International Congress Series 1261 (2004) 281–283

www.ics-elsevier.com

GATA C4 allele 17 as a marker for sub-Saharan origin of Y-chromosome lineages

Leonor Gusmão[a,*], Paula Sánchez-Diz[b], Cíntia Alves[a], Maria Brión[b], Sandra Beleza[a,b], Luísa Pereira[a], Alejandro Blanco[b], Maria João Prata[a,c], Angel Carracedo[b], António Amorim[a,c]

[a] *Population Genetics, IPATIMUP, Institute of Pathology and Immunology of University of Porto, Rua Dr. Roberto Frias, s/n 4200-465 Oporto, Portugal*
[b] *Institute of Legal Medicine, University of Santiago de Compostela, Spain*
[c] *Faculty of Sciences, University of Porto, Portugal*

Received 27 August 2003; received in revised form 27 August 2003; accepted 8 September 2003

Abstract. In 1020 males out of 13 population samples from Argentina, Brazil, Costa Rica, Macao, Mozambique, Portugal and Spain, GATA allele 17 was found exclusively in Mozambique [Forensic Sci. Int. 135 (2003) 158]. In the present work, allele 17 was further observed in samples from Angola (9.33%), Mozambique (10.79%) and S. Tomé e Príncipe (3.53%). This allele differs by at least two repeat units from all reported alleles: no instance of allele 18 was ever found. Allele 17 frequency and distribution, together with the absence of one-step-derived alleles, is compatible with a very low mutation rate and so potentially a good marker for African ancestry and haplogroup identification. We verified this hypothesis in two ways: (a) the sequencing of nine alleles 17 showed that the small number of uninterrupted repeats can explain the low mutation rate and (b) the simultaneous GATA C4 and SNP typing in 75 individuals from Angola, 85 from S. Tomé e Príncipe and 417 from Mozambique proved that, out of 56 samples carrying allele 17, 55 belong to B* paragroup and one to A*, both said to be ancient African lineages. Therefore, GATA C4 allele 17 most probably was originated as a unique event and its extreme association with AB* is compatible with its use as a marker for the assignment to this paragroup.

Keywords: Y chromosome; STR; GATA C4; SNP; Africa; Haplogroup

1. Introduction

The STR GATA C4 is one of the most polymorphic Y chromosome-specific markers currently used in population and forensic genetics. Until now, nine alleles were described in different populations and three different allelic sequence complex structures were

* Corresponding author. Tel.: +351-225570700; fax: +351-225570799.
E-mail address: lgusmao@ipatimup.pt (L. Gusmão).

0531-5131/
doi:10.1016/S0531-5131(03)01493-6

identified [1]. The most frequent allele described in African and Asian samples is allele 21, while 23 is the most frequent in European samples [2]. In the same way as for DYS392 and DYS438, GATA C4 presents a bimodal allele frequency distribution in European populations, with allele 21 as the second most frequent. In 1020 males out of 13 population samples from Argentina, Brazil, Costa Rica, Macao, Mozambique, Portugal and Spain, allele 17 was found exclusively in the Mozambique sample [2]. Apart from allele 17, the shortest allele until now described for GATA C4 was allele 19 and no instance of allele 18 was ever found (in Africans, Caucasians or Asians). The allele 17 frequency and distribution, together with the absence of one-step-derived alleles 16 or 18 is compatible with a very low mutation rate in these alleles and, therefore, it seems to be a good marker for African ancestry as well as for haplogroup identification. In order to check our hypothesis (a), we sequenced nine samples with this allele and (b) we studied the GATA C4 alleles association with haplogroups defined by SNPs.

2. Methods and materials

A sample of 75 individuals from Cabinda (Angola, African western coast), 85 from S. Tomé e Príncipe (Gulf of Guinea) and 417 individuals from Mozambique (African eastern coast), were typed for GATA C4 and SNPs. GATA C4 PCR amplification and sequence analysis were performed according to Sánchez-Diz et al. [2]. SRY10831a was typed according to Rosser et al. [3]. The remaining SNPs were typed by sequence analysis using the primers and PCR cycling condition described in Table 1. Sequencing reactions were performed using dRodamine Terminator Cycle Sequencing Ready Reaction Kit (AB, Applied Biosystems) following the manufacturer's instructions. Detection and analysis were undertaken in an ABI 3100 Genetic Analyser (AB, Applied Biosystems).

3. Results and discussion

GATA C4 allele 17 was found in three Sub-Saharan population samples from Angola (9,33%), Mozambique (10,79%) and S. Tomé e Príncipe (3,53%). In the same populations,

Table 1
SNP primer sequences and PCR conditions

SNP	Primers	PCR thermo cycling conditions
M168	5′AGT TTG AGG TAG AAT ACT GTT TGC T 3′ 5′AAA AAG CCA TGC AAT TAC CTG 3′	95 °C–2′
M112	5′ACT TTT TCC AAC AGT TAT TTT TGA 3′ 5′TAT ATT TCT TGA TGA TGA GAC CAA T 3′	32 cycles
M150	5′GCA GTG GAG ATG AAG TGA GAC 3′ 5′CCT ACT TTC CCC CTC TTC TG 3′	94 °C–30″
M108	5′AGA TGG AGC CAG CAG AAA G 3′ 5′ACA CAG ATG AAT TGA ATG ATG GT 3′	58 °C–30″
M109	5′GGG TAT CAA ATG TCT TCA ACC T 3′ 5′GGG AAT TTC CTG CTA CTT GC 3′	72 °C–45″

PCR amplification was carried out using 5 to 50 ng of genomic DNA in a 25 μl reaction volume containing 1.5 mM $MgCl_2$, 200 μM of dNTPs, 1 × Buffer and 0.5 U Taq DNA Polymerase Recombinant (Invitrogen).

Table 2
Haplogroup and GATA C4 alleles distribution inside AB* paragroup

Haplogroup	*N*	M168	SRY10831a	M112	M150	M108	M109	GATA C4
A*	1	C	A	✓	✓	✓	✓	17
B2a1	28	C	G	G	T	T	T	17
B2a* (xB2a1,2)	1	C	G	G	T	T	C	17
B2b	4	C	G	A	C	✓	✓	19
B2b	1	C	G	A	C	✓	✓	20
B2b	6	C	G	A	C	✓	✓	21
B2b	3	C	G	A	C	✓	✓	22
B* (xB2ab)	1	C	G	G	C	✓	✓	21

✓This SNP was not typed in this sample since a phylogenetic strategy was used to determine the paragroup.

this allele differs by at least two repeat units from the remaining alleles (20 to 24 in Angola, 19 to 24 in S. Tomé e Príncipe and 19 to 26 in Mozambique).

The same sequence was found for the nine allele 17 samples analysed, $(TCTA)_2[(TCTA)_2(TGTA)_2]_2(TCTA)_7$. In accordance with previous observations in other STRs, a low mutation rate is expected for allele 17, due to the low number of uninterrupted repeats.

All C4*17 samples ($n = 55$) were classified as belonging to the African AB* paragroup, in fact, just one was of A* paragroup. The GATA C4 allele distribution inside B* paragroup is shown in Table 2. These samples were further characterized using additional SNPs (Table 2). From the results, it was observed that all the samples carrying GATA C4 allele 17 belong to B2a* and no other alleles were found inside this paragroup. Paragroup B2b* was polymorphic for GATA C4, with alleles 19 to 22.

Although associations between STR haplotypes and SNP haplogroups do exist, most of the time STR alleles cannot be used to predict haplogroups, due to their high-STR mutation rates. However, the present results demonstrate that GATA C4 allele 17 most probably was originated as a unique event and, its extreme association with AB* is compatible with the use of GATA C4 as a marker for the assignment to paragroups A* or B2a*.

Acknowledgements

This work was supported by Fundacão para a Ciência e a Tecnologia, Programa Operacional Ciência, Tecnologia e Inovacão (POCTI).

References

[1] S. Beleza, C. Alves, A. González-Neira, M. Lareu, et al., Int. J. Leg. Med. 117 (2003) 27–33.
[2] P. Sánchez-Diz, L. Gusmão, S. Beleza, et al., Forensic Sci. Int. 135 (2003) 158–162.
[3] Z. Rosser, T. Zerjal, M.H. Hurles, et al., Am. J. Hum. Genet. 67 (2000) 1526–1543.