

# GENOME RESEARCH

## Evolutionary dynamics of segmental duplications from human Y-chromosomal euchromatin/heterochromatin transition regions

Stefan Kirsch, Claudia Münch, Zhaoshi Jiang, Ze Cheng, Lin Chen, Christiane Batz, Evan E. Eichler and Werner Schempp

*Genome Res.* 2008 18: 1030-1042; originally published online Apr 29, 2008;  
Access the most recent version at doi:[10.1101/gr.076711.108](https://doi.org/10.1101/gr.076711.108)

---

**Supplementary data**

"Supplemental Research Data"  
<http://www.genome.org/cgi/content/full/gr.076711.108/DC1>

**References**

This article cites 67 articles, 28 of which can be accessed free at:  
<http://www.genome.org/cgi/content/full/18/7/1030#References>

**Email alerting service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

---

**Notes**

---

To subscribe to *Genome Research* go to:  
<http://www.genome.org/subscriptions/>

---



# Evolutionary dynamics of segmental duplications from human Y-chromosomal euchromatin/heterochromatin transition regions

Stefan Kirsch,<sup>1,3</sup> Claudia Münch,<sup>1,3</sup> Zhaoshi Jiang,<sup>2</sup> Ze Cheng,<sup>2</sup> Lin Chen,<sup>2</sup> Christiane Batz,<sup>1</sup> Evan E. Eichler,<sup>2</sup> and Werner Schempp<sup>1,4</sup>

<sup>1</sup>Institute of Human Genetics, University of Freiburg, 79106 Freiburg, Germany; <sup>2</sup>Department of Genome Sciences, Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA

Human chromosomal regions enriched in segmental duplications are subject to extensive genomic reorganization. Such regions are particularly informative for illuminating the evolutionary history of a given chromosome. We have analyzed 866 kb of Y-chromosomal non-palindromic segmental duplications delineating four euchromatin/heterochromatin transition regions (Yp11.2/Yp11.1, Yq11.1/Yq11.21, Yq11.23/Yq12, and Yq12/PAR2). Several computational methods were applied to decipher the segmental duplication architecture and identify the ancestral origin of the 41 different duplcons. Combining computational and comparative FISH analysis, we reconstruct the evolutionary history of these regions. Our analysis indicates a continuous process of transposition of duplicated sequences onto the evolving higher primate Y chromosome, providing unique insights into the development of species-specific Y-chromosomal and autosomal duplcons. Phylogenetic sequence comparisons show that duplcons of the human Yp11.2/Yp11.1 region were already present in the macaque–human ancestor as multiple paralogs located predominantly in subtelomeric regions. In contrast, duplcons from the Yq11.1/Yq11.21, Yq11.23/Yq12, and Yq12/PAR2 regions show no evidence of duplication in rhesus macaque, but map to the pericentromeric regions in chimpanzee and human. This suggests an evolutionary shift in the direction of duplicative transposition events from subtelomeric in Old World monkeys to pericentromeric in the human/ape lineage. Extensive chromosomal relocation of autosomal-duplicated sequences from euchromatin/heterochromatin transition regions to interstitial regions as demonstrated on the pygmy chimpanzee Y chromosome support a model in which substantial reorganization and amplification of duplicated sequences may contribute to speciation.

[Supplemental material is available online at [www.genome.org](http://www.genome.org).]

Segmental duplications (SD) are euchromatic portions of DNA present at two or more locations in the human genome that satisfy the minimum requirement of 90% nucleotide sequence identity and are  $\geq 1$  kb in length (Eichler 2001). Initially recognized as a special feature of pericentromeric regions (Eichler et al. 1996, 1997), a broader distribution within subtelomeric and interspersed chromosomal regions was subsequently confirmed by genome-wide analyses (Bailey et al. 2001; Mefford and Trask 2002). Altogether, at least 5% of the human genome is composed of such duplicated sequences (Cheung et al. 2001, 2003; Bailey et al. 2002; She et al. 2004). Numerous studies show a strong association between the SD location and regions of genomic instability (Ji et al. 2000; Inoue and Lupski 2002; Bailey et al. 2004; Shaw and Lupski 2004; Sharp et al. 2005, 2006; Perry et al. 2006). The underlying mechanisms shaping the contemporary distribution pattern of human SDs has so far remained elusive. Over the last decade, it has emerged that SDs represent a basic feature of most animal genomes (Bailey and Eichler 2006). The apparent increase in interspersed SD content among primate genomes (Bailey and Eichler 2006; She et al. 2006) and its potential role in adaptive evolution (Johnson et al. 2001; Paulding et al. 2003; Birtle et al.

2005; Newman et al. 2005) is an important topic in primate genome evolution.

Among all human chromosomes, the Y chromosome has the highest SD content (Kuroda-Kawaguchi et al. 2001; Bailey et al. 2002; Bailey and Eichler 2003; Rozen et al. 2003; Skaletsky et al. 2003; She et al. 2006). Recently, we have cloned a previously unknown euchromatic island within the pericentromeric satellite 3 sequences of the euchromatin/heterochromatin transition region in Yq11.1/Yq11.21 (Kirsch et al. 2005). Whole-genome comparison of the assembled sequence revealed that it consisted exclusively of SDs. By inspecting the NCBI Y chromosome reference assembly, we found that all four euchromatin/heterochromatin transition regions of the human Y chromosome are characterized by the presence of SDs. Given the haploid nature of the Y chromosome and the fact that SDs are absent in both pseudo-autosomal regions, it can only participate in the genomic distribution process of SDs via duplicative transposition and/or translocation. Moreover, translocations between the Y chromosome and the autosomes are rare in primates (Wienberg 2005). This dramatically reduces the level of complexity in tracing the evolutionary history of Y-chromosomal SDs. In this context, we envisage the primate Y chromosome as a useful model to delineate the chromosomal and molecular evolution of other inter- and intrachromosomal SD regions.

In this study, we carried out a detailed molecular and cytogenetic evolutionary analysis for 866 kb of human Y-chromosomal

<sup>3</sup>These authors contributed equally to this work.

<sup>4</sup>Corresponding author.

E-mail [werner.schempp@uniklinik-freiburg.de](mailto:werner.schempp@uniklinik-freiburg.de); fax 49-761-2707041.

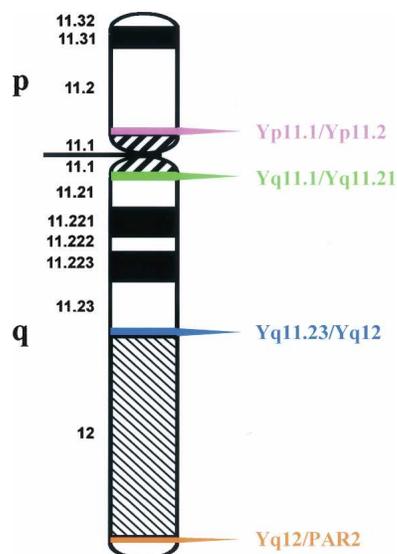
Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.076711.108>.

non-palindromic segmental duplications from the four euchromatin/heterochromatin transition regions in Yp11.2/Yp11.1, Yq11.1/Yq11.21, Yq11.23/Yq12, and Yq12/PAR2. We performed whole-genome sequence comparison simultaneously for human, common chimpanzee, and rhesus macaque. Owing to the under-representation of SDs in whole-genome sequencing assemblies, we extended our analyses by fluorescence in situ hybridization (FISH) within human and non-human primate species targeted for whole-genome sequence assembly (*Pan troglodytes*, *Gorilla gorilla*, *Pongo pygmaeus*, *Nomascus leucogenys*, *Macaca mulatta*, and *Callithrix jacchus*) (Eichler and DeJong 2002). To further refine species-specific variation and to provide a comprehensive evolutionary view of the evolution of these regions, we extended our analysis to include additional species representing different branches of the generally accepted phylogenetic tree of Goodman (1999) (*Pan paniscus*, *Hylobates lar*, *Hylobates muelleri*, *Papio hamadryas*, *Theropithecus gelada*, *Macaca nemestrina*, *Macaca silenus*, *Macaca fascicularis*, *Callithrix geoffroyi*, *Ateles geoffroyi*, and *Callicebus moloch*).

## Results

### Sequential composition of human Y-chromosomal transition regions

All four Y-chromosomal euchromatin/heterochromatin transition regions (Yp11.2/Yp11.1; Yq11.1/Yq11.21; Yq11.23/Yq12; Yq12/PAR2) are characterized by the presence of highly duplicated sequences. Both pericentromeric and the bordering AZFc-heterochromatin sequence assemblies are complete, whereas there still exists a gap between the heterochromatin and the contig extending into the PAR2 (Skaletsky et al. 2003). The location of each transition region is depicted along the human Y chromosome (Fig. 1). Precise positioning of the segmental duplications within the transition regions of the human genome se-



**Figure 1.** Location of segmental duplications in euchromatin/heterochromatin transition regions of the human Y chromosome. The region-specific color code of the lettering is retained in all consecutive diagrams.

**Table 1.** Position of human Y-chromosomal euchromatin/heterochromatin transition regions

TRANSITION REGION	CONTIG	POSITION	CLONES
Yp11.2/Yp11.1	NT_011878	634,291-749,306	RP11-160K17 (AC006987) RP11-108I14 (AC010970)
Yq11.1/Yq11.21	NT_113819	2,578-452,044	RP1-85D24 (AC140113) RP11-131M6 (AC134878) RP11-886I11 (AC134882) RP11-295P22 (AC134879)
Yq11.23/Yq12	NT_011903	4,597,035-4,832,695	RP11-557B9 (AC013734) RP11-428D10 (AC019099)
Yq12/PAR2	NT_091573	1-66,393	RP11-57J19 (AC025226)

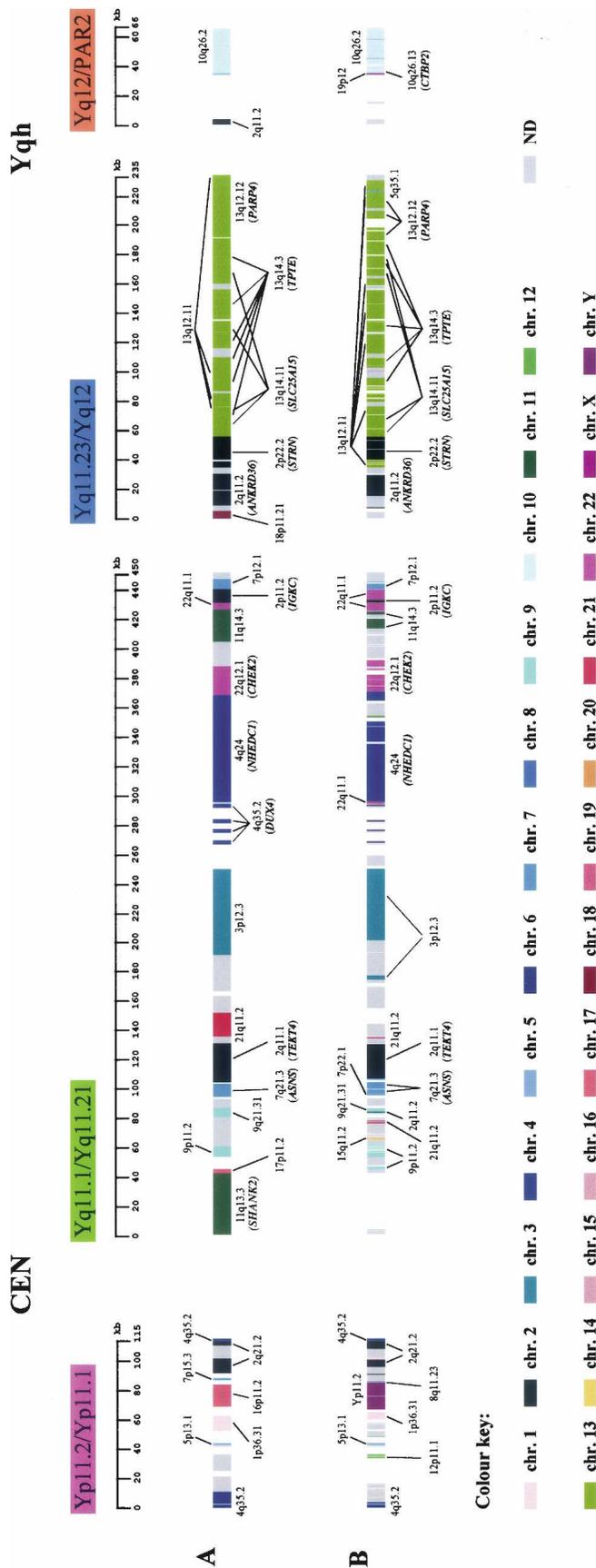
quence assembly are given according to NCBI Build 36.1, March 2006 (Table 1).

### Ancestral-duplicon architecture and gene content

The ancestral state for each of the four euchromatin/heterochromatin transition regions was determined by three computational methods: (1) the search for minimally shared segments by examination of all underlying pairwise alignments generated by whole-genome analysis comparison (Bailey et al. 2002); (2) the identification of segments with conserved genic structures (Horvath et al. 2005); and (3) the *DupMasker* analysis software, which is particularly useful for the detection of relatively short (<7 kb) duplicons (Jiang et al. 2007, 2008). Altogether, 41 duplicons originating from 18 different human chromosomes were identified (Fig. 2; Table 2). Twenty of these display discontinuous homology with the human Y chromosome. In total, 41.5% (17/41) show evidence of conserved exon-intron structure when compared to an ancestral locus. None, however, contained a complete transcription unit as revealed by *sim4* pairwise analyses ([http://www.bx.psu.edu/miller\\_lab](http://www.bx.psu.edu/miller_lab); Florea et al. 1998). Unlike the ancestral loci where a complete gene model could always be assigned, the Y chromosome gene structures likely correspond to unprocessed pseudogenes. A detailed description of the ancestral loci for each of the four regions may be found in Supplemental File 1.

### Phylogenetic analysis of human Y-chromosomal SDs

We performed phylogenetic analysis of non-coding sequences from human and two non-human primate species (*P. troglodytes*, *M. mulatta*) for a total of 78 duplication subunits representing 36 of 41 defined duplicons from the euchromatin/heterochromatin transition regions. The multiple sequence alignments ranged from 0.26 to 1.3 kb, depending on the size of the underlying duplication subunit. Discrete duplication subunits derived from the same duplicon (ancestral locus) always show highly similar neighbor-joining topologies. We, therefore, focused on the most extensive duplication subunits to reconstruct the evolutionary history of each duplicon. We calculated nucleotide substitution rates (see Methods) for each of the 28 duplicons where the ancestral duplicon in the macaque could be unequivocally determined (Supplemental Table 1). Calculating the times for duplication—seeding (for definition, see Eichler et al. 1997) and swapping (previously termed exchange; for definition, see Horvath et al. 2005)—events reveals that nine initial duplication events (1p36.31, 7p15.3, 15q11.2, 9q21.31, 7p22.1, 7q21.3, 11p14.3, 13q14.11, 13q14.3) most likely occurred before the Old World monkey and ape lineages diverged (23 million years ago [Mya]). The remaining 19 duplicons underwent duplication specifically within the ape lineage, ceasing at 8 Mya. Secondary duplica-



**Figure 2.** Ancestral-duplication determination of the human Y-chromosomal euchromatin/heterochromatin transition regions. The upper layer displays the chromosomal position and extension of the regions containing highly duplicated sequences with respect to the heterochromatin. The chromosomal location of a defined ancestral duplication is indicated by the color key. Duplicons are denoted according to the cytogenetic band position of their ancestral locus. Conserved genic structures are shown in parentheses. Gray blocks display sequences for which no ancestral state could be determined. (A) Duplication blocks accounting for a total of 866 kb were analyzed by phylogenetic analysis and comparative primate FISH to define their ancestral duplication (colored bars) state (see Methods). (B) Duplicated sequences simultaneously subjected to *DupMasker* analysis (Jiang et al. 2008) for prediction of ancestral duplicons are compared against experimentally determined loci.

**Table 2.** Ancestral-duplicon architecture and genic structures in human Y-chromosomal transition regions

Transition region	Ancestral location	Genic Structure	Duplicon name	
Yp11.2/Yp11.1	4q35.2		4q35.2	
	12p11.1		12p11.1	
	5p13.1		5p13.1	
	1p36.31		1p36.31	
	Yp11.2/16p11.2		Yp11.2/16p11.2	
	8q11.23	<i>CDC27</i>	8q11.23	
	7p15.3		7p15.3	
	2q21.2		2q21.2	
	4q35.2		4q35.2	
	Yq11.1/Yq11.21	11q13.3	<i>SHANK2</i>	<i>SHANK2</i>
17p11.2			17p11.2	
9p11.2			9p11.2	
15q11.2			15q11.2	
21q11.2			21q11.2	
2q11.2		<i>TEKT4</i>	<i>TEKT4</i>	
9q21.31			9q21.31	
7p22.1			7p22.1	
7q21.3		<i>ASNS</i>	<i>ASNS</i>	
2q11.2			2q11.2	
Yq11.1/Yq11.21	21q11.2		21q11.2	
	3p12.3		3p12.3	
	4q35.2	<i>DUX4</i>	<i>DUX4</i>	
	22q11.1		22q11.1	
	4q24	<i>NHEDC1</i>	<i>NHEDC1</i>	
	22q12.1	<i>CHEK2, TTC28</i>	<i>CHEK2</i>	
	11q14.3		11q14.3	
	22q11.1		22q11.1	
	2p11.2	<i>IGKC</i>	<i>IGKC</i>	
	Yq11.23/Yq12	18p11.21		18p11.21
2q11.2		<i>ANKRD36</i>	<i>ANKRD36</i>	
2p22.2		<i>STRN</i>	<i>STRN</i>	
13q12.11		<i>TPTE, SLC25A15</i>	13q12.11	
13q14.3		<i>TPTE, SLC25A15</i>	<i>TPTE<sup>a</sup></i>	
13q14.11		<i>TPTE, SLC25A15</i>	<i>SLC25A15</i>	
13q12.12		<i>PARP4</i>	<i>PARP4</i>	
5q35.1		<i>FAM58A</i>	5q35.1	
Yq12/PAR2		2q11.2		2q11.2
		19p12	<i>CTBP2</i>	19p12
	10q26.13	<i>CTBP2</i>	<i>CTBP2</i>	
	10q26.2		10q26.2	

Duplicons are primarily denoted according to the cytogenetic band position of their ancestral loci. In case a functional gene resides within the ancestral locus, the duplicon is named according to the corresponding gene symbol.

<sup>a</sup>*TPTE* is a positionally relocated gene—mouse synteny mapping identified 13q14.3 as the ancestral functional locus.

tion or swapping events occurred between 5.5 and 18.4 Mya with no correspondence between the date of the ancestral duplication and the onset of secondary duplications (Supplemental Table 1).

### Comparative FISH of human Y-chromosomal SDs in higher primates

To obtain further insight into the evolutionary dynamics of human Y-chromosomal SDs, we performed comparative FISH on male metaphase chromosomes from human and non-human primates. Nine large-insert genomic clones covering the four Y-chromosomal euchromatin/heterochromatin transition regions (Table 1) were used to determine the cytogenetic location and degree of conservation on the Y chromosomes (Table 3; Fig. 3). The paralogous multi-site signal pattern on the autosomes and

the X chromosome was compared for all higher primate species on the chromosomal-band level (Table 3). Specific FISH-signal patterns for each genomic clone are highlighted and discussed in Supplemental Figures 1–4, Supplemental Tables 2 and 3, and Supplemental File 2. A total of 756 discrete signals were analyzed, 85.3% of which (645 signals) mapped to euchromatin/heterochromatin transition regions.

Our FISH results demonstrate that these duplications first appeared on the primate Y chromosome after the separation of the Old World monkeys from the human/ape lineage (Fig. 3). We find that only two BAC clones corresponding to the human Yp11.2/Yp11.1 region and one BAC clone from the Yq11.1/Yq11.21 region show signals on the gibbon Y chromosome.

Analysis of the orangutan Y chromosomes discloses a complex signal distribution pattern involving the additional acquisition of duplicated sequences from the Yq11.1/Yq11.21 and Yq12/PAR2 transition regions (Fig. 3). This complex pattern is observed on the Y chromosomes for both orangutan subspecies. Nevertheless, striking differences in the pattern distinguish the Bornean and Sumatran orangutan Y chromosomes, indicating substantial chromosomal rearrangements after the split of the subspecies (Schempp et al. 1993).

In gorilla, we find further evidence of duplication corresponding to the Yq11.1/Yq11.21 and the Yq11.23/Yq12 transition regions (Fig. 3). The gorilla Y chromosome presents two distinctive features: a local concentration of duplicated sequences in the gorilla Yq11.2/Yq12.1 euchromatin/heterochromatin transition region and the first appearance of two copies of the future human Yq11.1/Yq11.21 transition region. One of these copies is a constituent of the gorilla Yq11.2/Yq12.1 euchromatin/heterochromatin transition region, while the other one is situated in distal gorilla Yp12. The distal gorilla Yp12 copy is preserved and corresponds to the single copy found at the orthologous chromosomal location in the *Homo–Pan* clade.

Interestingly, duplicated sequences from the human Yq11.23/Yq12 transition region are absent on both chimpanzee Y chromosomes, indicating the loss of these sequences since the divergence of human–chimpanzee, but before the divergence of common chimpanzee–pygmy chimpanzee lineages. Nevertheless, both chimpanzee Y chromosomes present a complex signal distribution pattern of duplicated sequences. The pygmy chimpanzee is particularly enriched for such sequences in its euchromatic long arm.

Among humans and great apes, comparative FISH analyses revealed multi-site signals corresponding to orthologous chromosomal locations (Supplemental Figs. 1–4). The level of signal intensity, however, varied particularly within centromeres and short arm regions of acrocentric chromosomes. Both orangutan subspecies showed identical multi-site patterns among the autosomes. Interestingly, genomic clones from the Yp11.2/Yp11.1, Yq11.1/Yq11.21, and Yq11.23/Yq12 regions detected prominent signals at evolutionary fusion sites (NLE 22qprox [HSA 2q24.3/6p21.1]; NLE 5pprox [HSA 1q32.1/13q12.13]; NLE 9pmed [HSA 1p32.3/4q13.2]) in the white-cheeked crested gibbon (Table 3; Supplemental Figs. 1–3; Supplemental File 2; Roberto et al. 2007). As neither cross-hybridization could be detected in the orthologous chromosomal regions in human and great apes nor in the rhesus macaque, these signals reflect de novo acquisitions of SDs in this gibbon species.

As pointed out earlier (Horvath et al. 2005), a marked decrease in copy number was noted as clones were hybridized to

**Table 3. Comparative FISH results for interchromosomal duplications**

TRANSITION REGION	HSA	PTR	PPA	GGO	PPY	NLE	MMU
<b>Yp11.1/Yp11.2</b>	1p36.3; 2cen; 2q21; 2q32; 3p13; 3cen; 3q26; 3q29; 4cen; 4q35; 8cen; 9cen; 9q13; 10p15; 10cen; 12p13; 12cen; 13*; 14*; 15*; 16cen; 16q11.1/q11.2; 20cen; 21*; 22*; Yp11.1/p11.2; Yq11.23	1p36.3; 2Bcen; 3p12; 4q35; 9cen; 9q34; 13*; 14*; 15p; 16p11; 16cen; 18*; 20cen; 21*; 22*; Yp11.2-q12.1	1p36.3; 2Bcen; 3p12; 4q35; 9cen; 9q34; 13p; 14p; 15*; 16p11; 16cen; 18*; 20cen; 21p; 22*; Yq11.3; Yq11.4; Yq11.4-q12.1	1cen; 1p36.3; 2Bcen; 3p12; 4q35; 8cen; 9cen; 12cen; 13*; 15*; 16p11; 16cen; 18cen; 20cen; 21p; 22p; Yp11.2; Ycen; Yq11.2-q12.1	1p36.3; 2A*; 2B*; 8cen; 9*; 12cen; 13*; 14*; 15*; 16p11; 16cen; 18*; 21*; 22*; bo: Yp11.1-p11.2; Yq12/q13; Yq13/q14; su: Yp13.1; Yp12; Yq11.1-q11.2; Yq13/q14.1; Yq14.2	20pprox (HSA2q21); 22qprox; 23p; 24p; 25p; Ycen; YNOR	10pprox
<b>Yq11.1/Yq11.21</b>	1p11/p12; 1q12; 1q12/q21; 1q44; 2cen; 2q14.3/q21; 3p12; 3cen; 4cen; 4q24; 4q35; 9p11/p12; 9cen; 9q12/q13; 10cen; 11q13; 13*; 14*; 15*; 16cen; 18*; 20cen; 21*; 22*; 20cen; 20q11.2; 21*; 22*; Yq11.1/q11.21	1q34; 2Acen; 2Bcen; 3p13; 4q15; 4q25; 9p12; 9p11; 9cen; 9q12; 10cen; 11q13; 13*; 14*; 15*; 16cen; 18*; 20cen; 21*; 22*; Yp11.2/p12; Yp11.2prox; Yp11.2-q11.2; Yq12.1/q12.2; Yq12.2	1q34; 2Acen; 2Bcen; 3p13; 4q25; 9p12; 9p11; 9cen; 9q12; 10cen; 11q13; 13*; 14*; 15*; 16cen; 18*; 20cen; 21*; 22*; Yp11.2/p12; Ycen; Yq11.3; Yq11.3/q11.4; Yq11.4; Yq12.1; Yq12.2	1cen; 1q34; 2Acen; 2Bcen; 3p13; 5cen; 9*; 9q12; 10cen; 11q13; 13*; 14*; 15*; 16cen; 18*; 21*; 22*; Yp12; Yq11.2; Yq12.2	1p36; 1q34; 2A*; 2B*; 3cen; 4q16; 4q25; 7p12; 7cen; 9q22; 9*; 11p16; 13*; 14*; 15*; 18*; 21*; 22*; bo: Yp12.2/p12.3; Yp11.2-q12; Yq13-q14; su: Yp13.1/p13.2; Yp11.1-q11.1; Yq13-q14.1; Yq14.1-q14.2	5pdist (HSA1q44); 5pprox (HSA13pprox); 5cen (HSA13p11.1); 7cen (HSA22q11.1); 9pmed; 9qmed (HSA4q24); 20cen (HSA2q21); 22qprox; 23p; 24p; YNOR	9qprox; 10qprox; 13qtel; 16pprox
<b>Yq11.23/Yq12</b>	1q12; 1q21; 1q25; 1q32; 2p22; 2cen; 2q21; 8p12/p21; 9p13/p21; 9q13; 9q22; 13cen; 13q12.3/q13; 15cen; 16cen; 21cen; 22cen; 22q13; Yq11.23/q12	1q22; 2Acen; 8p12/p21; 13cen; 13q12.3/q13; 16cen; 22cen; 22q13	1q22; 2Acen; 8p12/p21; 13cen; 13q12.3/q13; 16cen; 22cen; 22q12; 22q13	1q22; 2Acen; 8p12/p21; 9cen; 13*; 13q12.3/q13; 14*; 15*; 16cen; 18*; 21*; 22*; Yq11.2; Yq12.1/q12.2; Yq12.4/q12.5	1q22; 8p12/p21; 13cen; 13q12.3/q1321*; 22*; 18*; 21*; 22*; Yq11.2; Yq12.1/q12.2; Yq12.4/q12.5	5qprox (HSA13cen); 8pprox; 9pdist (HSA13q12.3/q13); 22qprox; 23p	8pmed (HSA8p12/p21); 10pprox (HSA22cen); 10qprox; 17ptel (HSA13cen); 17pprox (HSA13q12.3/q13)
<b>Yq12/PAR2</b>	1p21; 1q12; 2cen; 10q26.3; 12q12/q13; 13*; 14*; 15*; 21*; Xq28; Yq12/PAR	1p21; 10q26.3; Xq28; Yp11.1/p11.2; Yq11.1-q11.2	1p21; 10q26.3; Xq28; Yq11.2/q11.3; Yq11.4; Yq12.1/q12.2	1p21; 2Bcen; 10q26.3; 16p13; 18*; 21*; 22*; Xq28; Yp12; Yq11.2; Yq12.2	1p21; 2A*; 2B*; 9*; 10q26.3; 13*; 14p; 15p; 18*; 21*; 22p; Xq28; bo: Yp11.1-q12; Yq13/q14; su: Yq11.1-q13; Yq13/q14.1; Yq14.2	3ptel (HSA10q26.3); Xqtel (HSA Xq28)	1qmed (HSA1p21); 9qtel (HSA10q26.3); Xqtel (HSA Xq28)

BAC and PAC clones corresponding to each Euchr/Heterochr transition region were hybridized to metaphase spreads of human and non-human primate species. FISH results are listed for HSA (*H. sapiens*), PTR (*P. troglodytes*), PPA (*P. paniscus*), GGO (*G. gorilla*), PPYbo (*P. pygmaeus pygmaeus*), PPYsu (*P. pygmaeus abelii*), NLE (*N. leucogenys*), and MMU (*M. mulatta*). Broad signals comprising the centromere and short arm of an acrocentric chromosome are indicated with an asterisk (\*). The chromosomal designations for the great apes are given according to the human phylogenetic group (McConkey 2004), for NLE according to Roberto et al. (2007), and for MMU according to Rogers et al. (2006). The human orthologous regions for the latter two species are indicated in parentheses. Hybridization results for genomic clones from the Yq11.1/Yq11.21 region on human metaphase spreads were previously published in Kirsch et al. (2005).

white-cheeked crested gibbon and rhesus macaque. Identical results were obtained on metaphase spreads of all other Old World monkey species (*M. silenus*, *M. nemestrina*, *M. fascicularis*, *P. hamadryas*, *T. gelada*), thereby excluding species-specific variation. None of the human Y-derived probes indicated the presence of segmental duplications in the genome of New World monkeys (*C. jacchus*, *C. geoffroyi*, *A. geoffroyi*, *C. moloch*).

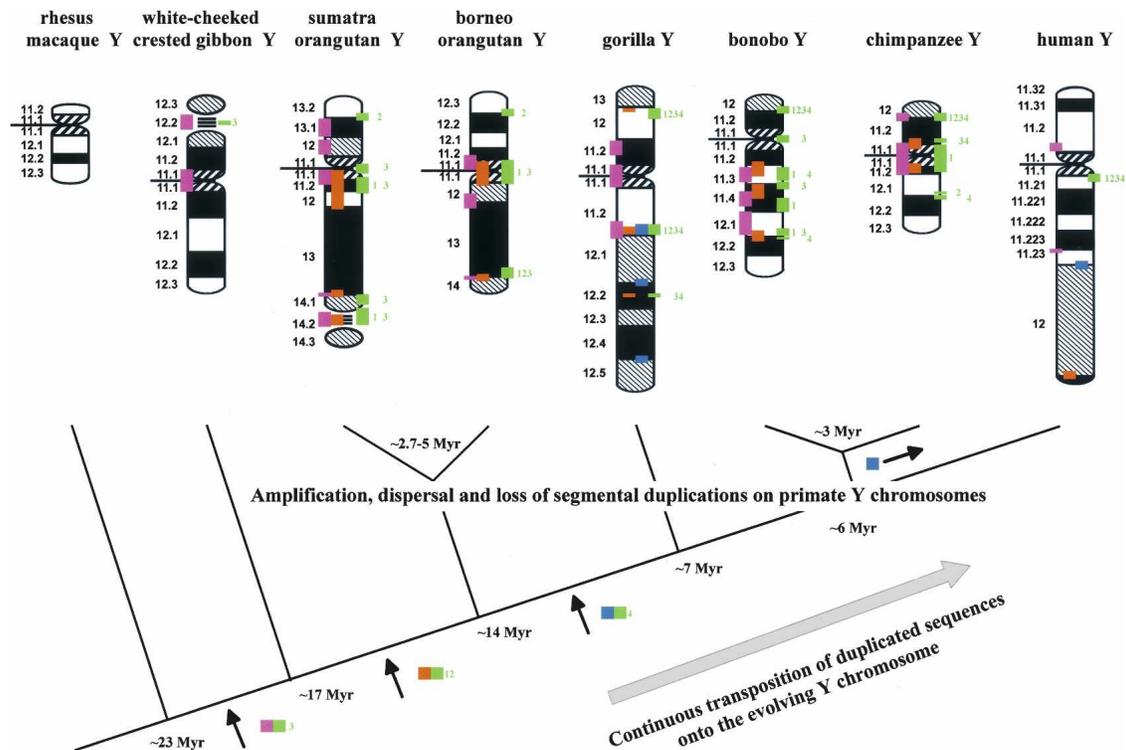
#### Direct comparison of *P. troglodytes* FISH and sequencing of SDs

The Y chromosome of the common chimpanzee (*P. troglodytes*) is the only non-human primate Y chromosome sequenced to date. To validate the comparative FISH experiments, we compared the sequence of human Y-chromosomal SDs from euchromatin/heterochromatin transition regions with the *P. troglodytes* chromosome Y reference assembly of NCBI Build 2.1.

#### Chimpanzee SDs to the human Yp11.2/Yp11.1 transition region

The human Yp11.2/Yp11.1 region exists in three complete copies on the chimpanzee Y, two of them (Yp: NW\_001252925; Yq: NW\_001252926) surrounding the Y centromere as components of a large inverted repeat structure (Supplemental Fig. 6A). The

separate copy (NW\_001252919) resides in distal chimpanzee Yp11.2. A further partial copy comprising the proximal 45 kb of the human Yp11.1/Yp11.2 region maps to proximal chimpanzee Yp11.2 (NW\_001252921). This distribution pattern is akin to the FISH signal pattern seen after hybridization of the BAC clones encompassing the human Yp11.2/Yp11.1 region to the chimpanzee Y chromosome (Fig. 3; Supplemental Fig. 1). The human Yp11.2/Yp11.1 region and all three homologous chimpanzee regions are enclosed by typical alpha-satellite DNA found near human chromosome centromeres (Supplemental Fig. 6B). Interestingly, the topology of the phylogenetic tree for the complete human and chimpanzee copies indicated the orthology of the human Yp11.2/Yp11.1 region to the distal Yp11.2 copy on the chimpanzee Y (Supplemental Fig. 6C). By molecular clock analysis, we calculated a rate of  $2.15 \times 10^{-9}$  nucleotide changes per base pair per year based on this orthology (assuming a divergence time of 6 million years [Myr] between chimpanzee and human lineages). We estimate that the first duplicative transposition occurred ~1.2 Mya with a second duplication inverting a larger genomic segment ~880,000 yr ago. The partial copy disassociated from the Yq centromeric copy ~780,000 yr ago. All duplication events are dated after the separation of *P. troglodytes* and *P. paniscus* and therefore reflect species-specific rearrangements.



**Figure 3.** A diagram delineating the evolutionary dynamics of Y-chromosomal segmental duplications. A multi-step process is depicted leading to the current distribution pattern of duplicated sequences on the Y chromosomes of the great and lesser apes. Serial intrachromosomal duplicative transpositions and chromosomal rearrangements generated a mosaic pattern on all great ape Y chromosomes analyzed. Continuous interchromosomal transfer of duplicated cassettes provided the basis to develop such a complex structure. The upper row outlines G-banded ideograms for each of the primate Y chromosomes analyzed. Each colored rectangle on a primate Y chromosome indicates the presence of a discrete SD region: (pink) Yp11.2/Yp11.1; (green) Yq11.1/Yq11.21 (green numbers refer to numbered BAC clones [(1) RP1-85D24, (2) RP11-131M6, (3) RP11-886I11, (4) RP11-295P22] spanning the human Yq11.1/Yq11.21 transition region); (blue) Yq11.23/Yq12; (orange) Yq12/PAR2. The phylogenetic tree indicates the divergence time in millions of years for each species: ~6 Mya for the *Homo*–*Pan* clade split, ~3 Mya for chimpanzee–bonobo split, ~7 Mya for the gorilla, ~14 Mya for the orangutans; ~17 Mya for the gibbon, ~23 Mya for the macaque (Goodman 2005), and ~2.7–5 Mya for the Bornean–Sumatran orangutan split (Steiper 2006). Colored rectangles intermediate to the evolutionary branching points indicate the period of interchromosomal addition or deletion of the respective duplicated sequences.

#### Chimpanzee SDs to the human Yq11.1/Yq11.21 transition region

Overlapping chimpanzee Y-chromosomal BAC clones CH251-830G14 (AC156805), CH251-119H06 (AC172374), CH251-307J23 (AC183800), and CH251-549O17 (AC185324) completely span the orthologous distal part of the human Yq11.1/Yq11.21 region (NT\_113819; bp 265,819–452,044). The position of the orthologous sequences in the *P. troglodytes* NCBI Build 2.1 (central Yp11.2) shows no concordance to the FISH signal pattern on the chimpanzee Y (Yp11.2/p12; Yp11.1; Yq12.1/12.2) generated by hybridization of BAC clones spanning the orthologous human region (Fig. 3; Supplemental Fig. 2). The segment of the human Yq11.1/Yq11.21 region covered by RP11-295P22 (AC134879) is completely included in the sequenced orthologous chimpanzee region. Hence the triple signal pattern of this BAC clone points toward, at least, partial duplication of the underlying sequences. One additional chimpanzee BAC clone harboring the proximal 45 kb of the orthologous human Yq11.1/Yq11.21 region (CH251-262O17 [AC158542]) is not assigned to a particular region of the chimpanzee Y chromosome. Independent nucleotide change calculations for both Yq11.1/Yq11.21 segments were averaged, resulting in a rate of  $2.45 \times 10^{-9}$  nucleotide changes per base pair per year. Whereas the nucleotide exchange rates are comparable for the Yq11.1/Yq11.21 and Yp11.1/Yp11.2 regions, the indel rate is an

order of magnitude higher for the Yq11.1/Yq11.21 region ( $9.5 \times 10^{-9}$  indels/bp per yr) relative to the Yp11.1/Yp11.2 region ( $1.7 \times 10^{-9}$  indels/bp per yr).

#### Chimpanzee SDs to the human Yq11.23/Yq12 and Yq12/PAR2 transition region

BAC clones encompassing the human Yq11.23/Yq12 region detected neither orthologous sequences nor orthologous signals. The single BAC clone extending into the human Yq12/PAR2 region presented FISH signals in both chimpanzee Y chromosome pericentromeric regions, but orthologous sequences were not detected in the current chromosomal assembly.

#### Comparative in silico analysis of human Y-chromosomal SDs

We used two complementary methods—WGAC (Bailey et al. 2001) and WSSD (Bailey et al. 2002)—to analyze the segmental duplication content within the four human Y-chromosomal euchromatin/heterochromatin transition regions on human (build35), chimpanzee (pantro2), and macaque (rhema2). In total, 96.5% (836.2/866.5 kb) of the sequence showed evidence of recent duplication (Supplemental Table 4). The sequences of all four euchromatin/heterochromatin transition regions are part of highly identical alignments (>94% sequence identity) and likely

arose recently during human genome evolution as a result of gene conversion or duplicative transposition. By analyzing the distribution of underlying pairwise alignments (Supplemental Table 4), we found that 62% (124/200) of the pairwise alignments map within 5 Mb of a centromere, suggesting that they are part of the interchromosomal burst of pericentromeric duplications that occurred after the great ape divergence from Old World monkeys (Horvath et al. 2001; Jiang et al. 2007).

#### SDs to the human Yp11.2/Yp11.1 transition region

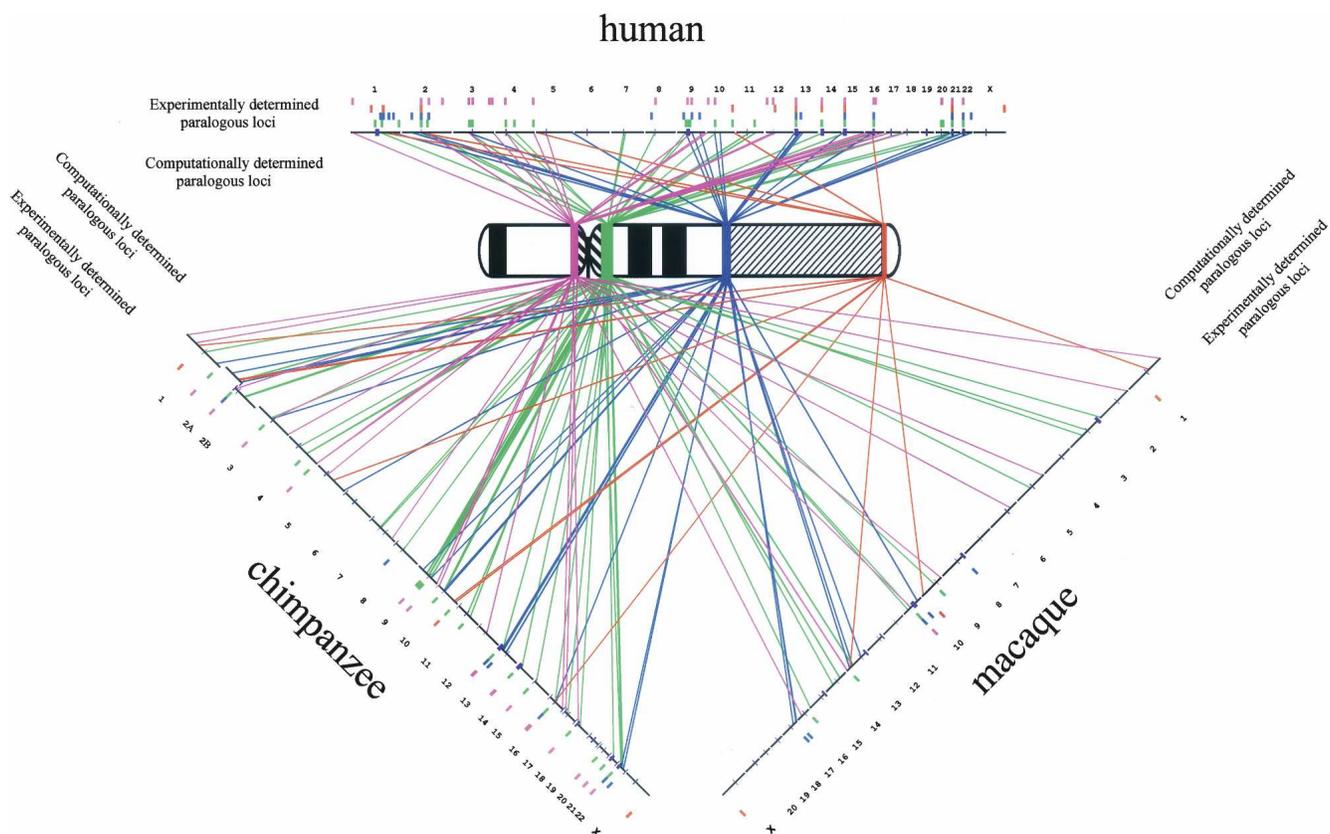
Among the four analyzed transition regions, the Yp11.2/Yp11.1 region is the only one not showing homologies to SDs on other chromosomes at high stringency conditions ( $\geq 95\%$  sequence identity and  $\geq 5$  kb in length). Reducing the stringency to  $\geq 90\%$  and  $\geq 1$  kb (low stringency conditions) led to the detection of SDs on chromosomes 1, 2, 3, 4, 8, 9, 10, 16, and 18 (Fig. 4; Supplemental Table 4). As expected, highly similar observations were obtained applying the same thresholds to the chimpanzee genome (Fig. 4; Supplemental Table 5). Different from the finding on the former genomes, fewer significant SD matches were found at low stringency conditions in the macaque genome on chro-

somes 1, 2, 5, 6, 9, 10, 13, and 16, totaling 49 kb (Fig. 4; Supplemental Table 6), which corresponds to 42% of the human Yp11.1/Yp11.2 region. Decreasing the sequence identity threshold in human-rhesus macaque comparative WSSD to  $\geq 83.5\%$  by taking into account the overall sequence identity between human and macaque (93%) (Gibbs et al. 2007) identified additional significant SD matches on chromosomes 1, 3, 5, 6, 8, 9, 15, and 19.

Taken together, the SDs of the human Yp11.1/Yp11.2 region are duplicated, on average, almost four times in the macaque genome and cluster in subtelomeric locations of chromosomes 1, 2, 5, 6, 8, 9, 13, and 15.

#### SDs to the human Y chromosome long arm transition regions

SDs corresponding to the human Yq11.1/Yq11.21 region were found to be largely consistent in the human and chimpanzee genomes (Fig. 4; Supplemental Tables 4, 5; Kirsch et al. 2005). Only three homologies to chromosomes 1, 11, and 14 could not be located in the chimpanzee genome under low stringency conditions. There is evidence within the macaque genome of much shorter SDs for syntenic regions on chromosomes 2, 3, 5, 10, 14, and 15 (Fig. 4; Supplemental Table 6). A total of 196 kb (43%) of



**Figure 4.** Comparative paralogous duplication pattern of human Y-chromosomal segmental duplications. An ideogram of the G-banded pattern of the normal human Y is shown in the *middle*. The colored boxes on the chromosome represent the euchromatin/heterochromatin transition regions composed of segmental duplications. The autosomes and the X chromosome are illustrated as small horizontal black lines *above* (human), *rightward* diagonally disposed (chimpanzee), and *leftward* diagonally disposed (macaque). Each chromosome is tagged by the corresponding chromosome number. Heterochromatic regions (constitutive heterochromatin, telomeric caps, and NORs) are indicated as tiny purple boxes on the horizontal lines. All diagonal lines represent pairwise alignments of  $\geq 1$  kb and  $\geq 90\%$  nucleotide identity identified by whole-genome sequence comparisons. The small colored boxes *in-between* the chromosomes and their corresponding numbers display the multi-site pattern from comparative FISH. Concerning the Yq11.1/Yq11.21, Yq11.23/Yq12, and Yq12/PAR2 region, experimental and computational results are 48% consistent for the human, 45% for the chimpanzee, and 30% for the macaque. For the Yp11.1/Yp11.2 region, the concordance decreases to 21% (human), 23% (chimpanzee), and 11% (macaque). Please note that paralogies detected by whole-genome sequence comparisons do not correspond to ancestral duplication locations.

aligned sequence of the Yq11.1/Yq11.21 region was identified at low stringency conditions. These SDs are generally located in the same Yq11.1/Yq11.21 region as those identified in human and chimpanzee.

Many (93%, 219.3/235.6 kb) of the Yq11.23/Yq12-region SDs map to human chromosomes 2, 9, 10, 13, 15, 16, 21, and 22 (Fig. 4; Supplemental Table 4). Similar results were obtained on the chimpanzee genome, with the exception of chromosomes 15 and 21 (Fig. 4; Supplemental Table 5). On the macaque genome, 76% (200.7/235.6 kb) of this region matched to three autosomal regions (chromosomes 10, 13, and 17), indicating that these SDs existed prior to macaque divergence (Fig. 4; Supplemental Table 6). The distal part of the Yq12/PAR2 region mapped to human and chimpanzee chromosome 10q26 and the syntenic region on macaque chromosome 9.

Taken together, SDs from the Yq11.1/Yq11.21, Yq11.23/Yq12, and Yq12/PAR2 regions show a preponderance to accumulate in pericentromeric regions of the chimpanzee and human genome, whereas they are predominantly interspersed in the macaque genome.

### Comparative FISH and comparative in silico analysis

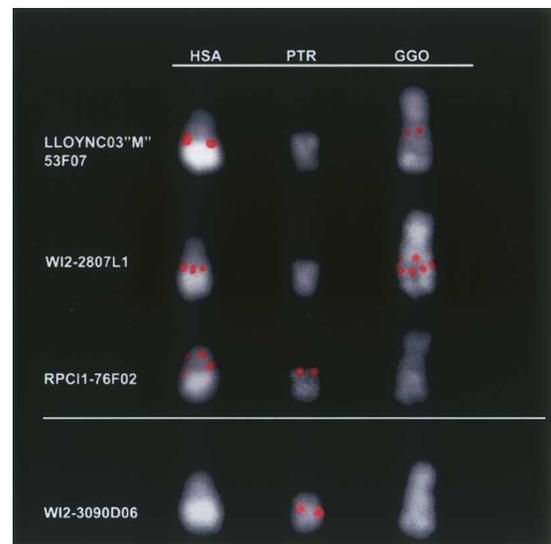
Contrasting the computational with the experimental results (Fig. 4) illustrates the complementary nature of these two approaches. In particular, 97 SD-carrying chromosomal regions were detected on the human genome, 72 on the chimpanzee genome, and 33 on the macaque genome. Although 41 of the human SD locations were identified by both methods, 35 SD locations were detected solely by FISH, and 21 locations could only be identified by computational methods. Similar results were observed in the chimpanzee (27/16/29) and macaque (8/5/20) genome with a shift in favor of the whole-genome comparison method.

### Duplication shadowing next to Y-chromosomal SDs

Since new lineage-specific SDs preferentially map near shared ancestral duplications (duplication shadowing; for definition, see Cheng et al. 2005 and Newman et al. 2005), we investigated the sequences flanking the non-palindromic SDs from all four human Y-chromosomal euchromatin/heterochromatin transition regions.

Only one (Yq11.23/Yq12) of the four transition regions showed indications of duplication shadowing. A genomic segment of 205 kb (NT\_011903: base pairs 4,391,308–4,597,035) proximal to the Yq11.23/Yq12 transition region displayed paralogy to HSA1 over its entire length, and a much smaller segment was paralogous to HSA15. The basic duplicon architecture consists of a 190-kb duplicon derived from HSA 1q32.1 fused to a 15-kb duplicon originating from HSA 1p36.2. A detailed description of the evolutionary history of this genomic segment is given in Supplemental File 3.

To track the evolutionary movements of both duplicons within the great ape lineage, we performed comparative FISH with the HSA1p36.2-fosmid WI2-2807L1 (G248P88697F1) and the HSA1q32.1-paralogous HSAYq11.23-cosmid LLNLYC03'M'53F07 on metaphase chromosomes of human and great apes (Fig. 5; Supplemental Table 7). Both genomic clones showed the expected signal pattern on human metaphase spreads, as deduced from the paralogous sequence variants. Slightly different numbers of duplication events were noted on great ape metaphase chromosomes (Supplemental Table 7), whereas the Y signal oc-



**Figure 5.** Comparative great ape FISH of individual duplicons. Y-chromosomal species-specific differences are shown on the extracted Y chromosomes of HSA (*H. sapiens*), PTR (*P. troglodytes*), and GGO (*G. gorilla*). Comparative FISH experiments were performed with probes of duplicons from 1q32.1 (LLOYNC03'M'53F07), 1p36.2 (WI2-2807L1), 1q43 (RPCI1-76F02) (Wimmer et al. 2002), and 21q21.3 (WI2-3090D6). Autosomal paralogous signals are summarized in Supplemental Table 7. In all experiments, the signals obtained on metaphase chromosomal spreads of the pygmy chimpanzee were concordant to those of the common chimpanzee.

currence showed a striking similarity to the human Yq11.23/Yq12 region. This Y-chromosomal hybridization pattern reflects the opposite pattern to a human Yq11.1-derived PAC clone RPCI1-76F02 presented in earlier observations (Wimmer et al. 2002).

To ascertain the existence of lineage-specific Y-chromosomal duplicons for the great apes, we investigated the *P. troglodytes* chromosome Y reference assembly with respect to sequences absent on the human Y chromosome. The BAC clone CH251-267E22 (AC147148) harbors part of a 133-kb duplicon present on human chromosomes 2, 13, 18, and 21, but not on Y. The HSA21 fosmid WI2-3090D6 (G248P800795B3) derived from the human ancestral 21q21.3 duplicon yielded no Y-signal on human and gorilla metaphase spreads although it was present on chimpanzee Y (Fig. 5). In all hybridizations, the signal pattern in pygmy and common chimpanzee was identical, and the orangutan showed no duplication events (Supplemental Table 7), although according to the calculated time of the primary duplication event (15.9 Myr), all great apes and the human should harbor the Y-derived duplicon (Supplemental Table 8). Surprisingly, phylogenetic analysis points toward two independent seeding events in the great ape lineage (Supplemental Fig. 8B). Nevertheless, in general, the calculated seeding and swapping times for the duplicons that we investigated are consistent with the occurrence of derived duplicons on the respective great ape chromosomes (Supplemental Fig. 8).

## Discussion

Basically, the male-specific region (MSY) of the human Y chromosome is a mosaic of heterochromatic sequences and three discrete classes of euchromatic sequences: X-transposed, X-

degenerate, and ampliconic (Skaletsky et al. 2003). Recently, we characterized an additional euchromatic sequence from the Yq11.1/Yq11.21 euchromatin/heterochromatin transition region that is almost completely composed of interchromosomal segmental duplications (Kirsch et al. 2005). Thorough inspection of the other three euchromatin/heterochromatin transition regions (Yp11.2/Yp11.1, Yq11.23/Yq12, Yq12/PAR2) revealed the existence of similarly complex structured sequences. Here we describe the comprehensive evolutionary analysis of this discrete class of Y-chromosomal euchromatic sequences: the non-palindromic segmental duplications. A combination of comparative computational and experimental techniques was applied to provide systematic information on the evolutionary origin, chromosomal distribution, and duplication timing of Y-chromosomal duplicons. Although the majority of original duplications can be assigned to a time period equivalent to the proposed hominoid burst of segmental duplications (10–20 Mya) (Horvath et al. 2005), we find that a substantial fraction originated before that time. Moreover, our evolutionary reconstruction identified species-specific duplicated sequences on the Y chromosome, especially among the African great apes.

Detailed analyses of pericentromeric regions of several human chromosomes (Eichler et al. 1996; Jackson et al. 1999; Horvath et al. 2000, 2003, 2005; Bailey et al. 2001, 2002; Crosier et al. 2002; She et al. 2004; Kirsch et al. 2005; Locke et al. 2005) have illuminated a general principle of human genome evolution: complex mosaics of segmental duplications originating from diverse euchromatic chromosomal regions are created by recurrent duplicative transpositions of diverse euchromatic segments into pericentromeric regions (Jackson et al. 1999; Guy et al. 2000, 2003; Horvath et al. 2000, 2005; Locke et al. 2005). Differential dispersal of larger duplicon cassettes among human and great ape pericentromeric regions subsequently leads to lineage-specific quantitative and qualitative differences. This phenomenon not only accounts for both pericentromeric regions of the human Y chromosome, but also for both of its heterochromatin-bordering regions. The segmental duplication architecture of all four regions was unraveled by both computational and experimental methods.

Whereas extensive concordance for both data sets was observed in pericentromeric and Yq12/PAR2 regions, differences were particularly noted in the Yq11.23/Yq12 region. This region is unique in its structure among the four euchromatin/heterochromatin transition regions, as it is composed of nested recurrent TPTE and SLC25A15 duplicons originating from different long arm regions of human chromosome 13 (Fig. 2). Intrachromosomal duplications of the ancestral duplicons might have created a genomic environment susceptible to excessive occurrence of paracentric inversions and/or gene conversions, thereby explaining the difficulties in accurately predicting ancestral duplicons in this region. Additionally, the frequent occurrence of genomic insertions and deletions in humans (Frazer et al. 2003; Korbel et al. 2007) may have led to the development of duplicon-specific structural variants in the intrachromosomal duplicates. This would result in the disruptions of larger homologous synteny blocks, which are prerequisites for the correct operating mode of the *DupMasker* software (Jiang et al. 2008).

Phylogenetic sequence comparison and comparative FISH analysis confirmed earlier observations that the majority of ancestral duplicons occurred between 10 and 25 Mya (Horvath et al. 2005). Nevertheless, we date almost one-third of all derivative duplicons (9/28) for which the ancestral duplicon in the rhesus

macaque could be unequivocally determined, to a pericentromeric seeding event that occurred earlier than 23 Mya. These more ancient duplications are enriched within the Yp11.2/Yp11.1 region and the proximal part of the Yq11.1/Yq11.21 region (Supplemental Table 1). Interestingly, the Yp11.2/Yp11.1-derivative duplicons are already amplified in the rhesus macaque genome and are predominantly located in subtelomeric regions. In contrast, duplicons from the Y long arm transition regions are primarily single-copy in the rhesus macaque and preferentially map to pericentromeric regions in chimpanzee and human.

What gave rise to this apparent shift in duplicative transposition events, that is, from subtelomeric in Old World monkeys to pericentromeric in the human/ape lineage? Although this remains largely an open question, one could imagine that crucial evolutionary changes in the chromatin configuration or epigenetic modifications might have altered the susceptibility of these regions to acquire duplicated sequences. In this respect, it is noteworthy that this accessibility shift from subtelomeric to pericentromeric regions coincides with two major changes in human chromosome evolution: the shortening of the subtelomeric regions (Gardner et al. 2007) and the development of higher-order alpha-satellites (Haaf and Willard 1998; Alkan et al. 2007). It is unclear if both processes happened independently or simultaneously to switch the preferred chromosomal region for the acquisition of duplicated sequences. This question illustrates the need for high-quality BAC-based sequences of these regions from humans and non-human primates, as valuable conclusions could be drawn regarding the evolutionary forces acting on chromosome evolution.

We note, however, that we cannot exclude the possibility of under-representation or absence of pericentromeric duplication regions within the whole-genome shotgun assembly of the macaque. However, both subtelomeric and pericentromeric duplications should be equally difficult targets for sequence assembly.

Our comparative genome-wide analysis clearly illustrates the complementary nature of FISH and genome-wide sequence comparisons to characterize segmental duplications in human and non-human primate genomes. Whereas WGAC/WSSD can detect duplications within an assembly, it requires the sequence to be represented at least once. FISH enables the detection of paralogous segments in as-yet-unsequenced regions of a given primate genome. Given the whole-genome shotgun assembly is currently favored over the BAC-based genome assembly, FISH will remain an important method to decipher the SD content of chromosomal regions particularly enriched for duplicated sequences (short arms of acrocentric chromosomes, euchromatin/heterochromatin transition regions, ancestral centromeres, evolutionary chromosomal fusion and fission sites). Such regions tend to be misassembled or absent within WGS and, therefore, under-represented by computational approaches. Recent attempts (Lyle et al. 2007) address this problem, thereby providing valuable evolutionary information on these regions. As such efforts rely on monochromosomal somatic cell hybrids, they do not take into account structural variation of these regions among different individuals.

Our combined cytogenetic and BAC-based sequence approach not only enabled us to reconstruct the evolutionary dynamics of non-palindromic SDs on the primate Y chromosomes (Fig. 3), but also identified species-specific acquisition of Y-chromosomal SDs in the great apes (Fig. 5). Our results do not contradict earlier observations (Archidiacono et al. 1998) as YAC clones used for comparative FISH on primate Y chromosomes

do not comprise the human Y-chromosomal euchromatin/heterochromatin transition regions (Affara et al. 1996; Kirsch et al. 1996). On all higher primate Y chromosomes, the propensity to accumulate SDs in euchromatin/heterochromatin transition regions is clearly visible. Interestingly, the gorilla Y with its fragmented heterochromatin shows SDs in six out of eight euchromatin/heterochromatin transition regions. The absence of SDs in two transition regions might be explained by the use of human Y-chromosomal BAC clones as probes. It is clear that not all SDs detected on great ape Y chromosomes are shared across all great ape Y chromosomes. The search for indications of duplication shadowing near the non-palindromic SDs supports the rapid divergence and substantial change in sequence content of primate Y chromosomes over short evolutionary times. The sequencing of the Y chromosomes among humans and other great apes will undoubtedly reveal additional Y-derivative duplicons that could serve as genomic markers for subspecies identification in the great apes. Human Y-chromosomal SDs, for example, are absent on the rhesus macaque Y chromosome as the Old World monkeys diverged at the onset of duplication seeding. However, one might expect to find a series of Old World monkey-specific duplicons if an Old World monkey Y chromosome was sequenced. Our BLASTN comparative sequence analyses favor this assumption as the rhesus macaque Y-chromosomal BAC clone CH250-164H2 (AC213058) carries a euchromatic segment paralogous to human 10p15.3, which is duplicated in the rhesus macaque, maps preferentially to subtelomeric locations, but is not duplicated in human or common chimpanzee. Evidence for other species-specific differences is supported by our comparative FISH analysis of closely related gibbon species (Supplemental Figs. 1, 5).

Such rare genomic changes based on structural variations of duplicated sequences might not only provide a valuable genomic tool to distinguish closely related species with highly similar autosomal complements, but also promises to shed some light on the evolution of our species. In the genus *Pan*, there has been substantial reorganization and amplification of duplicated sequences in the euchromatic portion of the long arm of the pygmy chimpanzee Y chromosome (Fig. 3). This may relate to evolutionary changes in transcriptional activity of fertility factors, thereby altering networks and pathways associated with male reproduction. Such changes could enhance the development of reproductive barriers during a time period of possible back hybridization in populations of diverging species and, in conjunction with geographic separation, could promote speciation (for review, see Hey et al. 2005).

## Methods

### Blood samples and cell lines

Blood samples of seven individuals of the chimpanzee (*P. troglodytes*), the pygmy chimpanzee (*P. paniscus*), and the lowland gorilla (*Gorilla gorilla gorilla*), three individuals of the Sumatran orangutan (*Pongo pygmaeus abelii*), and two individuals of the Bornean orangutan (*Pongo pygmaeus pygmaeus*) were obtained from the Zoologisch-Botanischer Garten Wilhelma Stuttgart (Germany), Twycross Zoo (United Kingdom), Zoo Leipzig (Germany), Apenheul (Netherlands), Tierpark Hellabrunn München (Germany), and Zoo Duisburg (Germany). Blood samples of three individuals of the white-handed gibbon (*H. lar*) were obtained from the Zoologisch-Botanischer Garten Wilhelma Stuttgart (Germany) and one sample of the Bornean gibbon (*H. muelleri*) was from the Zoo Münster (Germany). Blood samples of three

individuals of the rhesus macaque (*M. mulatta*), two individuals of the pig-tailed macaque (*M. nemestrina*), and one individual of the hamadryas baboon (*P. hamadryas*) were obtained from the Deutsches Primaten Zentrum Göttingen (Germany). Samples of two individuals of the Gelada baboon (*T. gelada*) were from the Zoologisch-Botanischer Garten Wilhelma Stuttgart (Germany). A blood sample from one individual each of the lion-tailed macaque (*M. silenus*) and the crab-eating macaque (*M. fascicularis*) were obtained from the Zoo Leipzig (Germany) and the Zoo Münster (Germany), respectively. A blood sample of the black-handed spider monkey (*A. geoffroyi*) was from the Zoologisch-Botanischer Garten Wilhelma Stuttgart (Germany). Samples from three individuals of the common marmoset (*C. jacchus*) were obtained from the Zoo Münster (Germany).

Lymphoblastoid cell lines of the white-cheeked crested gibbon (*N. leucogenys*) and the common marmoset (*C. jacchus*) were kindly provided by S. Müller (Munich).

Skin tissue of a dusky titi monkey (*C. moloch*) from the Zoologisch-Botanischer Garten Wilhelma Stuttgart (Germany) was used to establish a fibroblast cell line. A fibroblast cell line of a white-headed marmoset (*C. geoffroyi*) was kindly provided by S. Müller (Munich).

### Fluorescence in situ hybridization (FISH)

FISH analysis of chromosomal metaphase spreads derived from lymphocytes or lymphoblastoid and fibroblast cell lines was performed from unrelated human and non-human primate males. Prior to FISH, the slides were treated with RNase followed by pepsin digestion as described (Ried et al. 1992). FISH followed the method described by Schempp et al. (1995). Chromosome in situ suppression was applied to clones from human (RPC11, RPC111, WI-2) and Y chromosome-specific (LLNLOYCO3" M") genomic libraries. Human whole-chromosome painting (WCP) libraries (Jauch et al. 1992) were used to unequivocally assign hybridizing signals to syntenic regions in lesser apes and Old World monkeys. pMR100, a mouse-derived rDNA-containing plasmid, was used to mark the Old World monkey marker chromosome. After FISH, the slides were counterstained with DAPI (0.14 µg/mL) and mounted in Vectashield (Vector Laboratories). Preparations were evaluated using a Zeiss Axiophot epifluorescence microscope equipped with single-bandpass filters for excitation of red, green, and blue (Chroma Technologies). During exposures, only excitation filters were changed allowing for pixel-shift-free image recording. Images of high magnification and resolution were obtained using a black-and-white CCD camera (Photometrics Kodak KAF 1400; Kodak) connected to the Axiophot. Camera control and digital image acquisition involved the use of an Apple Macintosh Quadra 950 computer.

### Segmental duplication analysis

Duplicated sequences of 866 kb were compared to the human (build35), chimpanzee (pantro2), and macaque (rhema2) assemblies, whereas the threshold for segmental duplication detection (segdup) was set at 90% sequence identity with a length  $\geq 1$  kb. The results were subsequently parsed at different length and percent identity thresholds. Segmental duplications in the chimpanzee and macaque whole-genome assemblies were detected based on the depth of random shotgun sequencing read mapped to the query sequences. Simultaneously, the divergence among the mapped reads (5-kb gap-free, repeat-free intervals) were computed. WGAC and WSSD analyses have been run separately on the FASTA results. The results from these two processes were analyzed independently and then compared. The evolutionary ge-

netic distance for multiple substitutions was corrected using a two-parameter model (Kimura 1980).

### Duplicon identification

Ancestral duplicons were delineated by two independent methods. In the first approach, identification was conducted for each individual segmental duplication region by using RepeatMasker (RepeatMasker version 3.1.7; A. Smit and P. Green, <http://www.repeatmasker.org>) sequence as query against the nr/nt collection of GenBank. This procedure identified minimal evolutionary shared segments and conserved exon/intron structures characterized by different frequencies of occurrence and dispersal in the human genome. The alignment breakpoints between adjoining derivative duplicons were determined by non-masked query against the nr/nt collection of GenBank. For each duplicon, the genic paralogs were identified, and the paralog carrying the complete exon-intron complement defined as the ancestral duplication fragment. In those cases in which no derived gene structure could be detected, the most divergent paralog was defined as the ancestral duplicon. Supporting evidence was drawn from human-macaque alignments produced by BLASTN comparison of the putative human ancestral duplicons. As the majority of duplications found on the human Y have emerged since the divergence from the macaque lineage, this intra-order comparison facilitates this process. Once identified, the macaque loci were cross-referenced against the human genome (build 35) to verify the human ancestral duplicon delineation. As relatively short duplicons tend to remain undetected by these methods, we also performed *DupMasker* analysis (Jiang et al. 2008) to identify additional potential ancestral sequences for segmental duplications contained within the sequence. The software annotates each duplicon based on the duplicon database generated by Jiang et al. (2007).

### Phylogenetic analysis

FASTA formatted sequence files used to generate phylogenetic trees were extracted from the GenBank accession. Sequence alignments were built by using CLUSTALW (version 1.82) (Higgins et al. 1996), and neighbor-joining phylograms created by using MEGA (Molecular Evolutionary Genetic Analysis) v3.1 (<http://www.megasoftware.net>; Kumar et al. 2004). Neighbor-joining analysis was used with complete deletion parameters and bootstrap (1000 iterations) to provide confidence of each branching point in the phylogenetic trees. Similar phylogenetic trees were yielded by minimum evolution and maximum parsimony methods, but neighbor-joining methods were chosen as they are amenable to calculate divergence times between sequence taxa. We estimated the number of substitutions/site per year by correcting the divergence times for multiple substitutions using Kimura's two-parameter model (Kimura 1980). As the rates of nucleotide substitution vary for pseudogenic sequences, the rate of nucleotide substitution was calibrated based on orthologous sequence comparisons using a divergence of 23 Mya for macaque-human and 6 Mya for chimpanzee-human divergence. Duplication timing events were calculated by using the equation  $r = k/2T$  (Li 1997), where  $r$  is the rate of nucleotide changes per base pair per year,  $k$  is the distance calculated between the ancestral and chromosome Y sequences, and  $T$  is the time of divergence of the molecules.

### Acknowledgments

We thank Christine Hodler for technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft (Sche

214/8-1) in Germany. This work was supported by NIH grant GM058815 to E.E.E. and a Rosetta Inpharmatics Fellowship (Merck Laboratories) to Z.J. E.E.E. is an investigator of the Howard Hughes Medical Institute.

### References

- Affara, N., Bishop, C., Brown, W., Cokke, H., Davey, P., Ellis, N., Graves, J.M., Jones, M., Mitchell, M., Rappold, G., et al. 1996. Report of the Second International Workshop on Y Chromosome Mapping 1995. *Cytogenet. Cell Genet.* **73**: 33-76.
- Alkan, C., Ventura, M., Archidiacono, N., Rocchi, M., Sahinalp, S.C., and Eichler, E.E. 2007. Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data. *PLoS Comput. Biol.* **3**: e181. doi: 10.1371/journal.pcbi.0030181.
- Archidiacono, N., Storlazzi, C.T., Spalluto, C., Ricco, A.S., Marzella, R., and Rocchi, M. 1998. Evolution of chromosome Y in primates. *Chromosoma* **107**: 241-246.
- Bailey, J.A. and Eichler, E.E. 2003. Genome-wide detection and analysis of recent segmental duplications within mammalian organisms. *Cold Spring Harb. Symp. Quant. Biol.* **68**: 115-124.
- Bailey, J.A. and Eichler, E.E. 2006. Primate segmental duplications: Crucibles of evolution, diversity and disease. *Nat. Rev. Genet.* **7**: 552-564.
- Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J., and Eichler, E.E. 2001. Segmental duplications: Organization and impact within the current human genome project assembly. *Genome Res.* **11**: 1005-1017.
- Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., and Eichler, E.E. 2002. Recent segmental duplications in the human genome. *Science* **297**: 1003-1007.
- Bailey, J.A., Baertsch, R., Kent, W.J., Haussler, D., and Eichler, E.E. 2004. Hotspots of mammalian chromosomal evolution. *Genome Biol.* **5**: R23. <http://genomebiology.com/2004/5/4/R23>.
- Birtle, Z., Goodstadt, L., and Ponting, C. 2005. Duplication and positive selection among hominin-specific PRAME genes. *BMC Genomics* **6**: 120. doi: 10.1186/1471-2164-6-120.
- Cheng, Z., Ventura, M., She, X., Khaitovich, P., Graves, T., Osoegawa, K., Church, D., De Jong, P., Wilson, R.K., Pääbo, S., et al. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **437**: 88-93.
- Cheung, V.G., Nowak, N., Jang, W., Kirsch, I.R., Zhao, S., Chen, X.N., Furey, T.S., Kim, U.J., Kuo, W.L., Olivier, M., et al. 2001. Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* **437**: 69-87.
- Cheung, V.G., Estivill, X., Khaja, R., MacDonald, J.R., Lau, K., Tsui, L.C., and Scherer, S.W. 2003. Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biol.* **4**: R25. doi: 10.1186/gb-2003-4-4-r25.
- Crosier, M., Viggiano, L., Guy, J., Miscoco, D., Stones, R., Wei, W., Hearn, T., Ventura, M., Archidiacono, N., Rocchi, M., et al. 2002. Human paralogs of KIAA0187 were created through independent pericentromeric-directed and chromosome-specific duplication mechanisms. *Genome Res.* **12**: 67-80.
- Eichler, E.E. 2001. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet.* **11**: 661-669.
- Eichler, E.E. and DeJong, P.J. 2002. Biomedical applications and studies of molecular evolution: A proposal for a primate genomic library resource. *Genome Res.* **12**: 673-678.
- Eichler, E.E., Lu, F., Shen, Y., Antonacci, R., Jurecic, V., Doggett, N.A., Moyzis, R.K., Baldini, A., Gibbs, R.A., and Nelson, D.J. 1996. Duplication of a gene-rich cluster between 16p11.1 and Xq28: A novel pericentromeric-directed mechanism for paralogous genome evolution. *Hum. Mol. Genet.* **5**: 899-912.
- Eichler, E.E., Budarf, M.L., Rocchi, M., Deaven, L.L., Doggett, N.A., Baldini, A., Nelson, D.L., and Mohnenweiser, H.W. 1997. Interchromosomal duplications of the adrenoleukodystrophy locus: A phenomenon of pericentromeric plasticity. *Hum. Mol. Genet.* **6**: 991-1002.
- Florea, L., Hartzell, G., Zhang, Z., Rubin, G.M., and Miller, W. 1998. A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.* **8**: 967-974.
- Frazer, K.A., Chen, X., Hinds, D.A., Pant, P.V., Patil, N., and Cox, D.R. 2003. Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates. *Genome Res.* **13**: 341-346.
- Gardner, J.P., Kimura, M., Chai, W., Durrani, J.F., Tchakmakjian, L.,

- Cao, X., Lu, X., Li, G., Peppas, A.P., Skurnick, J., et al. 2007. Telomere dynamics in macaques and humans. *J. Gerontol. A Biol. Sci. Med. Sci.* **62**: 367–374.
- Gibbs, R.A., Rogers, J., Katze, M.G., Bumgarner, R., Weinstock, G.M., Mardis, E.R., Remington, K.A., Strausberg, R.L., Venter, J.C., Wilson, R.K., et al. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**: 222–234.
- Goodman, M. 1999. The genomic record of humankind's evolutionary roots. *Am. J. Hum. Genet.* **64**: 31–39.
- Goodman, M. 2005. Moving primate genomics beyond the chimpanzee genome. *Trends Genet.* **21**: 511–517.
- Guy, J., Spalluto, C., McMurray, A., Hearn, T., Crosier, M., Viggiano, L., Miolla, V., Archidiacono, N., Rocchi, M., Scott, C., et al. 2000. Genomic sequence and transcriptional profile of the boundary between pericentromeric satellites and genes on human chromosome arm 10q. *Genome Res.* **12**: 2029–2042.
- Guy, J., Hearn, T., Crosier, M., Mudge, J., Viggiano, L., Koczan, D., Thiesen, H.J., Bailey, J.A., Horvath, J.E., Eichler, E.E., et al. 2003. Genomic sequence and transcriptional profile of the boundary between pericentromeric satellites and genes on human chromosome arm 10p. *Genome Res.* **13**: 159–172.
- Haaf, T. and Willard, H.F. 1998. Orangutan alpha-satellite monomers are closely related to the human consensus sequence. *Mamm. Genome* **9**: 440–447.
- Hey, J., Fitch, W.M., and Ayala, F.J. 2005. Systematics and the origin of species: An introduction. *Proc. Natl. Acad. Sci.* **102** (Suppl. 1): 6515–6519.
- Higgins, D.G., Thompson, J.D., and Gibson, T.J. 1996. Using CLUSTALW for multiple sequence alignments. *Methods Enzymol.* **266**: 383–402.
- Horvath, J.E., Schwartz, S., and Eichler, E.E. 2000. The mosaic structure of human pericentromeric DNA: A strategy for characterizing complex regions of the human genome. *Genome Res.* **10**: 839–852.
- Horvath, J.E., Bailey, J.A., Locke, D.P., and Eichler, E.E. 2001. Lessons from the human genome: Transitions between euchromatin and heterochromatin. *Genome Res.* **11**: 2215–2223.
- Horvath, J.E., Gulden, C.L., Bailey, J.A., Yohn, C., McPherson, J.D., Prescott, A., Roe, B.A., de Jong, P.J., Ventura, M., Misceo, D., et al. 2003. Using a pericentromeric repeat to recapitulate the phylogeny and expansion of human centromeric segmental duplications. *Mol. Biol. Evol.* **20**: 1463–1479.
- Horvath, J.E., Gulden, C.L., Vallente, R.U., Eichler, M.Y., Ventura, M., McPherson, J.D., Graves, T.A., Wilson, R.K., Schwartz, S., Rocchi, M., et al. 2005. Punctuated duplication seeding events during the evolution of human chromosome 2p11. *Genome Res.* **15**: 914–927.
- Inoue, K. and Lupski, J.R. 2002. Molecular mechanisms for genomic disorders. *Annu. Rev. Genomics Hum. Genet.* **3**: 199–242.
- Jackson, M.S., Rocchi, M., Thompson, G., Hearn, T., Crosier, M., Guy, J., Kirk, D., Mulligan, L., Ricco, A., Piccininni, S., et al. 1999. Sequences flanking the centromere of human chromosome 10 are a complex patchwork of arm-specific sequences, stable duplications and unstable sequences with homologies to telomeric and other centromeric locations. *Hum. Mol. Genet.* **8**: 205–215.
- Jauch, A., Wienberg, J., Stanyon, R., Arnold, N., Tofanelli, S., Ishida, T., and Cremer, T. 1992. Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting. *Proc. Natl. Acad. Sci.* **89**: 8611–8615.
- Ji, Y., Eichler, E.E., Schwartz, S., and Nicholls, R.D. 2000. Structure of chromosomal duplicons and their role in mediating human genomic disorders. *Genome Res.* **10**: 597–610.
- Jiang, Z., Tang, H., Ventura, M., Cardone, M.F., Marques-Bonet, T., She, X., Pevzner, P.A., and Eichler, E.E. 2007. Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. *Nat. Genet.* **39**: 1361–1368.
- Jiang, Z., Hubley, R., Smit, A., and Eichler, E.E. 2008. *DupMasker*: A tool for annotating primate segmental duplications. *Genome Res.* (in press) doi: 10.1101/gr.078477.108.
- Johnson, M.E., Viggiano, L., Bailey, J.A., Abdul-Rauf, M., Goodwin, G., Rocchi, M., and Eichler, E.E. 2001. Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413**: 514–519.
- Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**: 111–120.
- Kirsch, S., Keil, R., Edelmann, A., Henegariu, O., Hirschmann, P., LePaslier, D., and Vogt, P.H. 1996. Molecular analysis of the genomic structure of the human Y chromosome in the euchromatic part of its long arm (Yq11). *Cytogenet. Cell Genet.* **75**: 197–206.
- Kirsch, S., Weiß, B., Miner, T.L., Waterston, R.H., Clark, R.A., Eichler, E.E., Münch, C., Schempp, W., and Rappold, G. 2005. Interchromosomal segmental duplications of the pericentromeric region on the human Y chromosome. *Genome Res.* **15**: 195–204.
- Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L., et al. 2007. Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**: 420–426.
- Kumar, S., Tamura, K., and Nei, M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5**: 150–163.
- Kuroda-Kawaguchi, T., Skaletsky, H., Brown, L.G., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Silber, S., Oates, R., Rozen, S., et al. 2001. The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat. Genet.* **29**: 279–286.
- Li, W. 1997. *Molecular evolution*. Sinauer Associates, Sunderland, MA.
- Locke, D.P., Jiang, Z., Pertz, L.M., Misceo, D., Archidiacono, N., and Eichler, E.E. 2005. Molecular evolution of the human chromosome 15 pericentromeric region. *Cytogenet. Genome Res.* **108**: 73–82.
- Lyle, R., Prandini, P., Osoegawa, K., ten Hallers, B., Humphray, S., Zhu, B., Eyraes, E., Castelo, R., Bird, C.P., Gagos, S., et al. 2007. Islands of euchromatin-like sequence and expressed polymorphic sequences within the short arm of human chromosome 21. *Genome Res.* **17**: 1690–1696.
- McConkey, E.H. 2004. Orthologous numbering of great ape and human chromosomes is essential for comparative genomics. *Cytogenet. Genome Res.* **105**: 157–158.
- Mefford, H.C. and Trask, B.J. 2002. The complex structure and dynamic evolution of human subtelomeres. *Nat. Rev. Genet.* **3**: 91–102.
- Newman, T.L., Tuzun, E., Morrison, V.A., Hayden, K.E., Ventura, M., McGrath, S.D., Rocchi, M., and Eichler, E.E. 2005. A genome-wide survey of structural variation between human and chimpanzee. *Genome Res.* **15**: 1344–1356.
- Paulding, C.A., Ruvolo, M., and Haber, D.A. 2003. The *Tre2* (*USP6*) oncogene is a hominoid-specific gene. *Proc. Natl. Acad. Sci.* **83**: 2934–2938.
- Perry, G.H., Tchinda, J., McGrath, S.D., Zhang, J., Picker, S.R., Cáceres, A.M., Iafra, A.J., Tyler-Smith, C., Scherer, S.W., Eichler, E.E., et al. 2006. Hotspots for copy number variation in chimpanzees and humans. *Proc. Natl. Acad. Sci.* **103**: 8006–8011.
- Ried, T., Baldini, A., Rand, T.C., and Ward, D.C. 1992. Simultaneous visualization of seven different DNA probes by in situ hybridization using combinatorial fluorescence and digital imaging microscopy. *Proc. Natl. Acad. Sci.* **89**: 1388–1392.
- Roberto, R., Capozzi, O., Wilson, R.K., Mardis, E.R., Lomiento, M., Tuzun, E., Cheng, Z., Mootnick, A.R., Archidiacono, N., Rocchi, M., et al. 2007. Molecular refinement of gibbon genome rearrangements. *Genome Res.* **17**: 249–257.
- Rogers, J., Garcia, R., Shelledy, W., Kaplan, J., Arya, A., Johnson, Z., Bergström, M., Novakowski, L., Nair, P., Vinson, A., et al. 2006. An initial genetic linkage map of the rhesus macaque (*Macaca mulatta*) genome using human microsatellite loci. *Genomics* **87**: 30–38.
- Rozen, S., Skaletsky, H., Marszalek, J.D., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., and Page, D.C. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**: 873–876.
- Schempp, W., Toder, R., Rietschel, W., Grützner, F., Mayerova, A., and Gauckler, A. 1993. Inverted and satellited Y chromosome in the orangutan (*Pongo pygmaeus*). *Chromosome Res.* **1**: 69–75.
- Schempp, W., Binkle, A., Arnemann, J., Gläser, B., Ma, K., Taylor, K., Toder, R., Wolfe, J., Zeitler, S., and Chandley, A.C. 1995. Comparative mapping of YRRM- and TSPY-related cosmids in man and hominoid apes. *Chromosome Res.* **3**: 227–234.
- Sharp, A.J., Locke, D.P., McGrath, S.D., Cheng, Z., Bailey, J.A., Vallente, R.U., Pertz, L.M., Clark, R.A., Schwartz, S., Segraves, R., et al. 2005. Segmental duplications and copy number variation in the human genome. *Am. J. Hum. Genet.* **77**: 78–88.
- Sharp, A.J., Cheng, Z., and Eichler, E.E. 2006. Structural variation in the human genome. *Annu. Rev. Genomics Hum. Genet.* **7**: 407–442.
- Shaw, C.J. and Lupski, J.R. 2004. Implications of human genome architecture for rearrangement-based disorders: The genomic basis of disease. *Hum. Mol. Genet.* **11**: 1–7.
- She, X., Horvath, J.E., Jiang, Z., Liu, G., Furey, T.S., Christ, L., Clark, R., Graves, T., Gulden, C.L., Alkan, C., et al. 2004. The structure and evolution of centromeric transition regions within the human genome. *Nature* **430**: 857–864.
- She, X., Liu, G., Ventura, M., Zhao, S., Misceo, D., Roberto, R., Cardone,

- M.F., Rocchi, M., NISC Comparative Sequencing Program, Green, E.D., et al. 2006. A preliminary comparative analysis of primate segmental duplications shows elevated substitution rates and a great-ape expansion of intrachromosomal duplications. *Genome Res.* **16**: 576–583.
- Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**: 825–837.
- Steiper, M.E. 2006. Population history, biogeography, and taxonomy of orangutans (Genus: *Pongo*) based on a population genetic meta-analysis of multiple loci. *J. Hum. Evol.* **50**: 509–522.
- Wienberg, J. 2005. Fluorescence in situ hybridization to chromosomes as a tool to understand human and primate genome evolution. *Cytogenet. Genome Res.* **108**: 139–160.
- Wimmer, R., Kirsch, S., Rappold, G.A., and Schempp, W. 2002. Direct evidence for the *Homo-Pan* clade. *Chromosome Res.* **10**: 55–61.

Received January 29, 2008; accepted in revised form April 7, 2008.