JASs Journal of Anthropological Sciences
Open Access
forum Vol. 99 (2021), pp. 183 - 185
doi. 10.4436/jass.99005

Uniparental markers and their role in the future of Molecular Anthropology

Francesc Calafell

Departament de Ciències Experimentals i de la Salut, Institut de Biologia Evolutiva (UPF-CSIC), Universitat Pompeu Fabra, Barcelona, Spain
e-mail:Francesc.calafell@upf.edu

Mitochondrial DNA and the Y chromosome share a property that has made them widely popular in Molecular Anthropology in the last 25 years: they escape recombination. Thus, save for the effects of recurrent mutation, perfect trees can be constructed from nucleotide variation at either genome region. They share also their uniparental inheritance, which incardinates them with the social and cultural aspects of relatedness and descent, as well as providing a sex-specific view of human migration (or lack thereof), in which measuring gene flow may be as easy as counting sequences with preassigned geographical origins. And, in the case of mtDNA, the high number of copies at which it is found means that it can be retrieved and sequenced from degraded or ancient samples where any similarly-sized autosomal region is long gone. A further advantage is its compactness: 16.5 Kb, with relatively minor structural variation, and only heteroplasmy and mitochondrial inserts in the nucleus (NUMTs) to worry about. Thus, the current situation is that tens of thousands of whole mtDNA sequences are available, all comfortably fitted into a universally accepted phylogenetic tree (www.phylotree.org) (van Oven and Kayser 2009)

The situation is not so ideal for the Y chromosome. Indeed, thousands of complete sequences have been produced and a few phylogenetic trees are available, but both come with their respective caveats. Complete sequence in Y-chromosome parlance means that only about 16% of the total sequence of the chromosome is deemed callable and reported, the rest being repetitive tracts that are beyond the capabilities of the usual massive parallel sequencing technologies. And the most comprehensive Y-chromosome phylogenies (https://isogg.org/tree/; https://www.yfull.com/tree/) lie beyond the realm of academia, having been produced by citizen scientists or by start-up companies, in order to help direct-to-consumer genomic company clients make heads or tails of their Y-chromosome results. Still, it is these trees that are often referred to by academic scientists when reporting their results.

Altogether, the application of unilinearly transmitted genome regions to Molecular Anthropology is a rather mature field, with seemingly little room for improvement. Still, I was not commissioned a eulogy, but a perspective for the future. So, what follows should be taken as my wish list for the field in the immediate future.

I mentioned above that very detailed phylogenetic trees exist both for mtDNA and the Y chromosome. Still, such detail does not extend to the geographic extent of each branch and tip. While for most, if not all, of the deepest branches a geographical notion of its range exists, as one moves towards the shallower branches, a population perspective is lost. Obviously, when one reaches the tips, that perspective is unreasonable, since most branch tips represent events that are so recent as to be represented by a few individuals. It would be highly desirable to add a population view to those phylogenies by sequencing population samples, both at a subcontinental level and enriching for understudied populations. This would also contribute to add complexity to the trees and to avoid the simplistic generalizations of systematically assigning haplogroups to particular population identifications or single prehistoric origins.

Initially, ancient DNA studies relied mostly on short mitochondrial sequences, or, if more DNA could be retrieved, then mtDNA coverage would be better than that for the rest of the genome, and thus more reliable. Next, the development of in-solution enrichment ("1240k capture") for a targeted set of 1,237,207, mostly autosomal SNPs (Lazaridis et al. 2014), as well as whole genome sequencing, shifted the field towards an almost exclusive autosomal-centric view (as it should, given that it is almost all of the genome), with mtDNA as an afterthought. Still, efforts are being made to retrieve and sequence mtDNA to high quality, with specific protocols (Mathieson et al. 2018) being recently developed. This opens the tantalizing possibility of diachronically watching the geographical and temporal mtDNA phylogeny unfold. Of course, then the main hurdle remains of finding the relevant samples, which are likely to be Palaeolithic or from warm climates that accelerate DNA degradation.

The Y chromosome has often been studied in a genealogical framework, by relating it to surname-tagged lineages, or to investigate the decoupling between biological and social lineages, as in false paternities (Calafell and Larmuseau 2017). As in ancient Roman law, unlike paternity, maternity is taken for granted. This should not necessarily be assumed, and at least one project (https://historiesvzw.be/projecten/mamamito/) is trying to remedy this situation. By sequencing mtDNA against a rich genealogical framework, it will allow a social insight on how often (and why) genetic links may not reflect the maternal genealogies, but may also provide empirical data on patterns and rates of mutation in mtDNA.

Steps are also being taken to produce actually complete Y-chromosome sequences (Kuderna et al. 2019). Although a technical tour de force was required to produce just one sequence, it is likely that further improvements will allow sequencing population samples or individuals of relevant position in the SNP-based Y-chromosome phylogeny. The goal is to incorporate structural diversity as well as sequence variation in so far inaccessible regions of the Y chromosome in the global phylogeny of the chromosome. And, although these technical improvements are not strictly necessary to that effect, the academic community should produce (or at least review) a solid, peer-reviewed Y-chromosome phylogeny.

Sampling properly is of paramount importance to any Molecular Anthropology project. Ideally, we select individuals with the relevant ancestral or geographical origins that match the research questions we have about particular populations; we often require all four grandparents to be born in the relevant populations. It is understandable, thus, that we as a field might be reluctant to engage with the many population genomics projects currently underway, which, having originated with a biomedical scope, are often rich in clinically-related metadata, but tend to lack geographical information about the volunteers' ancestors or even about themselves. Still, these projects are producing high-quality genomes, of which the autosomes may be analyzed in an anthropological sense (Leslie et al. 2015), but the unilinear markers tend to be woefully underexploited. I hope that in the future this is going to be corrected, and that, with all the caution that applies, we establish the relevant collaborations and approach these datasets as the treasure troves they might be.

References

Calafell F, Larmuseau MHD (2017) The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research. Hum Genet 136:559–573. https://doi.org/10.1007/s00439-016-1740-0

Kuderna LFK, Lizano E, Julià E, et al (2019) Selective single molecule sequencing and assembly of a human Y chromosome of African origin. Nat Commun 10:4. https://doi.org/10.1038/s41467-018-07885-5

Lazaridis I, Patterson N, Mittnik A, et al (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513:409–413. https://doi.org/10.1038/nature13673

Leslie S, Winney B, Hellenthal G, et al (2015) The fine-scale genetic structure of the British population. Nature 519:309–314. https://doi.org/10.1038/nature14230

Mathieson I, Alpaslan-Roodenberg S, Posth C, et al (2018) The genomic history of southeastern Europe. Nature 555:197–203. https://doi.org/10.1038/nature25778

van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30:E386–E394. https://doi.org/10.1002/humu.20921