

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/matrilocal. 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 synchronic method considers haplogroup frequencies observed in an admixed population as representative of a given geographical parental contributions (Herrnstadt et al., 2002; Kivisild et al., 2004; Karafet et al., 2008). Conversely, the diachronic approach quantifies the parental population contributions (Population Admixture, noted PA) to the

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; Msaidie 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 Arabic 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

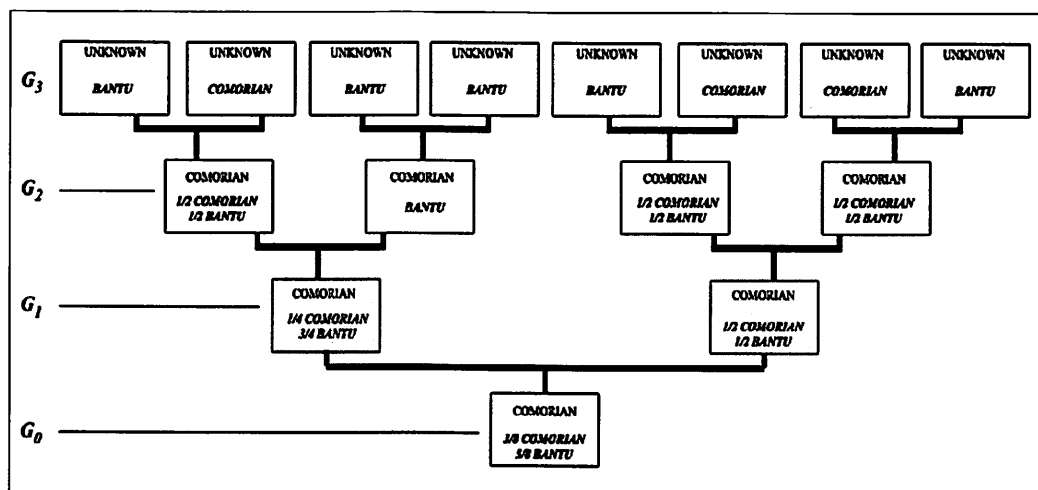


Fig. 1. Individual genealogy.

ancestry and all sampled individuals are “Comorian”. Now, considering that previous generation G3 is known (Fig. 1, bold italic). If five great grand-parents 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; Halder *et al.*, 2009; 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 p_1 and $(1 - p_1)$ 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 Hiorns, 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.

Parental Component	Autosomal	Y Chromosome	mtDNA
Bantu	65 to 95%	70 to 80%	70 to 90%
Arab	15 to 30%	20 to 30%	0 to 10%
Persia	0 to 15%	10 to 15%	0 to 10%
Asia	0 to 10%	0 to 5%	10 to 25%
<i>From Gourjon et al. (2011)</i>			

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; Msaidie *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 and 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.

Parental Component	Y Chromosome	mtDNA
Bantu	59,58% / 62,53%	88,91% / 84,75%
Arab/Persia	31,13% / 30,18%	0% / 4,67%
S-E Asia	6,34% / 10,24%	11,09% / 15,25%
From MSaidie <i>et al.</i> (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

- Alves-Silva J., Da Silva Santos M., Guimarães P.E., Ferreira A.C., Bandelt H.J., Pena S.D., Prado V.F. 2000. The ancestry of Brazilian mtDNA lineages. *Am. J. Hum. Genet.*, 67(2): 444-461.
- Benn-Torres J., Bonilla C., Robbins C.M., Waterman L., Moses T.Y., Hernandez W., Santos E.R., Bennett F., Aiken W., Tullock T., Coard K., Hennis A., Wu S., Nemesure B., Leske M.C., Freeman V., Carpten J., Kittles R.A. 2008. Admixture and population stratification in African Caribbean populations. *Ann. Hum. Genet.*, 72(Pt 1): 90-98.
- Benoist J. 1991. *Le métissage: biologie d'un fait social, sociologie d'un fait biologique. Linguistique et anthropologie.* L'Harmattan, Paris.
- Bertorelle G., Excoffier L. 1998. Inferring admixture proportions from molecular data. *Mol. Biol. Evol.*, 15(10): 1298-1311.
- Chakraborty R. 1985. *Gene identity in racial hybrids and estimation of admixture rates. Genetic differentiation in human and other animal populations.* Ahuja YR and Neel JV Ed., Delhi.
- Chikhi L., Bruford M., Beaumont M.A. 2001. Estimation of admixture proportions: a likelihood-based approach using Markov chain Monte Carlo. *Genetics*, 158(3): 1347-1362.
- Degioanni A., Gourjon G. 2010. Le mélange dans les populations humaines: modèles et méthodes d'estimation. *Anthropologie*, 48(1): 41-56.
- Falush D., Stephens M., Pritchard J.K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164(4): 1567-1587.
- Gourjon G. 2010. *L'estimation du mélange génétique dans les populations humaines.* Ph.D. Thesis, Aix-Marseille Université, Marseille.
- Gourjon G., Boëtsch G., Degioanni A. 2011. Gender and population history: Sex bias revealed by studying genetic admixture of Ngazidja population (Comoro Archipelago). *Am. J. Phys. Anthropol.*, 144(4): 653-660.
- Gourjon G., Degioanni A. 2008. Couleur de peau et classification biologique. In *Coloris Corpus.* CNRS Edition, Paris, 472 pp.
- Halder I., Yang B., Kranzler H.R., Stein M.B., Shriver M.D., Gelernter J. 2009. Measurement of admixture proportions and description of admixture structure in different U.S. populations. *Hum. Mutat.*, 30(9): 1299-1309.
- Herrnstadt C., Elson J.L., Fahy E., Preston G., Turnbull D.M., Anderson C., Ghosh S.S., Olefsky J.M., Beal M.F., Davis R.E., Howell N. 2002. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am. J. Hum. Genet.*, 70(5): 1152-1171.
- Karafet T., Mendez F., Meilerman M.B., Underhill P.A., Zegura S.L., Hammer M.F. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.*, 18(5): 830-838.
- Kivisild T., Reidla M., Metspalu E., Rosa A., Brehm A., Pennarun E., Parik J., Geberhiwot T., Usanga E., Villems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am. J. Hum. Genet.*, 75(5): 752-770.
- Klimentidis Y., Miller G., Shriver M.D. 2008. Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among Hispanics and Native Americans. *Am. J. Phys. Anthropol.*, 138(4): 375-383.
- Krieger H., Morton N., Mi M.P., Azevedo E., Freire-Maia A., Yasuda N. 1965. Racial admixture in north-eastern Brazil. *Ann. Hum. Genet.*, 29(2): 113-125.
- Lao O., Vallone P.M., Coble M.D., Diegoli T.M., van Oven M., van der Gaag K.J., Pijpe J., de Knijff P., Kayser M. 2010. Evaluating self-declared ancestry of U.S. Americans with autosomal,

- Y-chromosomal and mitochondrial DNA. *Hum. Mutat.*, 31(12): E1875-1893.
- Long J. 1991. The genetic structure of admixed populations. *Genetics*, 127(2): 417-428.
- Lopes Maciel L. G., Ribeiro Rodrigues E. M., Carneiro Dos Santos N. P., Ribeiro Dos Santos A., Guerreiro J. F., Santos S. 2011. Afro-derived amazonian populations: inferring continental ancestry and population substructure. *Hum. Biol.*, 83(5): 627-636.
- Marrero A., Das Neves Leite F.P., De Almeida Carvalho B., Peres L.M., Kommers T.C., Da Cruz I.M., Salzano F.M., Ruiz-Linares A., Da Silva Júnior W.A., Bortolini M.C. 2005. Heterogeneity of the genome ancestry of individuals classified as White in the state of Rio Grande do Sul, Brazil. *Am. J. Hum. Biol.*, 17(4): 496-506.
- Marrero A., Silva-Junior W.A., Bravi C.M., Hutz M.H., Petzl-Erler M.L., Ruiz-Linares A., Salzano F.M., Bortolini M.C. 2007. Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil. *Am. J. Phys. Anthropol.*, 132(2): 301-310.
- Masao F., Mutoro H. 1997. L'Afrique du VIIème au XIème siècle. In: *Histoire générale de l'Afrique*. Unesco (Ed.). Evreu, 560 pp.
- McKeigue P.M. 2005. Prospects for admixture mapping of complex traits. *Am. J. Hum. Genet.*, 76(1): 1-7.
- McKeigue P. M., Carpenter J. R., Parra E. J., Shriver M. D. 2000. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Ann. Hum. Genet.*, 64(Pt 2): 171-186.
- Msaidie S., Ducourneau A., Boetsch G., Longepied G., Papa K., Allibert C., Yahaya A.A., Chiaroni J., Mitchell M.J. 2011. Genetic diversity on the Comoros Islands shows early seafaring as major determinant of human biocultural evolution in the Western Indian Ocean. *Eur. J. Hum. Genet.*, 19(1): 89-94.
- Pritchard J., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155(2): 945-959.
- Roberts D., Hiorns R. 1965. Methods of analysis of the genetic composition of a hybrid population. *Hum. Biol.*, 37: 38-43.
- Shriner D., Adeyemo A., Ramos E., Chen G., Rotimi C.N. 2011. Mapping of disease-associated variants in admixed populations. *Genome Biol.*, 12(5): 223.
- Sousa V., Fritz M., Beaumont M.A., Chikhi L. 2009. Approximate bayesian computation without summary statistics: the case of admixture. *Genetics*, 181(4): 1507-1519.
- Wang J. 2003. Maximum-likelihood estimation of admixture proportions from genetic data. *Genetics*, 164(2): 747-765.