International Congress Series 1288 (2006) 82–84

www.ics-elsevier.com

Relative Y-STR mutation rates estimated from the variance inside SNP defined lineages

P.A. Soares [a,b,*], F. Pereira [a,c], M. Brion [d], C. Alves [a], M. Richards [b], A. Carracedo [d], A. Amorim [a,c], L. Gusmão [a]

[a] *IPATIMUP, Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Portugal*
[b] *Institute of Integrative and Comparative Biology, University of Leeds, United Kingdom*
[c] *Faculdade de Ciências da Universidade do Porto, Portugal*
[d] *Unidad de Genética Forense, Instituto de Medicina Legal, Univ. de Santiago de Compostela, Spain*

Abstract. Y specific microsatellites (STRs) have been widely used in forensic and population genetics in age estimates of human male lineages. Previously, estimates of mutation rates from father–son pairs have given quite variable results in different studies, essentially due to the rarity of mutations. We propose an indirect approach for determining relative mutation rates of Y-chromosome microsatellites based on STR allele size intra-lineage variance. Indeed, the present distribution of STR alleles offers us an insight into the mechanisms that have generated that diversity.

Keywords: Y-lineages; Y-STR; Mutation rate

1. Introduction

The non-recombining portion of the Y-chromosome is widely employed for defining male-specific history in human populations and detecting male exclusions in forensic cases. Microsatellites (or short tandem repeats, STRs) are important markers in this context since, due to their high variability, they narrow considerably the identification of lineages in forensics and allow ages estimation of specific SNP-defined lineages. Given this, determining specific mutation rates for STRs is crucial since it is the driving force in the age estimation and they are incorporated in exclusion calculations.

* Corresponding author. Institute of Integrative and Comparative Biology, University of Leeds, LS2 9JT, United Kingdom.
E-mail addresses: bgypas@leeds.ac.uk, pedros@ipatimup.pt (P.A. Soares).

0531-5131/
doi:10.1016/j.ics.2005.09.173

Mutation, even in the case of STRs, is a rare event to observe directly, so estimations of effective mutation rates through analysis of father–son pairs present very high confidence intervals. On the other hand, the present distribution of STR alleles reflects mutation processes and the different alleles are the result of mutation events that took place at some time in the past. Here we present a method to estimate relative mutation rates for the most commonly applied Y-STRs in forensics, taking into account the allele size variance (ASV) for the microsatellites within SNP defined lineages.

2. Material and methods

We performed simulations of the theoretical distribution of STR alleles through generations, starting with a frequency of 1 in one allele, using a stepwise mutation model (SMM) and different mutation rates. This situation is analogous to the emergence of a SNP allele and the consequent arising of STR variability within a SNP defined lineage.

3. Results and discussion

3.1. Relative mutation rate

Graphic 1 presents the variation of the ASV through time. The lines are all perfect linear correlations ($r=1$) which could be defined by the equation:

$$\mathrm{ASV} = \mathrm{mASV} \cdot \Delta t,$$

where mASV is the slope of the line and Δt is the elapsed time. When comparing ASVs for different STRs inside the same lineage, we are accessing a variation assembled in the same time interval (Δt), so, from the previous equation we obtain:

$$\Delta t = \frac{\mathrm{ASV}}{\mathrm{mASV}}.$$

Since for microsatellites A and B within the same lineage $\Delta t_A = \Delta t_B$,

$$\frac{\mathrm{ASV_A}}{\mathrm{mASV_A}} = \frac{\mathrm{ASV_B}}{\mathrm{mASV_B}} \Leftrightarrow \mathrm{mASV_A}\frac{\mathrm{ASV_A}}{\mathrm{ASV_B}} \cdot \mathrm{mASV_B}.$$

Since the slopes are directly obtained from the mutation rates the equation could be changed to

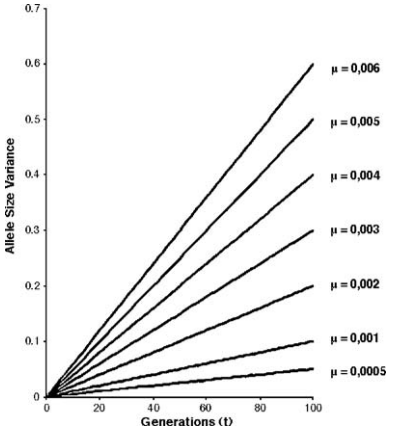
$$\mu_A = \frac{\mathrm{ASV_A}}{\mathrm{ASV_B}} \cdot \mu_B \; or \; \mu_A = R_{AB} \cdot \mu_B$$

where R_{AB} is the relative mutation rate between STRs A and B, obtained from the quotient between the ASV inside the lineage.

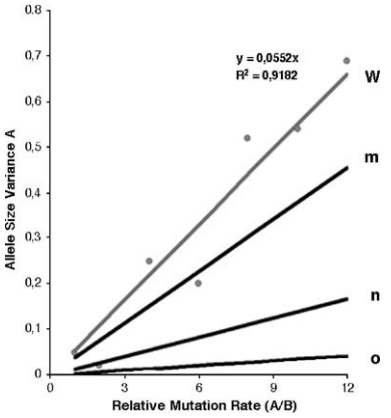
However lineages are not elements confined geographically with ideal growth, but are acted upon by several demographic processes, so the values of a single lineage could be misleading. The values of the ASVs inside a lineage could be biased in a way that are over-evaluated or under-evaluated in relation to another, but those biases should be randomly directional and the mean of the relative mutation rates of different lineages should be a good estimate of the real relation between Y-STR mutation rate.

3.2. Relative dating

When making a graphic relating the relative mutation rates of the Y-STRs in relation to one of them and the variance found at each of the microsatellites in a lineage we would theoretically obtain a perfect linear correlation (Graphic 2).



Graphic 1. Theoretical relation between allele size variance in a STR and elapsed time with seven different mutation rates.



Graphic 2. Relation between allele size variance of microsatellite A and relative mutation rate of A to B. w, m, n and o represent four lineages.

That correlation can be expressed as:

$$\mathrm{ASV_{Ai}} = L \cdot R_{\mathrm{AiB}}$$

where L represents the lineage's specific slope.

Since $\mathrm{ASV} = \mathrm{mASV} \cdot \Delta t$, we obtain: $\mathrm{mASV_{Ai}} \cdot \Delta t = L \cdot R_{\mathrm{AiB}}$.

And $\Delta\mathrm{ASV}$ and R_{AiB} are the same for one considered microsatellite in two different lineages m and n, so

$$\frac{m\mathrm{ASV_{Ai}}}{R_{\mathrm{AiB}}} = \frac{L}{\Delta t} \Leftrightarrow \frac{L_m}{\Delta t_m} = \frac{L_n}{\Delta t_n} \Leftrightarrow \Delta t_m = \frac{L_m}{L_n} \cdot \Delta t_n.$$

Concluding, the quotient between the two lineage specific slopes (L) indicates an estimate of a relative age between them. Since at this point we should be using mean values of relative mutation rates, the correlation will not be perfect (lineage w in Graphic 2), but the values will accommodate to the best fit.

3.3. Practical application

A relative mutation rate in itself is not very useful, but the values can be used to estimate an effective locus-specific mutation rate. The calibration can be done using either the STR with the highest number of analysed meioses (or the one that presents the more stable values with new experimental observations); or a mean value of relative mutation rate vs. a mean value of effective mutation rate. The plot of relative mutation rates can also be valuable for placing novel Y-STRs for which there is no estimate of mutation rate, as an alternative to accumulating a large number of meioses.

Acknowledgements

This work is supported by Fundação para a Ciência e Tecnologia POCTI, Programa Operacional Ciência, and PAS is supported by a Marie Curie EST fellowship in Advanced Genetic Analysis.