Italo Barraï^a

^a*Dipartimento di Biologia, Università di Ferrara, Via L. Borsari 46, I-44100 Ferrara, Italy*

^b*Centro de Medicina Experimental, Laboratorio de Genética Humana, IVIC, Caracas, Venezuela*

Received 11 February 2006
Available online 13 July 2006

Abstract

We studied the isonymic structure of Western Europe using the distributions of 26.2 million surnames in 8 countries, 125 regions and 2094 towns of the Subcontinent. We found that, for the whole of Western Europe, Nei's distance was correlated with geographic distance ($r = 0.610 \pm 0.009$). It was observed that at long geographic distances the isonymic distance stays below linearity and tends to become asymptotic, and this was attributed to long distance migration. A dendrogram of the 125 regions was built and the clusters identified by the dendrogram are almost exactly coincident with the nations of the Subcontinent. Random inbreeding calculated from isonymy, F_{ST} , was highest in Spanish regions, and lowest in France. The geographical distribution of α in 2094 towns, high in the Center and East of the Subcontinent and lower in Spain, is compatible with the settlement of subsequent waves of migrants moving from the West and from the South toward the centre of the Continent. The present surname structure of Western Europe is strictly linked to local languages.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Western Europe; Population structure; Surname distribution; Inbreeding levels by isonymy; Isolation by distance

1. Introduction

The deviations from panmixia, such as those due to limited number of ancestors, to sex, to preference for or against certain types of consanguineous marriages, and to limited migration in social or geographic space, constitute population structure. Although the relation between gene and genotype frequencies may be exceedingly complex, the deviation from panmixia may be indicated by the inbreeding coefficient F (Wright, 1921). Then, the estimation of F is a basic endeavour in the study of population structure.

To this purpose, a crude estimate is satisfactory, and this is provided by the concept of isonymy as defined by Crow and Mange (1965). They noted that the proportion of marriages isonymous by descent among all marriages with inbreeding coefficient F would be $I = 4F$ if all sex combinations of intermediate ancestors of the spouses

were equiprobable. This was applied to the estimation of the random component of inbreeding in populations, F_{ST} , so that $F_{ST} = (1/4)\sum p_i^2$, where p_i is the frequency of surname i in a population. This will roughly estimate the drift occurred to present in that population (Yasuda and Morton, 1967).

The simple and elegant formulation of Crow and Mange was repeatedly refined by several authors after 1965. It was noted explicitly that in the systems of surname attribution through the paternal line, surnames, although being a component of language which is culturally transmitted, simulate neutral alleles of a gene inherited only through the Y chromosome (Yasuda and Morton, 1967; Yasuda and Furusho, 1971; Yasuda et al., 1974), and therefore satisfy the expectations of the neutral theory of evolution (Cavalli-Sforza and Bodmer, 1971), which is entirely described by random genetic drift, mutation and migration (Kimura, 1983). The theoretical developments permitted a notable progress in the application of population genetics models in the past three decades, making isonymic studies popular

*Corresponding author. Fax: +1 39 0532 249761.

E-mail address: sc@dns.unife.it (C. Scapoli).

(for references see [Barrai et al., 2001](#) and [Rodriguez-Larralde et al., 2003](#)). These studies represent the link between genetic and cultural inheritance.

Surnames might be an appropriate material for testing hypotheses about migration, drift, and admixture, and the study of microevolution of human groups would be somewhat strengthened by the use of the large sample sizes which are promptly available for hypothesis testing, particularly when surnames correlate with haplotypes of Y-chromosomes ([Dipierri et al., 1998](#); [Jobling, 2001](#)).

2. Purpose of the present work

Using surnames, we have investigated the structure of several urban populations in Europe. The results we obtained indicate that in the Old Continent modern populations have settled for a time long enough to permit drift and some clinal dispersion of surnames, as indicated by the existence of isolation by distance. The countries studied have been relatively isolated from each other, and internal migration was large enough to create clinal dispersion, but not enough to counteract drift. If we were allowed to personalize surnames, we could say that after the surnames originated in European countries, they persisted over the areas for a time long enough for drift to act significantly. So, in European countries surnames originated, diffused, and had time to drift, as indicated by isolation by distance.

Our next consideration was then, what is the dynamic in countries where surnames immigrated recently? We studied Venezuela, where surnames came by migration much later than in Europe, and again found isolation by distance ([Rodriguez-Larralde et al., 2000](#)). However, the social structure of Venezuela in the past few centuries was not compatible with large internal movement of the population, so that, even if the time available for drift was less than in Europe, the scarce mobility of the population resulted in detectable isolation by distance. Subsequently, we looked for isolation by distance in a country where surnames are even more recent than in Venezuela, and where immigration was recent and important also in the 20th century, namely in the 48 conterminous United States of America ([Barrai et al., 2001](#)), and there we failed to evidence any relevant isolation by distance, isolation which was observed also in Argentina ([Dipierri et al., 2005](#)).

We now want to have a general and complete picture of the surname distributions in Western Europe, where we have available 26.2 million surnames. From these, we shall describe the isonymic structure of 2094 towns, 125 regions and 8 countries and the parameters which can be derived from isonymy. Another of the purposes of the present work is to report whether in the whole of Western Europe as a unit isonymic distance varies with geographic distance, as we observed in individual countries, in Venezuela and Argentina ([Barrai et al., 1997, 1999, 2000, 2002](#); [Rodriguez-Larralde et al., 1998a, b, 2000, 2003](#); [Dipierri et al., 2005](#); [Scapoli et al., 2005](#)). Eventually,

we hope to obtain indications on the direction of migration waves, studying the heterogeneity of surnames in the Subcontinent.

The present sample of 26.2 million people represents the agglomerate of our data from Austria (1.0 million, population 8.2 million), Belgium (1.1, 10.3), France (6.0, 60.4), Germany (5.2, 82.4), Italy (5.1, 58.0), Netherlands (2.4, 16.3), Spain (3.6, 40.3) and Switzerland (1.7, 7.5). The sample size used from each nation is very roughly proportional to its population, with the exception of Germany, which is underrepresented, and Switzerland, overrepresented. We can then study the quantities which can be derived from the surname distributions, particularly random inbreeding for the whole of the Subcontinent, for the 8 nations, for their 125 administrative regions, and for 2094 towns. We recall that in the hierarchical model of inbreeding ([Wright, 1951](#)), founder and descendant populations are related among them by a branching process. The inbreeding coefficient relative to the total population is F_{IT} ; the departure from panmixia within a descendant population is measured by F_{IS} ; the divergence of a descendant population from a founder population is measured by F_{ST} . The relation which links these components of inbreeding is:

$$F_{IT} = F_{ST} + F_{IS}(1 - F_{ST})$$

where F_{ST} is always positive (see [Barrai, 1971](#) for a particularly simple derivation of this relationship). F_{ST} results from division into subpopulations, so it is an appropriate metric to measure differentiation among towns, regions and countries. The main utility of F_{ST} in the history of a population is to indicate when most genetic drift is likely to have occurred. In the case of Europe, the difference between F_{ST} values in different areas might give indications about the relative population uniformity in the same areas. We recall that high values of F_{ST} from isonymy are possible when in a population or subgroup there are relatively few surnames, and low values when the number of surnames is large. As in the case of Y-linked alleles, drift of surnames is proportional to time, since at time t

$$F_t = 1 - e^{-t/N}$$

and then small F_{ST} will be indicative of recent immigration or settlement.

3. Materials and methods

3.1. Towns and regions

The data of the 2094 towns sampled for this work are readable and downloadable at <http://www.unife.it/projects/genetics/pdata> where we maintain our public data. Their positions are given in [Fig. 1](#).

The summary data of the 125 conterminous regions studied here are given in [Table 1](#). The area studied coincides with the traditional Western Europe, excluding Portugal and the Scandinavian Countries. The 26.2 million people are approximately a 9% sample from a total population of 283 million persons.

The 125 regions in which the countries are subdivided differ greatly for position, area, and population, from Galicia in western Spain to

Download English Version:

<https://daneshyari.com/en/article/4502804>

Download Persian Version:

<https://daneshyari.com/article/4502804>

[Daneshyari.com](https://daneshyari.com)