CHAPTER II:

Characterization Of The Intronless Human Sperm Protein 17 Gene Variant

CHAPTER II: Characterization of the Intronless Human Sperm Protein 17 Gene Variant

1. Abstract

Sequence analysis of human Sp17 cDNA revealed clones containing Sp17-like sequences without introns. This observation suggested the presence of an intronless Sp17 gene variant, Sp17-2. Sp17-2 was characterized in the human genome. RT-PCR analysis was employed to detect and differentiate between the Sp17 gene variants. Restriction enzyme digestion and southern blot analysis of human genomic DNA confirmed the existence of the Sp17-2 gene variant and suggested the presence of additional Sp17 gene variants. Sequence analysis of the Sp17-2 variant revealed features of a retroposed, non-functional pseudo-gene. Furthermore, subsequent analysis of Sp17-2 in the primate and rodent suggests that this retroposed pseudo-gene arose recently on the evolutionary timeline.

2. Introduction

The mammalian genome contains several models of related multigene families whose members are dispersed throughout the genome. Analyses of their sequence structure suggest that the gene members may be functional genes containing introns, or non-transcribed pseudogenes derived by retroposition (duplication through an RNA intermediate) (1). Moreover, the number of intronless pseudogenes and their locations within the genome vary among species (2). However, the high sequence conservation between many retroposed pseudogenes and their presumed parental genes suggest that retroposition is a recent evolutionary occurrence (3, 4, 5, 6).

The human Sp17 gene was previously described as an intron-containing gene (Chapter 1, 7). However, sequence analysis of the human intron-containing Sp17 gene revealed homology to clones containing Sp17-like sequences without introns (Chapter 1). This observation suggests the existence of intronless Sp17 gene variants in the human genome. Thus, a molecular biology approach was employed to identify and characterize Sp17 gene variants in the human Sp17 gene family.

In this study, RT-PCR and sequence analyses were performed to detect and differentiate between the Sp17-1 transcript and Sp17-2 gene variants. In addition, southern blot analysis was used to determine the existence of Sp17 gene variants in the human genome. Furthermore, PCR analysis was used to elucidate Sp17 gene variants in the primate and rodent.

3. Materials and Methods

3.1 RNA isolation, DNase treatment and Reverse transcription (RT)

Normal human testis tissue was obtained from the University of California, Davis (UCD) Donated Body Program (UCD biological use authorization #0657). Normal primate testis tissue was obtained from the UCD primate center (Animal Use Protocol # 8707). In addition, normal rat lung tissue was donated from a UCD laboratory. The tissues were homogenized (50-100 mg) in TRIzol reagent and the total RNA was

extracted, analyzed and stored as described in Chapter 1, section 3.1. The total RNA extracts (3ug) were treated with DNase I and converted into cDNA by reverse transcription (RT) as described in Chapter 1, section 3.2 (Chapter 1, Table 1.1).

3.2 Polymerase chain reaction (PCR)

The Sp17-2 nucleotide sequence was amplified in human testis cDNA and human genomic DNA (Chapter 1, section 3.8) by RT-PCR and PCR, respectively. Sp17 specific sense and anti-sense primers were synthesized by Biosynthesis, Inc and suspended to 100pmol in sterile water (Table 2.1).

The PCR reactions were prepared as described in Chapter 1, section 3.3 with the following primer and template modifications: 200nM sense primer, 200nM anti-sense primer, 4uL cDNA template or 1uL genomic DNA template. The PCR products were resolved by gel electrophoresis, visualized with UV light and photographed.

The appropriately sized testis PCR amplified DNA bands were excised, purified and subcloned into a TOPO vector as described in Chapter 1, section 3.6. The positive clones were purified and the DNA inserts were sequenced from the forward (M13R) and reverse (T7) directions using ABI BigDye Terminator chemistry (Applied Biosystems) by the DBS Automated DNA Sequencing Facility (University of California, Davis, CA). DNA sequence analysis, assembly, alignment and promoter analysis was performed as described in Chapter 1, section 3.7.

3.3 Southern blot analysis of human genomic DNA

Genomic DNA was extracted from human polymorphonuclear blood cells (pmbcs) as described in Chapter 1, section 3.8. A restriction map of the published Sp17 nucleotide sequence (Chapter 1, Figure 1.2) and proposed nucleotide sequence of the Sp17-2 variant was generated. Several restriction enzymes were selected and the predicted DNA fragment size that would hybridize to the Sp17 DNA probe was calculated.

The isolated human genomic DNA (4ug) was enzymatically digested using the restriction enzymes and corresponding buffers for EcoR1 (New England Biolabs (NEB), Beverly, MA), HindIII (NEB), BamH1 (NEB), SacII (NEB), Not1 (NEB), StuI (NEB), Bst-XI (Boehringer Mannheim GmBh, W. Germany) according to manufacturer guidelines. Double restriction digestion was performed using EcoR1 with HindIII and EcoR1 with BamH1. DNA fragments cut by Not1 were not predicted to hybridize to the Sp17 DNA probe. Each DNA digestion cocktail was prepared to a final volume 250uL containing 75 units restriction enzyme, 4ug human genomic DNA and 1X restriction enzyme buffer to volume. As required by the optimal enzyme conditions, 100mM BSA (25uL) was added to the enzyme reaction for BamH1 and Not1. All restriction digestions were incubated at the optimal enzyme temperature (37^{o}C or 45^{o}C) overnight.

At the completion of the enzymatic DNA digestion, the DNA fragments were extracted by ethanol precipitation and were resuspended in sterile deionized water.

The concentrated DNA fragments were mixed with 6X DNA gel loading dye, heated at 70^{o}C for 10 minutes and placed on ice. The DNA samples (total volume 25uL) were resolved by gel electrophoresis adjacent to a 1kb DNA molecular weight marker (Life Technologies, Inc.) on a 0.7% agarose gel. The agarose gel was run in 1X TAE at 85 volts for two hours. The gel was visualized with UV light and photographed next to a UV labeled ruler.

The photographed agarose gel was depurinated in 0.001% formic acid at room temperature for 10 minutes and washed in deionized water at room temperature for 10 minutes with constant agitation. The depurinated gel was transferred to Nylon paper (Micron Separations, Inc., Westborough, MA) by capillary movement of 0.4M Sodium hydroxide (NaOH) (8). The transfer was conducted at room temperature overnight.

The wells of the agarose gel and bands of the molecular weight marker were indicated on the nylon blot with a pencil. The nylon blot was washed with 2X SSC (300mM NaCl, 30mM Sodium citrate, pH 7.0) at room temperature for 15 minutes with constant agitation. The nylon southern blot was air dried at room temperature, cross-linked with UV light and vacuum dried at 80^{o}C for one hour. The southern blot was stored between two dry pieces of 3MM Whatman paper in a plastic bag at -20^{o}C.

The southern blot was pre-hybridized at 55^{o}C for 1 hour in hybridization solution (20% deionized foramide, 900mM NaCl, 6mM EDTA, 90mM Tris-HCl (pH 7.0), 1mM $Na_2P_2O_7$, 1% Sodium dodecyl sulfate (SDS), 0.5% non fat milk, 5ug/mL polyadenylic acid and 500ug/mL denatured salmon sperm DNA). The blot was hybridized with a human Sp17 coding region specific, ^{32}P-labelled nick-translated (Amersham Pharmacia Biotech, Inc., Piscataway, NJ) DNA probe (482bp; 0.5 x 10^{6} cpm/mL) in hybridization solution at 55^{o}C overnight (Table 2.2). The DNA probe was amplified from human testis by RT-PCR and spanned the Sp17 coding region sequence.

The blot was washed once with 2X SSC at room temperature for 30 minutes and once with 1X SSC (150mM NaCl, 15mM Sodium citrate, pH 7.0) containing 0.5% SDS at 65^{o}C for one hour. The blot was scanned with a STORM Imaging System (Molecular Dynamics, Sunnyvale, CA) and analyzed with Image Quant v1.2 computer software (Molecular Dynamics).

4. Results

4.1 PCR analysis

PCR analysis was used to amplify the Sp17-2 gene in the human genome (primer set Sg2full and Ag2full) (Figure 2.1). The Sp17-2 gene sequence was submitted to GenBank (Accession number AF335424). Sequence analysis of the Sp17-2 DNA fragment revealed homology to two clones, AC010997 and AL390034, which contained intronless Sp17-like sequences.

Sequence alignment of the Sp17-1 and Sp17-2 nucleotide sequences, over a 689bp reading frame stretch, revealed 91.1% homology (Figure 2.1). However, sequence homology was not detected between the Sp17-1 and Sp17-2 sequences in the upstream and downstream regions of the aforementioned reading frame. In addition, compared to the Sp17-1 nucleotide sequence, the Sp17-2 sequence exhibited 61 point mutations, 4 nucleotide insertions and 2 deletions leading to frame shifts and ORF stop codons, as close as 10 amino acids from the start codon. Thus, the Sp17-2 sequence is not predicted to encode a protein.

4.2 Sp17 promotor sequence analysis

Sequence analysis of the retroposed Sp17-2 pseudogene revealed that the upstream sequence lacks TATA and CAAT boxes which are characteristic of some mammalian RNA polymerase II promoters. However, several alternative consensus sequences were identified corresponding to two retroviral TATA boxes and to transcription factor binding sites including activator proteins (AP-1, Ap-2 and AP-4), the myeloid zinc finger gene 1 (MZF1) and an upstream stimulating factor (USF) (Figure 2.1).

RT-PCR analysis was used to detect a Sp17-2 transcript in normal human testis and to identify promoter sites in the Sp17-2 nucleotide sequence (Figure 2.2A). The Sp17-2 gene was amplified by PCR in reactions containing a genomic DNA template (primer sets Sgr2 and ACDS, Shu and Ag2, Sgr2 and Ag2, Sg2full and ACDS and Shu and A2full) (Figure 2.2B). Conversely, RT-PCR analysis demonstrated the absence of the Sp17-2 transcript in human testis cDNA, free of genomic DNA (Figure 2.2B).

4.3 Southern blot hybridization

A southern blot of restriction digested human genomic DNA was probed with a human Sp17 coding region specific ^{32}P-labelled nick-translated DNA probe. A strong Sp17 hybridization was detected to DNA cut with 8 of the 9 restriction enzyme combinations (Figure 2.3AB). In particular, Sp17 hybridization was observed in EcoR1 (13602kb and 11000 kb), BamH1 (6409kb), HindIII (8072kb and 6447kb), SacII (4900kb and 1200kb), StuI (10809kb, 6496kb, 2713kb and 1998kb), BstXI (7267kb, 5778kb and 3106kb), EcoR1 and BamH1 (10641kb, 3500kb and 1662kb), EcoR1 and HindIII (6447kb, 5100kb, 4100kb, 2600kb and 1662kb). Sp17 hybridization was not detected in DNA fragments cut with the restriction enzyme, NotI.

Sp17 hybridization to genomic DNA, restriction digested with BstXI, BamHI and HindIII, revealed bands corresponding to the predicted DNA fragment size for the Sp17-1 gene. Similarly, DNA hybridization bands from restriction digestion with StuI, EcoRI, EcoRI and BamHI, EcoRI and HindIII corresponded to the estimated DNA fragments expected for both the Sp17-1 and Sp17-2 genes. Furthermore, Sp17 hybridization with the DNA restriction digests from the EcoRI, EcoRI and BamHI, EcoRI and HindIII and SacII revealed unidentified bands that did not correspond to

any of the predicted Sp17 hybridization DNA fragment sizes, suggesting the existence of additional Sp17 gene variants.

4.4 Sp17-2 analysis in non-human mammals

The Sp17-1 transcript and Sp17-2 gene were examined in normal primate testis and normal rat lung by RT-PCR. DNA amplification from primate cDNA and genomic DNA templates revealed DNA bands at 326bp (primers Shu and ACDS) (Figure 2.4). Similarly, DNA amplification from the rat lung cDNA template revealed a DNA band at 473bp (primers Shu and Aspe5). However, DNA amplification was not detected from the rat lung genomic DNA template (primers Shu and Aspe5).

5. Discussion

Due to the presence of introns in genomic DNA, a PCR amplification from a human genomic DNA template using Sp17 coding region specific primers (SCDS and ACDS) was predicted to yield a 1kb DNA fragment. However, this PCR amplification resulted in a Sp17 DNA fragment of 239bp, suggesting the presence of an intronless Sp17 gene in the human genome.

Sequence comparison of the genomic amplified Sp17 coding region to the Sp17-1 nucleotide sequence (Chapter 1), revealed 91.1% homology. Conversely, sequence homology was not detected between the amplified 5'UTR Sp17-1 and Sp17-2 promoter regions. In addition, the Sp17 gene sequence was located on human chromosome 11, whereas the genomic DNA amplified Sp17 gene sequence was found on human chromosome 10. These findings strongly support the presence of an intron-containing (Sp17-1) and intronless (Sp17-2) gene within the same Sp17 gene family.

Although computer based analyses of the Sp17-2 sequence predict several Sp17-2 promoter sites, they appear to be untranscribed and nonfunctional. Moreover, Sp17-2 sequence homology was not detected in human ESTs and human testis cDNA sequences, suggesting that only the Sp17-1 gene is transcribed in human testis. Furthermore, as compared to the Sp17-1 transcript, the Sp17-2 gene sequence exhibits several point mutations, nucleotide insertions and nucleotide deletions. Thus, it is possible that mutational events causing stop codons may have occurred within the open reading frame rendering the Sp17-2 gene a non-functional pseudogene.

Retroposition is a phenomenon characterized by the lack of introns, absence of promoter sequences, presence of a remnant 3' polyadenylated sequence and flanking of short direct repeats within the gene sequence (4, 5, 6). Examples of retroposed genes include human Alu elements, SINEs (short interspersed DNA elements), LINEs (Long interspersed elements) and SCAND2 (SCAN domain-containing 2) (9, 10, 11, 12). In particular, SCAND2 was recently identified as a retroposon of the SM-20 gene, a gene found in skeletal and cardiac muscles. SCAND2 was located on human chromosome 15 and is thought to have evolved by the retroposition of SM-20 mRNA by an exon-shuffling mechanism (12).

Similarly, because the Sp17-2 gene variant exhibits hallmarks of retroposition, it may have evolved from the Sp17-1 gene by retroposition. Specifically, Sp17-2 may have evolved in the human genome by the reverse transcription of a spliced Sp17-1 mRNA with the subsequent integration in to human chromosome 10. However, it was not known whether other mammalian species exhibit the Sp17-1 gene and Sp17-2 gene variant.

Southern blot analysis confirmed the existence of the Sp17-1 and Sp17-2 genes and suggested the existence of additional Sp17 gene variants within the human genome. Similarly, a phylogenic southern blot analysis of a variety of mammalian species revealed Sp17 hybridization corresponding to the Sp17-1 gene in the rat, mouse, cow and rabbit (7). In addition, Sp17 southern hybridization in the rhesus macaque and dog revealed multiple unidentified bands suggesting the existence of Sp17 gene variants in non-human mammals. Furthermore, sequence analysis of the Sp17-2 nucleotide sequence revealed a polyadenylation stretch, which was not identified in the Sp17-1 transcript (Chapter 1). However, sequence analysis of baboon mRNA revealed a polyadenylation signal closely resembling the human Sp17-2 gene sequence (13). These findings suggest that the Sp17-2 gene variant is not unique to the human genome.

DNA amplification of Sp17 in cynomologous macaque and rhesus macaque testes revealed the presence of both the Sp17-1 transcript and Sp17-2 gene variant. Conversely, an analogous study of rat lung revealed Sp17-1 DNA amplification from the cDNA template but not Sp17-2 amplification from the genomic template. These DNA amplifications confirm the conservation of the Sp17-1 transcript among mammals. However, although the intronless Sp17-2 gene variant was detected in higher mammals (e.g. humans and primates), it may not be highly conserved in lower mammals (e.g. rodents). Thus, because of the close relation between primates and humans, it is possible that the retroposition of the Sp17-1 gene to generate the Sp17-2 pseudo-gene variant occurred recently on the evolutionary timeline.

6. Conclusion

There is strong evidence to support the existence of at least one Sp17 gene variant, Sp17-2. The characterization of the intronless Sp17-2 gene variant suggests that it arose as a result of the retroposition of the Sp17-1 gene. Furthermore, although the Sp17-1 transcript is highly conserved among mammals, it is likely that the putatively retroposed Sp17-2 pseudo-gene evolved in higher mammals more recently in evolutionary time. However, the role of the Sp17-2 gene variant is not yet known.

7. Tables and Figures

Table 2.1. Sp17 specific sense and anti-sense primers used in PCR to detect and differentiate between the Sp17-1 transcript and the Sp17-2 gene variant.

Sense (Shu)	5'	CGCGGATCCATGTCGATTCCATTCTCC	3'
Sense (Sgr2)	5'	GAGGTCAGAGGAGGAAGTAAGCTG	3'
Sense (Sg2full)	5'	CCAATGACTGATGGACACAGACTG	3'
Anti-Sense (aCDS)	5'	GAGTCTAAGATGGTGACTGATG	3'
Anti-Sense (Ag2full)	5'	GAGATGTGATAGCTGGCACTCATG	3'
Anti-Sense (Ag2)	5'	CCATGAGCCAAAATGGTGGTAGTC	3'
Anti-Sense (rAspe5)	5'	GGGTAAAACCTGTGGTCTCAATTG	3'

Table 2.2. Sp17 sense and anti-sense primers used to develop a Sp17 DNA probe for southern blot analysis.

Sense (Shu)	5'	CGCGGATCCATGTCGATTCCATTCTCC	3'
Anti-Sense (Ahu)	5'	CGGGGTACCAACCAGTGTCCTCACTTG	3'

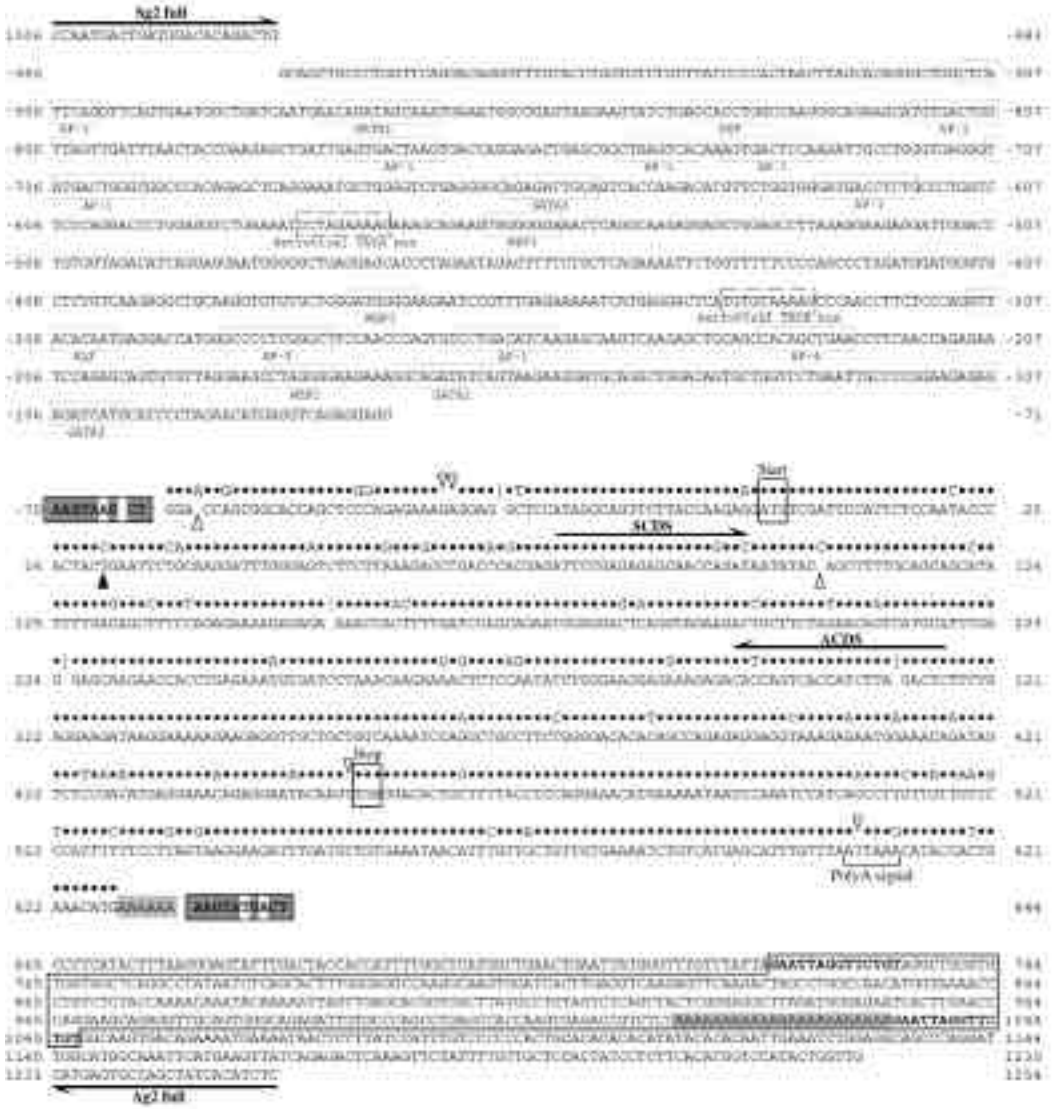

Figure 2.1. Nucleotide sequence of the intronless Sp17-2 retroposed pseudo-gene in human testis (Accession number AF335424). Characteristics of retroposition on the Sp17-2 gene are indicated and another conserved putative retroposon of unknown origin is boxed. Sp17-2 gene sequence alignment with the Sp17-1 transcript (dots) demonstrated 61 point mutations, 4 nucleotide insertions (down arrowheads) and 2 deletions (upward arrowheads). Sequence comparison was made with respect to the Sp17-1 transcript including exon-(intron)-exon boundaries (vertical lines), start and stop codons (solid boxes), polyadenylation signal (bracket), poly A track (shaded and bold), retroviral TATA box consensus sequences (broken line boxes), transcription factor binding sites (dotted boxed). Primer directions are indicated by arrows.

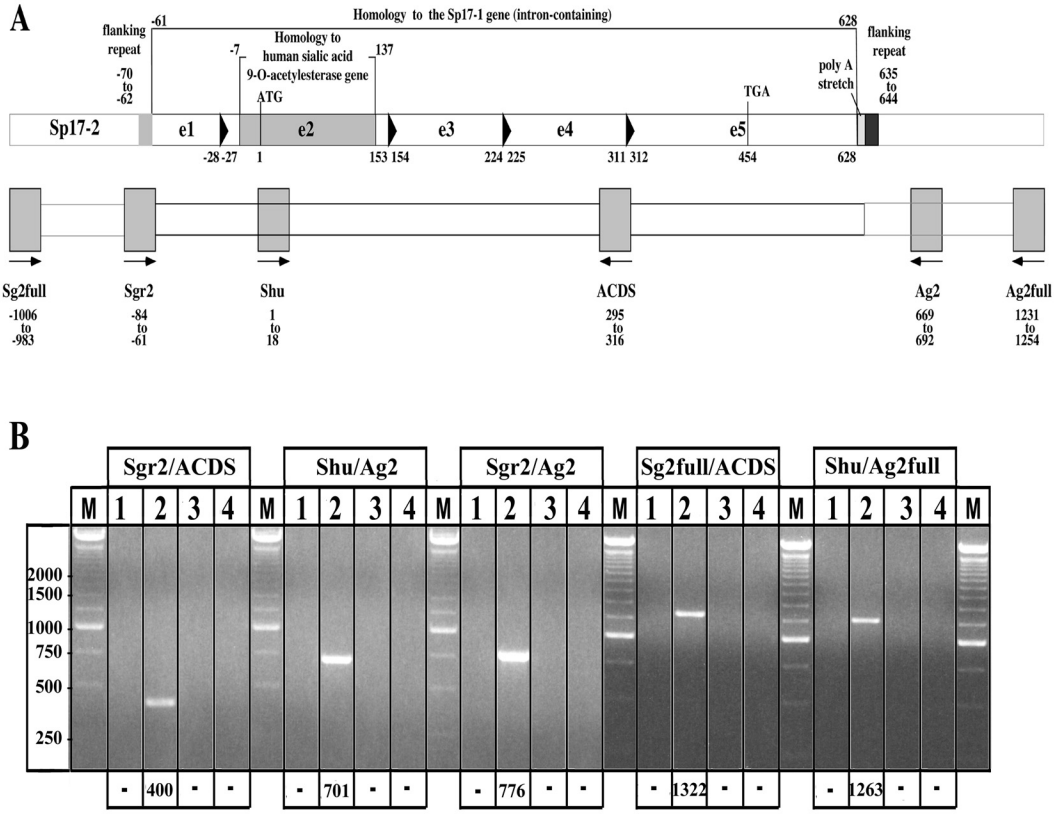

Figure 2.2. The detection of a Sp17-2 transcript in human testis by RT-PCR. **A.** Schematic diagram of the genomic organization of the human Sp17-2 gene flanked by direct repeats and the short poly(A) stretch. Sp17-1 sequence homology, exon-(intron)-exon boundaries (solid triangles) and start (ATG) and stop (TGA) codons (vertical bars) are indicated. Arrows denote primer directions and sequence orientation. **B.** RT-PCR analysis using Sp17-2 specific primers demonstrated the absence of an Sp17-2 transcript in human testis. The PCR cocktails were prepared with the following templates 1: cDNA; 2: gDNA; 3: RNA; 4: none (Chapter 1, Table 1.1). The PCR products were measured in bp adjacent to a DNA marker (M). All controls (3 and 4) were negative for genomic DNA.

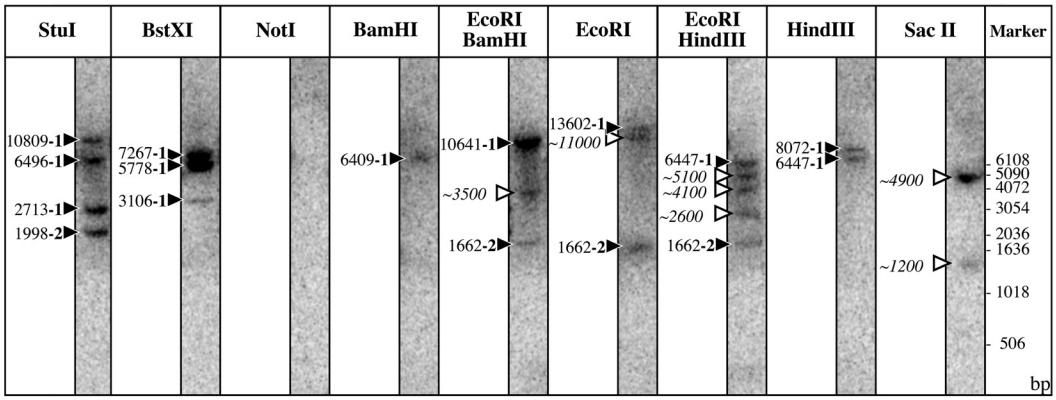

Figure 2.3. Southern blot analysis of human genomic DNA. Human genomic DNA was digested with the restriction enzymes Stu I, BstXI, Not I, BamHI, EcoRI/BamHI, EcoRI, EcoRI/HindIII, HindIII, and Sac II. The DNA fragments were separated on a 0.7% agarose gel, transferred onto a nylon membrane and hybridized with a probe for the open reading frame of Sp17 which including exon 2, 3, 4 and 5. Predicted fragment sizes are indicated with filled arrowheads, whereas unpredicted fragments are shown with open arrowheads in italic which indicate additional members of the Sp17 family. The DNA marker (kb) is shown at the right.

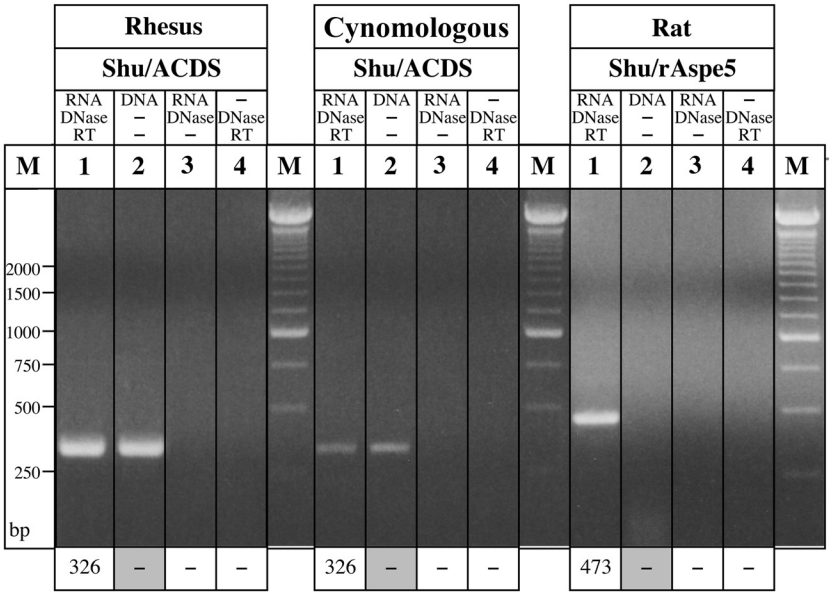

Figure 2.4. PCR analysis of the intronless Sp17-2 gene in the primate and rat. A Sp17-2 transcript was amplified from rhesus and cynomologous macaque testis cDNA. Similarly, the Sp17-2 gene was amplified from the genomic DNA of the rhesus macaque, cynomologous macaque and rat. However, the Sp17-2 transcript was not detected in rat lung cDNA. Sp17 specific primers were used and the PCR cocktails were prepared with the following templates 1: cDNA; 2: gDNA; 3: RNA; 4: none (Chapter 1, Table 1.1). The PCR products were measured in bp adjacent to a DNA marker (M). All controls (3 and 4) were negative for genomic DNA.

8. References

1 Mighell, A. J., N. R. Smith, P. A. Robinson and A. F. Markham. 2000. Vertebrate pseudogenes. Febs Letters. 468:109-14.

2 Piechaczyk, M., J. M. Blanchard, S. Riaad-El Sabouty, C. Dani, L. Marty and P. Jeanteur. 1984. Unusual abundance of vertebrate 3-phosphate dehydrogenase pseudogenes. Nature 312:469-471.

3 Brosius, J. and H. Tiedge. 1995. Reverse transcriptase: mediator of genomic plasticity. Virus Genes. 11:163-79.

4 Rogers, J. H., 1985. The origin and evolution of retroposons. Int. Rev. Cytol. 93:187-279.

5 Vanin, E. F., 1985. Processed pseudogenes: characteristics and evolution. Annu. Rev. Genet. 19: 253-272.

6 Weiner, A. M., P. L. Deininger and A. Efstratiadis. 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. 55:631-661.

7 Lea, I. A., R. T. Richardson, E. E. Widgren and M. G. O'Rand. 1996. Cloning and sequencing of cDNAs encoding the human sperm protein, Sp17. Biochim. Biophys. Acta. 1307:263-266.

8 Meinkoth J. and G. Wahl. 1984. Hybridization of nucleic acids immobilized on solid supports. Anal Biochem. 138:267-84.

9 Rowold D. J. and R. J. Herrera. 2000. Alu elements and the human genome. Genetica. 108:57-72.

10 Feschotte C., N. Fourrier, I. Desmons and C. Mouches. 2001. Birth of a retroposon: the Twin SINE family from the vector mosquito Culex pipiens may have originated from a dimeric tRNA precursor. Mol Biol Evol. 1:74-84.

11 Esnault C, Maestre J, Heidmann T. 2000. Human LINE retrotransposons generate processed pseudogenes. Nat Genet. 4:363-7.

12 Dupuy D, Aubert I, Duperat VG, Petit J, Taine L, Stef M, Bloch B, Arveiler B. 2000. Mapping, characterization, and expression analysis of the SM-20 human homologue, c1orf12, and identification of a novel related gene, SCAND2. Genomics. 69:348-54.

13 Adoyo, P.A., I. A. Lea, R. T. Richardson, E. E. Widgren and M. G. O'Rand. 1997. Sequence and characterization of the sperm protein Sp17 from the baboon. Mol. Reprod. Dev. 47:66-71.