Molecular Human Reproduction Vol.12, No.10 pp. 643–646, 2006
Advance Access publication August 3, 2006 doi:10.1093/molehr/gal063

Haplotypes, mutations and male fertility: the story of the testis-specific ubiquitin protease *USP26*

C.Ravel[1,2]*, B.El Houate[1,3]*, S.Chantot[1,2], D.Lourenço[1], A.Dumaine[1], H.Rouba[3], A.Bandyopadahyay[4], U.Radhakrishna[5], B.Das[6], S.Sengupta[6], J.Mandelbaum[2], J.P.Siffroi[2] and K.McElreavey[1,7]

[1]Reproduction, Fertility and Populations, Institut Pasteur, [2]EA1533, Université Pierre et Marie Curie-Paris 6, AP-HP, Hôpital Tenon, Laboratoire d'Histologie-Biologie de la Reproduction-Cytogénétique, Paris, France, [3]Department of Human Genetics, Institut Pasteur of Morocco, Casablanca, Morocco, [4]Department of Anthropology, University of Calcutta, Kolkata, West Bengal, India, [5]Department of Genetics, Boys Town National Research Hospital, Omaha, NE, USA and [6]Department of Anthropology, University of Dibrugarh, Dibrugarh, Assam, India

[7]To whom correspondence should be addressed at: Reproduction, Fertility and Populations, Institut Pasteur, 25 rue du Dr Roux, Paris, 75724 France. E-mail: kenmce@pasteur.fr

*Both authors contributed equally to this work.

Recently, mutations in the X-linked ubiquitin protease 26 (*USP26*) gene have been proposed to be associated with male infertility. In particular a 371insACA, 494T>C and 1423C>T haplotype, which results in a T123–124ins, L165S and H475Y amino acid change respectively, has been reported to be associated with Sertoli cell-only syndrome (SCOS) and an absence of sperm in the ejaculate. Here, we demonstrate that two of these changes actually correspond to the ancestral sequence of the gene and that the *USP26* haplotype is present in significant frequencies in sub-Saharan African and South and East Asian populations, including in individuals with known fertility. This indicates that the allele is not associated with infertility. The pattern of frequency distribution of the derived haplotype (371delACA, 494T), which is present at high frequencies in most non-African populations could be interpreted as either a result of migration followed by simple genetic drift or alternatively as positive selection acting on the derived alleles. The latter hypothesis seems likely, because there is evidence of strong positive selection acting on the *USP26* gene.

Key words: fertility/polymorphism/spermatogenesis/*USP26* haplotype/X chromosome

Introduction

Recently, it has been reported that mutations in the testis-specific ubiquitin protease 26 (*USP26*) gene may be associated with male infertility (Paduch *et al*., 2005; Stouffs *et al*., 2005). Polyubiquitin attachment targets many intracellular proteins for degradation by the proteasome, and monoubiquitination is often required for down-regulating plasma membrane proteins by targeting them to the vacuole. The *USP26* gene is present as a single exon on the X chromosome, and it encodes a predicted protease containing his and cys domains that are present in deubiquitinating enzymes (Wang *et al*., 2001). These enzymes are responsible for processing inactive ubiquitin precursors, proofreading ubiquitin–protein conjugates and removing ubiquitin from cellular adducts (Nijman *et al*., 2005). Ubiquitinated molecules can thus be rescued from degradation by the activity of deubiquitinating enzymes. The *USP26* gene was first reported by Wang *et al*. (2001), who demonstrated that both the mouse and human genes are expressed exclusively in testis. This led the authors to suggest that mutations involving this gene could contribute to human spermatogenic failure.

To test this hypothesis, Stouffs *et al*. (2005) sequenced the *USP26* gene in 42 patients with Sertoli cell-only syndrome (SCOS) and found 4 patients carrying three variants. The changes were 370–371insACA, 494T>C and 1423C>T, causing the amino acid changes T123-124ins, L165S and H475Y, respectively. All three variants are present on the same allele. These changes were not found in 10 control samples. An extended mutation screen of 69 patients with SCOS, using a restriction site created by the 494T>C variant, identified 4 more patients (7.2%) carrying this change (Stouffs *et al*., 2005). The haplotype was not observed in 32 patients with maturation arrest nor in 142 control samples using this method. The authors concluded that this haplotype (370–371insACA, 494T>C, 1423C>T) may either cause infertility or be a predisposing factor. Five of the eight patients carrying the haplotype were of Arabic origin (Middle East), two were Caucasians, and the ethnic origin of the remaining patient was unknown. An additional study (Paduch *et al*., 2005) found that 20 of 188 (10.6%) infertile males had amino acid changes in the *USP26* gene. No changes were found in a fertile control panel (n = 17). The 370-371insACA, 494T>C and 1423C>T haplotype was found in four (1.9%) patients. An additional individual carried the 371insACA, 494T>C variants without the 1423C>T substitution, and one patient had the 1423C>T variant without any other changes. In this study, the frequency of the 371insACA, 494T>C and 1423C>T haplotype was 4/44 (9.1%) in individuals with known histology. Several other variants were identified that were not observed in the control population (17 men of known fertility). These data are intriguing because they suggest that an X-linked *USP26* haplotype (371insACA, 494T>C and 1423C>T) is present in the general population, and this haplotype may either cause male

infertility or it may be a major predisposing factor. Unfortunately in both studies, DNA from other family members was unavailable for study, and therefore segregation of the haplotype in the pedigrees is unknown.

We decided to further investigate the potential role of the 371insACA, 494T>C and 1423C>T haplotype in male infertility by the analysis of a large series of individuals of known fertility or infertility from carefully defined ethnic groups and from different geographic origins. Our data show that the 371insACA, 494T>C, 1423C>T haplotype is compatible with fertility. Furthermore, analysis of the chimpanzee sequence indicated that two of the changes that define the haplotype are present in primates and actually represent the ancestral state rather than a derived mutation. Population studies demonstrated significant frequencies of the *USP26* haplotype in sub-Saharan and South and East Asian populations. The haplotype is virtually absent in other populations tested. These data suggest that the derived allele may be under positive selective pressure.

Materials and methods

Study populations

The initial study population of infertile individuals was of mixed ethnic background, and all sought infertility treatment at the Tenon Hospital in Paris. Written, informed consent was obtained for molecular investigations, and ethical approval was given by a local ethics committee in France. Each of these men was screened for Y chromosome microdeletions according to the European Academy of Andrology guidelines (Simoni *et al.*, 2004). Samples were also screened for gr/gr deletions (Repping *et al.*, 2003).

The study population of infertile men consisted of 99 individuals. They presented with either a complete absence of spermatozoa in the ejaculate (azoospermia; $n = 40$), extreme oligozoospermia [$<1 \times 10^6$ sperm/ml, ($n = 19$), of which 11 patients presented with oligoasthenospermia (OATS) and 2 individuals with cryptospermia], severe oligozoospermia [$1–5 \times 10^6$ sperm/ml ($n = 24$), of which 20 patients presented with OATS], moderate oligozoospermia [$5–10 \times 10^6$ sperm/ml ($n = 8$), of which 6 presented with OATS] or mild oligozoospermia [$10–20 \times 10^6$ sperm/ml ($n = 8$), of which 7 presented with OATS).

The control cohort consisted of a large human biodiversity panel where, in some cases, the fertility status or sperm count of the individual was known. The control cohort consisted of the entire Human Genome Diversity Project, Centre d'Etudes du Polymorphisme Humain (HGDP-CEPH) panel, comprising 1064 DNA samples from 52 worldwide populations (http://www.cephb.fr/HGDP-CEPH-Panel/). In addition, samples were included from a Tibeto-Burman speaking group ($n = 20$) and a Munda-speaking ($n = 25$) group from central and eastern India, respectively. The fertility status of these individuals is unknown. The biodiversity panel was further extended to include male individuals with either known fertility or who were normospermic. This consisted of 56 Indians (father of at least two children), 108 Moroccans (father of at least two children) and 61 ethnic French males of known fertility (father of at least one child; 20 individuals were normospermic). The total control cohort consisted of 1334 individuals, of which the fertility status or sperm count was known for 225 individuals.

Sequencing the USP26 gene

DNA was extracted from peripheral blood lymphocytes using standard techniques. PCR amplification of the *USP26* coding sequences was performed using the following combination of four primer pairs: USP26F1 5′-AATCCAAGGCAGATTGTTCA-3′, USP26R1 5′-TTCTTTGGGGAAGGTTGATG-3′, USP26F2 5′-CACAGCTGAACTTGCAAACA-3′, USP26R2 5′-TGCATGAAGATTCAC-3′, USP26F3 5′-ACACCAGTGGGTTTTCTTGC-3′, USP26R3 5′-ATTTTCGGGCACTGTTTGAA-3′, USP26F4 5′-TCAGTTCCACAAAGCTGGAG-3′, USP26R4 5′-CCATGGAGGAAGTGGTATCG-3′.

The conditions for PCR were as follows: (i) USP26F1/R1, 95°C for 5 min followed by 35 cycles of 95°C for 1 min, 60°C for 1 min and 72°C for 45 s; (ii) USP26F2/R2, 95°C for 5 min followed by 35 cycles of 95°C for 1 min, 60°C for 55 s and 72°C for 35 s; (iii) USP26F3/R3, 95°C for 5 min followed by 35 cycles of 95°C for 1 min, 60°C for 40 s and 72°C for 30 s; and (iv) USP26F4/R4, 95°C for 5 min followed by 35 cycles of 95°C for 1 min, 60°C for 1 min and 72°C for 35 s. An elongation step of 72°C for 5 min followed each amplification.

DNA sequence analysis was performed using at least 200 ng of purified DNA, 20 ng of primer and fluorescently labelled Taq DyeDeoxy terminator reaction mix (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. DNA sequence was determined using a 373 A automated DNA sequencer (Applied Biosystems). Restriction enzyme digestion of PCR products was performed using either the enzyme *Taq* I (New England Biolabs, Ipswich, MA, USA; amplicon USP26F2/R2) or *Fok* I (New England Biolabs; amplicon USP26F3/R3) according to the manufacturer's instructions. PCR products were visualized on a 2% agarose gel containing ethidium bromide. DNA samples from 2 chimpanzees (*Pan troglodytes*; a gift from luis Quintana-Murci, Institut Pasteur) were amplified and sequenced using the same conditions as described above.

Results and discussion

All patients were of mixed ethnic origin, but in each individual case the ethnicity of the individual was determined by a questionnaire. Initially, the study consisted of sequencing the entire *USP26* open-reading frame in a series of 30 patients with idiopathic infertility (20 azoospermic and 10 oligozoospermic). In all 30 cases, the open-reading frame was found to be identical to that of a normal male control. No polymorphic variants were observed. We therefore decided to extend this study to include a much larger group of patients. An extended screen for the 494T>C variant was performed on a panel of 99 patients using the restriction enzyme *Taq* I on the USP26F2/R2 amplicon in order to differentiate between a T or a C at position 494. By this technique, five individuals were identified who carried the 494T>C variant. This observation was confirmed by direct sequencing of the USP26F2/R2 amplicon in each of the five individuals (Figure 1A). The clinical details of these five cases are indicated in Table I.

We then investigated if these individuals also carried the two other changes associated with the 494T>C variant, namely the 371insACA and 1423C>T variants. All five individuals carried the 371insACA variant (Figure 1B), and digestion of amplicon USP26F3/R3 by the enzyme *Fok* I indicated that the 1423C>T variant was present in each case (data not shown). This was confirmed by direct sequencing of the PCR product (Figure 1C). Although four of these individuals were azoospermic, one of the individuals was observed to have oligozoospermia. Previous reports indicated that this haplotype is associated with SCOS, and no sperm have been observed in the ejaculate of men carrying this haplotype. One individual also had an *AZFc* deletion. Our observation either extends the phenotype associated with the haplotype or it could possibly indicate that the haplotype is actually a rare polymorphism. We noted that the five individuals carrying the haplotype were of West or North African origin, suggesting that the haplotype could be a polymorphism with limited geographical or ethnic distribution (Table I). To resolve this question, we performed two analyses. We determined the ancestral state of the variants by sequencing the orthologous gene in two chimpanzees (*P. troglodytes*). In parallel, we extended the genetic screen for the 494T>C variant to large biodiversity control panel which also consisted of individuals with known fertility or sperm count status. The results of the sequence of the chimpanzee DNA are shown in Figure 1D–F. These indicated that two of the 'mutated' variants, namely 494T>C and 370-371insACA, are actually the ancestral sequence. This was confirmed by BLAST analysis (data not shown).

A worldwide survey of 1334 DNA samples of defined ethnic/geographic origin indicated that the 494T (ancestral) variant is present at significant frequencies in sub-Saharan African populations (e.g. >20% in pygmy populations; Table II, Figure 2), and it is present in modest frequencies in North African populations. The ancestral sequence is

Figure 1. Worldwide distribution of 370-371insACA, 494T>C and 1423C>T *USP26* haplotype in healthy individuals. Each pie represents the population proportion of individuals carrying the haplotype (shown as the unshaded region). Populations numbered from 1 to 20 were analysed in this study and are described in detail in Table II.

Table I. Description of patients carrying the 371insACA, 494T>C and 1423C>T haplotype

Patient	Phenotype	Sperm count	Geographic origin
1	Azoospermia, deleted *AZFc*	–	North Africa
2	Azoospermia	–	Congo
3	Azoospermia	–	West Africa
4	Severe oligozoospermia	0.08×10^6/ml	West Africa
5	Azoospermia	–	Benin

also found in South and South-East Asian populations at significant frequencies. The frequency distribution data are consistent with a southern coastal out-of-Africa route that has been proposed for the origin of modern humans (Quintana-Murci *et al*., 1999). In each individual (n = 75) who carried the 494T>C variant, we determined the status of the 370-371insACA and 1423C>T variants by restriction site digestion or by direct sequencing. In all cases, the insertion 370-371insACA and the T variant at position 1423 were present. Of the 75 individuals carrying the 370-371insACA, 494T>C and 1423C>T haplotype, 9 were known to be fertile. These data suggest that the 370-371insACA, 494T>C and 1423C>T haplotype is not associated with infertility, and indeed two of these variants are the ancestral sequence. In the paper by Paduch *et al*. (2005), individuals who carried variants of this haplotype were identified. On the basis of this, it is possible to reconstruct the series of events leading to this haplotype (Figure 3).

Although our data indicate that the 370-371insACA, 494T>C and 1423C>T haplotype is unlikely to be associated with infertility, one should note that most individuals of non-Saharan African origin actually have the derived sequence (371delACA, 494T). Both of these variants

Table II. Populations studied for the presence of the 370-371insACA, 494T>C and 1423C>T haplotype, here referred to as the ancestral allele

Population (n)	Total alleles[§]	Ancestral allele (%)	Derived allele (%)
Pygmies (51)	56	21.4	78.6
Chinese (184)	184	9.2	90.8
Japanese (31)	39	2.6	97.4
Maghrebian (138[*])	138	3.6	96.4
Siberian (25)	32	9.4	90.6
Cambodian (11)	16	6.3	93.7
Israelian (148)	150	2.00	98.0
Pakistani (200)	200	10.00	90.0
New Guinea (39)	57	7.00	93.0
Kenyan (12)	13	7.7	92.3
Senegalese + Nigerian (49)	69	2.9	97.1
Namibian + South African (15)	15	13.3	86.7
Indian (101[†])	101	5.4	94.6
French (114[‡])	139	0.7	99.3
Italian (50)	69	0	100
Brazilian (45)	69	0	100
Colombian (13)	21	0	100
Mexican (50)	83	0	100
Russian (42)	61	0	100
Orkney Islanders (16)	25	0	100

*This includes 108 individuals of known fertility, 5 of these individuals carried the 370-371insACA, 494T>C and 1423C>T haplotype.
[†]Fifty-six of the DNA samples were obtained from men who were the father of at least 2 children. Of these 56 samples, 3 carried the 370-371insACA, 494T>C and 1423C>T haplotype.
[‡]One fertile (father of two children) individual carried the haplotype.
[§]Some samples were from 46,XX individuals and hence carried two copies of *USP26*.

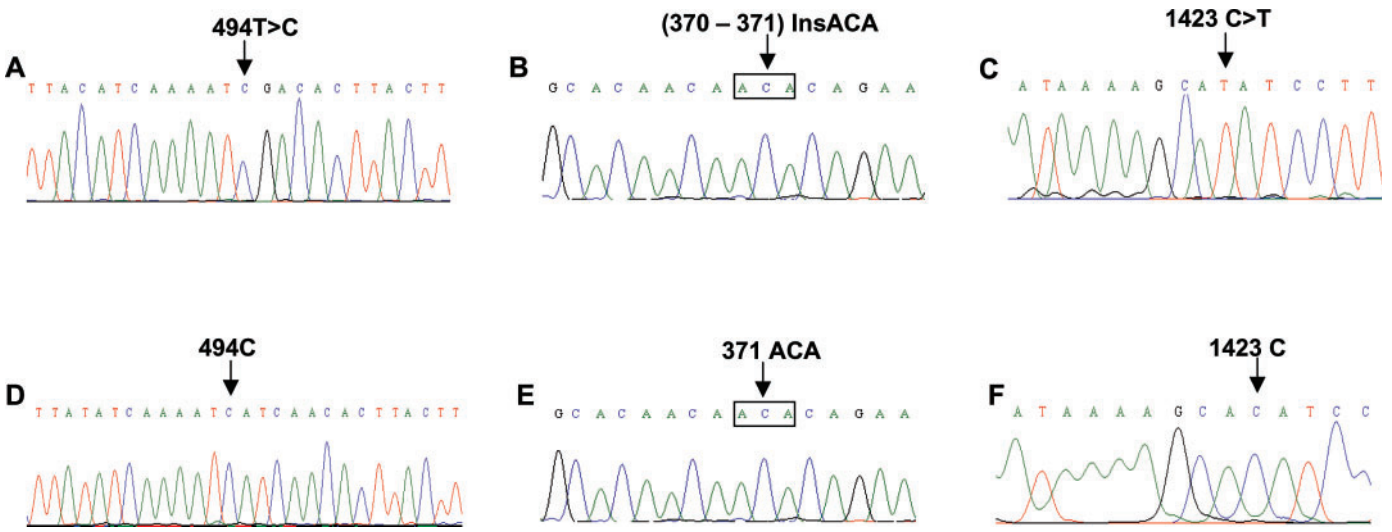

Figure 2. Polymorphic variants of the *USP26* gene. **A–C**, representative sequence of the human 370-371insACA, 494T>C, 1423C>T *USP26* haplotype showing the position of each of the changes. **D**–**F**, representative sequence of the corresponding region of the chimpanzee sequence indicating the ancestral sequence.

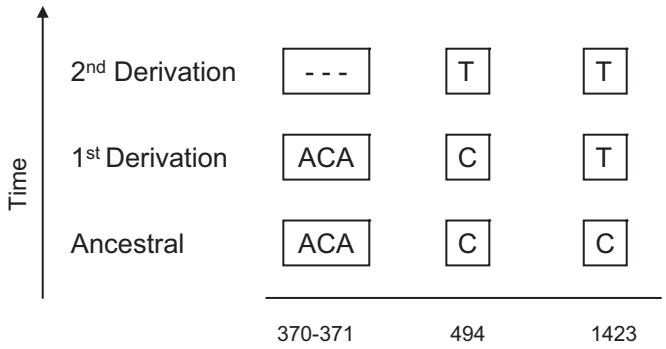

Figure 3. Schematic representation of the hypothetical *USP26* haplotype evolution. The ancestral sequence deduced from sequence analysis of the chimpanzee sequence. The first derivation haplotype has been reported by Paduch *et al*. (2005). The second derivation haplotype is common in most human populations and may be under positive selective pressure.

result in the amino acid changes T123-124del and S165L, respectively. The pattern of frequency distribution of the derived haplotype could be interpreted either as a result of migration followed by simple genetic drift or, alternatively, as positive selection acting on the derived allele (i.e. increased fertility associated with the derived allele). Recently, in another genetic context, one of us (K.M.) observed that the *CD209* gene, which encodes the DC-SIGN protein that interacts with *Mycobacterium tuberculosis*, as well as with other pathogens, shows an increased frequency of derived alleles in non-African populations as a result of host genetic adaptation to a longer history of exposure to tuberculosis (Barreiro *et al*., 2006). Indeed, in a survey for positive selection on 13 731 annotated genes from human to their chimpanzee orthologue, Nielsen *et al*. (2005) noted that the *USP26* gene exhibited the fifth highest positive selective value (human–chimpanzee likelihood ratio of 6.2 from the likelihood ratio test of d_N/d_S equals 1 versus d_N/d_S >1 in human–chimpanzee alignments). This observation strongly supports the hypothesis of positive selection acting on the derived 371delACA, 494T haplotype. Although the ancestral form of USP26, which is now limited in geographic distribution, is not associated with infertility, it may be associated with reduced fertility compared with the derived state.

This study highlights once again the crucial importance in genetic studies of choosing with care appropriately matched control populations that reflect not only the geographic origins of the case population but also the ethnic background.

Acknowledgements

B.E.H thanks the Direction des Affaires Internationales de l'Institut Pasteur, Paris, France for financial support for this study. This study was supported by the GIS-Institut des Malades Rare. The study also received funding from the European Science Foundation Collaborative Research Programme, the origin of man, language and languages (OMLL).

References

Barreiro LB, Neyrolles O, Babb CL, Tailleux L, Quach H, McElreavey K, Helden PD, Hoal EG, Gicquel B and Quintana-Murci L (2006) Promoter variation in the DC-SIGN-encoding gene CD209 is associated with tuberculosis. PLoS Med 3,e20.

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ *et al*. (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 3,e170.

Nijman SM, Luna-Vargas MP, Velds A, Brummelkamp TR, Dirac AM, Sixma TK and Bernards R (2005) A genomic and functional inventory of deubiquitinating enzymes. Cell 123,773–786.

Paduch DA, Mielnik A and Schlegel PN (2005) Novel mutations in testis-specific ubiquitin protease 26 gene may cause male infertility and hypogonadism. Reprod Biomed Online 10,747–754.

Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K and Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. Nat Genet 23,437–441.

Repping S, Skaletsky H, Brown L, van Daalen SK, Korver CM, Pyntikova T, Kuroda-Kawaguchi T, de Vries JW, Oates RD, Silber S *et al*. (2003) Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. Nat Genet 35,247–251.

Simoni M, Bakker E and Krausz C (2004) EAA/EMQN best practice guidelines for molecular diagnosis of y-chromosomal microdeletions. State of the art 2004. Int J Androl 27,240–249.

Stouffs K, Lissens W, Tournaye H, Van Steirteghem A and Liebaers I (2005) Possible role of USP26 in patients with severely impaired spermatogenesis. Eur J Hum Genet 13,336–340.

Wang PJ, McCarrey JR, Yang F and Page DC (2001) An abundance of X-linked genes expressed in spermatogonia. Nat Genet 27,422–426.

Submitted on May 4, 2006; resubmitted on June 23, 2006; accepted on June 26, 2006