ISSN 0026-8933, Molecular Biology, 2009, Vol. 43, No. 4, pp. 573–577. © Pleiades Publishing, Inc., 2009.
Original Russian Text © B.A. Malyarchuk, M.A. Perkova, M.V. Derenko, 2009, published in Molekulyarnaya Biologiya, 2009, Vol. 43, No. 4, pp. 623–627.

GENOMICS. TRANSCRIPTOMICS. PROTEOMICS

UDC 575.174

Different Instability of the CAG Microsatellite in Two Haplotype Groups of Human Mitochondrial DNA Polymerase Gamma

B. A. Malyarchuk, M. A. Perkova, and M. V. Derenko

Institute of Biological Problems of the North, Far East Branch of the Russian Academy of Sciences, Magadan, 685000 Russia; e-mail: malyarchuk@ibpn.ru.

Received October 16, 2008
Accepted for publication February 05, 2009

Abstract—Two single nucleotide polymorphisms of the mitochondrial DNA polymerase gamma gene (*POLG1*), rs2238296 (T/C) and rs758130 (T/C), were analyzed in individuals of different ethnicity (Russians and Buryats) with known genotypes of the CAG microsatellite located in the same gene. It was shown that microsatellite alleles with repeat numbers other than 10 were significantly more frequent within the TT haplotype. A phylogenetic analysis of human and chimpanzee *POLG1* intron 2 sequences suggested that the haplotype TT, which is more heterogeneous regarding the CAG repeat polymorphism, is evolutionally younger than the haplotype CC. These data may be useful in the further research of the association between the CAG microsatellite polymorphism of *POLG1* and male infertility.

DOI: 10.1134/S0026893309040074

Key words: human mitochondrial DNA polymerase γ (*POLG1*), microsatellite instability, linkage analysis, population polymorphism

INTRODUCTION

Human mitochondrial DNA (mtDNA) is replicated by mitochondrial DNA polymerase γ, which comprises a catalytic subunit encoded by *POLG1* on chromosome 15q25 and an accessory subunit encoded by *POLG2* on chromosome 17q23–24 [1]. The genetic variation of *POLG1* was studied mainly in patients with hereditary mitochondrial disorders and in control groups of individuals, primarily Caucasians [2–9]. This research led to the discovery of a number of *POLG1* mutations associated with numerous nucleotide deletions and substitutions in mtDNA of the patients' skeletal muscle and other tissues; the list of known mutations is available, for example, from the Human DNA Polymerase Gamma Mutation Database at www.tools.niehs.nih.gov/polg.

The best studied *POLG1* polymorphism is the CAG microsatellite repeat in exon 2 [8–12], with the number of repeats varying from 6 to 13. A screening study involving 1330 individuals from 12 ethnic groups established that the allele containing 10 CAG repeats ($(CAG)_{10}$ allele) was the most common one in all Eurasian populations: its frequency varied from 88% in Poland to 96% in Korea [12]. Among the alleles with the number of repeats other than 10 ($(CAG)_{non\text{-}10}$ alleles), the most frequent one contained 11 repeats (6.7%).

The polymorphism of the CAG repeat may be functionally significant, as it encodes a polyglutamine stretch in the enzyme, and a change in its length may affect the DNA synthesis fidelity and result in replication errors (single nucleotide deletions, insertions, or substitutions in mtDNA) [13]. It was reported that the frequency of the $(CAG)_{non\text{-}10}/(CAG)_{non\text{-}10}$ genotype was significantly higher in infertile men with azoospermia and in patients with seminoma of the testicle [8, 11]. Although other studies did not confirm an association of the $(CAG)_{non\text{-}10}/(CAG)_{non\text{-}10}$ genotype with azoospermia [9, 14–17], it may be present under certain conditions.

On the ethnic level, the $(CAG)_{non\text{-}10}/(CAG)_{non\text{-}10}$ genotype frequency varies from 0.5% in East Asian populations to 2.4% in European populations [12] and to 9–11% in populations of Africa [9]. As shown by the analysis of single nucleotide *POLG1* polymorphisms, there are two major groups of linked *POLG1* alleles (haplotypes) ([18], data from the HapMap project (www.hapmap.org)). For instance, in Tuvans, Altay, Yakuts, and Russians, the polymorphic *POLG1* loci rs758130 (T/C) and rs2238296 (T/C) are in a complete (or nearly complete in Tuvans) linkage disequilibrium [18]. However, the CAG microsatellite allele distribution within the two *POLG1* haplotype groups has not been studied so far, although such data may be helpful for clarifying the association between different microsatellite alleles and male infertility.

Our objective was to investigate two *POLG1* SNPs (rs758130 and rs2238296) in individuals of different ethnicity (Russians and Buryats) with known CAG

Table 1. The genotype distribution of three *POLG1* loci in Russians and Buryats

Loci			Frequency	
Number of CAG repeats	rs2238296	rs758130	Russians (n = 50)	Buryats (n = 94)
10/10	T/T	T/T	0.22 (11)	0.29 (27)
10/10	T/C	T/C	0.34 (17)	0.43 (40)
10/10	C/C	C/C	0.14 (7)	0.11 (10)
10/10	T/T	T/C	0	0.02 (2)
10/10	T/C	C/C	0.06 (3)	0.04 (4)
10/10	T/T	C/C	0	0.02 (2)
10/non-10	T/T	T/T	0.16 (8)	0.05 (5)
10/non-10	T/C	T/C	0.02 (1)	0.04 (4)
10/non-10	T/T	T/C	0.02 (1)	0
non-10/non-10	T/T	T/T	0.02 (1)	0
non-10/non-10	T/C	T/C	0.02 (1)	0

genotypes and to analyze the distribution of microsatellite variants within different SNP haplotypes.

EXPERIMENTAL

Subjects participating in this study were Russian residents of Belgorodskaya Oblast' (n = 50) and Buryats from the Buryat Republic (n = 94). The polymorphisms rs758130 (T/C) and rs2238296 (T/C) located in *POLG1* intron 2 1500 bp from each other were studied by RFLP analysis as described by Buikin et al. [18]. We also used our previously published data on the variation of the CAG repeat located in *POLG1* exon 2 in the above populations [12]. This region is located approximately 1500 bp in the 5' direction from rs2238296 and is available from the dbSNP database (www.ncbi.nlm.nih.gov/SNP) as rs28567406. Thus, the relative position of the loci is as follows: 5' rs28567406 (CAG microsatellite) rs2238296 ⟶ rs758130 3'. The numbering of nucleotide positions is given here according to the *POLG1* genomic sequence available from the GenBank database (www.ncbi.nlm.nih.gov/Genbank, Acc. No. Ac133637).

Nucleotide sequence variation in *POLG1* intron 2 was investigated between the loci rs28567406 (CAG repeat) and rs2238296. The 716 bp fragment was amplified with the primers 5'-ACAACCTGGACCAGCACTTC-3' and 5'-AAAGGCTGGGGATGCTAAAT-3' designed using the Primer3 software [19]. The PCR included 35 cycles of 30 sec at 94°C, 60 sec at 50°C, and 60 sec at 72°C. Amplification products were sequenced with an ABI Prism 3130 sequence analyzer (Applied Biosystems, United States) using a Big Dye Terminator kit (Applied Biosystems, v. 3.1). Nucleotide sequence alignment and analysis were performed with MEGA 3.1 software package [20].

Statistical analysis. The correspondence of the genotype distribution to the Hardy–Weinberg equilibrium, linkage disequilibrium and haplotype structures were analyzed using the Arlequin 3.01 software [21]. Haplotype frequencies were determined using the maximum likelihood-based EM algorithm [22]. The most probable haplotype structures were reconstructed from multilocus genotypes using a pseudobayesian ELB algorithm [21]. The divergence between the nucleotide sequences of *POLG1* intron 2 was calculated from p-distances using the MEGA 3.1 software package [20].

RESULTS AND DISCUSSION

We analyzed the genotype frequency distributions of rs758130 and rs2238296 *POLG1* SNPs in Russian and Buryats and found that they were similar (Table 1). The frequency of rs758130*C was 42% in Buryats and 38% in Russians, and the frequency of rs2238296*C was 36% in Buryats and 35% in Russians. In the studied population samples, the single-locus genotype distributions corresponded to the Hardy–Weinberg equilibrium. Our data suggest a nearly complete linkage disequilibrium between rs758130 and rs2238296 (r^2 = 0.844 in Russians, r^2 = 0.799 in Buryats, and D' = 1 and p = 0 in both populations). These results agree with the data from an earlier study [18] describing the *POLG1* genotype and haplotype distribution at rs758130 and rs2238296 in the populations of Siberia (in Tuvans, Altay, Yakuts, and Russians).

Table 2. The genotype distribution of three *POLG1* loci in Russians and Buryats

Loci			Frequency	
Number of CAG repeats	rs2238296	Rs758130	Russians ($2n$ = 100)	Buryats ($2n$ = 188)
10	T	T	0.47 (47)	0.54 (101)
10	C	C	0.35 (35)	0.36 (68)
10	T	C	0.04 (4)	0.05 (10)
non-10	T	T	0.13 (13)	0.05 (9)
non-10	C	C	0.01 (1)	0

In both populations, the majority of haplotypes were TT and CC (Table 2). The TT haplotype frequency was 60% in Russians and 59% in Buryats, and the CC frequency was 36% in both populations. The apparently recombinant haplotype TC was observed in 4% of Russians and in 5% of Buryats, which generally agrees with the data from an earlier study [18] reporting a 2% TC frequency in Tuvans.

Analyzing the CAG microsatellite genotype distribution, we found that the microsatellite locus was linked to rs758130 and rs2238296 ($D' > 0.8$, $p < 0.05$). Interestingly, both in Russian and Buryat populations, microsatellite alleles with repeat numbers other than 10 (non-10 alleles) were found mainly in combination with TT haplotypes (Table 2). Altogether, there were five haplotypes: $(CAG)_{10}TT$, $(CAG)_{10}CC$, $(CAG)_{10}TC$, $(CAG)_{non\text{-}10}TT$, and $(CAG)_{non\text{-}10}CC$. There was only one case of a CC haplotype including a $(CAG)_{non\text{-}10}$ allele (in a Russian, Table 2), whereas the TT haplotypes exhibited a complete range of microsatellite alleles (6 to 12 repeats).

To investigate the CAG allele distribution with respect to the SNP haplotypes, rs758130 and rs2238296 SNPs were typed in a group of individuals with genotypes $(CAG)_{10}/(CAG)_{non\text{-}10}$ and $(CAG)_{non\text{-}10}/(CAG)_{non\text{-}10}$ (n = 35) selected from a sample of 556 Russian individuals, residents of European Russia. There were six haplotypes: $(CAG)_{10}TT$, $(CAG)_{10}CC$, $(CAG)_{non\text{-}10}TT$, and $(CAG)_{non\text{-}10}CC$ described above, and two new ones, $(CAG)_{10}CT$ and $(CAG)_{non\text{-}10}TC$ (Table 3, 4). Evidently, $(CAG)_{non\text{-}10}$ alleles are extremely rarely found in combination with the CC haplotype: there were only two cases (3%) in contrast to 35 cases (50%) of TT haplotype combinations (Table 4). It should be mentioned that the varieties of microsatellite alleles found within the TT and CC haplotypes were partially overlapping: TT haplotypes included a complete range of 6 to 12 repeats, whereas CC haplotypes included 8 to 12 repeats. In general, our results suggest that instability of the CAG microsatellite as assessed by the number of non-10 alleles is significantly higher within the TT haplotype group.

The sequence variation of the 716 bp fragment of *POLG1* intron 2 lying between the CAG microsatellite and rs2238296 was studied in 20 individuals homozygous for rs2238296 and rs758130, and there were detected three SNPs registered in the dbSNP database as rs2283430, rs2239286, and rs2247233. The CC haplotype corresponded to the haplotype TCA of rs2283430, rs2239286, and rs2247233, while the TT haplotype group included two haplotypes, GGG and GGA. A comparative sequence analysis of this *POLG1* fragment in human and chimpanzee (NW_001225258 in the GenBank database) indicated that the TT haplotype is phylogenetically younger, because its sequence differs more from the chimpanzee haplotype (the average sequence divergence was 1.1% for the TT haplotype and 0.76% for the CC haplotype).

Table 3. The genotype distribution of three *POLG1* loci in Russian individuals with microsatellite genotypes $(CAG)_{10}/(CAG)_{non\text{-}10}$ and $(CAG)_{non\text{-}10}/(CAG)_{non\text{-}10}$

Loci			Sample requency (n = 35)
Number of CAG repeats	rs2238296	rs758130	
10/non-10	T/T	T/T	0.29 (10)
10/non-10	T/C	T/C	0.46 (16)
10/non-10	C/C	C/C	0.03 (1)
10/non-10	T/T	T/C	0.03 (1)
10/non-10	T/C	C/C	0.03 (1)
10/non-10	T/C	T/T	0.03 (1)
non-10/non-10	T/T	T/T	0.09 (3)
non-10/non-10	T/C	T/C	0.03 (1)
non-10/non-10	T/T	T/C	0.03 (1)

Table 4. The genotype distribution of three *POLG1* loci in Russian individuals with microsatellite genotypes $(CAG)_{10}/(CAG)_{non\text{-}10}$ and $(CAG)_{non\text{-}10}/(CAG)_{non\text{-}10}$

Loci			Sample frequency ($2n = 70$)
Number of CAG repeats	rs2238296	rs758130	
10	T	T	0.16 (11)
10	C	C	0.26 (18)
10	C	T	0.01 (1)
non-10	T	T	0.50 (35)
non-10	C	C	0.03 (2)
non-10	T	C	0.04 (3)

Our results also suggest that identical non-10(CAG) alleles originated independently within different *POLG1* haplotype groups. This is an important consideration for the molecular genetic research of male infertility, since azoospermia may be associated with $(CAG)_{non\text{-}10}/(CAG)_{non\text{-}10}$ genotypes within a particular *POLG1* variant. Probably, this may explain the controversial results concerning the association between male infertility and CAG microsatellite polymorphism of *POLG1*. This hypothesis should be verified by further investigation of *POLG1* polymorphism in healthy individuals and in those with disturbed spermatogenesis.

ACKNOWLEDGMENTS

We are grateful to I.K. Dambueva for the expert assistance provided in this work.

This work was supported by the Far East Branch of the Russian Academy of Sciences (projects no. 06-3-A-06-176, 09-3-A-06-221).

REFERENCES

1. Walker R.L., Anziano P., Meltzer P.S. 1997. A PAC containing the human mitochondrial DNA polymerase gamma gene (*POLG*) maps to chromosome 15q25. *Genomics*. **40**, 376–378.
2. van Goethem G., Dermaut B., Lofgren A., Martin J.J., van Broeckhoven C. 2001. Mutations of *POLG* is associated with progressive external ophthalmoplegia characterized by mtDNA deletions. *Nature Genet*. **28**, 211–212.
3. Barthelemy C., de Baulny H.O., Lombes A. 2002. D-loop mutations in mitochondrial DNA: Link with mitochondrial DNA depletion? *Hum. Genet*. **110**, 479–487.
4. Lamantea E., Tiranti V., Bordoni A., et al. 2002. Mutations of mitochondrial DNA polymerase gamma are a frequent cause of autosomal dominant and recessive progressive external ophthalmoplegia. *Ann. Neurol*. **52**, 211–219.
5. Del Bo R., Bordoni A., Sciacco M., Di Fonzo A., Galbiati S., Crimi M., Bresolin N., Comi G.P. 2003. Remarkable infidelity of polymerase gammaA associated with mutations in *POLG1* exonuclease domain. *Neurology*. **61**, 903–908.
6. Di Fonzo A., Bordoni A., Crimi M., Sara G., Del Bo R., Bresolin N., Comi G.P. 2003. *POLG* mutations in sporadic mitochondrial disorders with multiple mtDNA deletions. *Hum. Mutat*. **22**, 498–499.
7. Kollberg G., Jansson M., Perez-Bercoff A., Melberg A., Lindberg C., Holme E., Moslemi A.-R., Oldfors A. 2005. Low frequency of mtDNA point mutations in patients with PEO associated with *POLG1* mutations. *Eur. J. Hum. Genet*. **13**, 463–469.
8. Blomberg Jensen M., Leffers H., Petersen J.H., Daugaard G., Skakkebaek N.E., Rajpert-De Meyts E. 2008. Association of the polymorphism of the CAG repeat in the mitochondrial DNA polymerase gamma gene (*POLG*) with testicular germ-cell cancer. *Ann. Oncol*. **19**, 1910–1914.
9. Westerveld G.H., Kaaij-Visser L., Tanck M., van der Veen F., Repping S. 2008. CAG repeat length variation in the polymerase gamma (*POLG*) gene: Effect on semen quality. *Mol. Hum. Reprod*. **14**, 245–249.
10. Rovio A.T., Marchington D.R., Donat S., et al. 2001. Mutations at the mitochondrial DNA polymerase (*POLG*) locus associated with male infertility. *Nature Genet*. **29**, 261–262.
11. Rovio A.T., Abel J., Ahola A.L., et al. 2004. A prevalent *POLG* CAG microsatellite length allele in humans and African great apes. *Mammal. Genome*. **15**, 492–502.
12. Malyarchuk B.A., Papuga M., Grzybowski T., Wozniak M., Derenko M.V., Rogozin I.B., Rychkov S.Y., Czarny J., Zakharov I.A., Miscicka-Sliwka D. 2005. Low variability of the *POLG* $(CAG)_n$ repeat in North Eurasian populations. *Hum. Biol*. **77**, 355–365, 355–365.
13. Copeland W.C., Ponamarev M.V., Nguyen D., Kunkel T.A., Longley M.J. 2003. Mutations in DNA polymerase gamma cause error prone DNA synthesis in human mitochondrial disorders. *Acta Biochim. Polonica*. **50**, 155–167.
14. Krausz C., Guarducci E., Becherini L., Degl' Innocenti S., Gerace L., Balercia G., Forti G. 2004. The clinical significance of the *POLG* gene polymorphism in male infertility. *J. Clin. Endocrinol. Metab*. **89**, 4292–4297.
15. Aknin-Seifer I.E., Touraine R.L., Lejeune H., Jimenez C., Chouteau J., Siffroi J.P., McElreavey K., Bienvenu T., Patrat C., Levy R. 2005. Is the CAG repeat of mitochondrial DNA polymerase gamma (*POLG*) associated with male infertility? A multi-centre French study. *Hum. Reprod*. **20**, 736–740.
16. Brusco A., Michielotto C., Gatta V., et al. 2006. The polymorphic polyglutamine repeat in the mitochondrial DNA polymerase gamma gene is not associated with oligozoospermia. *J. Endocrinol. Invest*. **29**, 1–4.
17. Harris T.P., Gomas K.P., Weir F., Holyoake A.J., McHugh P., Wu M., Sin Y., Sin I.L., Sin F.Y. 2006. Molecular analysis of polymerase gamma gene and mitochondrial polymorphism in fertile and subfertile men. *Int. J. Androl*. **29**, 421–433.

18. Buikin S.V., Golubenko M.V., Pogrebenkova V.V., Tsimbalyuk I.V., Puzyrev V.P. 2006. Mitochondrial DNA polymerase γ gene (*POLG*): Frequency and linkage analysis of two single nucleotide polymorphisms in Siberian populations. *Mol. Biol*. **40**, 1081–1083.

19. Rozen S., Skaletsky H. 2000. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Eds Krawetz S., Misener S. Totow, NJ: Humana Press, pp. 365–386.

20. Kumar S., Tamura K., Nei M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform*. **5**, 150–163.

21. .Excoffier L., Laval G., Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform. Online*. **1**, 47–50.

22. Excoffier L., Slatkin M. 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol*. **12**, 921–927.