Ancestry and Admixture Rates: the Matter of Evolution in Admixture Study

G. Gourjon¹, A. Degioanni¹

¹Aix-Marseille Univ, UMR 7269 - CNRS – MMSH LAMPEA - BP 647 5 rue du Château de l'Horloge 13094 Aix-en-Provence cedex 2, France. E-mail: ggourjon@gmail.com

KEY WORDS: genetic admixture, genealogy, haplogroups.

Introduction

The admixture is an intricate matter in human populations (Gourjon, 2010), involving migratory flows and sociocultural relationships between populations and individuals, as directional mating, endogamy, or patri-matriilocality. It refers to the concept of population itself and also to genetic differences between populations (Benoist, 1991). Just after the admixture event, the admixed population presents allele frequencies proportional to parental contributions (Chakraborty, 1985). Successive generations of admixture, evolutionary forces, and recombination modify the genetic structure of newly admixed population and individuals’ genome. These dynamic and evolutionary concepts of the admixture event are of great importance when answering the fundamental question: “what is the ancestry?”

An individual can be defined by his historical ancestry (his parents and grand-parents origins), i.e. by all his genealogical ascendance (diachronic approach). But, the same individual can be also defined by his genomic ancestry, i.e. the proportions of his genome (Individual Admixture, noted IA) which is derived from either parental population (Pritchard et al., 2000; McKeigue, 2005) (synchronic approach).

From the population point of view, the common ancestral admixed population when admixture event occurred, taking into account some of evolutionary forces (Degioanni and Gourjon, 2010). Here, we detail these approaches to better understand the many-sided definition of ancestry. Relying on published works on the Comoro Archipelago located on the Indian Ocean (Gourjon et al., 2011; Msaide et al., 2011), we discussed significant differences in ancestry estimates depending on the approach used by the authors. Three main parental components have led to the current-day Comorian population (Masao and Mutoro, 1997): (i) African Bantu populations, (ii) populations from the Arabian Peninsula and from Persia (nowadays Iran), (iii) populations from the south-west Asia (mainly Indonesian).

Historical Ancestry

An individual can build his family tree, based on baptisms (birth), marriages, and death’s register. On traditional societies, the individual own knowledge and his related memories are substituted to reconstruct this genealogy, rising some self-identification problems (see below).

In example studies on Comoro population, unrelated individuals with four Comorian grand-parents were sampled as representative of the admixed population. Since no information about previous generations is available (Fig. 1, standard font: genealogy is known only for generations G0 to G2), “Comorian” is assumed to be an actual

![Fig. 1. Individual genealogy.](image-url)
ancestry and all sampled individuals are “Comorian”. Now, considering that previous generation G3 is known (Fig. 1, bold italic). If five great-grandparents appear to be Bantu while the other remain Comorian (G4 is still unknown), the grand-parents and parents’ origins become a mixing of Bantu and Comorian origins. Individual’s ancestry is expressed as a proportion of his ancestors’ origins (5/8 African and 3/8 Comorian) and his ancestry is no longer 100% “Comorian”. Going back into generations, the individual ancestry may vary at each generation, clearly showing that it depends on the number of generation for which information are available.

Frequently, phenotypic characteristics, and especially the skin pigmentation, are used to assign a given ancestry to individuals from admixed populations. Sampled individuals are self-identified or classified for example as Black, Indigenous, Hispanics (Alves-Silva et al., 2000; Benn-Torres et al., 2008; Klimentidis et al., 2008), on the basis of their skin color, their surnames/genealogies, or their memories. But, while individuals identified them as completely Black, Indigenous, or Hispanics, their genomes could exhibit a blend of origins if genetic markers are used (Marrero et al., 2005; Lao et al., 2010). See Gourjon and Degioanni (2008) for details on correlations between ancestries and skin pigmentation.

**Individual admixture**

In a population, each individual has his own genetic history. His genome is a combination of alleles originated from various ancestries, a mosaic created along generations by successive recombination/segregation processes (Pritchard et al., 2000). It is possible to assign sampled individual to one or more ancestral populations if he shares genetic markers with these populations. Then, one can assess the part of his genome that can be attributed to such or such a population: the Individual Admixture rates (i.e. IA). These IA have great interests in population genetics and in disease-association studies (Shriner et al., 2011).

In our example studies, data on individual’s genome were mostly unavailable and IA hasn’t been estimated for Comorian individuals.

Three main statistical approaches to estimate these IA have been used during this last decade: maximum likelihood method (Hanis et al., 1986), bayesian method (McKeigue et al., 2000), and bayesian clustering method (Pritchard et al., 2000; Falush et al., 2003). The third approach is currently the most used method and population admixture rates are inferred from the mean IA among all individuals. This gives a synchronic view of the genetic structure of the current-day population instead of parental contribution estimates (i.e. contributions of ancestral parental populations at the moment of the admixture event). Moreover, some important differences in IA estimates within an admixed population have been observed (Benn-Torres et al., 2008; Haldet et al., 2008; Lopes Maciel et al., 2011) because of different history in colonization’s processes among various regions.

**Population admixture**

Population genetic admixture classically refers to an instantaneous model which looks at two parental populations, evolving independently, before contributing in proportions p1 and (1 - p1) to the gene pool of an admixed population, during a very short period of time. Admixture coefficients or population admixture rates (PA) correspond to these parental population contributions at the moment of the admixture event (diachronic approach). Many estimators of PA have been developed in various methods (for a review, see Degioanni and Gourjon, 2010). The former methods (Krieger et al., 1965; Roberts and Hiers, 1965; Chakraborty, 1985) partially rely on Hardy and Weinberg equilibrium conditions, neglecting evolutionary forces and sampling error. To reduce bias, latter methods include some evolutionary forces as well as sampling error. For example, Long (1991), Chikhi et al. (2001), Wang (2003), or Sousa et al. (2009) introduce genetic drift and sampling error in their model, while coalescence approach of Bertorelle and Excoffier (1998) takes mutation into account.

Gourjon et al. (2011) estimated PA to the Comorian population with several estimators and several combinations of parental populations, using both autosomal and uniparental markers. Results are summarized in Tab. 1. The estimated PAs represent contributions of the four ancestral components to the new admixed genetic pool, 5-600 years ago. It is the only approach which provides a picture of the past origin of the whole admixed population.

<table>
<thead>
<tr>
<th>Parental Component</th>
<th>Autosomal</th>
<th>Y Chromosome</th>
<th>mtDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bantu</td>
<td>65 to 95%</td>
<td>70 to 80%</td>
<td>70 to 90%</td>
</tr>
<tr>
<td>Arab</td>
<td>15 to 30%</td>
<td>20 to 30%</td>
<td>0 to 10%</td>
</tr>
<tr>
<td>Persia</td>
<td>0 to 15%</td>
<td>10 to 15%</td>
<td>0 to 10%</td>
</tr>
<tr>
<td>Asia</td>
<td>0 to 10%</td>
<td>0 to 5%</td>
<td>10 to 25%</td>
</tr>
</tbody>
</table>

Tab. 1. *Population Admixture contributions to Comorian population.*

**Haplogroup frequencies**

Contributions of parental populations are commonly inferred from haplogroup frequencies, counting the number of individuals sharing a same haplogroup. This approach has been extensively applied to underline a sexual bias in admixture dynamics (Marrero et al., 2007; Msaide et al., 2011) because of a different gender introgression introduced by sociocultural and historical factors such as colonization process (Gourjon et al., 2011). Implicitly, haplogroup frequencies of an admixed population refer to a region where these haplogroups are present in high frequency rather than to an admixture contribution
of this region. Indeed, haplogroup frequencies give a picture of current-day population in a synchronic non-evolutionary approach. Conversely to the previous approach, the admixture event is obviously ignored. Moreover, some Hg are present in several parts of the world, as for J* which is found in South-East Asia but in India and Middle-East as well (YCC consortium, updated by Karafet et al. (2008). In Tab. 2, we present parental contributions to the Comorian population inferred from MSaidie et al. (2011) data. The first values indicate frequencies of haplogroups which characterize a particular geographical location. To get the second value, we added haplogroups found in several regions and consequently, which can’t be undoubtedly assigned to a given origin. We observe that no certitude about a Middle-East contribution to mtDNA gene pool can be demonstrated by this approach.

<table>
<thead>
<tr>
<th>Parental Component</th>
<th>Y Chromosome</th>
<th>mtDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bantu</td>
<td>59.58% / 62.53%</td>
<td>88.91% / 84.75%</td>
</tr>
<tr>
<td>Arab/Persia</td>
<td>31.13% / 30.18%</td>
<td>0% / 4.67%</td>
</tr>
<tr>
<td>S-E Asia</td>
<td>6.34% / 10.24%</td>
<td>11.09% / 15.25%</td>
</tr>
</tbody>
</table>

From MSaidie et al. (2011).

Tab. 2. Geographical origins of Comorian population based on Hg frequencies.

In comparison to admixture estimates for Y chromosome (Tab.1), the Bantu component is lower while the Asian is slightly increased. Conversely, for mtDNA, the Bantu component is significantly increased whereas the mean Asian and Middle-East contributions both decrease.

Origins and Genetic Admixture in human populations

Individual ancestry and population ancestry are frequently mixed in genetic and anthropological literature. An admixture rate is not a sum of the individual admixture rates which give a contemporary picture of genetic structure. It is the ancestral parental populations’ contributions to the genetic pool of an admixed population, just after the admixture event and before they evolved. To study an admixture event, one has to follow a diachronic path, going back to the initial event, and to consider ancestral populations (both admixed and parental) instead of current ones, to take into account the dynamics of the admixture event and the evolutionary forces until sampling procedure.

The haplogroup frequencies only provide a picture of the current-day genetic structure of the admixed population. Moreover, haplogroups are uninformative at a local geographical level, being exclusively representative of wide geographical regions of the world.

References


