Title: Theory and statistics of mutation rates: a mathematical framework reformulation for forensic applications

Author: N. Pinto C. Alves L. Gusmão A. Amorim

PII: S1875-1768(15)30102-5
DOI: http://dx.doi.org/doi:10.1016/j.fsigss.2015.09.053
Reference: FSIGSS 1050

To appear in:

Received date: 25-8-2015
Accepted date: 14-9-2015

Please cite this article as: N.Pinto, C.Alves, L.Gusmão, A.Amorim, Theory and statistics of mutation rates: a mathematical framework reformulation for forensic applications, Forensic Science International: Genetics Supplement Series http://dx.doi.org/10.1016/j.fsigss.2015.09.053

Theory and statistics of mutation rates: a mathematical framework reformulation for forensic applications

N. Pinto[a,b,c*], C. Alves[a,b], L. Gusmão[d], A. Amorim[a,b,e]

[a] *Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal*
[b] *Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Porto, Portugal*
[c] *Center of Mathematics of the University of Porto, Porto, Portugal*
[d] *Laboratório de Diagnóstico por DNA (LDD), Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, Brazil*
[e] *Faculdade de Ciências, Universidade do Porto, Porto, Portugal*

Abstract. Mutation is a fundamental topic in forensic genetics since the calculation of kinship likelihoods depends on estimates of germinal mutation rates. Germinal mutation rates have been computed by simply proportioning the number of true parent(s)-child genotypic configurations inconsistent with Mendelian segregation. For STRs, current technology is based on PCR fragment size determination and germinal mutations are detected when parent-child do not share alleles' length. For technical reasons, it is uncommon to obtain the sequence composition of STR alleles, and the identification of the allele that mutated is based on assumptions. It has been assumed that one-step are much more common than multi-step mutations. Whenever a genotypic parent-child configuration is compatible with Mendelian rules by a single-step mutation, this is the one assumed, despite other possible mutational events. Multi-step events are therefore evoked exclusively when a single-step cannot reconcile the observation with the true kinship, leading to an overestimation of single step mutation rates. For any mode of transmission other than purely haploid it is theoretically impossible to identify both the allele at "origin" (in parents) and at "destination" (in offspring). Therefore a more sophisticated statistical framework than the simple proportioning is then required to properly evaluate mutation rates. Moreover, the possibility of the occurrence of silent alleles should be simultaneously modeled since it has to be considered whenever an apparently homozygous individual is involved.

Keywords: kinship analyses, mutation, mutation rates, estimation of mutation rates

1. Introduction

Mutation is a genetic phenomenon characterized by a sudden change in the genome of somatic or germinal cells of the individual. Mutations that occur in the germinal line are susceptible of being transmitted to the offspring. In these cases the child exhibits an allele different from those present in both parents revealing what is designated as a Mendelian incompatibility. Moreover, in some cases, mutations can occur even if they do not translate into a Mendelian incompatibility [1], which are the so-called "hidden mutations" since they are undetectable through (current) genotyping techniques used in forensics based on the determination of the size of the alleles at STR markers.

Mutation rates have been estimated resorting to trios (or duos) parent(s)-child by the simple proportioning of the cases revealing Mendelian incompatibilities. When no incompatibilities are found the most parsimonious argument is used and the hypotheses that mutations have not occurred is assumed (see however [1]).

Some works have shown that mutation rates depend on several factors such as on: (a.) the type of markers, (b.) the system considered, (c.) the fragment sizes – and are thus allele specific, and (d.) the difference of repeats between the origin and mutated alleles – and are thus bi-allele specific, being the loss or gain of a single repeat considered much more frequent than multi-repeat gains or losses (see [2-5], for example). Indeed, whenever a genotypic configuration of a pair parent-child is compatible with the Mendelian rules of genetic transmission by assuming a single-step mutation, this is the one considered as having occurred. All other possibilities are disregarded, resulting in a vicious reasoning cycle that leads, necessarily, to a biased (overestimation) of single step mutation rates.

Moreover, it should be highlighted that, if for uniparental inheritance, as in Y-chromosomal transmission, there is no doubt which parental allele originated which allele at the offspring [6,7], the same does not occur for any other mode of transmission, which prevents the naïve proportioning method of being an appropriate approach for the estimation of bi-allelic mutation rates for autosomal and X-chromosomal analyses.

Finally, we want to highlight that some genotypic configurations incompatible with Mendelian segregation are also explainable by the presence of a silent allele, which is also a (hidden) source of bias and whose frequency cannot be estimated independently from mutation rates.

2. Methods