Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

Comparative analysis of human masculinity

B.K. Bhowmick, N. Takahata, M. Watanabe and Y. Satta

School of Advanced Sciences,
The Graduate University for Advanced Studies, Kanagawa, Japan
Corresponding author: B.K. Bhowmick
E-mail: bejon@soken.ac.jp/shilmud@animail.net

Genet. Mol. Res. 5 (4): 696-712 (2006)
Received March 3, 2006
Accepted September 4, 2006
Published November 23, 2006

ABSTRACT. To study rapidly evolving male specific Y (MSY) genes we retrieved and analyzed nine such genes. VCY, HSFY and RBMY were found to have functional X gametologs, but the rest did not. Using chimpanzee orthologs for XKRY, CDY, HSFY, PRY, and TSPY, the average silent substitution is estimated as 0.017 ± 0.006/site and the substitution rate is 1.42 x 10^{-9}/site/year. Except for VCY, all other loci possess two or more pseudogenes on the Y chromosome. Sequence differences from functional genes show that BPY2, DAZ, XKRY, and RBMY each have one pseudogene for each one that is human specific, while others were generated well before the human-chimpanzee split, by means of duplication, retro-transposition or translocation. Some functional MSY gene duplication of VCY, CDY and HSFY, as well as X-linked VCX and HSFX duplication, occurred in the lineage leading to humans; these duplicates have accumulated nucleotide substitutions that permit their identification.

Key words: Substitution rate, Divergence time, Pseudogene, Palindrome, Evolutionary mechanism, Maleness

INTRODUCTION

The "eccentric" Y chromosome is about half the size of X; it accommodates a few dozen genes. Some of them are house-keeping genes, and some are tissue specific. Previously, it was thought that human sex was determined by environmental factors, but now it is clear that X and Y distribution is the key factor for sex determination (Ohno, 1967). The finding of relics of pseudo-autosomal regions, X counterparts and distinct non-recombining (NRY) regions are explicit signs of stepwise Y evolution. As the Y is found only in the male, and it has distinct NRY regions, it can be easily deduced that NRY has some genes that have function only in males or are adverse for females. It is therefore of interest to know the origin of such genes. According to Sykes (1980), the Y chromosome will become extinct within 5,000 generations if the shrinking in size continues. Why is the Y short? Mutation is the agent; it accumulates in the NRY region, and consecutively some genes have lost functions; these nonselective constraints were knocked out by a series of deletion, resulting in the current puny Y. This mutation rate is theoretically constant lineage-to-lineage, and it is either globally or locally clock (Zuckerkandl and Pauling, 1965; Takahata and Satta, 1997) pursed. Using this concept and DNA sequence analysis, evolution can be speculated on more accurately. Relatively small and large numbers of sequence differences imply that divergence happened recently or a long time ago, respectively (Karin and Lahn, 2001). NRY constitutes about 95% of the length of the Y chromosome, and it harbors 12 novel genes, including seven that are testis specific (Jegalian and Lahn, 2001). Rozen et al. (2003) found abundant recombination in the NRY region, and they were therefore prompted to give a new name: male-specific region (MSY) inside the NRY. Skaletsky et al. (2003) found nine testis-specific genes in the MSY; all of them are in palindromes; however, Reijo et al. (2000) found one gene (DAZ) that was not palindromic. For evolutionary reasons, the Y chromosome had been thought to be a favored site for genes involved in spermatogenesis (Fisher, 1931). Genes that drive sperm production evolve unusually rapidly, presumably due to fierce rivalry (Metz and Palumbi, 1996; Ting et al., 1998; Wyckoff et al., 2000).

In order to understand the mode of evolution of MSY genes, we focused on nine such gene families, XKRY, VCY, CDY, HSFY, PRY, BPY2, DAZ, TSPY, and RBMY.

XKRY has two identical functional copies that are expected to be more centromeric in location (Skaletsky et al., 2003). The only known active X and Y gene homolog is VCY (Lahn and Page, 1999a). X contains four paralogs and Y contains two, suggesting X, Y clusters created by a duplication event (Fukumai et al., 2000). The CDY gene is testis specific, with multiple copies on Y and no X homolog (Skaletsky et al., 2003). No Y ortholog has been found in any non-primate or even in prosimians, suggesting that it moved to the primate Y recently (Lahn and Page, 1999b). Another study (Saut et al., 2000) found CDY1 and CDY2 to be two different genes, based on blot analysis. The HSFY gene is marked (Skaletsky et al., 2003) as MSY, with two functional copies, whereas Lahn and Page (2001) did not include this gene in their 12 novel NRY genes. They also did not find any X homolog. So, whether HSFY is MSY or not is still unknown. Three identical copies of PRY gene were found on the human Y, with an expectation of more telomeric copies (Lahn and Page, 2001). On the other hand, Vogt et al. (1996) found PRY1 and PRY2 to be two different genes in the AZFb region. While, according to Karin and Lahn (2001), PRY1 and PRY2 are alternatively spliced. Therefore, the position, number and phylogeny of this gene are still not resolved. Previously the BPY gene was recognized as the VCY gene. But now the BPY2 gene has been identified as a distinct gene, with its three nearly

identical copies. DAZ genes are considered as testis specific (Lahn and Page, 2001), with multiple copies (Stouffs et al., 2001); four copies were found by Saxena et al. (2000) and Skaletsky et al. (2003), organized into two clusters in the AZFc region of Y. Doubling of the DAZ genes took place approximately 50,000 to 200,000 years ago in the hominid Y (Agulnik, 1998; Howard, 2002). This gene was solely indicated as the candidate for the azoospermia factor. However, Agulnik (1998) found that DAZ has little or no role in spermatogenesis. They also indicated that exons and introns of DAZ are evolving with neutral genetic drift (Kimura, 1968) and without selective pressure. In the TSPY multi-gene family, which is a candidate for a factor that promotes gonadoblastoma formation, Arnemann et al. (1987) and Delbridge et al. (2004) found 35 copies in the amplicon region of this gene. RBMY consists of approximately 30 genes and pseudogenes, found on both arms of the Y chromosome. RBMX retains a widely found function, and it evolved with a male-specific function in spermatogenesis. Ma et al. (1993) and Saxena et al. (1996) found six functional copies in the pal 3 and inverted repeat regions and 26 pseudogenes. Delbridge and Graves (1999) suggested that RBMY and RBMX evolved from a gene on the mammalian proto-X and -Y pair, at least 130 million years ago, before the divergence of eutherian and metatherian mammals.

MATERIAL AND METHODS

Human copies were retrieved from the NCBI BLAST and the Pseudogene.org databases. Orthologous chimpanzee copies were obtained from NCBI Homologene; whenever no deposited sequence was found in NCBI, then orthologous bac clones were obtained from the DDBJ data bank. Clustal W (Thomson et al., 1994) and Graph alignment (http://darwin.nmsu.edu/cgi-bin/graph_align.cgi) were used to align homologous regions. MEGA-2 software (Kumar et al., 2001) and the neighbor joining method (Saitou and Nei, 1987) were used to estimate distances and develop the phylogenetic tree. The EMBL databank was used for the expression record. The human-chimp divergence time (DT) was obtained from the fossil record (Stewart and Disotell, 1998). By using orthologous distances and the fossil record (DT), average rates for specific genes were estimated individually by the formula: rate = (Dis (distance)/2T (time))/site/year. By using these average rates, DT of human paralogs was estimated by the formula DT = (Dis/2 x rate) million years ago (mya). Gene Atlas and Gene Card were used to obtain the specific function of the genes. ABI Online analysis tools (EBI) for GC content observation (Williams, 2000) were used to find SNPs between nearly identical sequences and tandem repeats analysis (Benson, 1999). The Hovergen family (http://pbil.univ-lyon1.fr/cgi-bin/acnuc-link-ac2fam?db=Hoverprot&query=O14609) was used to retrieve coding sequence (CDS) regions from different lineages. Dot plot analysis (Pustell and Kafatos, 1982) was performed for large-scale sequence comparison. Window analysis (Maizel and Lenk, 1981) was done for region by region sequence comparison. BLAST2 (Altschul et al., 1997), Gene.MIT (Burge and Karlin, 1997), Grail (http://darwin.nmsu.edu/cgi-bin/graph_align.cgi) and NCBI model makers were used to obtain putative gene structures.

RESULTS

All the genes had orthologous Y copies in the chimp, except BPY2, DAZ and RBMY. Chimp orthologs had 0.017 ± 0.006 substitutions/site on average difference compared to those of humans for MSY genes. This estimation is very similar to that of Ebersberger et al. (2002),

who used only non-repeated regions of NRY. Based on the average difference, we attempted to estimate rate, giving 1.4 x 10^{-9} substitutions/site/year. As DAZ, BPY2 and RBMY do not have chimp orthologs, we used the average rate from the present study and the estimation of the study of Kumar and Hedges (1998) to estimate their DT. In other cases, to estimate the DT of the paralogs for individual genes, we used their specific rate (Table 1). DT estimation for all human MSY paralogs are shown in Table 2. Human paralogs of XKRY, CDY, HSFY, BPY2, PRY, DAZ, TSPY, and RBMY gave 12, 60, 15, 16, 53, 0, 11, and 20% (functional to pseudogene) divergence on average, respectively. DT (mya) of the above order of genes ranged from 0 to 42, 0 to 0, 0 to 251, 0 to 153, 0 to 162, 0 to 35, 0 to 0.17, 2 to 31, and 0 to 221 (Table 2). Thus, we can infer that XKRY splitting started off at the mouse-rat divergence period, VCY and DAZ within the human lineage, CDY and RBMY at the Aves-Crocodilia divergence period, HSFY and BPY2 at the human-marsupial divergence period, and PRY and TSPY at the mouse-rat divergence period, according to the molecular time scale of Nachman et al. (1998). We retrieved functional and pseudogene copy numbers for all MSY genes (Table 1). Correlation between physical distances and pair-wise distances was very weak, except for the VCY gene (Table 1). Among pseudogenes, maximum was for reduced length and shuffled exons. Some of them are predicted as processed pseudogenes (Figure 1). Some palindromic regions were found as new and some were considered as off-target (Figure 2). Some MSY genes showed tendency for rearrangement towards non-Y chromosomes to maintain maleness despite Y extinction. Selective pressure was found to be absent among different mammalian lineages (only available in NCBI and Homologene and DDBJ databases (accessed April 18, 2005)).

DISCUSSION

XKRY

No homologous functional copy was found either in humans or in other primates. So, we can think that this gene may be human and Y specific. Lahn and Page (2001) found two functional copies, whereas we found only one. One copy found by Lahn and Page (2001) we found to be a pseudogene, due to a single-base insertion. Identical copies are adjacent; these were inferred to be originated by a duplication event. One pseudogene, which was isolated and distantly located, but identical, apparently was a translocation after duplication. We found two functional copies outside of palindromes 4 and 5, while Lahn and Page (2001) found them on palindromes 4 and 5. We found eight pseudogenes, while Lahn and Page (2001) found six.

VCY

The functional copies are identical and oppositely directed, so, it can be inferred that they were duplicated and inverted very recently. We found three X homologs, different what were found by Fukumai et al. (2000). X copies are tandemly repeated (VCX-2r, VCX-8r and VCX-10r). Another hypothetical X homolog (Loc401578) was found which has 6 repeats, and alternatively spliced with VCX-10r. All X copies showed around 1% divergence among them on average and 1.7% from Y copies (silent sites are only considered). Using Nachman et al. (1998) rate (who estimated the rate by using seven human X gene introns), the DT of VCX copies was estimated (Table 2). Two Y copies showed 100% identity in their genic region as well as in their

Table 1. Observation of MSY gene families and their evolutionary pattern.

Gene name	Y (F/P)	Main functions	X (F/P)	Dis with Chimp orthologs	Avg. rate (x 10^{-9} sb/s/y)	Y/X Avg Dis (ortholog)	Avg Ds-Dn (Human)	Avg Ds-Dn (Mus)	Avg Ds-Dn (Macc)
XKRY	1/8	Kell blood related, fertilization	0/0	0.017 ± 0.005	1.475 ± 0.006	NF	ND	NF	NF
VCY	2/0	Variable charging, spermatogenesis	3/0	0.0158 ± 0.009	1.31 ± 0.009	0.183	0.0 ± 0.00	NF	NF
CDY	2/14	Sex ratio distortion, chromo do protein production, sperma-togenesis	0/1	0.0154 ± 0.009	1.283 ± 0.009	NF	0.310 ± 0.026	-0.021 ± 0.03	0.32 ± 0.02
HSFY	2/5	Heat shock determining, spermatogenesis	2/0	0.0133 ± 0.005	1.1 ± 0.005	NF	0.0 ± 0.00	NF	NF
PRY	3/3	PTP production, sperm ejaculation related	0/0	0.0174 ± 0.005	1.45 ± 0.005	NF	0.030 ± 0.06	NF	NF
BPY2	1/3	Germ cell development, male infertility related	0/0	NF	NF	NF	ND	NF	NF
DAZ	2/1	RNA-binding protein that plays an essential role in spermatogenesis	0/0	NF	NF	NF	-0.007 ± 0.013	NF	NF
TSPY	1/5	Unknown, may be involved in sperm differentiation and proliferation (Gene card)	1/0	0.025 ± 0.006	2.15 ± 0.006	NF	0.058 ± 0.043	0.0725 ± 0.051	0.001 ± 0.061
RBMY	2/9	Spermatogenesis	2/0	NF	NF	0.043	0. 329 ± 0.02	0.378 ± 0.041	ND

Data are reported as means ± SEM. Dis: average (Avg) substitutions (sb)/site (s) in human; F: functional; P: pseudogene; Ds-Dn = Ds means distances in synonymous sites and Dn means distances in non-synonymous sites; NF: not found in the database; ND: not done; P: distances of human paralogs.

Table 2. Distance (Dis) and divergence time (DT) estimates of the human MSY paralogs.

CDY

Dis/DT	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16
A1		24.94	242.01	243.18	228.76	228.76	239.28	239.28	239.28	24.94	247.07	247.07	247.03	247.03	251.36	251.36
A2	0.064 ± 0.01		239.14	240.45	226.03	226.03	239.28	239.28	239.28	239.28	248.63	248.63	248.63	248.63	251.36	251.36
A3	0.621 ± 0.02	0.614 ± 0.02		8.96	68.97	68.97	74.43	74.43	74.43	74.43	219.40	219.40	220.97	220.97	229.92	229.92
A4	0.624 ± 0.02	0.617 ± 0.02	0.023 ± 0.00		66.25	66.25	77.16	77.16	77.16	77.16	217.84	217.84	219.40	219.40	227.20	227.20
A5	0.587 ± 0.02	0.580 ± 0.02	0.177 ± 0.02	0.170 ± 0.02		0	79.50	79.50	79.50	79.50	212.78	212.78	213.95	213.95	226.03	226.03
A6	0.587 ± 0.02	0.580 ± 0.02	0.177 ± 0.02	0.170 ± 0.02	0.000 ± 0.00		79.50	79.50	79.50	79.50	212.78	212.78	213.95	213.95	226.03	226.03
A7	0.614 ± 0.02	0.614 ± 0.02	0.191 ± 0.02	0.198 ± 0.02	0.204 ± 0.02	0.204 ± 0.02		0	0	0	224.47	224.47	226.03	226.03	231.09	231.09
A8	0.614 ± 0.02	0.614 ± 0.02	0.191 ± 0.02	0.198 ± 0.02	0.204 ± 0.02	0.204 ± 0.02	0.000 ± 0.00		0	0	224.47	224.47	226.03	226.03	231.09	231.09
A9	0.614 ± 0.02	0.614 ± 0.02	0.191 ± 0.02	0.198 ± 0.02	0.204 ± 0.02	0.204 ± 0.02	0.000 ± 0.00	0.000 ± 0.00		0	224.47	224.47	226.03	226.03	231.09	231.09
A10	0.064 ± 0.02	0.614 ± 0.02	0.191 ± 0.02	0.198 ± 0.02	0.204 ± 0.02	0.204 ± 0.02	0.000 ± 0.00	0.000 ± 0.00	0		224.47	224.47	226.03	226.03	231.09	231.09
A11	0.634 ± 0.02	0.638 ± 0.02	0.563 ± 0.02	0.559 ± 0.02	0.546 ± 0.02	0.546 ± 0.02	0.576 ± 0.02	0.576 ± 0.02	0.576 ± 0.00	0.576 ± 0.02		0	1.169	1.169	115.35	115.35
A12	0.634 ± 0.02	0.638 ± 0.02	0.563 ± 0.02	0.559 ± 0.02	0.546 ± 0.02	0.546 ± 0.02	0.576 ± 0.02	0.576 ± 0.02	0.576 ± 0.02	0.576 ± 0.02	0.000 ± 0.00		1.169	1.169	115.35	115.35
A13	0.634 ± 0.02	0.638 ± 0.02	0.566 ± 0.02	0.563 ± 0.02	0.549 ± 0.02	0.549 ± 0.02	0.580 ± 0.02	0.580 ± 0.02	0.580 ± 0.02	0.580 ± 0.02	0.003 ± 0.00	0.003 ± .000		0	0	115.35
A14	0.634 ± 0.02	0.638 ± 0.02	0.566 ± 0.02	0.563 ± 0.02	0.549 ± 0.02	0.549 ± 0.02	0.580 ± 0.02	0.580 ± 0.02	0.580 ± 0.02	0.580 ± 0.02	0.003 ± 0.00	0.003 ± 0.00	0.000 ± 0.00		115.35	115.35
A15	0.645 ± 0.02	0.645 ± 0.02	0.590 ± 0.02	0.583 ± 0.02	0.580 ± 0.02	0.580 ± 0.02	0.593 ± 0.02	0.593 ± 0.02	0.593 ± 0.02	0.593 ± 0.02	0.296 ± 0.02	0.296 ± 0.02	0.000 ± 0.02	0.296 ± 0.02		0
A16	0.645 ± 0.02	0.645 ± 0.02	0.590 ± 0.02	0.583 ± 0.02	0.580 ± 0.02	0.580 ± 0.02	0.593 ± 0.02	0.593 ± 0.02	0.593 ± 0.02	0.593 ± 0.02	0.296 ± 0.02	0.296 ± 0.02	0.296 ± 0.02	0.296 ± 0.02	0	

XKRY

Dis/DT	A1	A2	A3	A4	A5	A6	A7	A8	A9
A1		0	40.33	40.33	40	40	40	41.69	41.69
A2	0.000 ± 0.00		40.33	40.33	40	40	40	41.69	41.69
A3	0.119 ± 0.01	0.119 ± 0.01		0	40.67	40.67	40.67	40	40
A4	0.119 ± 0.01	0.119 ± 0.01	0.000 ± 0.00		40.67	40.67	40.67	40	40
A5	0.118 ± 0.00	0.118 ± 0.00	0.120 ± 0.01	0.120 ± 0.01		0	0	42.03	42.03
A6	0.118 ± 0.00	0.118 ± 0.00	0.120 ± 0.01	0.120 ± 0.01	0.000 ± 0.00		0	42.03	42.03
A7	0.118 ± 0.00	0.118 ± 0.00	0.120 ± 0.01	0.120 ± 0.01	0.000 ± 0.00	0.000 ± 0.00		42.03	42.03
A8	0.123 ± 0.01	0.123 ± 0.01	0.118 ± 0.00	0.118 ± 0.00	0.124 ± 0.01	0.124 ± 0.01	0.124 ± 0.01		0
A9	0.123 ± 0.01	0.123 ± 0.01	0.118 ± 0.00	0.118 ± 0.00	0.124 ± 0.01	0.124 ± 0.01	0.124 ± 0.01	0.00 ± 0.00	

PRY

Dis/DT	A1	A2	A3	A4	A5	A6
A1		33.79	0	33.79	33.10	35.17
A2	0.098 ± 0.01		33.79	0	0.344	21.03
A3	0.000 ± 0.00	0.098 ± 0.01		33.79	33.10	35.17
A4	0.098 ± 0.01	0.000 ± 0.00	0.098 ± 0.01		0.344	21.03
A5	0.096 ± 0.01	0.001 ± 0.00	0.096 ± 0.01	0.001 ± 0.00		20.344
A6	0.102 ± 0.01	0.061 ± 0.01	0.102 ± 0.01	0.061 ± 0.01	0.059 ± 0.01	

Continued on next page

Table 2. Continued.

TSPY

Dis/DT	A1	A2	A3	A4	A5	A6
A1		2.32	26.51	30.23	27.67	23.25
A2	0.010 ± 0.00		26.97	30.46	27.90	25.58
A3	0.114 ± 0.01	0.116 ± 0.01		25.58	22.55	24.65
A4	0.130 ± 0.01	0.131 ± 0.01	0.110 ± 0.01		28.37	30.93
A5	0.119 ± 0.01	0.120 ± 0.01	0.097 ± 0.00	0.122 ± 0.01		29.30
A6	0.100 ± 0.01	0.110 ± 0.01	0.106 ± 0.01	0.133 ± 0.01	0.126 ± 0.01	

*BPY2

Dis/DT	A1	**A2**	A3	A4
A1		0	0.331	161.58
A2	0.000 ± 0.00		0.331	161.58
A3	0.001 ± 0.01	0.001 ± 0.00		161.92
A4	0.488 ± 0.01	0.488 ± 0.01	0.489 ± 0.01	

*DAZ

Dis/DT	**A1**	**A2**	A3
A1		0.165	0
A2	0.0005 ± 0.00		0.165
A3	0.000 ± 0.00	0.0005 ± 0.00	

VCY

Dis/DT	**A1**	**A2**
A1		0
A2	0.000	

ζ VCX

Dis/DT	**S**	**R**	**Q**
S		1.92	3.84
R	0.005 ± 0.00		5.76
Q	0.010 ± 0.00	0.015 ± 0.00	

Continued on next page

Table 2. Continued.

ζ RBMX

Dis/DT	**A1**	**A2**
A1		263.88
A2	0.380 ± 0.01	

HSFY

Dis/DT	**Y1**	**Y2**	A	B	C	D	E
Y1		0	41.81	41.81	69.54	69.54	123.64
Y2	0.000 ± 0.00		41.81	41.81	69.54	69.54	123.64
A	0.092 ± 0.00	0.092 ± 0.00		0	31.82	31.82	124.55
B	0.092 ± 0.00	0.092 ± 0.00	0.000 ± 0.00		31.82	31.82	124.55
C	0.153 ± 0.00	0.153 ± 0.00	0.070 ± 0.01	0.070 ± 0.01		0	153.18
D	0.153 ± 0.00	0.153 ± 0.00	0.070 ± 0.01	0.070 ± 0.01	0.000 ± 0.00		153.18
E	0.272 ± 0.00	0.272 ± 0.00	0.274 ± 0.01	0.274 ± 0.01	0.337 ± 0.02	0.337 ± 0.02	

ζ HSFX

Dis/DT	**A1**	**A2**
A1		4.16
A2	0.006 ± 0.00	

RBMY

Dis/DT	**A**	**B**	C	D	E	F	G	H	I	J	K
A		0	46.83	0	46.83	14.78	219.37	191.90	46.83	46.83	46.83
B	0.000 ± 0.00		46.83	0	46.83	15.03	219.37	191.90	46.83	46.83	46.83
C	0.133 ± 0.01	0.133 ± 0.01		46.83	0	44.37	216.90	194.37	0	0	0
D	0.000 ± 0.00	0.000 ± 0.00	0.133 ± 0.01		46.83	14.79	219.79	219.37	191.90	46.83	46.83
E	0.133 ± 0.01	0.133 ± 0.00	0.000 ± 0.00	0.133 ± 0.01		44.34	216.90	194.37	0	0	0
F	0.042 ± 0.01	0.042 ± 0.01	0.126 ± 0.00	0.040 ± 0.01	0.126 ± 0.01		216.90	196.48	44.37	44.37	44.37
G	0.623 ± 0.02	0.623 ± 0.02	0.616 ± 0.02	0.623 ± 0.02	0.616 ± 0.02	0.616 ± 0.020		221.13	216.90	216.90	216.90
H	0.545 ± 0.02	0.545 ± 0.02	0.552 ± 0.02	0.545 ± 0.02	0.552 ± 0.02	0.558 ± 0.020	0.628 ± 0.02		194.37	194.37	194.37
I	0.133 ± 0.01	0.133 ± 0.01	0.000 ± 0.00	0.133 ± 0.01	0.000 ± 0.00	0.126 ± 0.010	0.616 ± 0.02	0.552 ± 0.02		0	0
J	0.133 ± 0.01	0.133 ± 0.01	0.000 ± 0.00	0.133 ± 0.01	0.000 ± 0.00	0.126 ± 0.014	0.616 ± 0.02	0.552 ± 0.02	0.00 ± 0.00		0
K	0.133 ± 0.01	0.133 ± 0.01	0.000 ± 0.00	0.133 ± 0.01	0.000 ± 0.00	0.126 ± 0.010	0.616 ± 0.02	0.552 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	

Blue bold: functional copies; *: used average rate to estimate DT as no available orthologs found; ζ: Nachman et al. (2000) rate was used to estimate DT.

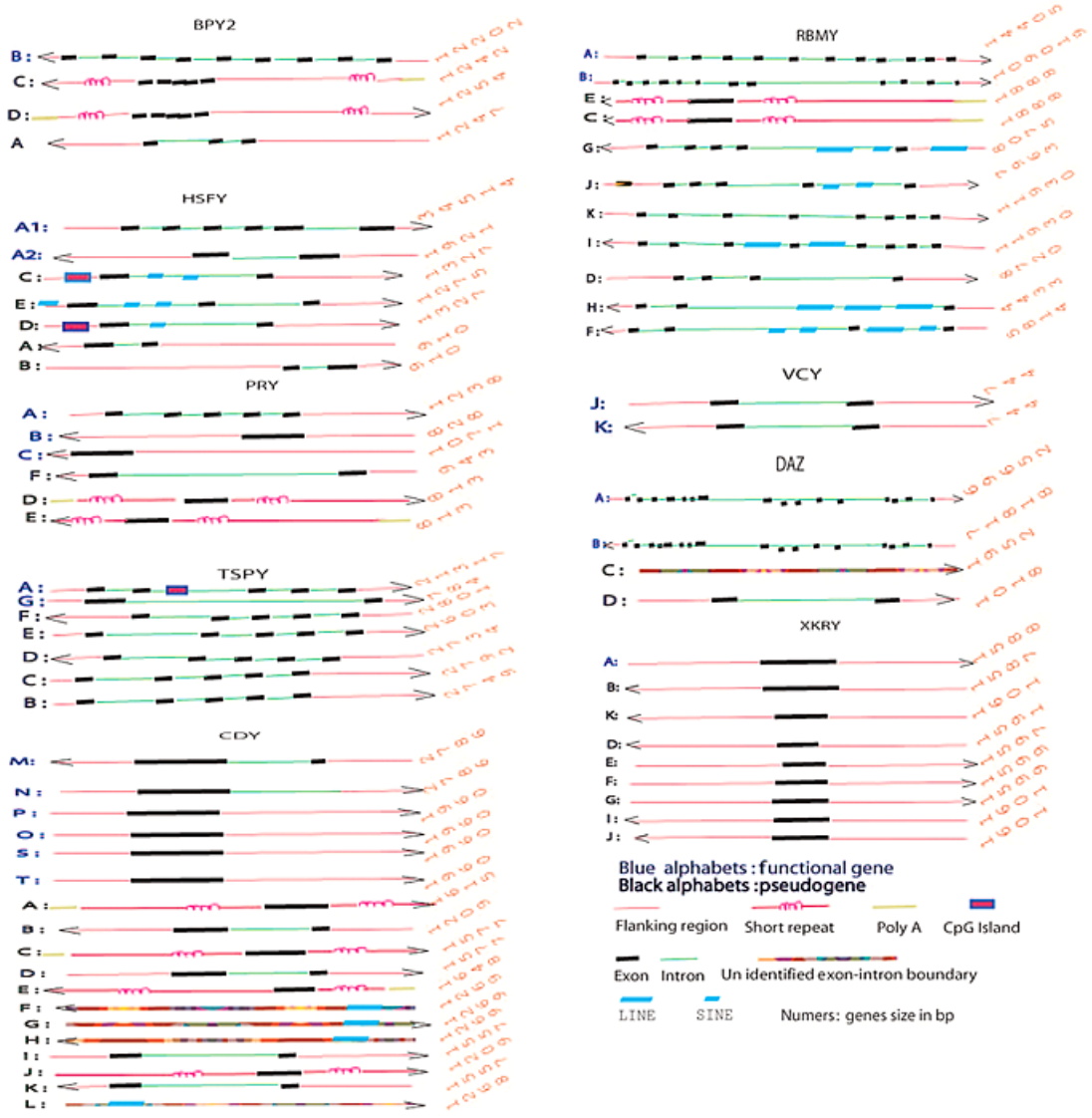

Figure 1. Exon-intron structure with respect to inferred mechanism.

long flanking regions, indicating region duplication. Based on this long identity and inverted orientation, this region can also be considered a new palindrome (about 15 to 16 MB). As VCX and VCY copies have different substitution rates for evolutionary reasons, we refrained from making estimations of when X and Y clusters are diverged.

CDY

According to the present study, CDY1 and CDY2 have a 6.4% difference in the non-coding region, and their functions are also different. CDY1 has four functional copies located on

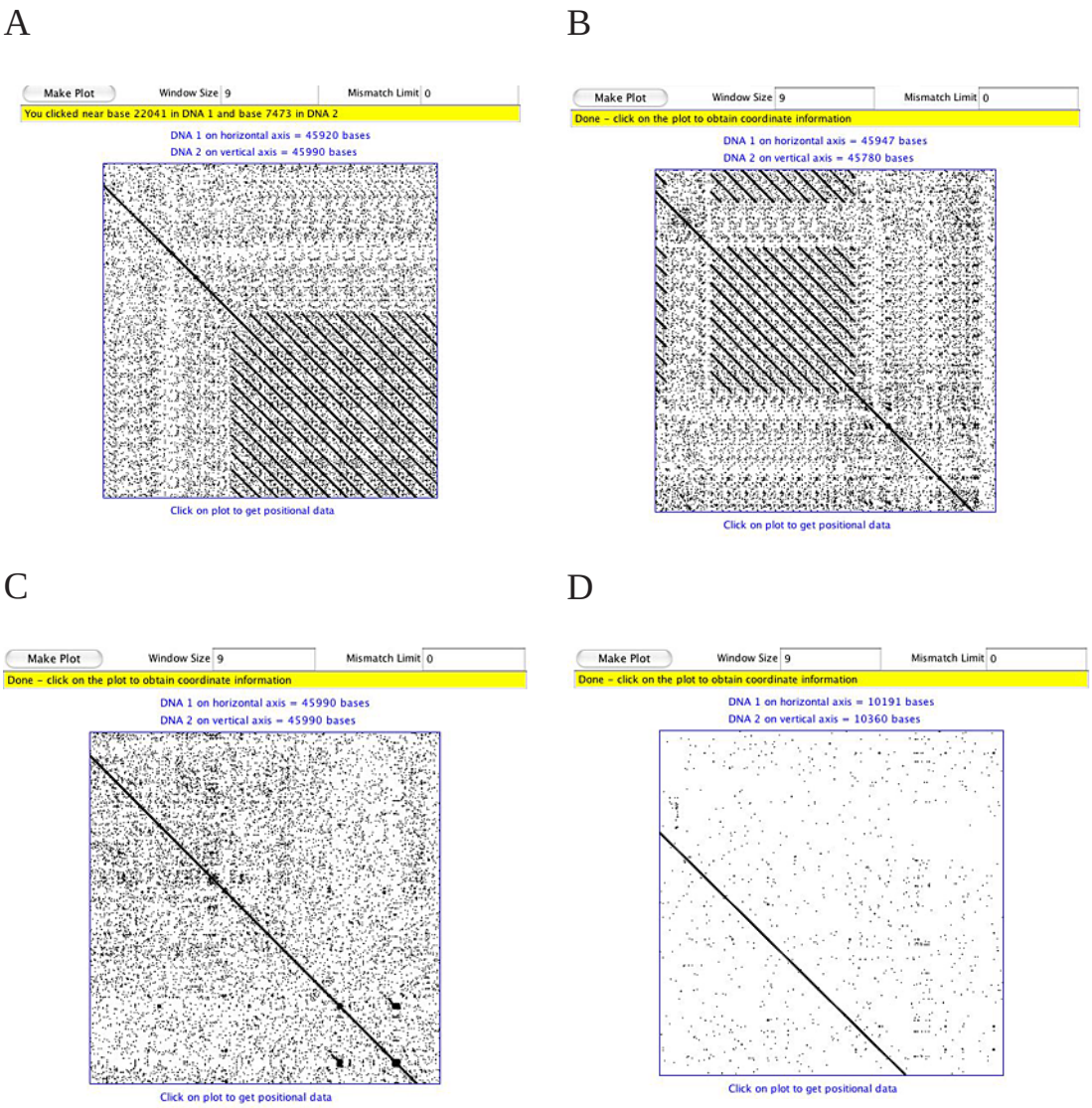

Figure 2. Dot plot analysis of questionable palindromes 1 and 3.

palindromes four and one. CDY2 has two copies and both are located on palindrome four. This finding differs from that of Lahn and Page (2001) who found CDY as one gene, with four functional copies and located on palindromes 5 and 1. Our CDY1 copies differ from the original copy only by non-coding differences, but their CDSs were 100% identical and they were located on palindromes 4 and 1. CDY2 has two 100% identical copies on palindrome 4. We retrieved 14 pseudogenes, whereas Lahn and Page (2001) found 27 pseudogenes. Saut et al. (2000) found two different CDY genes, which was also concluded in the present study. However, they did not indicate any pseudogenes. Based on the DT (Table 2), human Y paralog

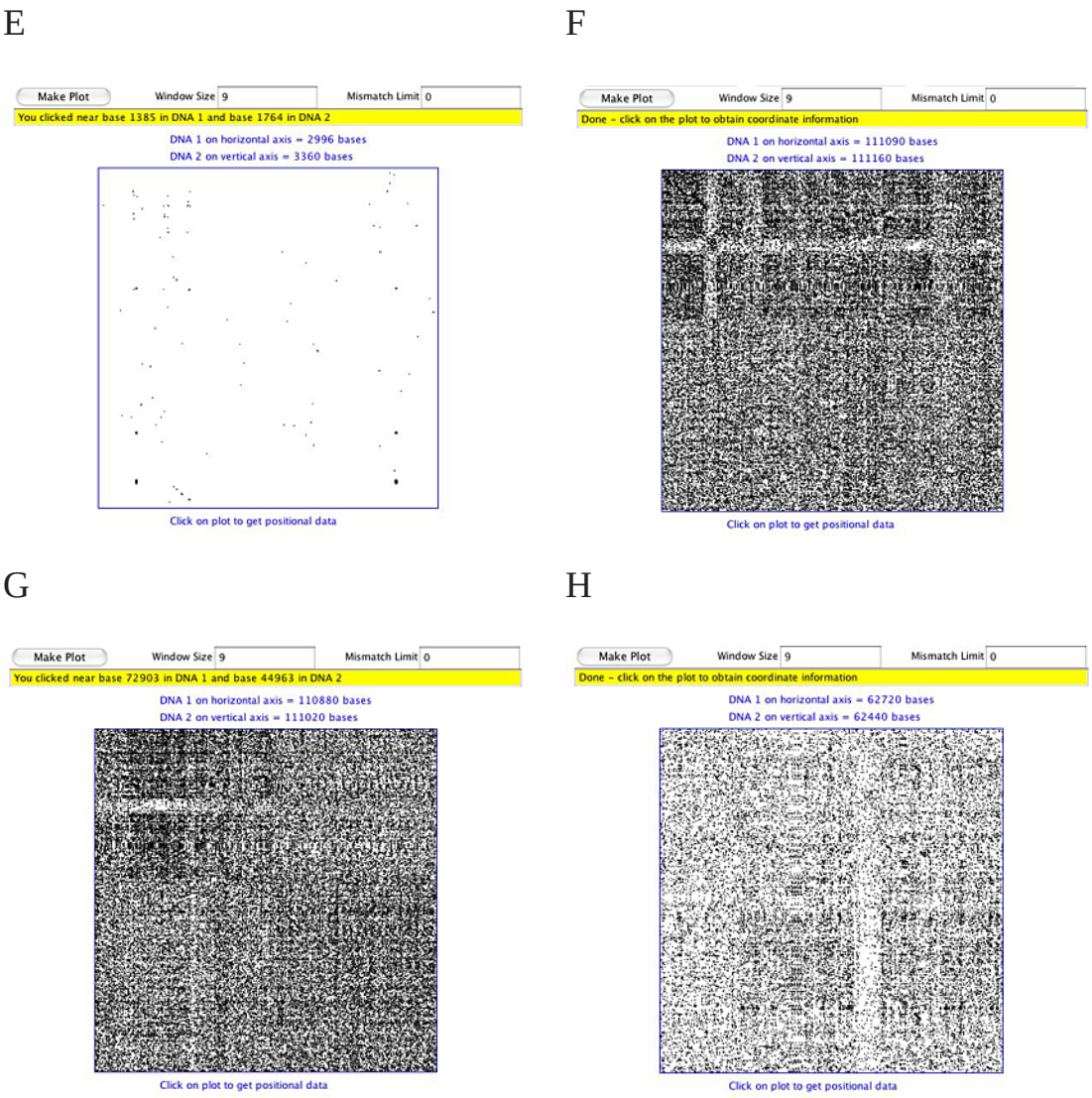

Figure 2. Continued.

duplication started during the Aves-Crocodilia divergence period. We found some functional domains in bacterial lineages, different from what was found by Lahn and Page (2001). They also indicated that two Y copies have no introns, but we found that CDY1 has an intron surrounded by two exons and that CDY2 has only one exon. We found two functional copies that diverged 4.6 mya. One X copy was found to be a pseudogene. It was observed that autosomal copies are multi-exonic, whereas Y copies are mono-exonic and di-exonic, indicating that the CDY gene translocated in Y by exon shuffling or by other unknown exon-reducing events. In addition, autosomal copies are ubiquitously expressed, whereas Y copies are only testis ex-

pressed. So, obviously, it can be said that CDY is attracted by Y, which is beneficial for male or harmful for female.

HSFY

Two Y copies were found outside of palindrome 4, while Skaletsky et al. (2003) found two copies on palindrome 4. They also stated that there were no pseudogenes for this gene, but we found five. HSFY1 is 100% identical with HSFY2 up to its 2nd exon. But they are not alternatively spliced. Four pseudogenes are adjacent, so they may have originated by a duplication event. One copy is isolated; it can be assumed to be a translocation event.

PRY

We found functional copies in the palindrome 1, IR region and the edge of palindrome 3. So the lack of complimentary copies in the respective palindromes, palindromic regions of palindromes 1 and 3, which was reported by Lahn and Page (2001), is questionable. We also found different numbers of functional and pseudogene copies.

BPY2

Two functional copies were found by Lahn and Page (2001), while we found them to be pseudogenes. They are located on the right arm of palindrome 1, and the functional copy is on the left arm of palindrome 1. Again continuously spreaded palindrome 1 is confusing.

DAZ

Reijo et al. (1995) and Chen and Li (2001) found one pair on palindrome 2 and another pair on palindrome 1. We found that DAZ4 and DAZ1 are functional genes and that they are oppositely directed. Two functional DAZ genes are 100% identical, and they are located on the left arm of palindrome 1, inversely, and two pseudogenes that we found are located on the right arm of palindrome 1. Orientation of both couples again raised the question of a contiguous region of palindrome 1.

TSPY

Arnemann et al. (1987) found 35 functional copies in the ampliconic region, with five pseudogenes. We found two functional copies near the IR3 region and five pseudogenes. Pseudogenes are discretely located, although distant copies have higher similarity than close neighbors. So, the mechanism of their origin requires further observations.

RBMY

We retrieved two functional copies (in palindrome 3 and in the IR region) and nine pseudogenes of RBMY in the human Y, different from what was found by Inglis et al. (1993). We also found two chimp functional copies but they are showing to be highly diverged from

human-chimp DT. So, we did not consider them as orthologs to estimate evolutionary rates. As we could not consider chimp copies as orthologs, we used the average rate to estimate the DT of human paralogs. As palindromes 1 and 3 seem to be confusing, we analyzed these regions by dot plot to determine similarity, showing that very small regions have long diagonals, and that long regions are not continuously palindromic (Figure 2).

Generally speaking, we found that MSY genes are evolving in a neutral manner in their non-coding regions. Based on all available homologous CDS copies from human, chimp, macaw, and the mouse it was found that except for human DAZ, all accessible lineages show either neutral evolution or accelerated synonymous substitution (Table 1). In the case of CDY, *Mus muculus* shows some positive selection, but due to statistical error we could not make specific conclusions. MSY genes have been found to have somewhat greater sequence differences than X and autosomals (Chen and Li, 2001). This can be explained as male-driven evolution (Miyata et al., 1987).

However, when we examined chimp orthologs, it was found that VCX and RBMX genes have greater differences than those of their Y homologs. This is because VCX copies are repeated by 10, 8, 6, and 2 nucleotides, along with increased GC content. The RBMX gene has 78 T repeats, 2 TTGC repeats and 3 GA repeats, along with 3 SNPs within its very short sequence. Differences in GC contents might also result in larger sequence differences. Palindromic regions were considered from the NCBI map, according to Lahn and Page (2001), which was specially designed for the MSY region (P1: 23.30 to 26.25 M, P2: 23.0 to 23.30 M, P3: 21.85 to 22.60 M, P4: 18.32 to 18.74 M, P5: 17.32 to 18.32 M, P6: 16.07 to 16.33 M, P7: 15.79 to 15.81 M, P8: 13.90 to 13.98 M). All of the genes that we examined (only functional copies) gave displaced positions (Figure 3) compared to the previous study made by Lahn and Page (2001). Some pseudogenes did not have introns, but their consonant functional copy/copies did, which can be explained by pseudogene processing caused by retro-transposition. But, as most annotations were found in mRNA, and some exons were split in the BLAST hit, we cannot yet classify them as pseudogenes. We could not classify XKRY, CDY, BPY2, PRY, TSPY, and RBMY genes as MSY. In other cases, some have either X (VCX, HSFX, CDX) or autosomal (CDYL, DAZL) homologs, or both. VCX and DAZL have greatest expression in the testis. Therefore, we classified VC and DAZ genes as male specific, but not MSY. The function of the HSFX copies is still unknown. So, we could not classify HSFY. In all cases, except for XKRY, pseudogene-pseudogene DT was less than that of functional-pseudogenes. Most pseudogenes may have originated from other pseudogenes, either through duplication or conversion, or by other unknown events. Positions and orientation of human paralogs are shown in Figure 4. Furthermore, functional-functional copies have less divergence. So, it can be presumed that functional genes produce functional genes and pseudogenes produce pseudogenes more frequently than functional to pseudogene production (Table 2). At almost all loci, we found that pseudogenes are reduced in length when compared to their functional copies (Figure 1). According to Varsity (http://www.varsity.co.uk/index.php?option=com_content&task=view&id=7324&Itemid=27), the Y chromosome will be extinct within about 125,000 years. So, how will males continue? In the rodent, the Mole Vole, there is no Y, but the male still exists. So, if homologous MSY copies are acquired by other chromosomes and these promote male functions, male's existence is possible. That is why we examined whether MSY genes have male-specific expression on their non-Y counterparts in humans. We also found that VCX copies are male specific, and no mammalian lineages have VCX homologs. This can be inferred as VCX

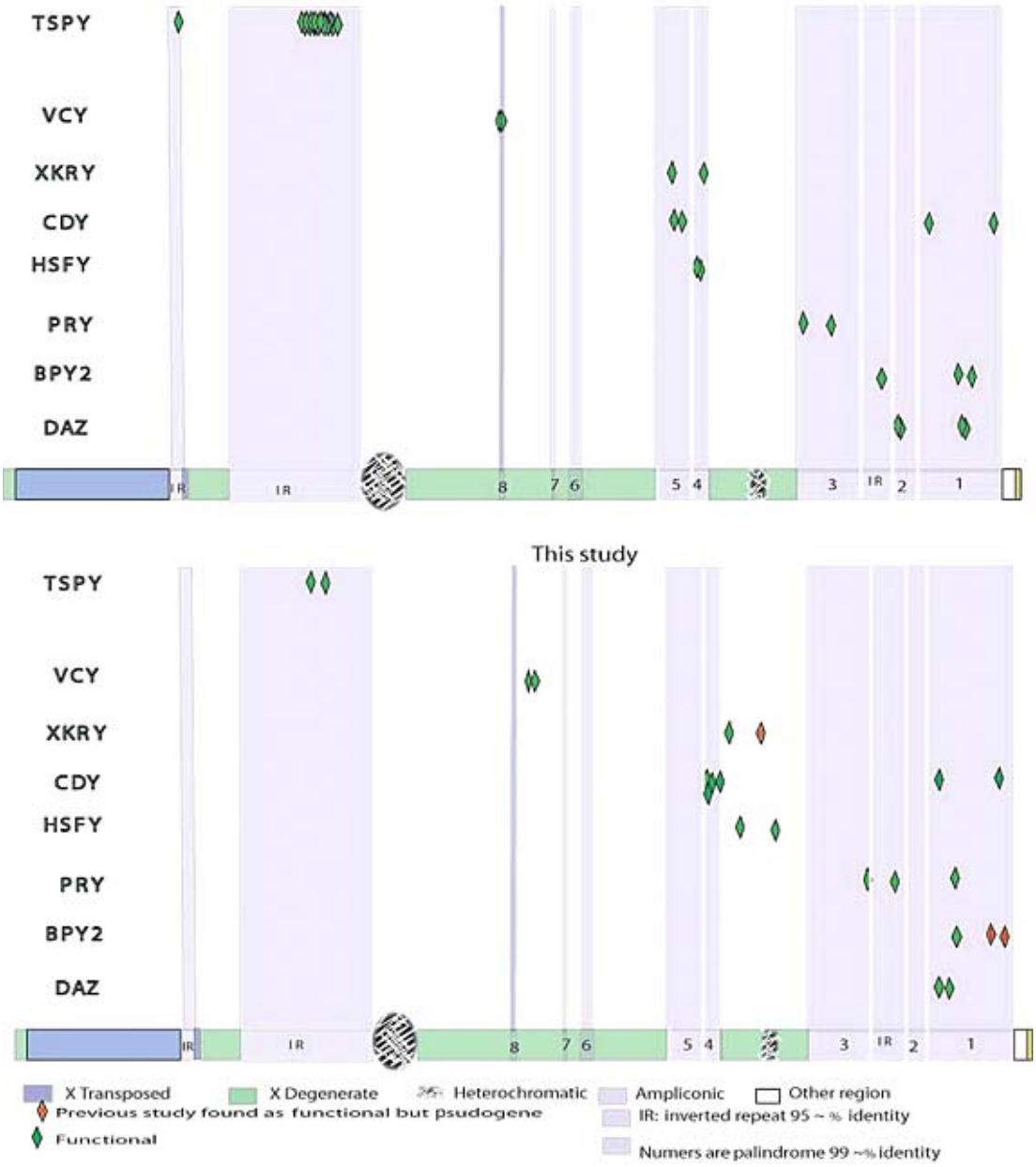

Figure 3. Locations of MSY paralogs and their comparison with previous study by Skalestsky et al., 2003.

copies are relocated from Y. To provide evidence, we estimated the DT of VCX and VCY copies by using the X rate from Nachman et al. (1998) and the Y rate from our study, along with the human autosomal rate from Chen and Li (2001). In all cases, the prediction was that VCX and VCY copies originated within the primate lineage. But as we did not find any VCX homologs in mammals, we propose that this copy translocated in the human X to maintain maleness in humans, in the face of Y extinction. We also checked the flanking regions of VCX and

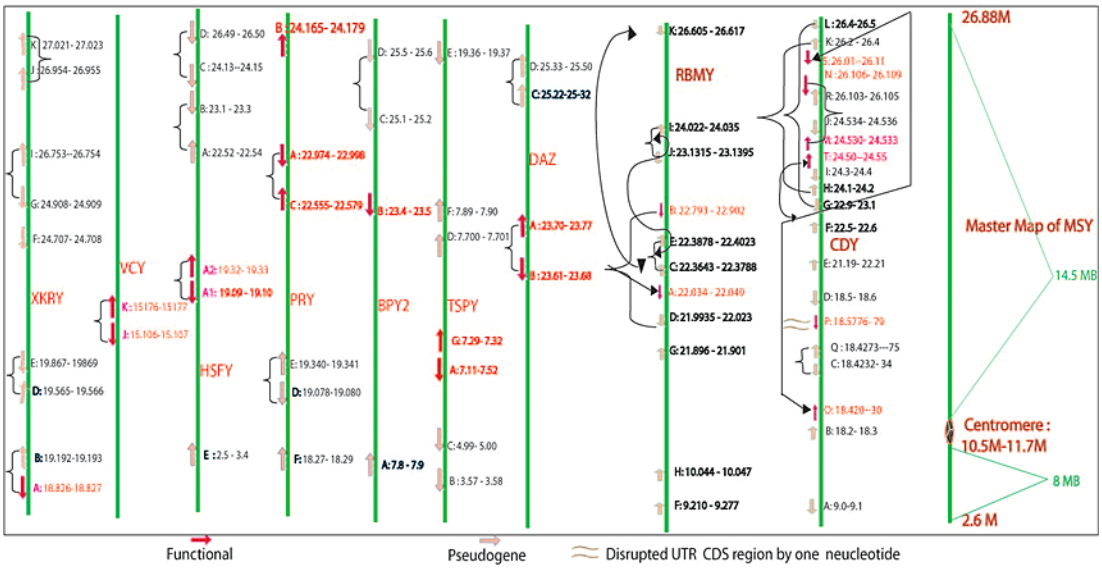

Figure 4. Locus map of human MSY paralogs.

VCY copies; but they were very dissimilar, as VCY flanks were found to be 100% similar, while corresponding X copies had no similarity. So, we cannot consider that translocation occurred through unequal crossing over. If only the gene was translocated, our inference is more logical. This assumption is a bit different from that of Fukumai et al. (2000). We think that duplication occurred in Y and then translocated onto X to maintain male existence. But, Fukumai et al. (2000) suggested that X, Y copies originated by a duplication event.

The function of HSFX gene copies is still unknown. But several copies are found on chromosomes 16, 17, 21, 4, 6, 7, and 8 (partial CDS), successively. These non-Y copies might have male-specific functions, supporting our argument. Some TSPY homologs, TSPYL1 to L6 are located on chromosomes 6, X, 20, 6, 8, and 2, respectively, in humans, with yet unknown expression. We also found some TSPYL testis-expressed homologs in the mouse, rat and chimp (partial CDS). In the chimp, L5 is also on chromosome 8, as in humans, in the Norway rat on chromosome 7, and in the mouse on chromosome 15. L1 is found on chromosome 10 in the mouse, whereas it is on chromosome 6 in humans. So, except for L5 in the chimpanzee, other homologs show a tendency towards chromosome changing. As all of the non-Y-linked TSPY members seem to be functional (partial CDS but no stop codon), at least some of them might have testis expression (expression yet unknown). DAZL, which is testis specific, is located in human on chromosome 3. Another DAZ homolog (LOC 460209), which was isolated from chimpanzee chromosome 3, is also testis specific. Chromosomes 10 and 17 in the mouse and, chromosomes 4 and 7 in the rat have testis-specific DAZ orthologs. So, non-Y chromosomes also have testis-expressing copies and have a propensity for position changing. So, it can be presumed that it is possible that other chromosomes will acquire male-specific properties after the extinction of Y. As little information is currently available, it is nearly impossible to make concrete conclusions about whether male-specific characters change with location. However, in the case of VC, DAZ and TSP, we strongly suggest that such changes occur.

Although the prediction of processed pseudogenes is difficult, we considered four processed pseudogenes in CDY, two in BPY2, two in PRY and two in RBMY (Figure 2), whereas Lahn and Page (2001) gave no evidence for processed pseudogenes for MSY genes. We also observed exon shuffling and length reduction in pseudogenes, in comparison with their functional copies. We speculate that the MSY region is still being reduced.

Apparently, exon shuffling, gene duplication, retro-transposition, and chromosomal rearrangements all have played important roles in the evolution of the MSY gene.

ACKNOWLEDGMENTS

Research supported by a graduate student scholarship from the Japanese Ministry of Education. The authors are very grateful to Drs. Naoyoka Takahata, Yoko Satta and Masakatsu Watanabe for their valuable guidance, comments and inspiration.

REFERENCES

Agulnik AII (1998). Evolution of the DAZ gene family suggests that Y-linked DAZ plays little, or a limited, role in spermatogenesis but underlines a recent African origin for human populations. *Hum. Mol. Genet.* 7: 1371-1377.

Altschul SF, Madden TL, Schaffer AA, Zhang J, et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.

Arnemann J, Epplen JT, Cooke HJ, Sauermann U, et al. (1987). A human Y-chromosomal DNA sequence expressed in testicular tissue. *Nucleic Acids Res.* 15: 8713-8724.

Benson G (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27: 573-580.

Burge C and Karlin S (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268: 78-94.

Chen F and Li W (2001). DNA variability and recombination rates at X-linked loci in humans. *Am. J. Hum. Genet.* 68: 444-456.

Delbridge ML, Lingenfelter PA, Disteche CM and Graves JA (1999). The candidate spermatogenesis gene RBMY has a homologue on the human X chromosome. *Nat. Genet.* 22: 223-224.

Delbridge ML, Longepied G, Depetris D, Mattei MG, et al. (2004). TSPY, the candidate gonadoblastoma gene on the human Y chromosome, has a widely expressed homologue on the X - implications for Y chromosome evolution. *Chromosome Res.* 12: 345-356.

Ebersberger I, Metzler D, Schwarz C and Paabo S (2002). Genome wide comparison of DNA sequences between humans and chimpanzees. *Am. J. Hum. Genet.* 70: 1490-1497.

Fisher RA (1931). The evolution of dominance. *Biol. Rev.* 6: 345-368.

Fukami M, Kirsch S, Schiller S, Richter A, et al. (2000). A member of a gene family on Xp22.3, VCX-A, is deleted in patients with X-linked nonspecific mental retardation. *Am. J. Hum. Genet.* 67: 563-573.

Howard JM (2002). "Mitochondrial Eve", "Y Chromosome Adam", testosterone, and human evolution. *Riv. Biol.* 95: 319-325.

Jegalian K and Lahn BT (2001). Why the Y is so weird? *Sci. Am.* 284: 56-61.

Karin J and Lahn BT (2001). Why Y is so Weird? *Sci. Am.*: 1561.

Kimura M (1968). Evolutionary rate at the molecular level. *Nature* 217: 624-626.

Kumar S and Hedges SB (1998). A molecular timescale for vertebrate evolution. *Nature* 392: 917-920.

Kumar S, Tamura K, Jakobsen IB and Nei M (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245.

Lahn BT and Page DC (1999a). Four evolutionary strata on the human X chromosome. *Science* 286: 964-967.

Lahn BT and Page DC (1999b). Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. *Nat. Genet.* 21: 429-433.

Lahn BT and Page DC (2001). Functional coherence of the human Y chromosome. *Science* 278: 675-680.

Maizel JV Jr and Lenk RP (1981). Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proc. Natl. Acad. Sci. USA* 78: 7665-7669.

Ma K, Inglis JD, Sharkey A, Bickmore WA, et al. (1993). A Y chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis. *Cell* 75: 1287-1295.
Metz EC and Palumbi SR (1996). Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein binding. *Mol. Biol. Evol.* 13: 397-406.
Miyata T, Hayashida H, Kuma K, Mitsuyasu K, et al. (1987). Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb. Symp. Quant. Biol*. 52: 863-867.
Ohno S (1967). Sex chromosome and sex-linked genes. Springer-Verlag, Berlin.
Pustell J and Kafatos FC (1982). A high speed, high capacity homology matrix: zooming through SV40 and polyoma. *Nucleic Acids Res.* 10: 4765-4782.
Reijo R, Lee TY, Salo P, Alagappan R, et al. (1995). Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nat. Genet*. 10: 383-395.
Reijo RA, Dorfman DM, Slee R, Renshaw AA, et al. (2000). DAZ family proteins exist throughout male germ cell development and transit from nucleus to cytoplasm at meiosis in humans and mice. *Biol. Reprod*. 63: 1490-1496.
Rozen S, Skaletsky H, Marszalek JD, Minx PJ, et al. (2003). Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* 423: 873-876.
Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. 4: 406-425.
Saut N, Terriou P, Navarro A, Levy N, et al. (2000). The human Y chromosome genes BPY2, CDY1 and DAZ are not essential for sustained fertility. *Mol. Hum. Reprod*. 6: 789-793.
Saxena R, Brown LG, Hawkins T, Alagappan RK, et al. (1996). The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nat. Genet*. 14: 292-299.
Saxena R, de Vries JW, Repping S, Alagappan RK, et al. (2000). Four DAZ genes in two clusters found in the AZFc region of the human Y chromosome. *Genomics* 67: 256-267.
Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423: 825-837.
Stewart CB and Disotell TR (1998). Primate evolution - in and out of Africa. *Curr. Biol*. 8: R582-R588.
Stouffs K, Lissens W, Van Landuyt L, Tournaye H, et al. (2001). Characterization of the genomic organization, localization and expression of four PRY genes (PRY1, PRY2, PRY3 and PRY4). *Mol. Hum. Reprod*. 7: 603-610.
Sykes B (1990). Y are doomed. Accessed http://www.varsity.co.uk/index.php?option=com_content&task=view&id=7324&Itemid=27.
Takahata N and Satta Y (1997). Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. *Proc. Natl. Acad. Sci. USA* 94: 4811-4815.
Thompson JD, Higgins DG and Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22: 4673-4680.
Ting CT, Tsaur SC, Wu ML and Wu CI (1998). A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282: 1501-1504.
Vogt PH, Edelmann A, Kirsch S, Henegariu O, et al. (1996). Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum. Mol. Genet*. 5: 933-943.
Williams G (2000). Genomics Research Wellcome Trust Genome Campus. Hinxton, Cambridge.
Wyckoff GJ, Wang W and Wu CI (2000). Rapid evolution of male reproductive genes in the descent of man. *Nature* 403: 304-309.
Zuckerlandl E and Pauling L (1965). Evolving genes and proteins. Academic Press, New York.