

International Congress Series 1239 (2003) 57-60

An evaluation of the proportion of identical Y-STR haplotypes due to recurrent mutation

L. Pereira^{a,b,*}, M.J. Prata^{a,b}, A. Amorim^{a,b}

^aInstituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), R. Dr. Roberto Frias, s/n, 4200 Porto, Portugal ^bFaculdade de Ciências da Universidade do Porto, Praça Gomes Teixeira 4050 Porto, Portugal

Abstract

We present an approach to the evaluation of the probability of Y-STR haplotypes identity by state (IBS, in opposition to identity by descent, IBD) as an alternative to the empirical one for which it is necessary to have simultaneous information on STRs and SNPs, which is rare in the forensic field. It is based on the mismatch distribution analyses for the number of repeat unit differences between pairs of Y-STR haplotypes. The estimates of the IBS by both methods in a sample where STR and SNP information are available are compared and also the relevance of population structure in determining widely different proportions of IBS vs. IBD. © 2003 Elsevier Science B.V. All rights reserved.

Keywords: Recurrence; Identity by state; Identity by descent; Mismatch distribution

1. Introduction

Searching for haplotype matching is the main tool used in forensics to evaluate how significant is the observation of a particular haplotype in a certain sample. A neglected phenomenon that can be an important bias source to this kind of analysis is the possibility of recurrence, due to the high mutation rate of STRs, and hence, the rise of haplotypes identical by state, rather than by descent.

One approach that has been used to evaluate the proportion of these classes among identical haplotypes is to assess haplotype identity using SNPs, the same Y-STR haplotype in two different Y-SNP haplogroups pointing to identity by state and not by descent. De

^{*} Corresponding author. Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), R. Dr. Roberto Frias s/n, 4200 Porto, Portugal. Tel.: +351-225570700; fax: +351-225570799.

E-mail address: lpereira@ipatimup.pt (L. Pereira).

Knijff [1] estimated an IBS proportion of 0.2% in a sample of 275 Dutch screened for eight Y-STRs and four SNPs.

Unfortunately, combined information for both Y-STRs and Y-SNPs in the same sample is very rare, and it is therefore impossible to measure the proportion of IBS, at a large scale, using the above-described approach.

We suggest an approach based upon mismatch distributions/haplotype pairwise comparison that can at least indicate the proportion of Y-STR haplotypes in which mutation can originate IBS.

2. Material and methods

A total of 123 different Y-STR haplotypes for 229 Dutch were collected from the Y-STR Haplotype Reference Database (http://ystr.charite.de). The samples of Holland (N=87), Friesland (N=44), Groningen (N=48) and Limburg (N=50) were considered together. The Y-STRs analysed were DYS19, DYS389I and II, DYS390, DYS391,



Fig. 1. Mismatch distributions for the number of loci (A) and number of repeat unit (B) pairwise differences for Y-STRs in a Dutch sample.

DYS392 and DYS393. Mismatch distributions were calculated in Arlequin 2.0 [2] software and the input files were: (1) the microsatellite input file, considering the number of repeat units observed in each locus, for obtaining the mismatch distribution for differences in number of locus; (2) a DNA input file, where each repeat unit is considered as an ambiguous position (N) and differences in the number of repeat units are considered as insertion (N) or deletion (-), for estimating the mismatch distribution for differences in repeat units between pairs of haplotypes.

3. Results

The two mismatch distributions obtained for the 229 Dutch are displayed in Fig. 1. Graph A refers to the mismatch distribution for differences per locus and graph B for differences in repeat units.

It is noteworthy that 86.8% (6.38%/7.35%) of the differences between haplotypes differing at one locus consists in a single repeat unit.

The probability of a certain Y-STR haplotype becoming identical by state to another one inside the class of those that differ by a single repeat can be roughly estimated as $m(1-m)^{2n-1}$, where *n* is the number of loci defining the haplotype (7 in our case) and *m* the average mutation rate $(3.17 \times 10^{-3} [3])$. Applying it to the proportion of the class (6.38%) it results, for this sample, in a value of 0.0194%.

We have not considered the rare cases of IBS resulting from more complex phenomena since these will add an almost insignificant contribution (for instance, two-step mutation rate is roughly 10 times lower [3]).

4. Discussion

Table 1

STRs have been the markers of choice in forensics due to its high polymorphism. SNPs, although presenting high stability, are much less polymorphic, becoming less informative in the forensic field, while more advanced technologies allowing its easy and

reperious of puls of higher differing in o (n) and r (n) repeat and and estimates of identity by state (inb)						
h_0 (%)	h_1 (%)	IBS (%)	Population	h ₀ (%)	h_1 (%)	IBS (%)
1.85	6.70	0.020	The Netherlands (229)	2.05	6.38	0.019
1.69	5.77	0.018	Poland (596)	1.51	5.83	0.018
0.92	3.02	0.009	Estonia (133)	1.31	3.68	0.011
0.98	3.45	0.010	Latvia (145)	0.93	4.53	0.014
0.43	2.18	0.007	Lithuania (151)	1.17	4.73	0.014
1.11	4.26	0.013	Russia (85)	2.24	6.44	0.020
0.83	3.18	0.010	Norway (300)	1.41	3.75	0.011
1.55	5.61	0.017	Buenos Aires (100)	1.23	2.87	0.009
	h_0 (%) 1.85 1.69 0.92 0.98 0.43 1.11 0.83 1.55	$\begin{array}{c cccc} h_0 (\%) & h_1 (\%) \\ \hline h_0 (\%) & h_1 (\%) \\ \hline 1.85 & 6.70 \\ \hline 1.69 & 5.77 \\ \hline 0.92 & 3.02 \\ \hline 0.98 & 3.45 \\ \hline 0.43 & 2.18 \\ \hline 1.11 & 4.26 \\ \hline 0.83 & 3.18 \\ \hline 1.55 & 5.61 \\ \hline \end{array}$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	h_0 (%) h_1 (%)IBS (%)Population1.856.700.020The Netherlands (229)1.695.770.018Poland (596)0.923.020.009Estonia (133)0.983.450.010Latvia (145)0.432.180.007Lithuania (151)1.114.260.013Russia (85)0.833.180.010Norway (300)1.555.610.017Buenos Aires (100)	h_0 (%) h_1 (%) IBS (%) Population h_0 (%) 1.85 6.70 0.020 The Netherlands (229) 2.05 1.69 5.77 0.018 Poland (596) 1.51 0.92 3.02 0.009 Estonia (133) 1.31 0.98 3.45 0.010 Latvia (145) 0.93 0.43 2.18 0.007 Lithuania (151) 1.17 1.11 4.26 0.013 Russia (85) 2.24 0.83 3.18 0.010 Norway (300) 1.41 1.55 5.61 0.017 Buenos Aires (100) 1.23	h_0 (%) h_1 (%)IBS (%)Population h_0 (%) h_1 (%)1.856.700.020The Netherlands (229)2.056.381.695.770.018Poland (596)1.515.830.923.020.009Estonia (133)1.313.680.983.450.010Latvia (145)0.934.530.432.180.007Lithuania (151)1.174.731.114.260.013Russia (85)2.246.440.833.180.010Norway (300)1.413.751.555.610.017Buenos Aires (100)1.232.87

Proportions of pairs of haplotypes differing in $0 (h_0)$ and $1 (h_1)$ repeat unit and estimates of identity-by-state (IBS)

Populations deposited in the Forensic database were grouped by country (data extracted in 11/04/2001; only Zeeland was not used).

fast typing are not available. So, extensive forensic databases available for matching are now (and will be maintained for more years) exclusively based on STRs.

It seems therefore that the approach presented here can be useful in the forensic field, for the evaluation of the significance and evidential value of Y-STR haplotype matches.

Another important issue is the use of large-scale databases. It has been claimed that, since there is no significant population substructuring for STRs in Europe (contrary to SNPs [4]) matching against the whole database is legitimate. We demonstrate that this claim is at least debatable, since the opportunity for IBS varies a lot across Europe, as judged from the mismatch distributions (Table 1), in a way congruent with the well-defined SNPs haplogroups' gradients [5]: higher SNPs diversity in south-central Europe associated with lower proportion of haplotype pairs prone to IBS; and decreasing SNPs diversity towards west and north associated with higher risk of recurrence.

Acknowledgements

This work was partially supported by a research grant (PRAXIS XXI BD/13632/97) from Fundação para a Ciência e a Tecnologia and IPATIMUP by Programa Operacional Ciência, Tecnologia e Inovação (POCTI), Quadro Comunitário de Apoio III.

References

- P. De Knijff, Y chromosome shared by descent or by state, in: C. Renfrew, K. Boyle (Eds.), Archaeogenetics: DNA and the Population Prehistory of Europe, McDonald Institute Monographs, Oxbow Books, Cambridge, 2000, pp. 301–304.
- [2] S. Schneider, D. Roessli, L. Excoffier, Arlequin ver. 2.0: A Software for Population Genetic Data Analysis, Genetics and Biometry Laboratory, University of Geneva, Switzerland, 2000.
- [3] M. Kayser, L. Roewer, M. Hedman, L. Henke, J. Henke, S. Brauer, C. Krüger, M. Krawczak, M. Nagy, T. Dobosz, R. Szibor, P. De Knijff, M. Stoneking, A. Sajantila, Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs, Ann. Hum. Genet. 66 (2000) 1580–1588.
- [4] Z.H. Rosser, et al., Y-chromosomal diversity within Europe is clinical and influenced primarily by geography rather than language, Am. J. Hum. Genet. 67 (2000) 1526–1543.
- [5] L. Pereira, I. Dupanloup de Ceuninck, Z.H. Rosser, M.A. Jobling, G. Barbujani, Y-chromosome mismatch distributions in Europe, Mol. Biol. Evol. 18 (2001) 1259–1271.