nature genetics

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

Comparative analysis of chimpanzee and human Y chromosomes unveils complex evolutionary pathway

Yoko Kuroki[1], Atsushi Toyoda[1], Hideki Noguchi[1,2], Todd D Taylor[1], Takehiko Itoh[3], Dae-Soo Kim[4], Dae-Won Kim[4,5], Sang-Haeng Choi[4], Il-Chul Kim[4], Han Ho Choi[4], Yong Sung Kim[4], Yoko Satta[6], Naruya Saitou[7], Tomoyuki Yamada[2], Shinichi Morishita[2], Masahira Hattori[1,8], Yoshiyuki Sakaki[1], Hong-Seog Park[4,5] & Asao Fujiyama[1,9]

The mammalian Y chromosome has unique characteristics compared with the autosomes or X chromosomes. Here we report the finished sequence of the chimpanzee Y chromosome (PTRY), including 271 kb of the Y-specific pseudoautosomal region 1 and 12.7 Mb of the male-specific region of the Y chromosome. Greater sequence divergence between the human Y chromosome (HSAY) and PTRY (1.78%) than between their respective whole genomes (1.23%) confirmed the accelerated evolutionary rate of the Y chromosome. Each of the 19 PTRY protein-coding genes analyzed had at least one nonsynonymous substitution, and 11 genes had higher nonsynonymous substitution rates than synonymous ones, suggesting relaxation of selective constraint, positive selection or both. We also identified lineage-specific changes, including deletion of a 200-kb fragment from the pericentromeric region of HSAY, expansion of young Alu families in HSAY and accumulation of young L1 elements and long terminal repeat retrotransposons in PTRY. Reconstruction of the common ancestral Y chromosome reflects the dynamic changes in our genomes in the 5–6 million years since speciation.

The completion of the sequencing of the human genome has demonstrated the importance of comparative studies for deciphering the information and functions written in our genomes[1,2]. As chimpanzees are evolutionarily the closest living species to humans, various studies[3–11] have aimed to understand the genetic basis of the similarities and differences between the two species. Sequence-based genome-wide studies have defined with precision the nucleotide difference between human and chimpanzee as 1.23% (refs. 4,5,8). Owing to this high similarity in the sequences, it was essential to produce high-quality sequence data in order to delineate species-specific changes accumulated after the species diverged[10].

The mammalian sex chromosomes are thought to have originated from a pair of autosomes that ultimately evolved into the X and Y chromosomes through (i) the incorporation of genes controlling sex determination and (ii) the acquisition of mechanisms to prevent recombination between the X and Y chromosomes and compensate for the difference in gene dosage of the X chromosome–encoded genes between male and female[12–17]. Thus, unlike other chromosomes, modern-day Y chromosomes have undergone numerous structural changes without correction through recombinatorial processes[17–21].

We have reported previously that chimpanzee BAC clones mapped poorly onto the human Y chromosome compared with other chromosomes[4]. Thus, we produced a high-quality sequence of the Y chromosome for the same male chimpanzee that we had previously analyzed for the chimpanzee chromosome 22 (PTR22; ref. 10). In this study, we sequenced approximately 12.7 Mb of PTRY, which corresponds mostly to the X-degenerate region of HSAY and partly to the chimpanzee ampliconic region (**Fig. 1**). Precise alignments produced from the PTRY sequences and the corresponding sequences of HSAY allowed us to carry out comparative analyses with HSAY[19], HSAX[20] and the only other finished human-chimpanzee chromosome complement of HSA21 and PTR22 (ref. 10).

RESULTS

Analysis of the PTRY sequence

We isolated clones from the whole-genome BAC library and Y-specific BAC and fosmid libraries originating from a single male chimpanzee (named Gon) that we had previously analyzed for PTR22 (ref. 10). The minimum tiling path, at the time of data freeze for this analysis, consisted of 96 minimally overlapping clones forming four clone contigs covering 905,338 bp of PTRYp and 10,074,253 bp of PTRYq, including 271 kb of the Y-specific pseudoautosomal region 1 (PAR1) sequence at the end of PTRYq (**Figs. 1** and **2**, **Table 1**, **Supplementary Table 1** and **Supplementary Methods** online). The accuracy of the

[1]RIKEN Genomic Sciences Center, Yokohama 230-0045, Japan. [2]Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa 277-0882, Japan. [3]Mitsubishi Research Institute, Tokyo 100-8141, Japan. [4]Genome Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-333, Korea. [5]University of Science and Technology, Daejeon 305-333, Korea. [6]The Graduate University for Advanced Studies, Hayama 240-0193, Japan. [7]National Institute of Genetics, Mishima 411-8540, Japan. [8]Kitasato University, Sagamihara 228-8555, Japan. [9]National Institute of Informatics, Tokyo 101-8430, Japan. Correspondence should be addressed to H.-S.P. (hspark@kribb.re.kr) or A.F. (afujiyam@nii.ac.jp).

Received 18 July 2005; accepted 30 November 2005; published online 1 January 2006; corrected after print 29 January 2006 (details online); doi:10.1038/ng1729

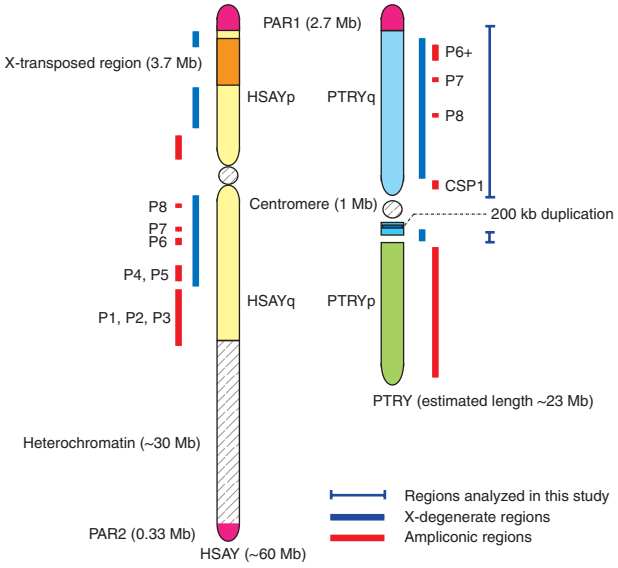

Figure 1 Schematic view of the sequenced area and major landmarks on chimpanzee chromosome Y. Although in the text we followed a previously reported method[21] for the designation of the p- and q-arms of HSAY and PTRY, in this figure, for simplicity, PTRY is inverted, and the PAR1 regions of each chromosome are aligned. From FISH imaging and from fluorescence values obtained from flow karyotyping, we estimated the size of PTRY as 23 Mb (data not shown). P1–P8, P6+ and CSP1 indicate the positions of the palindromic structures.

sequence was estimated at 99.9998%. One of two internal clone gaps (between PTB-143L19 and PTB-101K22) was subsequently closed after the data freeze. The remaining gap, sized as 15 kb by fiber–fluorescent *in situ* hybridization (FISH), lies within a palindromic structure on PTRYq (between PTFY-072J06 and PTFY-037C19). The PTRYq sequence now covers 10,243,335 bp including additional sequences from the chimpanzee ampliconic region.

Owing to the limited availability of chimpanzee cDNA sequences, the positions of the chimpanzee genes were inferred from the manually annotated HSAY genes and were mapped onto PTRY using the HSAY-PTRY sequence alignments (9,006,686 bp in total; see Methods). Another gene map, including annotation inferred from HSAY at the National Center for Biotechnology Information (NCBI), is provided separately (**Supplementary Fig. 1** online). Altogether, we identified 32 genes (19 protein-coding genes and 13 non–protein-encoding RNAs) and 59 pseudogenes in the analyzed regions of PTRY (**Supplementary Table 2** online). All possible PTRY transcripts in these regions corresponding to annotated HSAY pseudogenes were also identified as pseudogenes (that is, their ORFs were disrupted). We mapped 19 protein-coding genes on our sequence, whereas 20 coding genes were identified on the corresponding regions of HSAY (**Table 2**). This difference comes from the lack of a complete *CD24L4* gene sequence in PTRY. Using comparative PCR analysis, we further

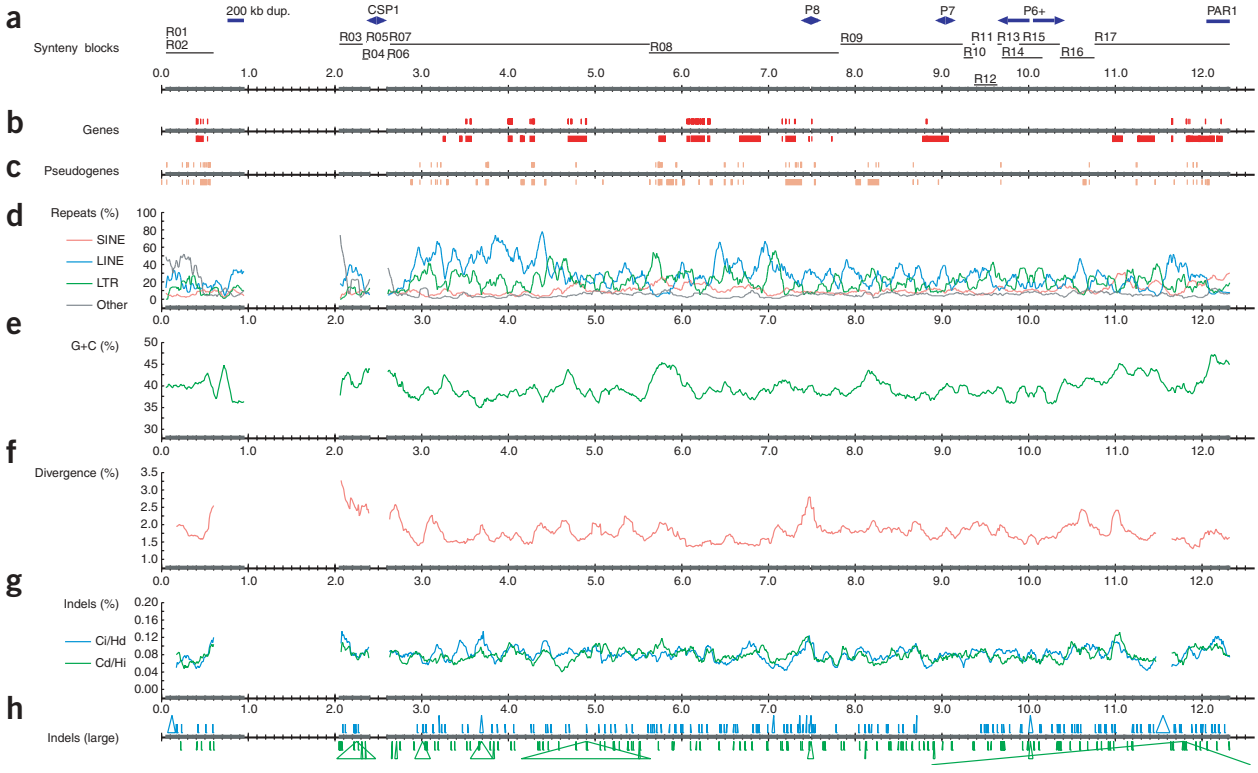

Figure 2 Features of the PTRY sequence. (**a**) Alignment blocks with human chromosome Y. (**b**) Mapped genes according to the VEGA HSAY annotation. Genes on the plus and minus strands are located above and below the axis, respectively. (**c**) Pseudogenes according to the VEGA HSAY annotation. (**d**) Interspersed repetitive sequence content. Window size: 100 kb. (**e**) G+C content. Window size: 100 kb. (**f**) Divergence rate with human chromosome Y. Window size: 100 kb. (**g**) Indel (<10 bp) rate with human chromosome Y. Window size of 100 kb. Ci/Hd indicates chimpanzee insertion or human deletion. Cd/Hi indicates chimpanzee deletion or human insertion. (**h**) Large indels (>250 bp) are indicated by rectangles. The length of the base of each triangle indicates indel size.

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

Table 1 Statistics on HSAY[a] and PTRY

		HSAY				PTRY			
		All		Aligned regions		All		Aligned regions	
Size (bp)[b]		24,871,691		9,678,597		10,950,101		10,197,059	
G+C content (%)		39.85		39.48		39.58		39.61	
Repeats		Number	Length (bp)	Number	Length (bp)	Number	Length (bp)	Number	Length (bp)
SINEs	All	10,089	2,484,596	3,965	993,015	4,327	1,083,963	4,083	1,023,286
	AluY	1,288	366,989	457	131,117	534	153,863	490	141,103
	AluYa5,AluYa8,AluYb8,AluYb9	97	29,022	46	13,619	2	526	2	526
LINEs	All	6,332	6,280,931	2,519	2,730,875	2,769	2,986,580	2,622	2,820,027
	L1HS, L1PA2	115	303,854	39	123,879	59	175,985	57	170,953
LTRs		4,701	4,393,986	2,150	1,858,604	2,428	2,103,312	2,307	2,007,340
DNAs		1,487	407,677	759	199,472	821	217,370	781	204,604
RNA		55	5,424	17	1,892	17	1,857	15	1,707
Satellites		1,161	1,110,631	170	188,879	309	270,957	200	197,966
Others		15	17,013	7	7,671	9	9,196	7	8,725
Simple repeats/low complexity regions		7,842	660,735	2,899	229,318	3,268	246,199	3,037	232,178
Total		31,682	15,360,993	12,486	6,209,726	13,948	6,919,434	13,052	6,495,833
(%)			61.76		64.16		63.19		63.70

[a]HSAY data was taken from ref. 19. Both the whole and aligned sequences were analyzed using the same parameters (see Methods). [b]Ns are not included.

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

confirmed that *CD24L4* was inserted in the human lineage after the speciation of human and chimpanzee (**Supplementary Fig. 2** and **Supplementary Methods** online).

Many genes on the Y chromosome are reported to have evolved rapidly and to have undergone positive selection[22]. We tested this hypothesis for the 19 PTRY protein-coding genes by comparison of nonsynonymous and synonymous substitutions between HSAY and PTRY (**Table 2**). All genes analyzed in this study were found to have at least one nonsynonymous substitution in their coding regions with the exception of *GYG2* isoform 2, which has none (**Supplementary Table 3** online).

Comparative analysis of the PTRY sequence

We first constructed a whole base-to-base sequence alignment between PTRY and HSAY (see Methods) to compare structural differences. The overall substitution level between PTRY and HSAY is 1.78% (excluding insertions or deletions (indels)). This value decreases to 1.66% when repetitive sequences are excluded. In contrast, the values are 1.83% and 1.91% in L1 long interspersed nuclear elements (LINEs) and Alu short interspersed nuclear elements (SINEs), respectively (**Table 1** and **Supplementary Table 4** online). In all cases, the nucleotide divergence between HSAY and PTRY exceeds that of HSA21 and PTR22 (1.44%, excluding indels; 1.38%, L1/LINEs; 1.81%, Alu/SINEs)[10]. This higher substitution rate between the Y chromosomes confirms the accelerated evolutionary rate of the Y chromosome, as previously reported[23–25]. This is presumably a consequence of the large number of cell divisions in the male germ line (**Supplementary Note** online).

Notably, the correlation coefficient between base substitutions and indels (about 17,400 sites; most of them are short, as described below) is 0.40, which is considerably higher than that between substitutions and G+C content (0.26; **Supplementary Fig. 3** online). Because a molecular clock exists for substitutions and indels[26], this correlation might indicate that the evolutionary or mutation rate for both substitutions and indels varies from region to region of the non-recombining Y chromosome. To examine this hypothesis, we compared the observed distribution of nucleotide differences (within 10-kb windows) and expected differences under the Poisson distribution (**Supplementary Fig. 3**). The results show that the observed distribution is considerably more varied than the Poisson expectation. This discrepancy suggests that the evolutionary or mutation rate varies depending on the region of the Y chromosome sequence, although there is essentially no recombination. If this finding applies to other chromosomes, it will have significant implications on many evolutionary analyses, such as correlation of substitution and recombination rates and estimation of ancestral population size.

As the PTRY sequence analyzed in this study originated from a single male animal, Gon[10] (PTB library), we compared intraspecies nucleotide diversity using the sequences in the public database[51] (**Supplementary Table 5** online). The result shows relatively lower nucleotide diversity in a pairwise comparison with two other animals, Clint (CHORI-251 library) and Donald (RP-43 library), 0.0422 ± 0.0006% (s.e.m.) and 0.0343 ± 0.0012%, respectively. This extent of diversity is lower than the result (0.067%) from a previous study[27]. The most likely explanation to this difference is that the current analysis is most likely from one subspecies, *Pan troglodytes verus* (*P. t. verus*), whereas the former result included the two other subspecies (**Supplementary Note**).

From the intra- and interspecies comparisons, we detected the presence of several blocks around the boundary of PAR1 and the male-specific region of the Y chromosome (MSY), in which nucleotide differences between HSAY and PTRY were significantly higher than in other regions (**Fig. 3**). It has been reported recently[28] that there is a gradient of silent substitution rates in the human PAR1 region. It would be interesting if such an L1-rich region were to function as an insulator that prevented homologous recombination between the X and Y chromosomes proceeding beyond the PAR1-MSY boundary (**Supplementary Note**).

Identification of lineage-specific events

Indels are one of the major types of events that have caused structural differences between the human and chimpanzee genomes[1,3,8–10,29–31]. In this study, we compared PTRY and HSAY using sequence

Table 2 Comparison of HSAY and PTRY genes located in the male-specific regions

				Human		Chimp		Position					
Gene	Isoform	CDS[c]	Variant	Length (nt)	Length (aa)	Length (nt)	Length (aa)	Start	End	Strand	*Ks*	*Ka*	*Ka/ks*
RPS4Y2[a]	1	C	1	792	263			3255247	3280582	–	0.00401	0.01256	3.1322
EIF1AY	1	C	1,3	435	144			3443800	3460706	–	0.00705	0.00348	0.4936
	2	C	2	384	127			3443275	3460706	–	0.00790	0.00397	0.5025
JARID1D	1	C	1,3	4620	1539	4608	1535	4005775	4044929	+	0.01272	0.00356	0.2799
	2	C	2	4449	1482	4437	1478	4005937	4044924	+	0.01192	0.00370	0.3104
	3	P	4	4429	1475	4417	1471	4005775	4044270	+	0.01192	0.00344	0.2886
	4	P	5	880	292	868	288	4041398	4043150	+	0.01308	0.00471	0.3601
	5	P	6	1276	424			4005987	4029385	+	0.02328	0.00419	0.1800
CYorf15B	1	C	1,2	546	181	144	47	4149214	4162932	–	0.00000	0.00000	inf
CYorf15A	2	C	2	396	131			4164948	4188018	–	0.01478	0.02197	1.4865
CD24L4[b]	1	C	1	243	80	–	–						
AC002992.5	1	C	1,2	192	63			5747072	5813062	–	0.02818	0.01691	0.6001
(*GYG2*-like)	2	c	3	195	64			5772349	5812998	–	0.02818	0.00000	0.0000
	3	c	4	360	119			5794893	5813062	–	0.04083	0.01202	0.2944
	4	c	5	291	96			5798765	5812978	–	0.02549	0.00540	0.2118
USP9Y	1	c	1,2	7668	2555	5601	1866	6110319	6264476	+	0.00822	0.01151	1.4002
	2	p	3	713	236			6250660	6263247	+	0.00832	0.00768	0.9231
DDX3Y	1	c	1,3,4	1983	660			6308114	6322604	+	0.00608	0.00764	1.2566
	2	p	2	673	223			6308124	6317902	+	0.00000	0.01377	inf
	3	p	5	750	249	753	250	6309742	6318698	+	0.00421	0.01239	2.9430
UTY	1	c	1,2	4044	1347			6675101	6909521	–	0.01040	0.00885	0.8510
	2	c	3	3723	1240			6724040	6909518	–	0.01131	0.00805	0.7118
	3	c	4	3240	1079			6749116	6909518	–	0.01371	0.00989	0.7214
	4	c	5	624	207			6822120	6908834	–	0.01440	0.01580	1.0972
TMSB4Y	1	c	1,2	135	44			7163239	7165658	+	0.06752	0.02001	0.2964
VCY1B	1	c	1,2	378	125	375	124	7473929	7474912	–	0.01784	0.05604	3.1413
VCY	1	c	1	378	125	375	124	7506529	7507516	+	0.01784	0.05604	3.1413
NLGN4Y	1	c	1	1947	648			8783257	9077185	–	0.03518	0.00830	0.2359
	2	c	2,4	2451	816			8785183	9076187	–	0.03373	0.00894	0.2650
	3	c	3	2622	873			8785183	9076187	–	0.03285	0.00950	0.2892
	4	p	5	710	235	776	257	8893729	8986249	–	0.02625	0.01030	0.3924
PRKY	1	c	1,2	834	277			10975071	11083898	–	0.02292	0.03046	1.3290
BL1Y	1	c	1,2,3,4	1569	522	414	137	11262846	11452127	–	0.05368	0.01320	0.2459
AMELY	1	c	1,3	579	192			11651925	11660023	+	0.00545	0.00871	1.5982
	2	c	2	621	206			11653343	11659868	+	0.00507	0.00808	1.5937
ZFY	1	c	1,2	2406	801			11847686	11897737	–	0.00815	0.00356	0.4368
	2	p	3	274	90			11868852	11897513	–	0.00000	0.01070	inf
RPS4Y	1	c	1,2	792	263			11964546	11989729	–	0.02016	0.00155	0.0769
	2	p	3	787	261			11964640	11989295	–	0.02024	0.00157	0.0776
SRY	1	c	1,2	615	204			12047711	12048555	+	0.00504	0.01568	3.1111

inf, infinite. [a]*RPS4Y2*, ribosomal protein S4, Y-linked 2, is defined as a pseudogene in NCBI Entrez Gene. [b]*CD24L4* is a member of the *CD24* family (NCBI Entrez Gene). The other four CD24 family members are found in both human (CD24, 6q21; CD24L1, 1p36; CD24L2, 15q21-q22; and CD24L3, chr20; NCBI Entrez Gene) and chimpanzee (exact positions unclear). [c]c: complete coding sequence (CDS); p: partial coding sequence.

alignments (excluding the ampliconic region), emphasizing both the difference between the two Y chromosomes and the difference between the autosomes and the Y chromosomes. For this purpose, we extracted all the gaps from the sequence alignment as potential indel sites (Methods). Most of the indels were found to be small and to account for 88% (1–10 bp) or 98% (1–100 bp) of the total gaps. The largest event identified was a 3.7-Mb insertion into HSAY. This is known as the X-transposed region and is thought to have occurred after the speciation of humans and chimpanzees[19,20].

We and others have previously reported that a copy of a 200-kb paralogous fragment is missing from the pericentromeric region of PTR22q[10,32,33]. This 200-kb segmental duplication originated from the HSA21q22.1 region through transposition to the pericentromeric regions of HSA18p11, HSA13q11, HSA21q11.1 and the ancestral chromosome(s) of HSA2q and the corresponding regions in the great apes[32,33]. It has been reported that the 200-kb fragment is not found on HSAY and PTRY; however, through sequence comparison, we identified another copy of this fragment at the pericentromeric region of PTRYp (**Figs. 1** and **2**). In contrast, the corresponding sequence was not found on HSAY. A neighbor-joining tree showed essentially the same tree topology as previous reports except for an acceleration in the substitution rate (that is, a longer branch) of the PTRY copy (**Supplementary Fig. 4** online). As the copy on PTRY shares a greater abundance of base substitutions with the copies on PTR22 and HSA21q22.1 than with those of other pericentromeric copies (98 versus 2, respectively), it is most likely that the PTRY copy

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

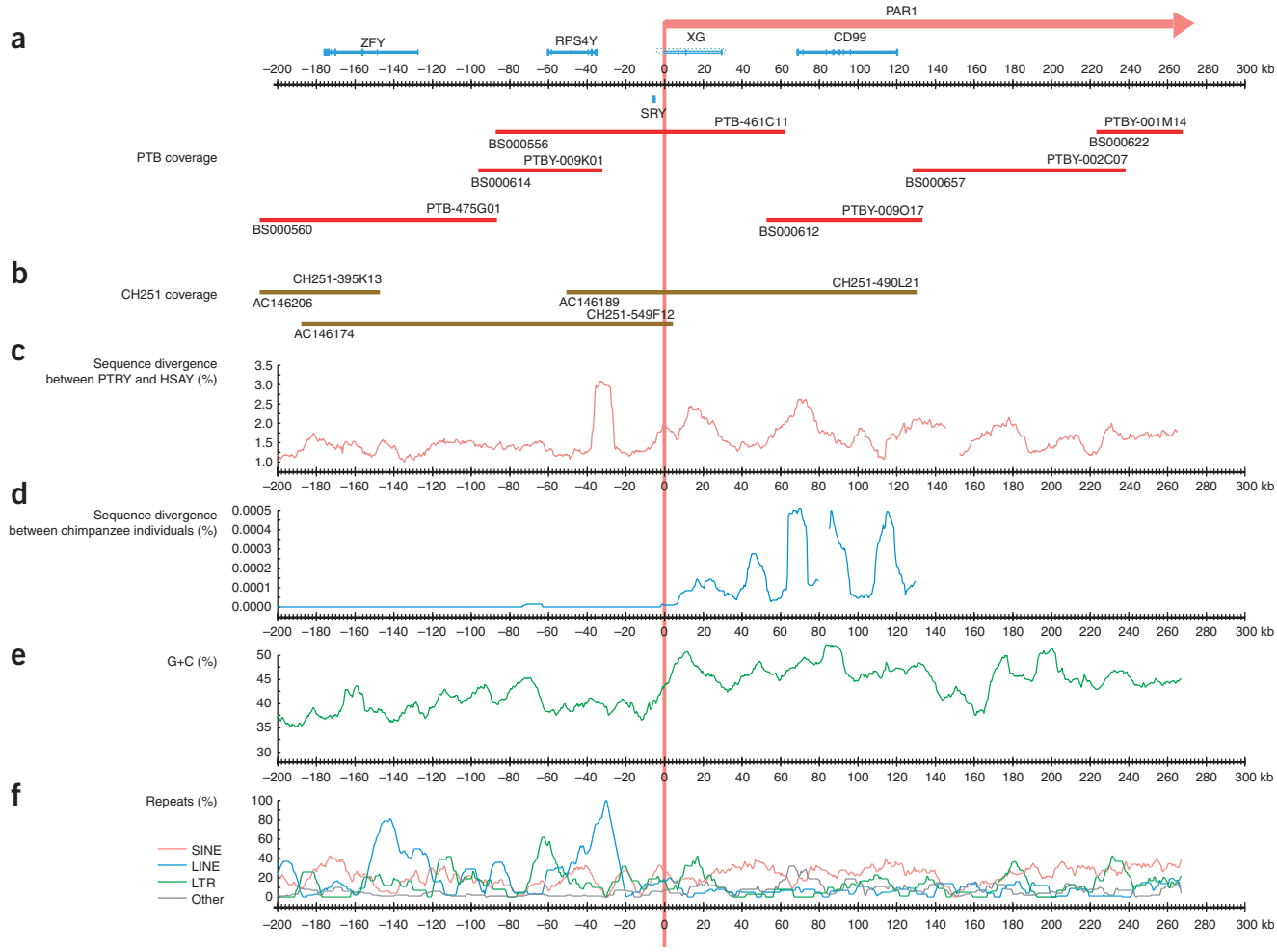

Figure 3 Comparison of the PAR1 and MSY regions of PTRY. The boundary between PAR1 and MSY is indicated by a red vertical line. The region to the right of the boundary is PAR1. (**a**) PTB clone coverage. (**b**) CH251 clone coverage. (**c**) Divergence between PTRY (PTB clones) and HSAY. Window size: 10 kb. (**d**) Divergence between PTB clones and CH251 clones. Window size: 10 kb. (**e**) G+C content. Window size: 10 kb. (**f**) Interspersed repetitive sequence content. Window size: 10 kb.

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

was duplicated from the ancestral PTR22/HSA21q22.1 copy. The presence of the 200-kb fragment on PTRY was further confirmed by FISH analysis (**Supplementary Fig. 4**). Through comprehensive sequence comparison, we concluded that it is most likely that duplication of the HSA21q22.1/PTR22 copy occurred on the ancestral Y chromosome before the speciation of humans and chimpanzees and that the copy on HSAY was lost thereafter (**Supplementary Note**).

LINEs and long terminal repeat (LTR) retrotransposons comprise the major component of repetitive sequences on both PTRY and HSAY[19] (**Table 1**). In an extreme case, we found that the LTR12B/D retrotransposon is repeated 36 times within a 260-kb insertion in HSAY at the site corresponding to the 3.7 Mb position of the PTRY sequence (**Fig. 2**). Furthermore, we found at least 14 integration sites, ranging from 4.5 kb to 9.0 kb in size, of chimpanzee endogenous retroviruses on PTRY[34–36]. We were unable to find such sequences in the human genome (cutoff value of identity was 90%). Twelve of the 14 integration sites share sequence identity greater than 95% and are likely to belong to the same viral group (CERV1 and CERV2) previously reported on chimpanzee chromosome 7. The remaining two copies, located at internal ends of the P8 palindrome arms, were found to be completely identical, suggesting conversion of sequences between the palindrome arms.

By comparing HSA21 and PTR22, we have shown previously that the human genome experienced almost ten times more Alu insertions than the chimpanzee genome, and that the mode of Alu insertion is distinctly different between the two species; that is, the observed ratio of the insertion frequencies between AluYa to AluYb is approximately 1:2 (ref. 10). To explore whether or not such phenomena are also observed on the Y chromosomes, we selected 62 young Alu elements from the sequence alignments for further analysis (Methods and **Supplementary Table 6** online). In contrast to HSA21, the AluYa5 subfamily forms a larger cluster than the AluYb subfamily on HSAY (**Fig. 4a**). Thus, it seems likely that the insertions in autosomes have a higher chance of removal, possibly through recombination, although we cannot rule out the possibility that the insertion frequency was higher in the Y chromosomes. It is also notable that HSAY contains a larger number of young AluY subfamilies than HSA21 (at approximately a 2:1 ratio) although the total content of SINEs is the same (approximately 10%) among these chromosomes[10] (also reported in this study), suggesting that roughly half of the young Alu elements have been removed from HSA21. In contrast, the number of AluYa

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

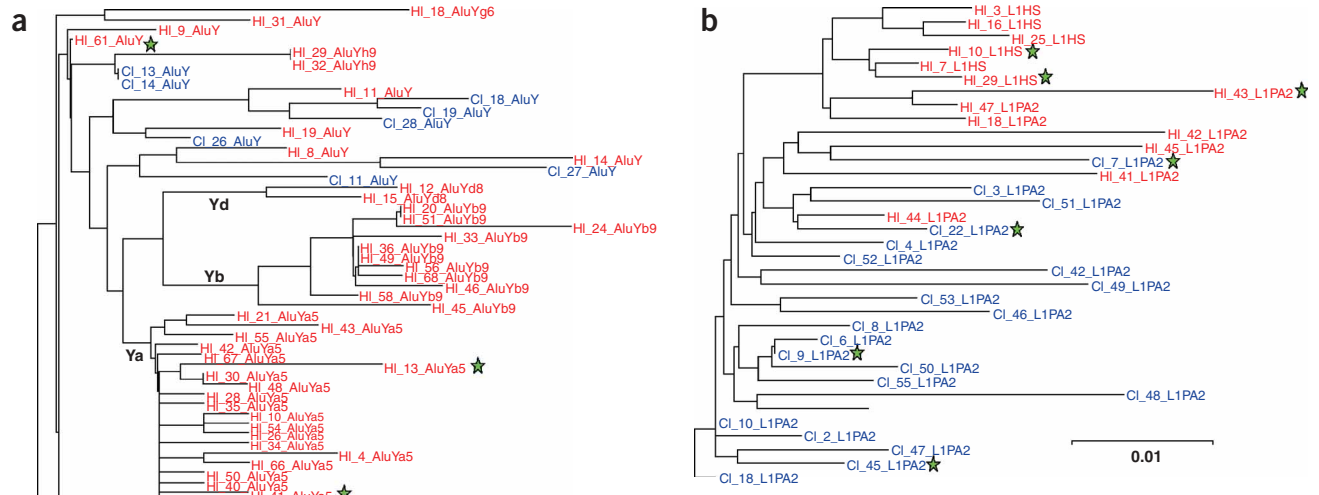

Figure 4 Evolutionary relationship between inserted repetitive elements. (**a**,**b**) Neighbor-joining tree of (**a**) young Alu families and (**b**) young L1 families inserted into PTRY (blue) and HSAY (red) after speciation. The scale indicates an evolutionary distance of 0.01. Sequences with green stars indicate confirmation of lineage specificity (either insertion into HSAY or PTRY) through comparative PCR analysis (experimental details can be found in **Supplementary Methods**). Ya, Yb and Yd designate clusters of specific Alu subfamilies in human.

and AluYb subfamilies is markedly lower in PTRY (two in PTRY versus 46 in HSAY; **Table 1**), confirming previous reports on expansion of these specific Alu subfamilies in the human lineage[10,31,37,38].

The human and chimpanzee Y chromosomes are enriched for LINEs (25% and 27% of the total sequence, respectively (ref. 19 and this study)), as is the X chromosome (29%; ref. 20), whereas LINEs make up only 17% of the total sequence in HSA21 and PTR22 (ref. 10). Although most of these LINEs must have been inherited from a common ancestor, we confirmed that the content of young L1 elements (L1PA2 and L1HS) within the entire LINE sequences was higher in both HSAY and PTRY (4.5% and 6.1%, respectively, in contrast to 1.4% for both HSA21 and PTR22; refs. 1,10). However, we found that the L1 elements integrated into PTRY were mostly from the L1PA2 subfamily, whereas they were mostly from the L1HS subfamily in HSAY, suggesting that there was a period of time when L1PA2 actively expanded in the chimpanzee lineage (**Fig. 4b**). We also carried out comparative PCR analysis to identify the lineage specificity of these events (**Fig. 4** and **Supplementary Methods**). As we reported previously[10], most of the insertions were found to be related to repetitive sequences, and, in contrast, most of the deletions were not (**Supplementary Note**).

Global structural changes

As most of the PTRY sequence analyzed in this study corresponds to the X-degenerate region, and the structure of the X chromosome is strongly conserved among mammalian species[20], evolution of the region could be described as a process of structural changes from the corresponding regions of the current human X chromosome[20]. Consequently, we inferred the evolutionary process through computational calculation (see Methods) and reconstructed the common ancestral Y chromosome shared by human and chimpanzee (**Fig. 5** and **Supplementary Fig. 5** online).

A majority of the structural differences between HSAY and PTRY can be explained by single inversion events that occurred in the ancestral Y chromosomes of each lineage. In the case of HSAY, we speculate that the inversion event occurred around the *NLGN4Y-STS* region (**Fig. 5c**), which concurrently split synteny blocks 5 and 7 into 5-h1/h2 and 7-h1/h2, respectively. This model is in agreement with a previous report[39]. Similarly, an inversion event that occurred in PTRY split synteny blocks 1 and 4′ into 1-p1/p2 and 4′-p1/p2, respectively. As a result, these rather simple inversion events can explain the fragmentation of the syntenic regions, found with different orders and orientations, between HSAY and PTRY.

We identified counterparts of the three large-scale palindromes (P6, P7 and P8) of HSAY[19] in the PTRY sequence (we also found a P5-like sequence in PTRY, but not in the contiguously sequenced region of the minimum tiling path). In addition, we identified one chimpanzee-specific palindrome (CSP1; **Fig. 1**, **Supplementary Fig. 6** and **Supplementary Note**). The HSAX sequences corresponding to P4 and P5 do not form palindromes, and the absence of P4 in chimpanzee has been previously reported[40]. P7 is relatively small and has nearly a identical structure in human and chimpanzee. In contrast, we observed significant differences between the human and chimpanzee forms of P6 and P8 (**Fig. 6**). Chimpanzee P6, at 760 kb, is larger than

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

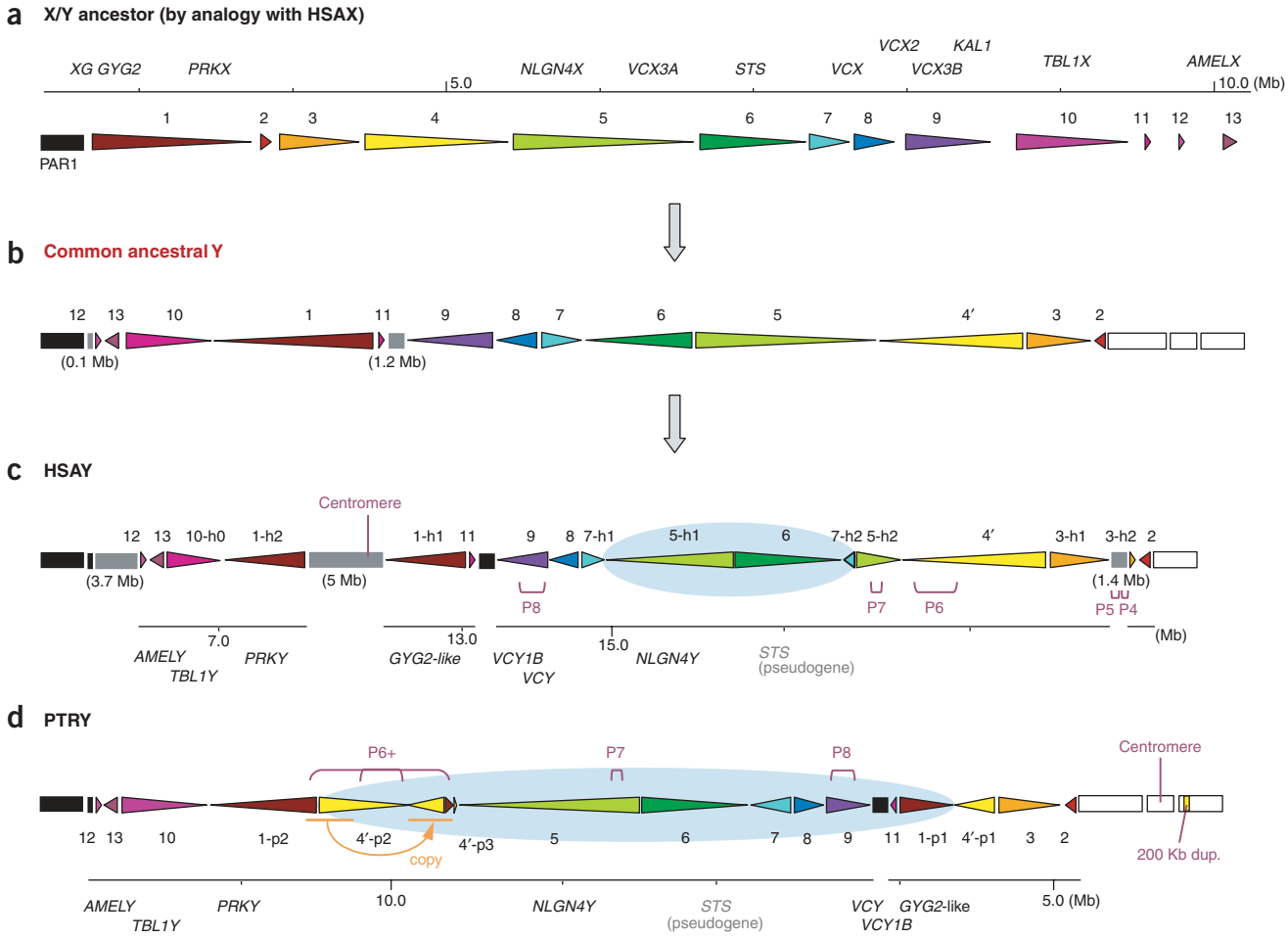

Figure 5 Synteny blocks in the sequenced region of PTRY and comparison with HSAY. Synteny blocks in PTRYq and HSAY identified through comparison with the corresponding regions of HSAX are shown as color-coded triangles. The scale (Mb) in the figure is variable from region to region. (**a**) Synteny blocks from HSAX, which is regarded as the ancestor of the X and Y chromosomes, are shown. Correlation with the block numbering system reported previously[20] is shown in **Supplementary Table 7** online. Representative genes and positions based on NCBI build 35 of HSAX are shown at top. The structures of blocks 4 and 4′ are substantially different. (**b**) Simulated configuration of the synteny blocks of the predicted common ancestral Y chromosome of human and chimpanzee. Gray boxes indicate insertion blocks. (**c**) Syntenic structure of HSAY. P4 to P8 indicate palindromes. The blue shading around *NLGN4Y-STS* indicates an inversion. (**d**) Syntenic structure of PTRY. A large segmental inversion that involves multiple blocks from 4′-p2 to 1-p1 is shaded in blue.

its human counterpart (280 kb); thus, it was designated as P6+ (**Figs. 5** and **6**). The difference in size is due to an extension that occurred in chimpanzee P6 after the inversion that includes both synteny block 4′ and part of synteny block 1-p2. Human P6 is likely to be the ancestral type (**Fig. 5**).

DISCUSSION

The Y chromosomes of human and chimpanzee show substantial differences that have accumulated independently during the past 5–6 million years. In order to assess the neutrality of the differences between the orthologous gene pairs of human and chimpanzee, we estimated the number of synonymous (*Ks*) and nonsynonymous (*Ka*) substitutions per nucleotide. Eleven (56%) out of 19 genes showed *Ka*/*Ks* values >1 for at least one splicing variant. *VCY, VCY1B, RPS4Y2* and *SRY* had *Ka*/*Ks* values >3.0 (**Table 2**). This pattern for HSAY and PTRY is significantly different from that for HSA21 and PTR22, for which only a small proportion ($<10\%$) of genes showed *Ka*/*Ks* >1 (ref. 10). It should be noted, however, that *Ka* was significantly greater than *Ks* (at the 0.1% level) for the *USP9Y* gene (**Supplementary Table 3**). There is insufficient genomic sequence data to examine the difference of the evolutionary patterns of protein-coding genes between the Y chromosomes and the autosomes. Therefore, it is unclear whether the preponderance of nonsynonymous substitutions for Y chromosome–encoded genes is indicative of positive selection. Further comparative genomic studies are necessary to clarify these points (**Supplementary Note**).

The extent of genetic variation within *P. t. verus* also caused a relatively short time of the most recent common ancestor (TMRCA) and a small effective population size of male (N_{em}). When we assume the species divergence time between humans and chimpanzees to be 6 million years, the mutation rate on the Y chromosome becomes 1.5×10^{-9} per site per year ($0.0178 / (2 \times 6 \times 10^6)$). To estimate the TMRCA, the divergence rates of the two pairs of chimpanzee individuals[51] were averaged ($(0.0422 + 0.0343)/2 = 0.0383$). Because this average divergence can be considered representative for a randomly chosen pair of sequences, the corresponding divergence time is

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

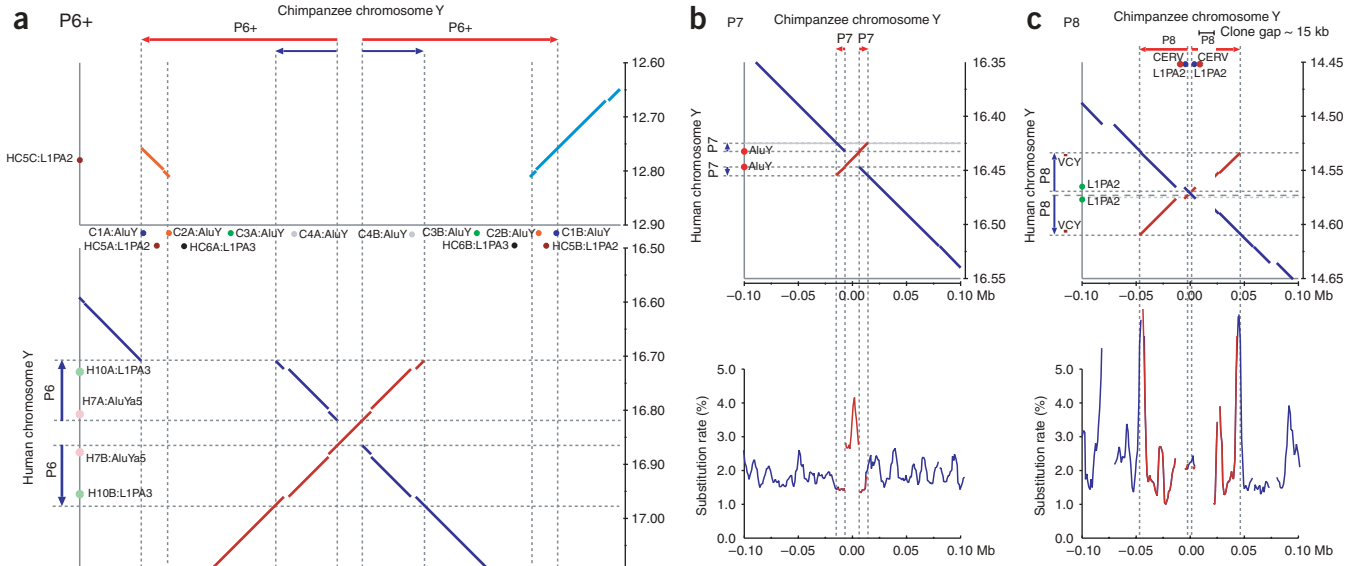

Figure 6 Comparison of the palindromic structures between PTRY and HSAY. Shown are dot plot analysis between the corresponding palindromes of PTRY and HSAY (upper plots in each panel) and the nucleotide substitution rates between the palindrome arms, spacer regions or flanking regions of the palindromes (lower plots in each panel). In addition, positions of the repetitive elements found in both palindrome arms are indicated by colored circles. (**a**) Comparison between P6+ in PTRY and P6 in HSAY. (**b**) Comparison of P7 between PTRY and HSAY. (**c**) Comparison of P8 between PTRY and HSAY. Higher nucleotide substitution rates were observed at the boundaries between the palindrome arms and the outside of the palindrome. Both human and chimpanzee *VCY* and *VCY1B* genes are located on the palindrome arms. The clone gap (approximately 15 kb) in PTRY P8 is indicated above the PTRY P8 arm.

a half of the TMRCA for many sampled sequences[41]. Then the TMRCA becomes 260,000 ± 20,000 years (2×0.000383 / ($2 \times 1.5 \times 10^{-9}$)). Because the expected value of nucleotide diversity, π, is calculated as $2N_{em}\mu_g$, where μ_g is the mutation rate per site per generation, we obtained $\mu_g = 2.25 \times 10^{-8}$ per site per generation ($1.5 \times 10^{-9} \times 15$) and estimated N_{em} as 8,400 ± 760, assuming a generation time of 15 years. Note that even these small TMRCA and N_{em} values are close to those in the extant human Y chromosomes (TMRCA = ~40,000 to ~140,000 years and N_{em} = ~5,000 to ~8,100; see ref. 42 and references therein).

We also compared the genetic variability of the Y chromosome with that of autosomal data. In the case of PTR22, we obtained diversity rates of 0.172 ± 0.010% (CHORI-251) and 0.149 ± 0.003% (RP-43) between the same two chimpanzee individuals[51]—substantially higher values than for PTRY. In the case of the Y chromosome, because of differences in the mode of inheritance and in ploidy, the extent of intraspecies nucleotide divergence is expected to be one-quarter of that of autosomes, assuming that the mutation rate is the same for the Y chromosome and the autosomes, and that the sex ratio is equal to 1. As mentioned in the previous section, because the mutation rate of the Y chromosome is likely to be α times higher than that of autosomes, the ratio of intraspecies divergence of the Y chromosome to that of an autosome should be $\alpha/4$ and should tend towards 1, as long as α is not small. However, we observed a ratio of approximately 0.25, which is smaller than this expectation. This suggests two possibilities. First, natural selection has an important role in reducing intraspecies divergence. Owing to the nonrecombining nature of the Y chromosome, both negative and positive selection can clearly reduce the divergence of the entire chromosome, a mechanism known as background selection or selective sweep. Second, this may reflect some behavioral characteristics in chimpanzees, such as reproductive behavior in chimpanzee society or differences in migration rate (or pattern) between males and females. For example, if there is a small number of dominantly reproducing males, the effective size of the male population is reduced. Large-scale population studies should be able to address this problem, and we believe that our data and the HSAY data[19] provide additional support for further population studies of the Y chromosome[27,43–45].

The integration of endogenous viral DNAs is an additional insertion event that occurred in the chimpanzee lineage. As the 14 endogenous viral sequences still retain the typical gene arrangement of the retrovirus, *LTR-gag-pol-env-LTR*, these insertions must have occurred relatively recently through the expansion of a particular type of LTR retrotransposon. Or alternatively, they may be the relics of an epidemic viral infection. In total, we detected approximately 100 copies of these insertions in the chimpanzee draft sequence (data not shown), the same as reported previously[36]. Considering the proportion of the sequence analyzed in this study, approximately 0.4% of the total genome, extrapolation of the insertion frequency in PTRY (14 copies) to the entire genome predicts 3,500 copies of these insertions, which is much higher than the observed value of 100. This suggests that the removal of deleterious events from the autosomes was by recombination or other repair mechanisms or, less likely, that a Y-specific insertion mechanism was the cause.

The simulated structure of the common ancestral Y chromosome suggests several large-scale insertion and deletion events. In the human lineage, for example, a 1.4-Mb insertion containing two palindromes (P4 and P5) split synteny block 3 into parts 3-h1 and 3-h2 (**Fig. 5c**) in the well-known X-transposed region located near the PAR1-MSY boundary[18–20]. The largest insertion (5 Mb) splits synteny block 1 into parts 1-h1 and 1-h2. Analysis of the surrounding regions suggests it is most likely that the 5-Mb insertion, which contains the centromere, is the result of a translocation. We also identified a relatively short deletion (approximately 150 kb) in human synteny block 10 (10-h0). In contrast, only two insertion blocks, of 0.1 Mb and 1.2 Mb, were observed on the common ancestral chromosome (**Fig. 5b**). Notably, no significant indels were observed in the corresponding regions of PTRY (synteny blocks 12 to 4) except for the segmentally duplicated parts of synteny blocks 1 and 4 introduced via a palindromic formation (**Fig. 5b,d**).

The *VCY* and *VCY1B* genes are located at opposite ends of the P8 palindrome in opposing orientations with the promoters on the inside, although in their counterpart on the X chromosome, *VCX* does not form a palindromic structure. Although the configuration of *VCY* and *VCY1B* is identical in human and chimpanzee, both of the human genes contain a single 30-bp unit encoding a glutamic acid–rich domain, whereas both of the chimpanzee genes contain eight 30-bp repeating units. The chimpanzee genes are unlikely to be translated into proteins owing to the presence of a stop codon upstream of the 30-bp repeats.

Future studies need to address the reason for the conservation of the structure of seemingly inactive chimpanzee *VCY* genes and other palindromic structures. However, we can suggest several possibilities, such as a gene conversion–like mechanism between the arms of the palindrome[40,46] or a mismatch repair mechanism using two palindrome arms as templates. Comparison of the nucleotide substitution rates between PTRY and HSAY P6+ (P6), P7 and P8 also supports this idea (**Fig. 6** and **Supplementary Note**). We identified at least five highly conserved ampliconic structures on PTRY. Because of their complex nature, we could only partially sequence these regions, but we were still able to clarify the chimpanzee-specific palindromic structure (CSP1) near the centromere of PTRYq and find evidence for an intrachromosomal duplication between the arms of the chromosome at regions near the centromere (**Supplementary Fig. 6** and **Supplementary Note**).

There are at least 12 major haplotype groups in the current human Y chromosome population[45]. We have previously reported differences in spermatogenic ability among males with different Y haplotypes. This introduces the possibility that higher spermatogenic ability could have acted favorably in the past; that is, Y chromosomes having genes advantageous to male fitness could increase their frequency in the population within a relatively short time[47,48]. The accelerated rate of evolution of the Y chromosome along with recent changes in the population size of great apes and changes in human society related to reproductive behavior may impose other effects on the population of their Y chromosomes.

The Y chromosome has often been regarded as a genetic wasteland; however, we believe that the structural characteristics presented in this study provide us with a realistic model of the dynamic changes that our genomes have undergone during the past 5–6 million years, since the separation of human and chimpanzee. Further comparative studies like this one will lead to a better understanding of the mechanisms through which our modern-day genomes and chromosomes have evolved from those of our ancestors.

METHODS

Mapping and sequencing of chimpanzee Y chromosome. Seed clones for construction of clone contigs were initially selected using a PCR screening system with the three-dimensionally arranged PTB1 library[4] using existing or newly designed STS information from the sequence of HSAY[1,19,49]. These clones were then subjected to clone-end or draft sequencing, from which new chromosome walking primers were designed for further screening and examination of clone overlaps. Several representative clones were also examined cytogenetically to confirm their localization on the chromosome by FISH analysis. In addition, we constructed Y chromosome–specific BAC (PTBY) and fosmid (PTFY) libraries for extending clone contigs and gap closing[10]. Sequence accuracy was calculated from 5.6 Mb of overlapping clone data.

Alignment of human and chimpanzee sequences. The PTRY sequence was first screened for matches with the published HSAY sequence[19] using the NCBI BLAST program. All the matched pairs were then subjected to a global alignment process using the LAGAN program[50]. The alignments were then refined by manual inspection. The alignment data is available from the web site of the RIKEN Genomic Sciences Center.

Gene annotation. We used manually curated HSAY annotations provided by the HAVANA group of The Wellcome Trust Sanger Institute. Positions of the human genes were converted to those on PTRY through the sequence alignments (**Supplementary Table 2**). Finally, all the genes mapped on the corresponding region of HSAY in the Vertebrate Genome Annotation (VEGA) and NCBI RefSeq databases were confirmed in the finished sequence except for *CD24L4*.

Analysis of the species-specific expansion of repetitive elements. We chose young Alu and L1 elements for further analysis (**Fig. 4**) if the indels contained the entire unit or if divergence was <3% (**Supplementary Table 6**).

Analysis of synteny blocks. Synteny blocks in the X-degenerate regions of HSAY and PTRY were surveyed through local alignments obtained from homologous regions between HSAX-HSAY and HSAX-PTRY using the BLASTN program. Regions longer than 30 kb and those that showed similarity to the region between PAR1 (ref. 20) and the *AMELX* gene of HSAX (from 2.7 Mb to 11.2 Mb) were used to construct the synteny map (**Fig. 5**). We then assessed the process of structural changes from HSAX (used as a hypothetical ancestor of the region) to the sequenced region of PTRYq and the corresponding region of HSAY as the most parsimonious series of inversions through computational calculation. After large indels were integrated, a common ancestral Y chromosome was defined as the branch point of the two pathways from HSAX to PTRYq (**Fig. 5** and **Supplementary Fig. 5**).

Accession codes. GenBank/DDBJ/EMBL: AY692036, AY692037, BS000531–BS000572, BS000574–BS000581, BS000583, BS000585, BS000587–BS000682. See **Supplementary Table 1** for a full list of the clone names and accession numbers. In addition, the following as yet unpublished BACs, listed in **Supplementary Table 5**, were generated and deposited in GenBank by R.K. Wilson (Washington University School of Medicine) and D.C. Page (Whitehead Institute): AC142320, AC144378, AC145769, AC145782, AC146011, AC146175, AC146245, AC146268, AC147117, AC147148, AC147343, AC147657, AC147670 and AC151848.

URLs. Manually annotated HSAY genes, http://vega.sanger.ac.uk/Homo_sapiens/; NCBI, http://www.ncbi.nlm.nih.gov/; NCBI BLAST, http://www.ncbi.nlm.nih.gov/BLAST/; RIKEN-Genomic Sciences Center, http://stt.gsc.riken.jp/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We are grateful to all the technical staff of RIKEN Genomic Sciences Center (RIKEN-GSC) and Genome Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB). In addition, we wish to thank Y. Nakahori (Tokushima University), T. Matsuzawa (Primate Research Institute, Kyoto University), T. Ishida (The University of Tokyo), N. Collier (National Institute of Informatics) for the supply of the human and chimpanzee samples, and discussion. We also thank the Osaka Municipal Tennoji Zoo (Osaka, Japan)

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

for discussions on the subspecies of chimpanzee. We would also like to thank A. Frankish, E. Hart and J. Harrow from the HAVANA group, the Wellcome Trust Sanger Institute, for providing annotation of the human Y chromosome. We are also thankful for D.C. Page (Whitehead Institute), R.K. Wilson (Washington University School of Medicine) and the Chimpanzee Sequencing and Analysis Consortium for the availability of the sequences in the public databases. This work was supported in part by: Special Fund for RIKEN-GSC; Grant-in-Aid for Scientific Research on Priority Areas "Genome Science" from the Ministry of Education, Culture, Sports, Science and Technology, Japan; The Ministry of Science & Technology, Korea, and Special Fund of KRIBB, Korea.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics

1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
3. Gagneux, P. & Varki, A. Genetic differences between humans and great apes. *Mol. Phylogenet. Evol.* **18**, 2–13 (2001).
4. Fujiyama, A. *et al.* Construction and analysis of a human-chimpanzee comparative clone map. *Science* **295**, 131–134 (2002).
5. Ebersberger, I., Metzler, D., Schwarz, C. & Paabo, S. Genomewide comparison of DNA sequences between humans and chimpanzees. *Am. J. Hum. Genet.* **70**, 1490–1497 (2002).
6. Chou, H.H. *et al.* Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc. Natl. Acad. Sci. USA* **99**, 11736–11741 (2002).
7. Enard, W. *et al.* Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).
8. Sakaki, Y. *et al.* Human versus chimpanzee chromosome-wide sequence comparison and its evolutionary implication. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 455–460 (2003).
9. Olson, M.V. & Varki, A. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nat. Rev. Genet.* **4**, 20–28 (2003).
10. The International Chimpanzee Chromosome 22 Consortium. DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature* **429**, 382–388 (2004).
11. Ptak, S.E. *et al.* Fine-scale recombination patterns differ between chimpanzees and humans. *Nat. Genet.* **37**, 429–434 (2005).
12. Ohno, S. *Sex Chromosomes and Sex-Linked Genes* (Springer, Berlin, 1967).
13. Bishop, C. *et al.* Extensive sequence homologies between Y and other human chromosomes. *J. Mol. Biol.* **173**, 403–417 (1984).
14. Graves, J.A. The origin and function of the mammalian Y chromosome and Y-borne genes–an evolving understanding. *Bioessays* **17**, 311–320 (1995).
15. Lahn, B.T. & Page, D.C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).
16. Waters, P.D. *et al.* The human Y chromosome derives largely from a single autosomal region added to the sex chromosomes 80–130 million years ago. *Cytogenet. Cell Genet.* **92**, 74–76 (2001).
17. Lahn, B.T., Pearson, N.M. & Jegalian, K. The human Y chromosome, in the light of evolution. *Nat. Rev. Genet.* **2**, 207–216 (2001).
18. Page, D.C., Harper, M.E., Love, J. & Botstein, D. Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. *Nature* **311**, 119–123 (1984).
19. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
20. Ross, M.T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).
21. Archidiacono, N. *et al.* Evolution of chromosome Y in primates. *Chromosoma* **107**, 241–246 (1998).
22. Gerrard, D.T. & Filatov, D.A. Positive and negative selection on mammalian Y chromosomes. *Mol. Biol. Evol.* **22**, 1423–1432 (2005).
23. Miyata, T. *et al.* Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 863–867 (1987).
24. Erlandsson, R., Wilson, J.F. & Paabo, S. Sex chromosomal transposable element accumulation and male-driven substitutional evolution in humans. *Mol. Biol. Evol.* **17**, 804–812 (2000).
25. Makova, K.D. & Li, W.H. Strong male-driven evolution of DNA sequences in humans and apes. *Nature* **416**, 624–626 (2002).
26. Saitou, N. & Ueda, S. Evolutionary rate of insertions and deletions in non-coding nucleotide sequences of primates. *Mol. Biol. Evol.* **11**, 504–512 (1994).
27. Stone, A.C. *et al.* High levels of Y-chromosome nucleotide diversity in the genus *Pan*. *Proc. Natl. Acad. Sci. USA* **99**, 43–48 (2002).
28. Filatov, D.A. A gradient of silent substitution rate in the human pseudoautosomal region. *Mol. Biol. Evol.* **21**, 410–417 (2004).
29. Bailey, J.A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
30. Liu, G. *et al.* Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res.* **13**, 358–368 (2003).
31. Hedges, D.J. *et al.* Differential Alu mobilization and polymorphism among the human and chimpanzee lineages. *Genome Res.* **14**, 1068–1075 (2004).
32. Orti, R. *et al.* Conservation of pericentromeric duplications of a 200-kb part of the human 21q22.1 region in primates. *Cytogenet. Cell Genet.* **83**, 262–265 (1998).
33. Golfier, G. *et al.* The 200-kb segmental duplication on human chromosome 21 originates from a pericentromeric dissemination involving human chromosomes 2, 18 and 13. *Gene* **312**, 51–59 (2003).
34. Bonner, T.I. *et al.* Molecular cloning of a family of retroviral sequences found in chimpanzee but not human DNA. *J. Virol.* **43**, 914–924 (1982).
35. Kjellman, C., Sjogren, H.O. & Widegren, B. The Y chromosome: a graveyard for endogenous retroviruses. *Gene* **161**, 163–170 (1995).
36. Yohn, C.T. *et al.* Lineage-specific expansions of retroviral insertions within the genomes of African great apes but not humans and orangutans. *PLoS Biol.* **3**, e110 (2005).
37. Gibbons, R. *et al.* Distinguishing humans from great apes with AluYb8 repeats. *J. Mol. Biol.* **339**, 721–729 (2004).
38. Han, K. *et al.* Under the genomic radar: the stealth model of Alu amplification. *Genome Res.* **15**, 655–664 (2005).
39. Schwartz, A. *et al.* Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Hum. Mol. Genet.* **7**, 1–11 (1998).
40. Rosen, S. *et al.* Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**, 873–876 (2003).
41. Li, W.-H. *Molecular Evolution* 237–267 (Sinauer, Sunderland, Massachusetts, 1997).
42. Thomson, R., Pitchard, J.K., Shen, P., Oefner, P.J. & Feldman, M.W. Recent common ancestry of human y chromosomes: Evidence from DNA sequence data. *Proc. Natl. Acad. Sci. USA* **97**, 7360–7365 (2000).
43. Underhill, P.A. *et al.* Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**, 947–949 (1997).
44. Underhill, P.A. *et al.* Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**, 358–361 (2000).
45. The Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**, 339–348 (2002).
46. Kittler, R. *et al.* Apparent intrachromosomal exchange on the human Y chromosome explained by population history. *Eur. J. Hum. Genet.* **11**, 304–314 (2003).
47. Kuroki, Y. *et al.* Spermatogenic ability is different among males in different Y chromosome lineage. *J. Hum. Genet.* **44**, 289–292 (1999).
48. Nakahori, Y. The Y chromosome-a hypothesis on the evolution of sex chromosomes. in *Sexual Differentiation and Maturation. Frontier in Endocrinology 17* (Hibi, I. & Tanaka, T., eds.) 1–14 (Ares-Serono Symposia Publications, Rome, 1996).
49. Yamada, T. & Morishita, S. Computing highly specific and noise tolerant oligomers efficiently. *J. Bioinform. Comput. Biol.* **2**, 21–46 (2004).
50. Brudno, M. *et al.* LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13**, 721–731 (2003).
51. Hughes, J.F. *et al.* Conservation of Y-linked genes during human evolution revealed by comparative sequencing. *Nature* **437**, 100–105 (2005).

Corrigendum: Comparative analysis of chimpanzee and human Y chromosomes unveils complex evolutionary pathway

Y Kuroki, A Toyoda, H Noguchi, T D Taylor, T Itoh, D-S Kim, D-W Kim, S-H Choi, I-C Kim, H H Choi, Y S Kim, Y Satta, N Saitou, T Yamada, S Morishita, M Hattori, Y Sakaki, H-S Park & A Fujiyama
Nat. Genet. **38, 158–167 (2006).**

The following paper, which was published during the revision process of this manuscript, should have appeared in the reference list and should have been cited as below. The authors wish to apologize to the readers and the original data producers for this mistake and inconvenience.

51. Hughes, J.F. ***et al.*** **Conservation of Y-linked genes during human evolution revealed by comparative sequencing.** *Nature* **437, 100–105 (2005).**

On page 160, in the first full paragraph in the right-hand column, the first sentence should read, "As the PTRY sequence analyzed in this study originated from a single male animal, Gon[10] (PTB library), we compared intraspecies nucleotide diversity using the sequences in the public database[51] (**Supplementary Table 5** online)."

On page 164, in the first full paragraph in the right-hand column, the third sentence should read, "To estimate the TRMCA, the divergence rates of the two pairs of chimpanzee individuals[51] were averaged ((0.0422 + 0.0343)/2 = 0.0383)."

On page 165, in the first full paragraph in the left-hand column, the second sentence should read, "In the case of PTR22, we obtained diversity rates of 0.172 ± 0.010% (CHORI-251) and 0.149 ± 0.003% (RP-43) between the same two chimpanzee individuals[51]—substantially higher values than for PTRY."

In addition, the following as-yet-unpublished BACs, listed in **Supplementary Table 5**, were generated and deposited in GenBank by R.K. Wilson (Washington University School of Medicine) and D.C. Page (Whitehead Institute): AC142320, AC144378, AC145769, AC145782, AC146011, AC146175, AC146245, AC146268, AC147117, AC147148, AC147343, AC147657, AC147670 and AC151848.

Finally, in line 7 of the Acknowledgments, T.-n. Zoo (Osaka, Japan) should have been listed as Osaka Municipal Tennoji Zoo (Osaka, Japan).

The errors have been corrected in the HTML and PDF versions of the article. This correction has been appended to the PDF version.

© 2006 Nature Publishing Group http://www.nature.com/naturegenetics