ZOOLOGICAL SCIENCE **14**: 609–614 (1997)

Nucleotide Sequence of cDNA and the Gene Expression of Testis-Specific Protein Y in the Japanese Monkey

Heui-Soo Kim, Takashi Kageyama, Shin Nakamura
and Osamu Takenaka*

Department of Cellular and Molecular Biology, Primate Research Institute, Kyoto University, Inuyama, Aichi 484, Japan

ABSTRACT—We cloned the cDNA for Japanese monkey Testis-Specific Protein Y (TSPY). The cDNA contained an open reading frame of 246 amino acids. This coding fragment shared 89% nucleotide sequence identity and 81% amino acid sequence identity with the homologous fragment of previously isolated human TSPY cDNA. Monkey TSPY was assumed to be a molecular mass of 28 kDa and an isoelectric point of pH 5.35. This protein was hydrophilic and contained an Arg and Lys-rich region which was a potential DNA binding site. Expression of the TSPY gene examined by reverse transcription PCR showed that the transcript was detectable only in testis, suggesting that TSPY plays an important role in spermatogenesis of primates.

INTRODUCTION

Testis-specific protein Y (TSPY) is known to be a product of the Y-chromosome specific gene. The expression of the TSPY gene has been shown to be restricted to testicular tissue (Arnemann *et al.*, 1991) and appeared to be confined to germ cells of the spermatogonial and early spermatocyte stages in adult human males (Chandley and Cooke, 1994; Schnieders *et al.*, 1996). Although the exact function of TSPY is still unknown, this protein might play a role in DNA replication (Schnieders *et al.*, 1996). A cDNA clone for human TSPY was isolated from an adult human testis cDNA library (Arnemann *et al.*, 1991). A genomic clone for human TSPY was also isolated successively (Zhang *et al.*, 1992). The gene contained six exons with five introns and was estimated to be approximately 2.7 kb.

Human TSPY gene-related sequences are organized as constitutive parts of DYZ5 repeat units (Manz *et al.*, 1993) which are located on the short arm of the Y chromosome (Tyler-Smith *et al.*, 1988). DYZ5 sequences have been shown to be conserved on the Y chromosome of the great apes by Southern blot and *in situ* hybridization (Guttenbach *et al.*, 1992). Using the chromosomal *in situ* hybridization technique, Schempp *et al.* (1995) showed that TSPY gene-related sequences are conserved and Y chromosome specific in hominoids. The number of TSPY genes and related genes are highly amplified especially in primates (Kim and Takenaka, 1996; Kim *et al.*, 1996). We have sequenced exon 1, exon 2, and the first intron of the TSPY gene of great apes and the baboon and determined phylogenetic relationship among them (Kim and Takenaka, 1996). In our succeeding report, we compared restriction patterns and chromosomal localizations of TSPY genes in man, gibbons, and Old World monkeys, and found variations of gene structures among them (Kim *et al.*, 1996). Since in primates it is possible that TSPY genes evolve some structural differentiations and cause specificities in reproductive systems, it is necessary to clarify the structures of TSPY genes in various primates. To date, however, the complete structure of a TSPY gene is not known except for the human gene.

In the present study, we cloned and sequenced a full-length of the TSPY cDNA from Japanese monkey testis RNA, described its molecular characterizations, and examined the expression of the TSPY gene in various tissues.

MATERIALS AND METHODS

Isolation of total RNA from monkey tissues

Tissues were collected from a 6-year-old Japanese monkey (*Macaca fuscata*) immediately after death by exsanguination via bilateral carotid arteries under deep anesthesia with ketamine hydrochloride and sodium pentobarbital, in accordance with the guideline of the Primate Research Institute, Kyoto University. Total RNA was extracted by TRIZOL reagent (BRL).

Cloning of the cDNA for Japanese monkey TSPY

Full-length TSPY cDNA was prepared by RT-PCR. First, TSPY mRNA was reverse transcribed into single-stranded cDNA by AMV reverse transcriptase (BRL) using a primer, P356 (5'-CCTTGAGAATG-TTTATTTTTCATTCC-3'). Following this synthesis, cDNA was amplified by PCR using primers P356 and P459 (5'-CCAAGGAGGGCACC-GCCTTC-3'). Primers P356 and P459 were designed based on published sequences of human TSPY cDNA (Zhang *et al.*, 1992). Their locations in TSPY cDNA are shown in Fig. 1. The PCR was performed with a thermocycler manufactured by Perkin Elmer Cetus (Model 9600) as follows. After the initial denaturation step at 94°C for 3 min, DNA

* Corresponding author: Tel. +81-568-63-0577;
FAX. +81-568-62-9557.

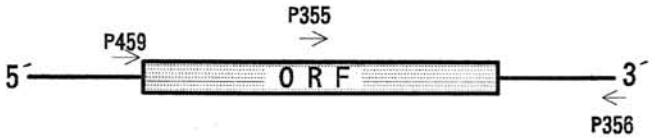

Fig. 1. Primer locations for RT-PCR. All primers were designed based on published sequences of human TSPY gene. The open box represents the open reading frame (ORF).

was amplified for 40 cycles at 94°C for 1 min, 60°C for 1 min and 72°C for 1.5 min. The PCR products were analyzed with agarose gel electrophoresis, purified by QIAEX gel extraction kit (QIAGEN), and cloned into T-vector which was prepared by modification of the Holton and Graham (1991) method.

Sequencing and data analyses

The nucleotide sequence was determined on both strands of the insert DNA using the dideoxy chain termination method (Sanger *et al.*, 1977). At least three cloned fragments from each of the PCR-amplified DNAs were sequenced. The various analyses of nucleotide sequences and encoded amino acid sequences were done with the aid of the GENETYX system (Ver. 9, SDC, Tokyo). Sequence similarity was searched for in the protein database from SWISS-PRO and the hydropathy value was calculated using the method of Eisenberg *et al.* (1984). The pairwise distance of the number of nucleotide substitutions was estimated using the method of Tajima and Nei (1984). The number of synonymous and nonsynonymous nucleotide substitutions were obtained by the MEGA program (Ver. 1.01, USA).

Detection of TSPY mRNA by RT-PCR

Expression of the TSPY gene in various tissues was analyzed by RT-PCR with specific primers P355 (5'-CAGATGTCAGCCCTGATCACTG-3') and P356 (5'-CCTTGAGAATGTTTATTTTTCATTCC-3') using *Taq* DNA polymerase. The locations of these primers in TSPY cDNA are shown in Fig. 1. The size of the PCR product was expected to be 630 bp. As a standard control for this method, we also examined the expression of the human G3PDH gene with the primers GPD-S (5'-ACCACAGTCCATGCCATCAC-3') and GPD-AS (5'-TCCACCACCCTGTTGCTGTA-3') using r*Tth* DNA polymerase (TOYOBO). The parameters included 1 cycle at 60°C for 30 min, followed by 94°C for 2 min, then 40 cycles at 94°C for 1 min and 60°C for 1.5 min, with an additional step of 60°C for 7 min. The PCR products were electrophoresed on 1.5% agarose gel, stained with ethidium bromide, and photodocumented on a UV transilluminator.

RESULTS AND DISCUSSION

Molecular cloning of TSPY cDNA and structure analysis

The nucleotide and deduced amino acid sequences of monkey TSPY cDNA are shown in Fig. 2. The nucleotide sequences were the same in three independent clones of RT-PCR products, being almost certain that the sequence was free from the mutated ones during amplification. The cDNA consisted of 976 nucleotides, which included an open reading frame of 741 nucleotides encoding 246 amino acids. The ATG triplet indicated as nucleotides 1-3 was an initiation codon of the open reading frame and stop codon appeared in nucleotides 739-741 as a TGA triplet. A polyadenylation site, AATAAA, was not included in the sequence since we used an oligonucleotide containing this site as a PCR primer. The analysis of the deduced amino acid sequence indicates that the gene product is a slightly acidic protein with an isoelectric point of pH 5.35 and has a molecular mass of 28 kDa. Amino acids contained at high levels in monkey TSPY were Glu (10.6%), Ala (9.4%), and Arg (7.3%). The sum of Glu and Asp was higher than that of basic residues such as Arg and Lys being consistent with the monkey TSPY which was an acidic protein. One of the characteristic features of this protein was its hydrophilic nature, and three hydrophilic regions (residues 42-51, 79-105, and 202-207) were clearly identified (Fig. 3). Interestingly, the 26-residue segment (residues 79-105) had an abundance of basic residues such as Arg and Lys. The region was well conserved between monkey, human, and bovine TSPY, and positively charged residues were concentrated in the lined box (Fig.5). This finding allows us to speculate that these residues may be implicated in the DNA binding property as a structural evidence of the Arg and Lys-rich region. Such a feature was already appeared in the high mobility group (HMG) box, a main functional domain, of the sex determining region Y (SRY) (Jantzen *et al.*, 1990). The putative nucleic-acid-binding motif within SRY and its testis specific expression is consistent with that of SRY which has a role in the developmental regulation of the testis (Sinclair *et al.*, 1990). TSPY may represent such a function in monkey testis. The Arg and Lys-rich sequence might have other roles such as that serving as a nuclear localization signal (Schnieders *et al.*, 1996). The exact function of this characteristic sequence remains to be clarified.

Comparison of TSPY sequences

The similarity of nucleotide sequences between monkey and human TSPY cDNAs was 85.7% as a whole, and 88.6% and 76.3% in the coding and 3' untranslated regions, respectively. The deduced amino acid sequence of monkey TSPY cDNA revealed an identity of 81.4% with the human TSPY (Fig. 5). Twenty-one deletions in nucleotide sequence (deletion of 7 amino acids in protein sequence) in the monkey coding region and eighteen insertions in the monkey 3' untranslated region were noted. The deleted sequence in monkey TSPY were corresponded to Pro-Arg-Glu at position 51-53 and Ser-Pro-Asp-Arg at position 234-237 of human TSPY. A fragment of bovine TSPY gene was isolated by Jakubiczka *et al.* (1993), and was shown to have high similarity (74.8%) with the human TSPY gene. The corresponding nucleotide sequences in monkey TSPY cDNA were slightly higher in similarity (78%) (Table 1). Amino acid sequences between monkey, human, and bovine TSPYs showed that the similarity between monkey and human TSPYs (81.4%) were much higher than those between monkey and bovine TSPYs (57.5%) (Fig. 5). Since nucleotide sequences between monkey and bovine TSPYs were showed high similarity, the high numbers of amino acid substitutions between them were caused by the high ratio of substitutions in the first or second codons between the triplets of the coding regions of monkey and bovine TSPY genes.

In order to clarify TSPY gene evolution, we calculated the pairwise distance of the number of nucleotide substitu-

```
                                                                    -3 AGG
        10        20        30        40        50        60        70        80        90
ATGGAGTCTCTACAGGAGGGGGAGGCCGGGGCGCAGAGCGAGCAGGTAGCTTTGGGGGAGGAGGCGGTGCTGGGAGCGGATGACATAATG
M  E  S  L  Q  E  G  E  A  G  A  Q  S  E  Q  V  A  L  G  E  E  A  V  L  G  A  D  D  I  M
       100       110       120       130       140       150       160       170       180
GCGGAGGTGGAGGTGGTGGCCCACCAGGAAGCCGACGAGAAGCGGCAGGAGCAGGTCCAGCGGGCACAGCCTGGCCCTGGGCCCATGAGC
A  E  V  E  V  V  A  H  Q  E  A  D  E  K  R  Q  E  Q  V  Q  R  A  Q  P  G  P  G  P  M  S
       190       200       210       220       230       240       250       260       270
CCAGAGTCTGCACTGGAGGAGCTGCTGGCCGTTCAGGTGGAGCTGGAGCCGGTTAATTCCCGAGCCAGGAAGGCCTTTTCTCAGCAGAGG
P  E  S  A  L  E  E  L  L  A  V  Q  V  E  L  E  P  V  N  S  R  A  R  K  A  F  S  Q  Q  R
       280       290       300       310       320       330       340       350       360
GAAAAGATGGAGCGGAGGCGCAAGCCCCACCTGGACCGCAGAGGCGCCATCATCCAGAGCATGCCTGGCTTCTGGGCCAATGTTATTGCA
E  K  M  E  R  R  R  K  P  H  L  D  R  R  G  A  I  I  Q  S  M  P  G  F  W  A  N  V  I  A
       370       380       390       400       410       420       430       440       450
AACCACCCTCGGATGTCGGCCCTGCTCACTGACCAAGATGAAGACATGCTGAGCTACATGATCAACTTGGAGGTGAAAGAAGCGAAGCAT
N  H  P  R  M  S  A  L  L  T  D  Q  D  E  D  M  L  S  Y  M  I  N  L  E  V  K  E  A  K  H
       460       470       480       490       500       510       520       530       540
CCCGTTCATCTCTGCCAGATCATGTTGTTCTTTCGGAGTAACCCCTACTTCCAGAATAAAGTGATTACCAAGGAATATCTCGTGAACGTC
P  V  H  L  C  Q  I  M  L  F  F  R  S  N  P  Y  F  Q  N  K  V  I  T  K  E  Y  L  V  N  V
       550       560       570       580       590       600       610       620       630
ACAGAATACAGGGCTTCTCATTCCACTCCAATTCAGTGGTGTCAGGATTATGAAGTTGAGGCCTATCGCCGCAGACACAACAACAGCGGT
T  E  Y  R  A  S  H  S  T  P  I  Q  W  C  Q  D  Y  E  V  E  A  Y  R  R  R  H  N  N  S  G
       640       650       660       670       680       690       700       710       720
CTTAACTTCTTCAACTGGTTTTCTGACCACAACTTCGCAGGATCCAATAGGATTGCTGAGTCCTATGTAAGGACCTGTGGCGCAATCCCC
L  N  F  F  N  W  F  S  D  H  N  F  A  G  S  N  R  I  A  E  S  Y  V  R  T  C  G  A  I  P
       730       740       750       760       770       780       790       800       810
TGCAATACTACAGGAGGATGAAGCCACCTGAAGAGGAAACAGAGATTTCAGGGAACGCGCAGATGTTGGGTTGAATATGATGGAGCATCG
C  N  T  T  G  G  *
       820       830       840       850       860       870       880       890       900
GACACAGGTGCGTGTTCACCTAACACGGCAGAACTCCTGAAAACTTACTACAGTATGCAGGATGTCAGCACTCAGCATGGTCTTGTGCAC
       910       920       930       940       950       960       970
AGGAACTAAAGGACAAGCAGATCCAGTCACAGCAAATGAAGACAGGAAGCGGGGTGGGGGAATTGAATTGTAT
```

Fig. 2. Nucleotide sequences of the monkey TSPY cDNA (accession no. AB001421) together with the translation of its open reading frame. Amino acids are shown in single-letter codes. The initiation codon (ATG) is underlined. The termination codon (TGA) is indicated by an asterisk.

tions per site (K_N) and the rate of nucleotide substitutions per site per year (V_N) using divergence times from paleontological data (Gingerich, 1984; Pilbeam, 1986). As shown in Table 2, the K_N and V_N values in the coding region between monkey and human TSPY cDNAs were 0.096 and 1.6-2.4 $\times 10^{-9}$ / site / year, respectively. Whereas, the K_N and V_N values in the 3' untranslated region between them were 0.199 and 3.3-5.0 $\times 10^{-9}$ / site / year, respectively. These values are almost the

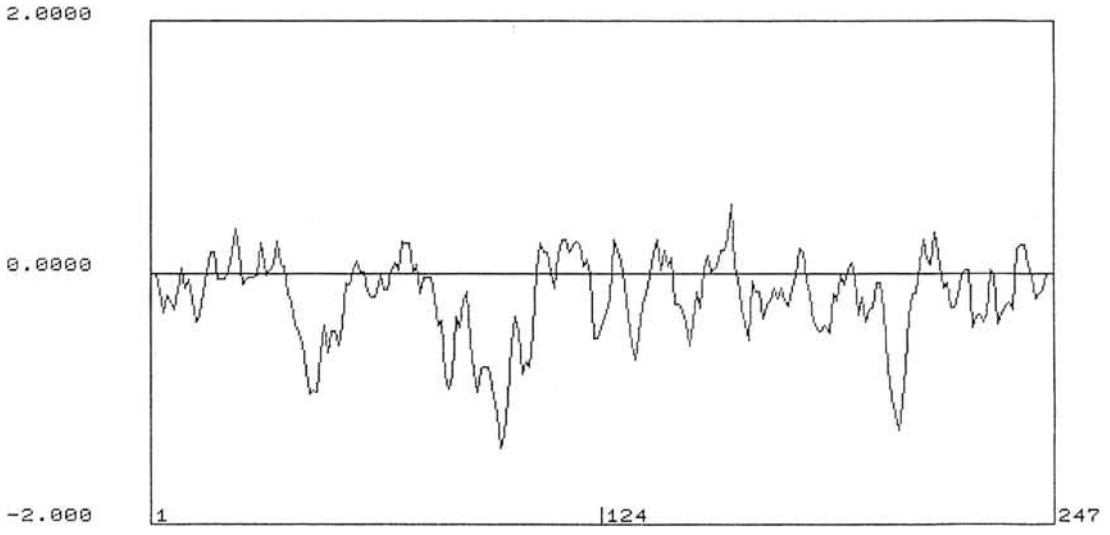

Fig. 3. Hydrophilicity and hydrophobicity plot of putative monkey testis-specific protein. The y axis shows the hydrophilicity and hydrophobicity values (–2.0000 to 2.0000) for TSPY as residues shown on the x axis. Values below the 0.0000 mark represent hydrophilic regions; those above are hydrophobic regions. Amino acid positions 42-51, 79-105, and 202-207 show mainly hydrophilicy.

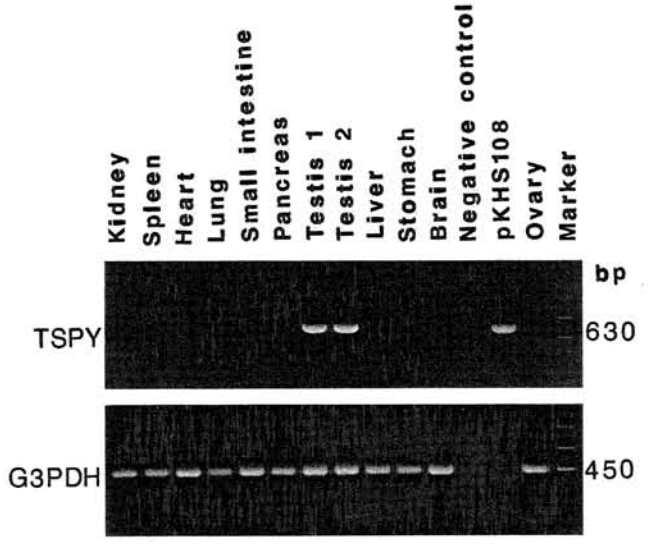

Fig. 4. Expression analyses of the monkey TSPY gene by RT-PCR. The TSPY-specific DNA fragment was amplified by 40 cycles of RT-PCR using 1 μg each of total RNAs from kidney (lane 1), spleen (lane 2), heart (lane 3), lung (lane 4), small intestine (lane 5), pancreas (lane 6), testes (lanes 7, 8, from different individual sources), liver (lane 9), stomach (lane 10), brain (lane 11), testis (lane 12, without *Taq* DNA polymerase), and ovary (lane 14). The pKHS108 (TSPY cDNA containing plasmid) was used as positive control (lane 13). As the marker, *Taq I*-digested pUC118 was loaded in lane 15. The expected fragment size is 630 bp for transcript-specific RT-PCR product. RT-PCR of G3PDH mRNA was used for the quantifications of RNA level in each tissue.

same as that of the first intron of monkey and human TSPY genes cited in our previous report (Kim *et al.*, 1996). In comparison with the coding and 3' untranslated region of the TSPY gene, V_N of the 3' untranslated region was twofold higher than that of the coding region. Furthermore, comparing the sequence of the TSPY genomic DNA between great apes and humans, the rates of nucleotide substitutions per site per year were higher in the TSPY intron than in the TSPY exon (Kim and Takenaka, 1996). Therefore, noncoding regions of the TSPY gene evolved more rapidly than the coding region.

We calculated the level of synonymous substitutions per site (Ks) and nonsynonymous substitutions per site (Ka) in the TSPY gene (Table 3). As nonsynonymous substitutions are more strongly influenced by selection than synonymous substitutions, their ratio is a good indicator of selection. Comparing the human-monkey (762 bp) and monkey-bovine (123 bp) TSPY, the values of Ks were 12.6 and 29.8, while those of Ka / Ks were 0.68 and 0.84, respectively. The number of synonymous substitutions were higher than the number of nonsynonymous substitutions in the TSPY gene. Therefore, directional selection has not occurred in the TSPY gene.

RNA expression analyses

Expression of the TSPY gene was examined by RT-PCR analyses in various tissues including the testis (Fig. 4). Using the specific primers (P355 and P356), a fragment of 630 bp was generated from only testes RNA (lane 7, 8), and no expression was detectable in the other tissues tested. No PCR product was observed in testis without reverse transcriptase treatment (lane 12). When we used plasmid DNA containing monkey TSPY cDNA as a template for PCR, the 630-bp fragment was clearly observed (lane 13). These results showed that the expression of the TSPY gene is specific for testes as suggested by Zhang *et al.* (1992).

To date, the TSPY, SRY, RNA binding motif (RBM), previously called the YRRM, Y-located RNA recognition motif, and deleted in azoospermia (DAZ) genes are Y-specific although there are many homologous genes between the X and Y chromosomes in man, such as, the ribosomal protein S4

Fig. 5. Comparison of the amino acid sequences of monkey testis-specific protein with those of the human (Zhang *et al.*, 1992) and bovine (Schnieders *et al.*, 1996). The numbering of each residue is based on the human sequence. Although the complete sequence of bovine TSPY has been reported (Schnieders *et al.*, 1996), the N-terminal 53 residues sequence is not shown in Fig. 5. The nucleotide sequence of the bovine genomic DNA fragment is known (Jakubiczka *et al.*, 1993) and its transcribed amino acid sequence corresponds to residue 107 to residue 147. The nucleotide and deduced amino acid sequences of this fragment were used for comparison with those of monkey and human sequences, respectively. Arg and Lys residues are shown by black background.

Table 1. Percent similarity of nucleotide sequences in the TSPY gene

Species pair	Coding region	3' Untranslated region
Human-Monkey	88.6 (762)	76.3 (232)
Human-Bovine	74.8 (123)	–
Monkey-Bovine	78.0 (123)	–

The number in parenthesis represent base pairs compared with each other. The same system is used in Table 2 and 3.

Table 2. Mean ± standard error of the number of nucleotide substitutions per site in the TSPY gene

Species pair	Coding region	3' Untranslated region
Human-Monkey	0.096 ± 0.012 (762)	0.199 ± 0.035 (232)
Human-Bovine	0.332 ± 0.070 (123)	–
Monkey-Bovine	0.275 ± 0.060 (123)	–

Table 3. Synonymous substitutions per site (Ks), nonsynonymous substitutions per site (Ka), and their ratio (Ka/Ks) in the TSPY gene

Species pair	Ks × 100	Ka × 100	Ka/Ks
Human-Monkey	12.6 ± 2.9 (762)	8.6 ± 1.3 (762)	0.68
Human-Bovine	45.1 ± 17.1 (123)	27.4 ± 6.1 (123)	0.61
Monkey-Bovine	29.8 ± 12.8 (123)	25.0 ± 5.8 (123)	0.84

(RPS4Y), zinc finger protein Y (ZFY), amelogenin (AMELY), and steroid sulphatase (STS-Y). Our results from RT-PCR analysis have demonstrated that the expression of monkey TSPY is confined to the testis. Expression of both TSPY and RBM genes were confined to germ cells of the spermatogonial and early spermatocyte stages of adult human testis in RNA in situ hybridization (Chandley and Cooke, 1994). Testis-specific and germ cell-specific expression of these genes allow us to speculate their specific roles in spermatogenesis.

ACKNOWLEDGMENTS

This research was supported by Grants-in-Aid for Scientific Research on Priority Areas (No. 06273217 to OT; No. 07640900 to TK) from the Japan Ministry of Education, Science, Sports and Culture, and partially funded by the Sasakawa Scientific Research Grant (No. 7-161 to KHS) from the Japan Science Society.

REFERENCES

Arnemann J, Jakubiczka S, Thüring S, Schmidtke J (1991) Cloning and sequence analysis of a human Y chromosome derived, tescular cDNA, TSPY. Genomics 11: 108–114

Chandley AC, Cooke HJ (1994) Human male fertility-Y-linked genes and spermatogenesis. Hum Mol Genet 3: 1449–1452

Eisenberg D, Weiss RM, Terwilliger TC (1984) The hydrophobic moment detects periodicity in protein hydrophobicity. Proc Natl Acad Sci USA 81: 140–144

Gingerich PD (1984) Primate evolution: evidence from the fossil record, comparative morphology, and molecular biology. Yearbook Phys Anthropol 27: 57–72

Guttenbach M, Müller U, Schmid M (1992) A human moderately repeated Y-specific DNA sequence is evolutionary conserved in

the Y chromosome of the great apes. Genomics 13: 363–367
Holton TA, Graham MW (1991) A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucl Acids Res 19: 1156
Jakubiczka S, Schnieders F, Schnidtke J (1993) A bovine homologue of the human TSPY gene. Genomics 17: 732–735
Jantzen H-M, Admon A, Bell SP, Tjan R (1990) Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. Nature 344: 830–836
Kim H-S, Hirai H, Takenaka O (1996) Molecular features in TSPY gene of gibbons and Old World monkeys. Chrom Res 4: 500–506
Kim H-S, Takenaka O (1996) A comparison of TSPY genes from Y-chromosomal DNA of the great apes and humans: sequence, evolution, and phylogeny. Am J Phys Anthropol 100: 301–309
Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetics analysis, version 1.01. The Pennsylvania State University, University Park, PA 16802
Manz E, Schnieders F, Brechlin AM, Schmidtke J (1993) TSPY-related sequences represent a microheterogeneous gene family organized as constitutive elements in DYZ5 tandem repeat units on the human Y chromosome. Genomics 17: 726–731
Pilbeam D (1986) Distinguished lecture: hominoid evolution and hominoid origins. Am Anthropol 88: 295–312
Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467
Schempp W, Binkele A, Arnemann J, Glaser B, Ma K, Taylor K, Toder R, Wolfe J, Zeitler S, Chandley AC (1995) Comparative mapping of YRRM-and TSPY-related cosmids in man and hominoid apes. Chrom Res 3: 227–234
Schnieders F, Dörk T, Arnemann J, Vogel T, Werner M, Schmidtke J (1996) Testis-specific protein, Y-encoded (TSPY) expression in testicular tissues. Hum Mol Genet 11: 1801–1807
Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf A-M, Lovell-Badge R, Goodfellow PN (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346: 240–244
Tajima F, Nei M (1984) Estimation of evolutionary distance between nucleotide sequences. Mol Biol Evol 1: 269–285
Tyler-Smith C, Taylor L, Müller U (1988) Structure of a hypervariable tandemly repeated DNA sequence in the short arm of the human Y chromosome. J Mol Biol 203: 837–848
Zhang JS, Yang-Feng TL, Muller U, Mohandas TK, de Jong PJ, Lau Y-FC (1992) Molecular isolation and characterization of an expressed gene from the human Y chromosome. Hum Mol Genet 1: 717–726

(Received March 21, 1997 / Accepted April 18, 1997)