Supplement to the Cechova, Vegesna, et al.

SUPPLEMENTAL NOTES

Supplemental Note S1. Developing a classifier and evaluating it.

We developed a classifier to determine which assembly scaffolds are ampliconic and which are X-degenerate. Ideally, Y ampliconic regions would have a higher copy number in the reference genome than X-degenerate regions, however, Y ampliconic regions are often collapsed in next-generation-sequencing-based assemblies (1). Our classifier therefore combines the copy count in the reference with mapping read depth information, since collapsed Y ampliconic regions will have a higher number of mapping reads than single-copy X-degenerate regions, in whole-genome-sequencing datasets originating from male individuals. We proceeded to classify scaffolds in our orangutan, bonobo, and gorilla Y assemblies (such annotations are already available in the human and chimpanzee Y assemblies (2, 3)). To test classification performance, we examined scaffolds carrying known X-degenerate and ampliconic genes (Table SN1A). The classification was successful in the vast majority of cases (accuracy was 85%, 100%, and 88% for bonobo, gorilla, and orangutan, respectively, Table SN1A).

Table SN1A. The number of X-degenerate and ampliconic genes that were confidently mapped to scaffolds classified as X-degenerate or ampliconic, per our classifier.

Species	Gene type	Classified as X-degenerate	Classified as ampliconic
GORILLA	ampliconic	1	**6**
	X-degenerate	**11**	2
BONOBO	ampliconic	1*	**5**
	X-degenerate	**9**	0
ORANGUTAN	ampliconic	1	**2**
	X-degenerate	**5**	0

*BPY is present in a single copy in gorilla

Methods

To classify scaffolds as either ampliconic or X-degenerate (PAR regions have been filtered out in our assemblies), each assembly was divided into 5-kb windows with 2-kb overlaps. First, for each window, we calculated copy number using a modified version of AmpliCoNE (4) that can handle scaffolds instead of a single continuous reference. AmpliCoNE takes into account mappability (calculated using the GEM mappability program (5) for k=101), repetitive element content as provided by RepeatMasker (Open-4.0.7, RepBaseRepeatMaskerEdition-20181026), and performs GC correction. X-degenerate regions are expected to have similar copy number estimates (i.e. estimates based on sequencing depth), whereas copy number estimates for ampliconic regions are expected to be higher and vary greatly, depending on the copy number of the underlying amplicons in the genome and in the assembly. Fully resolved amplicons, such as palindrome arms, are present multiple times in the assembly, and thus each window carrying them maps equally well to multiple places in the assembly. Thus, all windows were mapped back to the assembly and, after excluding secondary alignments, all multi-mapping windows were flagged as potentially ampliconic (windows with MAPQ=0 as produced by bwa mem (6)). If a window was both nonrepetitive (non-zero copy number by AmpliCoNE) and multi-mapping, it was assigned as ampliconic. Additionally, all uniquely mapping windows (MAPQ=60) with high copy number for species-specific thresholds (see Table SN1B) were also assigned as ampliconic. Uniquely mapping windows (MAPQ=60) with copy number below the species-specific thresholds were assigned as X-degenerate. This led to the majority of windows classified as either ampliconic or X-degenerate. Two types of windows remained unassigned; those that map uniquely but contain no copy-number information from AmpliCoNE, and those that contain copy-number information but have mapping quality below 60. All such windows had their assignment interpolated. As these windows were internally represented as missing values (NA), we used na.approx function from the R package zoo (7). Each scaffold was then classified as either ampliconic or X-degenerate based on the majority vote of all underlying windows. A small subset of our assemblies (0.12, 0.33, and 0.71 Mb in bonobo, gorilla and orangutan) remained unclassified.

Table NS1B. The copy-number thresholds for X-degenerate (<) and ampliconic (≥) windows, the size of the classified ampliconic and X-degenerate regions, and the corresponding number of scaffolds.

	Gene type	threshold	size [Mb]	# scaffolds
GORILLA	ampliconic	0.85	4.2	76
	X-degenerate		9.8	156
BONOBO	ampliconic	0.5	10.8	2,521
	X-degenerate		12.5	967
ORANGUTAN	ampliconic	0.75	2.2	670
	X-degenerate		14.5	358

For validation, we mapped coding sequences of X-degenerate and ampliconic genes from chimpanzee, gorilla, and Sumatran orangutan to verify that the scaffolds carrying these genes were classified correctly. We required strict mapping to avoid false hits; first we used blat (v. 36) with default parameters and then required at least thirty matches (matches≥30), the recovery of at least 20% of the original coding sequence query ((matches/qSize)≥0.2), and 99% of matching bases ((matches/(matches+misMatches))≥0.99). All accompanying scripts are available from https://bitbucket.org/biomonika/assembly_classification/ and in-house scripts *run_copy_number.sh* and evaluate.sh that output the annotation of scaffolds as .gff file.

Supplemental Note S2. Male mutation bias.

Using branch-specific values for autosomal and Y-chromosomal substitution rates from Fig. 1, we obtained the following values of α, or the male-to-female substitution rate ratio (8):

Species	Substitution rate on the Y	Substitution rate on autosomes (A)	Y/A	α
Human	0.0098	0.0066	1.48	2.88
Chimp	0.0037	0.0021	1.76	7.40
Bonobo	0.0044	0.0025	1.76	7.33
Gorilla	0.0155	0.0089	1.74	6.74
Orang	0.0582	0.0268	2.17	∞
BC	0.0088	0.0046	1.91	22.0
BCH	0.0046	0.002	2.30	∞

The species-specific estimates of α we obtained above follow the trend observed in our previous study (see Table 1 in (9)) based on a comparison of substitution rates at a much shorter genetic region (~10 kb) homologous between chromosomes Y and 3. Namely, we also observe that α is lower in human than in bonobo or gorilla. Note that our estimates are derived from closely related species and thus might be inaccurate because of ancient genetic polymorphism (9). This phenomenon is difficult to correct for in branch-specific estimates. However, it can be accounted for in pairwise estimates.

Focusing on pairwise comparisons we presented in the Results, we computed α using both uncorrected and corrected by ancestral polymorphism autosomal rate estimates. In primates, the Y chromosome has much lower diversity than autosomes do (10), and thus we did not correct for it. The results are shown below:

	Y	A	Y/A	α	Corrected A*	Corrected Y/A	Corrected α
Gorilla-human	0.0299	0.0175	1.71	5.86	0.01592	1.88	15.41
Gorilla-chimp	0.0326	0.0176	1.85	12.54	0.01602	2.03	∞
Gorilla-bonobo	0.0333	0.0180	1.85	12.33	0.01642	2.03	∞

*We subtracted 0.00158 -- the diversity estimated from gorilla populations (11) -- from our autosomal substitution rate estimates.

Again, consistent with the data presented in Results, we observed larger differences between the Y chromosomal and autosomal substitution rate estimates for gorilla-chimpanzee and gorilla-bonobo comparisons, than for the gorilla-human comparisons.

We also evaluated the potential effect of ancient genetic polymorphism on the ratio of pairwise estimates of autosomal substitution rates we present in Results. We found that this effect is minimal.

	Y	A uncorrected	A ratio	A corrected*	Corrected ratio to gorilla-human comparison
Gorilla-human	0.0299	0.0175		0.0159	
Gorilla-chimp	0.0326	0.0176	1.006	0.0160	1.0063
Gorilla-bonobo	0.0333	0.018	1.029	0.0164	1.0314

*We subtracted 0.00158 - the diversity estimated from gorilla populations (11) -- from the autosomal substitution rate.

Supplemental Note S3. AUGUSTUS gene predictions

AUGUSTUS (12) predicted 219 genes on the bonobo Y assembly, of which 25 complete or partial genes represent homologs of known human protein-coding genes. In the case of Sumatran orangutan, AUGUSTUS predicted 90 genes of which 33 complete or partial genes represent homologs of known human protein-coding genes. After implementing requirements of gene predictions (1) to have start and stop codons, and (2) be present on contigs that align to human or chimpanzee Y, we did not find any novel genes on the Y chromosome of orangutan, however we found two candidates—*SUZ12* and *PSMA6*—which have >95% identity and >90% coverage to the gene homologs on the autosomes of bonobo.

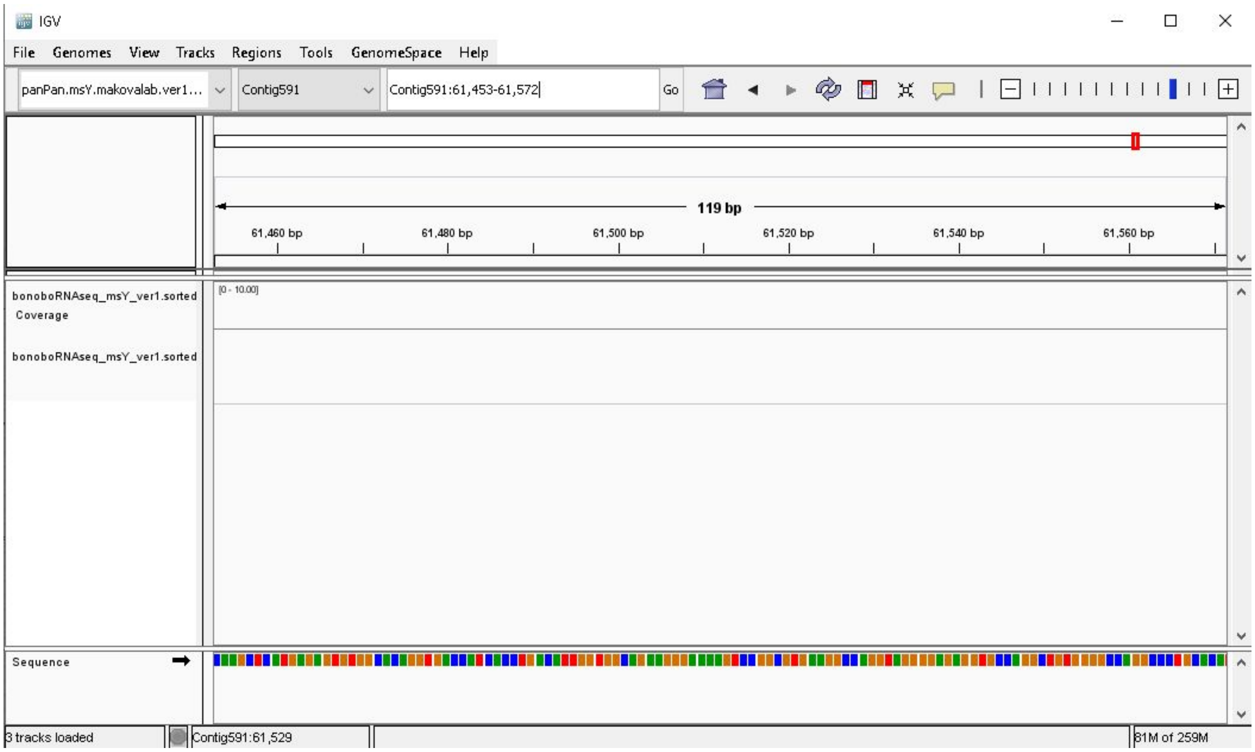
A possible transposition of the autosomal *SUZ12* gene (located on human chromosome 17) onto the bonobo Y chromosome was predicted (Table SN3) based on the limited number of introns (one intron), in contrast to its autosomal homolog, which has 15 introns (NM_015355). The *SUZ12* gene has no matches to human and gorilla Y, however the first 121 bp of its predicted sequence align to chimpanzee (palindromes C2, C11 and C15) and orangutan Y. However, when we aligned testis RNA-seq data to the predicted *SUZ12* gene on the bonobo Y chromosome, the first exon with the start codon was not expressed (Fig. SN3A), whereas the second exon was expressed (Fig. SN3B). The single nucleotide variants in the RNA-seq reads mapping to the second exon are consistent with the variants of the *SUZ12* gene present on chromosome 17. Thus, we concluded that the translocated *SUZ12* was pseudogenized on bonobo Y.

The *PSMA6* gene was also predicted in bonobo, which shared 99.6% identity with its homolog on the chimpanzee Y and 97.8% identity on human Y. However, there were no homologous sequences in the orangutan and gorilla Y assemblies. The *PSMA6* gene on human Y was annotated as a pseudogene in the Entrez database (Gene ID: 5687). Therefore, we concluded that *PSMA6* is also a pseudogene on the bonobo. Thus, no novel genes, as compared to the human Y chromosome genes, were found on the bonobo and orangutan Y chromosomes.

Table SN3. Gene annotation of the *SUZ12* homolog on the bonobo Y, as predicted by AUGUSTUS

Sequence name	Feature	Start	End	Strand
Contig591	gene	74	61572	-
Contig591	transcript	74	61572	-
Contig591	stop_codon	74	76	-
Contig591	CDS	74	2353	-
Contig591	CDS	61453	61572	-
Contig591	start_codon	61570	61572	-

Figure SN3A. The IGV (13) **view of the first exon of the SUZ12 homolog on the bonobo Y.** Testis-specific RNA-seq reads (SRA ID: SRR306837) were mapped to bonobo Y assembly using BWA MEM (6).


panPan.msY.makovalab.ver1...

Contig591

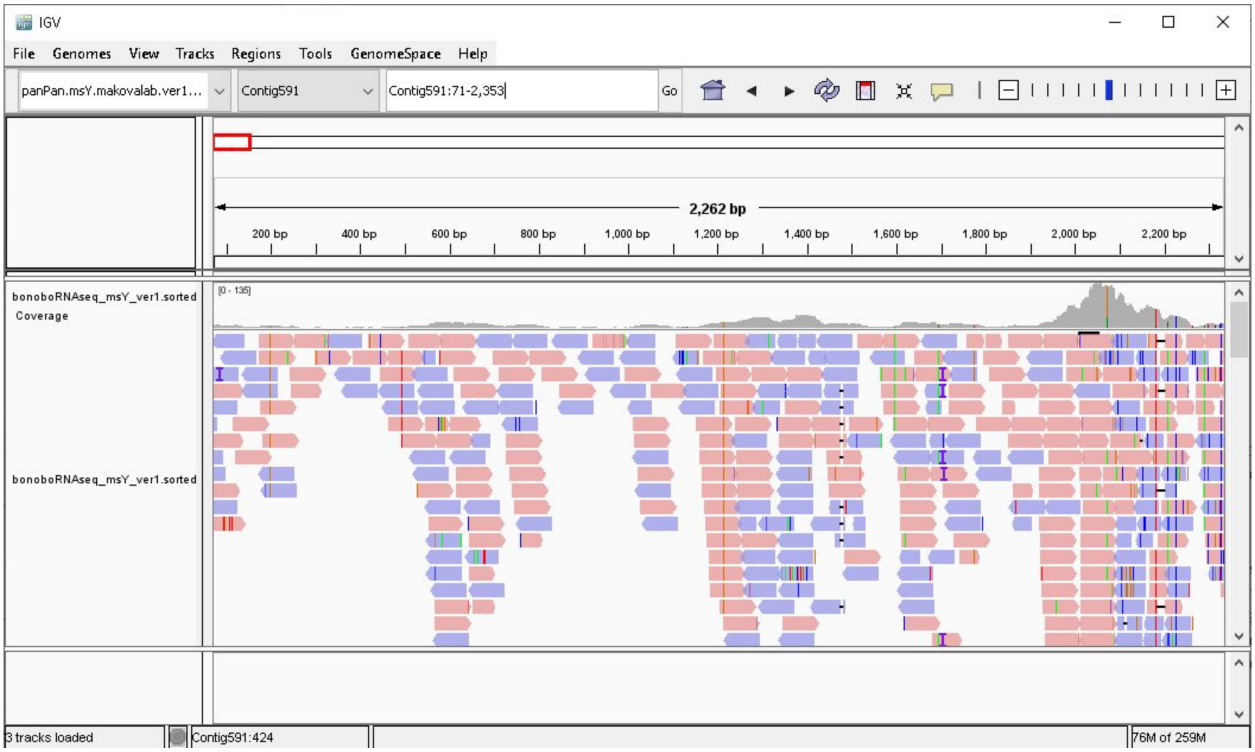
Contig591:61,453-61,572

3 tracks loaded

Contig591:61,529

81M of 259M
IGV
File Genomes View Tracks Regions Tools GenomeSpace Help
Go
119 bp
61,460 bp
61,480 bp
61,500 bp
61,520 bp
61,540 bp
61,560 bp
bonoboRNAseq_msY_ver1.sorted Coverage
[0 - 10.00]
bonoboRNAseq_msY_ver1.sorted
Sequence

Figure SN3B. IGV view of second exon on SUZ12 homolog on bonobo Y. Testis specific RNASeq reads (SRA ID: SRR306837) were mapped to bonobo Y assembly using BWA MEM (6).


panPan.msY.makovalab.ver1...

Contig591

Contig591:71-2,353

3 tracks loaded

Contig591:424

76M of 259M
IGV
File Genomes View Tracks Regions Tools GenomeSpace Help
Go
2,262 bp
200 bp
400 bp
600 bp
800 bp
1,000 bp
1,200 bp
1,400 bp
1,600 bp
1,800 bp
2,000 bp
2,200 bp
bonoboRNAseq_msY_ver1.sorted Coverage
[0 - 135]
bonoboRNAseq_msY_ver1.sorted

Supplemental Note S4. Evolutionary scenarios for palindromes.

Palindrome 4. We used two different approaches to obtain information about the conservation of human and chimpanzee palindromes across great apes. First, we used the multiple sequence alignment of Y chromosome assemblies to obtain the coverage of these palindromes. In the case of P4 we observed 23,175 bp of alignment in the bonobo assembly (**Table S5**). This analysis gave us the percentage of P4 present in bonobo Y assembly, however it did not infer that there is a continuous 23-kb block of palindrome P4 on bonobo Y. P4 could be highly fragmented due to Y chromosome degradation or rearrangements and the multiple sequence alignment can still capture such homologous fragments of the palindromes. The longest blocks in the Y multiple sequence alignment that overlap with P4 are 2-4 kb in length and these alignment blocks included sequences from gorilla and human Ys. In the remaining species, sequences homologous to P4 are mostly represented by gaps. Second, to identify the copy number of P4 homologs present in other species, we used alignments generated with LASTZ (14) based alignments. Non-overlapping 1-kb windows of the assembly were aligned to human P4 using LASTZ. The read depth of windows with >80% identity to P4 was used to estimate the copy number of P4. However, we did not find any 1-kb window in bonobo Y which maps to human palindrome P4 with >80% identify. Since we did not find high-confidence windows aligning to P4 in bonobo, we concluded that it is highly fragmented in this species.

Evolution of sequences homologous to human and chimpanzee palindromes. *Common ancestor of great apes.* Partial sequences of all human and chimpanzee palindromes were present. P1, P2, P4, P5, P8, C2, C3, C4, and C17 were in multi-copy state (Tables SN4A-B). P3, P6, and C1 might have been in single- or multi-copy state (Tables SN4A-B).

Orangutan. Increase in copy number of P1,P2, P5, C3, and C4 (Fig. 2B, Table SN4A-B). Loss of C3 (≈25% loss in coverage Table S6) and P2 (≈27% loss in coverage Table S5) segments when compared to other great apes.

Common ancestor of gorilla, human, bonobo, and chimpanzee. P3, P6, and C1 are in a multi-copy state either at this node or in the common ancestor of great apes (Tables SN4A-B).

Gorilla. Loss of copy number for P8, C2, and C17 compared to bonobo and orangutan (Fig. 3, Tables SN4A-B). C3 and C4 had more than two copies in human, chimpanzee and orangutan, however only two copies in gorilla (Tables SN4A-B). For C2, bonobo and orangutan have higher copy number when compared to gorilla. Loss of segments in C1 (≈15% loss in coverage) and C19 (≈40-60% loss in coverage) when compared to other great apes (Table S6).

Common ancestor of human, bonobo, and chimpanzee. All the palindromes are assumed to be multi-copy with the exception of C5, which could have been in a single-copy state (Tables SN4A-B). Gorilla and orangutan have more than two copies of P4 which is in two copies or lost in human, bonobo and chimpanzees (Tables SN4A-B).

Human. Palindrome P3 has ≈30-35% covered in other great apes (Table S5), so we assume that the remaining portion of P3 is human-specific. Humans also lost most of the sequences homologous to palindrome C2; we observe some sequence homologous to C2 on the human Y, however they are degraded and not visible on an alignment in human and chimpanzee Y dot plot (3). Therefore, we assume C2 was deleted human.

Common ancestor of bonobo and chimpanzee. Gain of C2 segment, both bonobo and chimpanzee share 85% coverage whereas the other great apes cover 20-30% of C2, which implies a *Pan* genus specific gain of C2 sequences (Table S6). C1 and C5 groups are present in more than two copies in both chimpanzee (3) and bonobo, where as other species have two or fewer copies of these palindromes (Fig. 3A). The *Pan* genus lost P4, we observe that sequence homologous to P4 are present in bonobo and chimpanzee Y, however they are degraded and not visible as an alignment in the human and chimpanzee Y dot plot (3). Therefore, we assume that P4 was deleted.

Bonobo. Bonobo lost copies of P1, P2, P6, P7, C3, C4, C18, and C19 (Fig. 2B, Table SN3A-B). It also experienced loss of segments in C18 (≈30-60% loss in coverage Table S6) and P7 (≈60% loss in coverage Table S5) when compared to other great apes.

Chimpanzee. Chimpanzee gained copies of P1, P2, and P5 as these palindromes share homology with C3 and C4, which are in multi-copy in chimpanzee (3). Chimpanzee also gained a segment of C1, a palindrome which has <50% coverage with the other great ape Y chromosomes (Table S6).

Table SN4A. Reconstructing human palindrome evolution using maximum parsimony. The values from extant species were taken from Table S7 and rounded to the following numbers of copies: <1.34 => “1”, “1.33-1.66” =>”1-2”, “1.66-2.5” => “2”, “>2.5” => “M” (“more than two”).

	P1	P2	P3[#]	P4	P5	P6	P7	P8
Bonobo	1-2	1-2	1-2	0	2	1-2	1	2
Chimpanzee	M	M	1[$]	0	M	2	2	2
BC*	1-2-M	1-2-M	1	0	2-M	2	1-2	2
Human	2	2	2	2	2	2	2	2
BCH**	2	2	1-2	2	2	2	2	2
Gorilla	2	2	2	M	2	1-2	1	1
BCHG***	2	2	2	2-M	2	2	1-2	1-2
Orangutan	M	M	1	M	M	1	1	2
GA****	2-M	2-M	1-2	M	2-M	1-2	1	2

*Bonobo-chimpanzee common ancestor

**Bonobo-chimpanzee-human common ancestor

***Bonobo-chimpanzee-human-gorilla common ancestor

****Common ancestor of great apes

[$] We conservatively assigned the copy number of P3 as 1 in chimpanzee to make sure we do not inflate its combined copy number.

[#] We can conservatively assume that P3 became multi-copy in BCHG, but instead it might have been multi-copy in the common ancestor of great apes and lost its multi-copy status in orangutan instead

Table SN4B. Reconstructing chimpanzee palindrome evolution using maximum parsimony. The values from extant species were taken from Table S7 and rounded to the following numbers of copies: <1.34 => “1”, “1.33-1.66” =>”1-2”, “1.66-2.5” => “2”, “>2.5” => “M” (“more than two”).

	C1 group	C2 group	C3 group	C4 group	C5 group	C17	C18	C19
Chimpanzee	M	M	M	M	M	2	2	2
Bonobo	M	M	1	1-2	M	2	1	1
BC*	M	M	1-M	M	M	2	2	1-2
Human	2	0	M	M	1	2	2	2
BCH**	2-M	M	M	M	1-M	2	2	2
Gorilla	2	2	2	2	1	1	1	1
BCHG***	2	2-M	2-M	2-M	1	1-2	1	1-2
Orangutan	1	M	M	M	1	2	1	1
GA****	1-2	M	M	M	1	2	1	1

*Bonobo-chimpanzee common ancestor

**Bonobo-chimpanzee-human common ancestor

***Bonobo-chimpanzee-human-gorilla common ancestor

****Common ancestor of great apes

Supplemental Note S5. Analysis of X-Y gene conversion and selection

Gene conversion. Prior to analyzing selection, we had to perform an analysis of X-Y gene conversion, as this process can interfere with selection detection. We studied the incidence of X-Y gene conversion between 12 X-degenerate genes and their homologs on the X chromosome (from 16 X-degenerate genes present on the human Y we excluded *CYorf15A*, *CYorf15B*, *RPS4Y1*, and *RPS4Y2,* as parts of these genes have repeats on the Y and homologs on the X, making detection of gene conversion difficult). Gene conversion was examined using a multiple-sequence alignment of the X and Y chromosomes.

We softmasked the X and Y chromosomes of great apes using RepeatMasker (15) (RepeatMasker -pa 63 -xsmall -species Primates ${assembly}.rmsk.fa). ProgressiveCactus (16) was used to align the chromosomes. A guide tree which pairs the X and Y chromosomes from the same species list was used (((chimpY, chimpX),(bonoboY, bonoboX),(humanY, humanX),(gorillaY, gorillaX),(sorangY, sorangX))). The resulting alignment output from ProgressiveCactus was converted to MAF format by hal2maf (17) using human Y as a reference (--noAncestors --refGenome humanY --maxRefGap 100 --maxBlockLen 10000). We then parsed the alignment blocks (retaining blocks longer than >50bp; range of gene conversion tracts we observed in human was 55-290 bp, as was observed in previous studies (18, 19)) which fall within the coordinates of X-degenerate genes. We did not perform additional filtering based on repeat content within the alignment blocks. For each block, the alignments which constitute the sequences from both the X and Y chromosomes for a species were collected in a FASTA file. We used GENCONV (20) (/w9 /lp -nolog) to identify gene conversion events based on multiple sequence alignment files. By default, the output of GENCONV constitutes a global list of high-confidence gene conversion events after multiple testing for each sequence pair. We also used /lp parameters with which a second list of significant gene conversions for all possible pairwise comparisons GENCONV performed was generated. We used a p-value cutoff of 0.05, as GENCONV provides p-values after correcting for multiple comparisons. We used pairwise comparisons to address gene conversion in cases where a chromosome was represented by more than one sequence in the alignment. From the GENCONV output, we parsed events that constitute gene conversion between the X and Y chromosome from the same species and retained events which are longer than 50 bp.

In total, we detected 143 candidate gene conversion events up to 50-410 bp in length (minimal length of 50 bp was used for detection), including 46 high-confidence ones (Table S5NA). Among these, most events were observed in the genes from younger strata—30 in *PRKY* (stratum 5), nine in *NLGN4Y* (stratum 4), four in *AMELY* (stratum 4), and two in *TBL1Y* (stratum 4). Higher homology between these genes and their X homologs, as opposed to between genes from older strata and their X homologs, is expected to facilitate gene conversion, which work most efficiently with homology >92% (18). No gene conversion events >50 bp in size were found in the exonic regions of X-degenerate genes, thus this process should not affect our selection analysis.

Table S5NA. Gene conversion between X and Y chromosomes of great apes using GENECONV. The values represent the number of high-confidence gene conversion events with significant *p*-values (<0.05) which are corrected for multiple comparisons across all sequence pairs. The values in brackets represent the total number of gene conversion events with significant *p*-values (<0.05) for comparisons corrected for the length of the alignment.

Gene (Stratum)	Bonobo	Chimpanzee	Human	Gorilla	Orangutan	Total
***AMELY* (4)**	1 (1)	1 (1)	1 (1)	1 (1)	0	4 (4)
***DBY* (3)**	0 (1)	0 (1)	0	0	0 (1)	0 (3)
***EIF1AY* (3)**	0	0	0	0	0	0
***KDM5D* (2)**	0	0	0	0	0	0

***NLGN4Y* (4)**	2 (6)	2 (7)	3 (7)	1 (8)	1 (3)	9 (31)
***PRKY* (5)**	8 (25)	10 (27)	4 (17)	8 (17)	0	30 (86)
***SRY* (1)**	0	0	0	0	0	0
***TBL1Y* (4)**	1 (4)	0 (2)	1 (7)	0 (3)	0 (2)	2 (18)
***TMSB4Y* (3)**	0	0	0	0	0	0
***USP9Y* (3)**	0	0	0	0	0	0
***UTY* (3)**	0	0	0	0	0	0
***ZFY* (3)**	0	0	0	0	1 (1)	1(1)
Total	12 (37)	13 (38)	9 (32)	10 (29)	2 (7)	46(143)

Selection. We used the codeml module of PAML (version 4.8) (21) to detect branch-specific differences in the nonsynonymous-to-synonymous rate ratios and to test for positive selection acting on X-degenerate genes (excluding pseudogenes *CYorf15A* and *CYorf15B* (*TXLNGY*) in human, chimpanzee, bonobo, gorilla, Sumatran orangutan, Bornean orangutan, and macaque). Selection at ampliconic sequences was not analyzed because their sequences are still incomplete even for high-quality Y chromosome assemblies (e.g., of the chimpanzee Y (3)) and the sequences of their multiple copies remain undeciphered except for the human Y (22). Coding sequences of Y-chromosomal X-degenerate genes were retrieved from GenBank or deciphered in this study and aligned using ClustalW (23). The phylogenies were generated with the Neighbor-Joining method (24) (with 1,000 bootstrap replicas) as implemented in Mega7 (25). First, for each gene, the one-ratio model (assuming the same nonsynonymous-to-synonymous rate ratio ω for the entire tree) was compared with the two-ratio model (assuming that the branch-specific ratio ω_s is different from the background ratio ω_o). When the difference between the two models was significant, this indicated that the synonymous-to-nonsynonymous rate ratio was different for the branch tested. In these cases, to test for positive selection, the model assuming the foreground ratio ω to be fixed at 1 (neutral evolution) was compared against an alternative model with branch-specific $\omega > 1$ (positive selection).

We found a total of seven gene-branch combinations that had foreground nonsynonymous-to-synonymous rate ratio significantly different than the background nonsynonymous-to-synonymous rate ratio (Table S5NB). We observed significantly different nonsynonymous-to-synonymous rate ratios on the chimpanzee and bonobo ancestor than other lineages for three X-degenerate genes (*DDX3Y*, *EIF1AY*, and *PRKY*). We also detected significantly different nonsynonymous-to-synonymous rate ratios in bonobo for two X-degenerate genes (*DDX3Y* and *EIF1AY*), in chimpanzee for *ZFY*, and in human for *RPS4Y2*. However, none of these ratios was significantly higher than one, providing no evidence for positive selection.

Table S5NB. Gene-branch combinations with significantly higher branch-specific than background nonsynonymous-to-synonymous rate ratio.

Gene	Branch	Background ω_o	Branch-specific ω_s	P-value for testing $\omega_s > \omega_o$	P-value for testing $\omega_s > 1$
DDX3Y	Bonobo	0.19	1.11	0.02	0.54
DDX3Y	BC*	0.18	1.44	0.004	0.48

EIF1AY	Bonobo	0.02	>1 (division by 0)	0.03	0.45
EIF1AY	BC	0.02	>1 (division by 0)	0.03	0.44
PRKY	BC	0.18	1.64	0.03	0.11
RPS4Y2	Human	0.16	>1 (division by 0)	0.0002	1.00
ZFY	Chimpanzee	0.04	>1 (division by 0)	0.04	0.44

*BC: bonobo-chimpanzee common ancestor

Supplemental Tables

Table S1. The statistics for *de novo* bonobo, gorilla, and orangutan Y chromosome assemblies and for the human and chimpanzee reference Y chromosomes (2, 3)**.**

Species	Assembly length (Mb)	NG50[1] (in bp, using G=8.5 Mb)	N50[2] (in bp)	Number of scaffolds	Ns (in Mb)
Orangutan	17.4	1,388,499	773,523	1,178	1.3
Gorilla	14.3	150,017	95,534	268	0.008
Bonobo	23.4	153,556	32,114	3,590	0.8
Chimpanzee[1]	26.4	-	-	1	1.1
Human[2]	57.2	-	-	1	33.6

[1] NG50: the size of the scaffold for which half of the conserved X-degenerate regions (set to 8.5 Mb based on estimates in human (2)) is in scaffolds that are equal to or larger than this size.

[2] N50: the size of the scaffold for which half of the assembly is in scaffolds that are equal to or larger in size.

Table S2. Alignment statistics.

(A) Portion of a species' Y chromosome assembly aligning to each other species, in ProgressiveCactus (16) multi-species alignments. Percentage shown is the portion of bases in the column-species Z that has any pairwise alignment to the row-species W. Denominator is the non-N count of the column-species Z. **(B)** Portion aligning to each other species, in LASTZ (14) pairwise alignments. **(C)** Average identity in ProgressiveCactus (16) alignment blocks containing all five species. **(D)** Average identity in LASTZ (14) pairwise alignments.

A. Portion of a species aligning to each other species, in **progressiveCactus multi-species** alignments.

proportion of aligning to	Human Y	Chimpanzee Y	Bonobo Y	Gorilla Y	Orangutan Y
Human Y	—	77.22%	47.52%	75.72%	62.89%
Chimpanzee Y	66.39%	—	52.69%	66.86%	60.65%
Bonobo Y	61.74%	82.98%	—	64.77%	57.76%
Gorilla Y	57.79%	58.47%	36.76%	—	52.51%
Orangutan Y	54.93%	63.26%	38.51%	61.20%	—

B. Portion aligning to each other species, in **LASTZ pairwise** alignments.

proportion of aligning to	Human Y	Chimpanzee Y	Bonobo Y	Gorilla Y	Orangutan Y
Human Y	—	92.13%	64.59%	89.61%	86.40%
Chimpanzee Y	84.26%	—	65.58%	85.19%	85.85%
Bonobo Y	84.05%	95.99%	—	86.50%	85.34%
Gorilla Y	78.21%	83.21%	56.05%	—	79.45%
Orangutan Y	75.53%	83.57%	55.67%	79.44%	—

C. Average identity in **progressiveCactus multi-species** alignment blocks containing all five species.

identity of aligning to	Human Y	Chimpanzee Y	Bonobo Y	Gorilla Y	Orangutan Y
Human Y	—	97.90%	97.85%	97.20%	93.72%
Chimpanzee Y	97.89%	—	99.22%	96.98%	93.58%
Bonobo Y	97.80%	99.17%	—	96.90%	93.47%
Gorilla Y	97.20%	96.99%	96.95%	—	93.55%
Orangutan Y	93.59%	93.45%	93.37%	93.42%	—

S2D. Average identity in **LASTZ pairwise** alignments.

identity of aligning to	Human Y	Chimpanzee Y	Bonobo Y	Gorilla Y	Orangutan Y
Human Y	—	95.76%	95.58%	95.81%	92.47%
Chimpanzee Y	95.60%	—	97.84%	94.53%	91.95%
Bonobo Y	95.19%	98.27%	—	94.30%	91.83%
Gorilla Y	95.01%	93.99%	93.98%	—	91.85%
Orangutan Y	92.46%	91.91%	92.28%	92.71%	—

Table S3. The estimated number of substitutions (after correcting for multiple hits). (**A**) between gorilla and chimpanzee, and between gorilla and human; (**B**) between gorilla and bonobo, and between gorilla and human, considering autosomes and Y chromosome separately, with corresponding χ^2 statistics and *p*-value showing an elevation in the *Pan* lineage. We used a test similar to the relative rate test used in (26). The last column shows the alignment length.

A

Autosomes	**N subst. gorilla -> chimp**	**N subst. gorilla -> human**	**ratio**	**Difference from 1 χ^2, *p*-value**	**Alignment length**
	40,594,903	40,364,251	1.006	657.1; $<1\times10^{-5}$	2,306,528,605 bp
Y	**N subst. gorilla -> chimp**	**N subst. gorilla -> human**	**ratio**	**χ2**	**Alignment length**
	154,937	142,105	1.090	554.5; $<1\times10^{-5}$	4,752,665 bp

B

Autosomes	**N subst. gorilla -> bonobo**	**N subst. gorilla -> human**	**ratio**	**χ2**	**Alignment length**
	41,517,515	40,364,251	1.029	16,243.7; $<1\times10^{-5}$	2,306,528,605 bp
Y	**N subst. gorilla -> bonobo**	**N subst. gorilla -> human**	**ratio**	**χ2**	**Alignment length**
	158,264	142,105	1.114	869.7; $<1\times10^{-5}$	4,752,665 bp

Table S4. Gene birth and death rates (in events per millions of years) on the Y chromosome of great apes, as predicted using the Iwasaki and Takagi gene reconstruction model (27)**.**
BC - common ancestor of bonobo and chimpanzee; BCH - common ancestor of bonobo, chimpanzee, and human; BCHG - common ancestor of bonobo, chimpanzee, human, and gorilla. GA - common ancestor of great apes.

	X-degenerate genes		**Ampliconic genes**	
Branch	**Birth rate**	**Death rate**	**Birth rate**	**Death rate**
Bonobo-BC	$1.00x10^{-4}$	$1.00x10^{-4}$	$1.00x10^{-4}$	0.182
Chimpanzee-BC	$1.00x10^{-4}$	$8.00x10^{-2}$	$1.00x10^{-4}$	$1.00x10^{-4}$
Human-BCH	$1.90x10^{-5}$	$1.90x10^{-5}$	$1.90x10^{-5}$	$1.90x10^{-5}$
Gorilla-BCHG	$1.43x10^{-5}$	$1.43x10^{-5}$	$1.43x10^{-5}$	$1.43x10^{-5}$
Orangutan-GA	$7.14x10^{-6}$	$2.06x10^{-2}$	$7.14x10^{-6}$	$7.14x10^{-6}$
Macaque-Root	$3.45x10^{-6}$	$3.45x10^{-6}$	$3.45x10^{-6}$	$9.92x10^{-3}$
BC-BCH	$2.35x10^{-5}$	$4.89x10^{-2}$	$2.35x10^{-5}$	$9.54x10^{-2}$
BCH-BCHG	$5.71x10^{-5}$	$5.71x10^{-5}$	$5.71x10^{-2}$	$5.71x10^{-5}$
BCHG-GA	$1.43x10^{-5}$	$1.43x10^{-5}$	$1.43x10^{-5}$	$1.43x10^{-5}$
GA-Root	$6.67x10^{-6}$	$4.04x10^{-3}$	$6.67x10^{-6}$	$6.67x10^{-6}$

Table S5. The sequence coverage (percentage) of human palindromes P1-8 (arms) by non-human great ape Y assemblies.

The repeats annotated by RepeatMasker (15) were excluded from the calculations.

Human palindrome	Length, bp					Coverage, percentage			
	Chimpanzee	Bonobo	Gorilla	Orangutan	Human*	Chimpanzee	Bonobo	Gorilla	Orangutan
P1	362895	340679	334030	249906	608650	59.62	55.97	54.88	41.06
P2	50379	50144	49092	29106	76387	65.95	65.64	64.27	38.10
P3	64460	62335	68254	55671	179793	35.85	34.67	37.96	30.96
P4	77345	76543	83948	72509	93979	82.30	81.45	89.33	77.15
P5	158548	158405	157040	148243	166929	94.98	94.89	94.08	88.81
P6	35555	35436	34535	32729	36832	96.53	96.21	93.76	88.86
P7	4439	1134	4385	4134	5414	81.99	20.95	80.99	76.36
P8	15027	13698	12688	12783	16728	89.83	81.89	75.85	76.42
Total	768648	738374	743972	605081	1184712	64.88	62.33	62.80	51.07

*Palindrome arm length

Table S6. The sequence coverage (percentage) of chimpanzee palindromes C1-19 (arms) across great apes.

The repeats annotated by RepeatMasker (15) were excluded from the calculations. The palindromes were clustered into five homology groups: C1 (C1+C6+C8+C10+C14+C16), C2 (C2+C11+C15), C3 (C3+C12), C4 (C4+C13), and C5 (C5+C7+C9). The numbers in bold represent the palindrome with highest coverage within each group, which we used as the representative coverage of that homology group.

Chimpanzee palindrome	**Length, bp**					**Coverage, percentage**			
	Human	**Bonobo**	**Gorilla**	**Orangutan**	**Chimpanzee***	**Human**	**Bonobo**	**Gorilla**	**Orangutan**
C1	36118	35368	25476	37460	66916	**53.98**	**52.85**	**38.07**	**55.98**
C2	35233	111830	27957	38673	141680	24.87	78.93	19.73	27.30
C3	58885	59610	57211	38076	82274	**71.57**	**72.45**	**69.54**	**46.28**
C4	119067	120124	119349	116986	140401	84.80	85.56	85.01	83.32
C5	113353	107475	113915	110252	136579	**82.99**	**78.69**	**83.41**	**80.72**
C6	16015	15745	6517	17378	58166	27.53	27.07	11.20	29.88
C7	45695	45809	46167	45408	137290	33.28	33.37	33.63	33.07
C8	21783	21047	12228	22766	64725	33.65	32.52	18.89	35.17
C9	46890	46809	47450	46382	139044	33.72	33.66	34.13	33.36
C10	16096	15957	6926	17487	59322	27.13	26.90	11.68	29.48
C11	24198	105870	27840	38567	123692	19.56	85.59	22.51	31.18
C12	46539	47451	46260	26584	76418	60.90	62.09	60.54	34.79
C13	117844	118277	117312	116043	132034	**89.25**	**89.58**	**88.85**	**87.89**
C14	19998	19943	11045	21285	47853	41.79	41.68	23.08	44.48
C15	23050	95239	25619	36280	111841	**20.61**	**85.16****	**22.91**	**32.44**
C16	40950	33549	28037	41545	76391	53.61	43.92	36.70	54.38
C17	15364	15072	12961	12978	17516	**87.71**	**86.05**	**74.00**	**74.09**
C18	6686	2582	4741	6304	6888	**97.07**	**37.49**	**68.83**	**91.52**
C19	155747	156318	57117	122860	168341	**92.52**	**92.86**	**33.93**	**72.98**
Total	426178	488431	303092	383879	637282	72.96	81.67	57.36	66.48

*Palindrome arm length

**In the case of cluster C2, different palindromes had highest coverage for different species and we considered C15 as a representative because it had the highest coverage for more than one species, while other palindromes had highest coverage for only one species.

Table S7. The copy number for sequences homologous to (A) human and (B) chimpanzee palindromes. The numbers for bonobo, gorilla, and orangutan were obtained based on median read depth of 1-kb windows homologous to human or chimpanzee palindromes. The copy number for chimpanzee and human were obtained by examining the dotplot of human and chimpanzee Y(3). In brackets, we indicate the known homologs of chimpanzee and human palindromes in human and chimpanzee, respectively (3).

A

Palindrome/ Species	**P1**	**P2**	**P3***	**P4**	**P5**	**P6**	**P7**	**P8**
Human	2	2	2	2	2	2	2	2
Chimpanzee	>2(C3+C4)	>2(C3)	>2(C1 parts)	0	>2(C4)	2(C19)	2(C18)	2(C17)
Bonobo	1.64	1.46	1.62	0	2.13	1.42	0.70	1.98
Gorilla	2.19	2.13	1.87	2.74	2.04	1.42	1.16	1.19
Orangutan	6.52	13.13	1.05	5.67	7.29	1.06	1.04	1.80

*Note: We assume that some parts of human palindrome P3 in chimpanzee are multi-copy (those that share homology with C1), while others are single-copy.

B

Palindrome/ Species	**C1 group**	**C2 group**	**C3 group**	**C4 group**	**C5 group**	**C17**	**C18**	**C19**
Human	2(P3 parts)	0	>2(P1,P2)	>2(P5,P1)	1	2 (P8)	2(P7)	2(P6)
Chimpanzee	>2	>2	>2	>2	>2	2	2	2
Bonobo	13.29	8.42	1.18	1.41	9.31	2.27	0.80	1.25
Gorilla	2.28	2.48	2.13	2.07	1.30	1.19	1.15	1.28
Orangutan	1.02	6.07	12.13	7.85	1.07	1.87	1.02	1.02

Table S8. Summary of Y chromosome species-specific sequences.
The copy number of species-specific blocks in multi-species alignments. Only >100-bp blocks were analyzed to compute the percentages (shown in parentheses) from the total species-specific sequence. CN - copy number.

Species	Total length	In blocks with CN≤1.33 (i.e. CN≈1)	In blocks with 1.33<CN≤1.66 (CN between 1 and 2, uncertain)	In blocks with 1.66<CN≤2.5 (i.e. CN≈2)	In blocks with CN>2.5 (i.e. CN>2)
Bonobo	9,473,169 bp	1,180,899 bp (12.47%)	630,172 bp (6.65%)	1,673,854 bp (17.67%)	5,976,284 bp (63.09%)
Gorilla	1,632,223 bp	364,390 bp (22.32%)	340,282 bp (20.85%)	253,831 bp (15.55%)	663,754 bp (40.66%)
Orangutan	3,520,105 bp	2,247,989 bp (63.46%)	136,390 bp (3.85%)	190,607 bp (5.41%)	945,119 bp (26.68%)

Table S9. The number of observed chromatin contacts, followed by the chromatin interactions weighted by their probability (multi-mapping reads are allocated probabilistically (28)**).** The density of chromatin interactions is higher in palindromes. The table is based on human iPSC data (29). See also Fig. S8.

HUMAN	total number of chromatin interactions	weighted number of chromatin interactions	region length [Mb]	**density of weighted interactions [per Mb]**
palindrome interactions	82,619	14,362	6	**2,489**
mixed interactions	70,533	12,111	-	-
other interactions	181,178	26,064	24	**1,086**
sum	334,330	52,536	30	

Table S10. The generated-here and previously unpublished sequencing datasets used for assembly generation and classification, including their summary information.

Species	Read type	Source	Individual ID	Insert size (bp)	Read count	Read length (bp)	Sequencing technology
Bonobo	paired-end	Whole-genome	PR00251	1,000	319,637,420	251	Illumina HiSeq2500
Bonobo	mate-pair	Whole-genome	PR00251	8,000	120,181,539	100	Illumina HiSeq2500
Bonobo	long reads	Y flow-sorted	Ppa_MFS	>7,000	818,697	variable	PacBio
Orangutan	paired-end	Whole-genome	1991-0051	1,000	303,928,291	251	Illumina HiSeq2500
Orangutan	synthetic long reads	Whole-genome	AG06213	~350 bp	437,361,894	151	10X Genomics
Orangutan	mate-pair	Whole-genome	1991-0051	8,000	120,384,643	151	Illumina HiSeq2500

*Before any trimming, adapter removal or quality filtering. For paired reads, the number of read pairs is listed.

Table S11. The number of variants before and after the polishing step.
The haploid variants reported by FreeBayes (30) were filtered to retain only high quality calls (QUAL≥ 30). The polishing step reduces the number of variants present in the assembly.

ORANGUTAN	Before the polishing	After the polishing
All called variants	294,544	286,782
Called variants with QUAL≥30	38,597	4,537

BONOBO	Before the polishing	After the polishing
All called variants	414,617	408,580
Called variants with QUAL≥30	52,158	5,318

Table S12. The coordinates of palindromes on panTro6 and hg38 Y chromosome.
The coordinates were obtained from (31) and updated to the current version of chimpanzee Y, panTro6.

A

Palindrome	Start	End	The end of left arm (approximated)
P1	23359067	26311550	24822577
P2	23061889	23358813	23208197
P3	21924954	22661453	22208730
P4	18450291	18870104	18640356
P5	17455877	18450126	17951255
P6	16159590	16425757	16269541
P7	15874906	15904894	15883575
P8	13984498	14058230	14019652

B

Chimpanzee palindrome	Start	End	Amplicon color following (3)	Approximate arm length (half of palindrome)	The end of left arm (approximated)
C1	1759451	2,053,069	Pink	146809	1906260
C2	2298081	2,984,818	Blue	343368.5	2641450
C3	3587737	3,925,944	Red	169103.5	3756841
C4	4669973	5,444,881	Gold	387454	5057427
C5	8651546	9,099,775	Turquoise	224114.5	8875661
C6	9099776	9,383,863	Pink	142043.5	9241820
C7	9383864	9,832,093	Turquoise	224114.5	9607979
C8	9832094	10,116,181	Pink	142043.5	9974138
C9	10116182	10,564,411	Turquoise	224114.5	10340297
C10	10564412	10,851,248	Pink	143418	10707830
C11	11060051	11,674,261	Blue	307105	11367156
C12	12651797	12,963,129	Red	155666	12807463
C13	13340000	14,084,660	Gold	372330	13712330
C14	14798116	15,024,316	Pink	113100	14911216
C15	15493746	16,056,815	Blue	281534.5	15775281
C16	16477310	16,791,733	Pink	157211.5	16634522
C17	21591500	21,671,300		39900	21631400
C18	23517939	23,546,819		14440	23532379
C19	23807052	24,577,396		385172	24192224

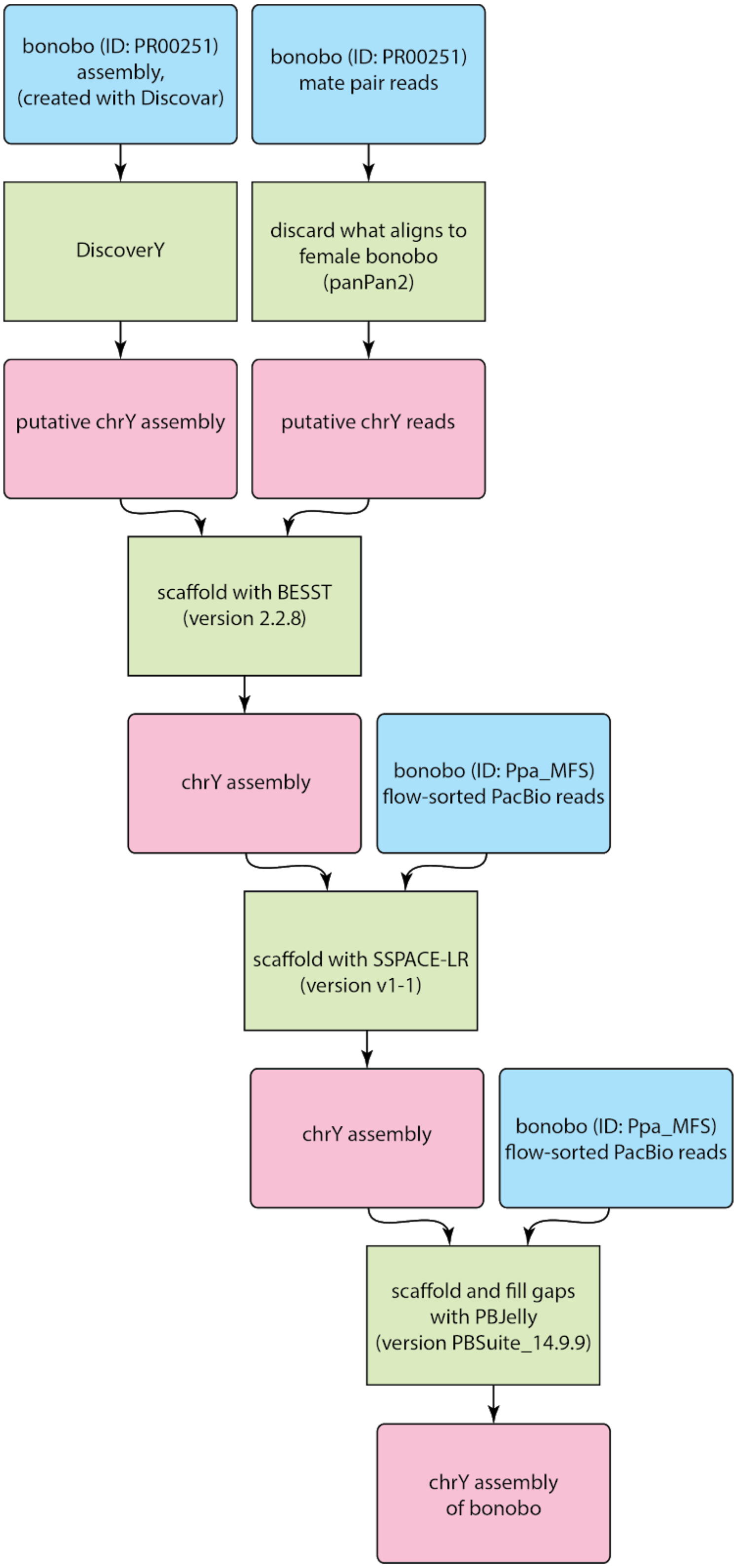
Table S13. The list of ENCODE datasets analyzed in the search for regulatory elements in P6 and P7.

Experiment	Unfiltered BAM	Target	Tissue	Link
ENCSR000ALB	ENCFF735TGN	H3K27Ac	HUVEC	https://www.encodeproject.org/experiments/ENCSR000ALB/
ENCSR000AKL	ENCFF322MOQ	H3K4me1	HUVEC	https://www.encodeproject.org/experiments/ENCSR000AKL/
ENCSR000ALG	ENCFF261CBZ	Control	HUVEC	https://www.encodeproject.org/experiments/ENCSR000ALG/
ENCSR000EOQ	ENCFF042VZB	Dnase-seq	HUVEC	https://www.encodeproject.org/experiments/ENCSR000EOQ/
ENCSR112ALD	ENCFF319GEZ	CREB1	HepG2	https://www.encodeproject.org/experiments/ENCSR112ALD/
ENCSR136ZQZ	ENCFF807LQS	H3K27Ac	Testis	https://www.encodeproject.org/experiments/ENCSR136ZQZ/
ENCSR956VQB	ENCFF077NRU	H3k4me1	Testis	https://www.encodeproject.org/experiments/ENCSR956VQB/
ENCSR215WNN	ENCFF511LQO	Control	Testis	https://www.encodeproject.org/experiments/ENCSR215WNN/
ENCSR729DRB	ENCFF639PHQ	Dnase-seq	Testis	https://www.encodeproject.org/experiments/ENCSR729DRB/

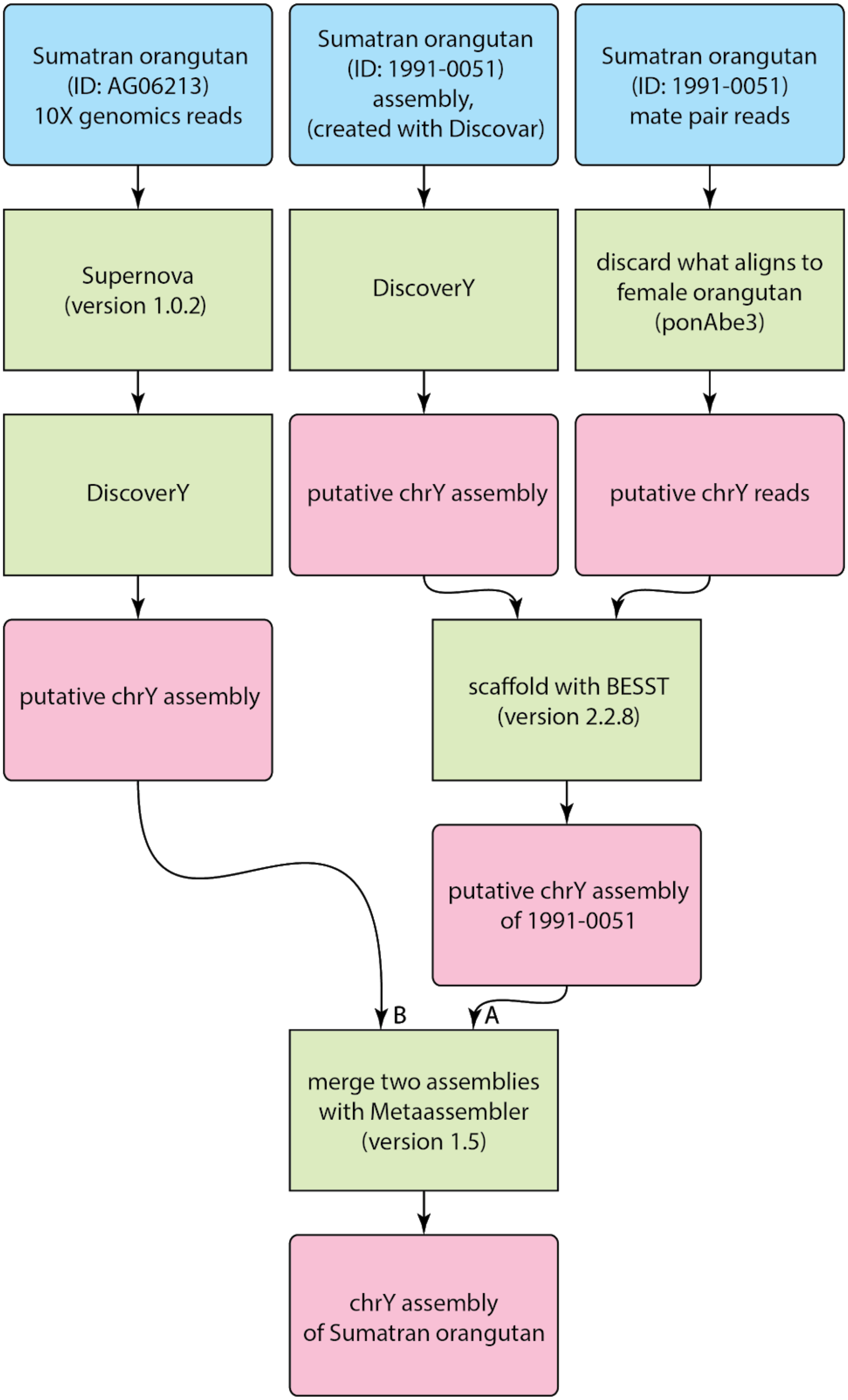
Supplemental Figures

Figure S1. Flowcharts for the assemblies of (A) bonobo, (B) Sumatran orangutan, and (C) gorilla. Blue: input datasets, green: software tools, pink: processed datasets.

A. Bonobo


bonobo (ID: PR00251) assembly, (created with Discovar)
bonobo (ID: PR00251) mate pair reads
DiscoverY
discard what aligns to female bonobo (panPan2)
putative chrY assembly
putative chrY reads
scaffold with BESST (version 2.2.8)
chrY assembly
bonobo (ID: Ppa_MFS) flow-sorted PacBio reads
scaffold with SSPACE-LR (version v1-1)
chrY assembly
bonobo (ID: Ppa_MFS) flow-sorted PacBio reads
scaffold and fill gaps with PBJelly (version PBSuite_14.9.9)
chrY assembly of bonobo

B. Sumatran orangutan



Sumatran orangutan (ID: AG06213) 10X genomics reads
Sumatran orangutan (ID: 1991-0051) assembly, (created with Discovar)
Sumatran orangutan (ID: 1991-0051) mate pair reads
Supernova (version 1.0.2)
DiscoverY
discard what aligns to female orangutan (ponAbe3)
DiscoverY
putative chrY assembly
putative chrY reads
putative chrY assembly
scaffold with BESST (version 2.2.8)
putative chrY assembly of 1991-0051
B
A
merge two assemblies with Metaassembler (version 1.5)
chrY assembly of Sumatran orangutan

C. Gorilla

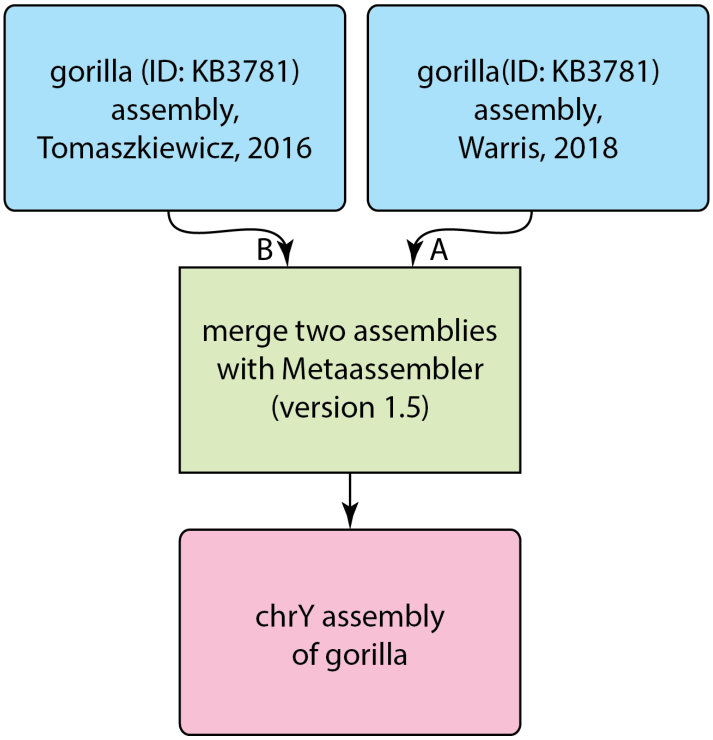

gorilla (ID: KB3781) assembly, Tomaszkiewicz, 2016
gorilla(ID: KB3781) assembly, Warris, 2018
B
A
merge two assemblies with Metaassembler (version 1.5)
chrY assembly of gorilla

Figure S2A. Sequence comparisons between human (hg38) and bonobo, gorilla and orangutan Y chromosome assemblies.

The order of scaffolds and/or contigs was modified in each assembly to match the human Y chromosome using Mauve (32) v.2015-02-25. The heterochromatic portion of the *q* arm of the human Y chromosome was omitted. Dot plots were generated with Gepard v.1.40 (33) using word length 50.

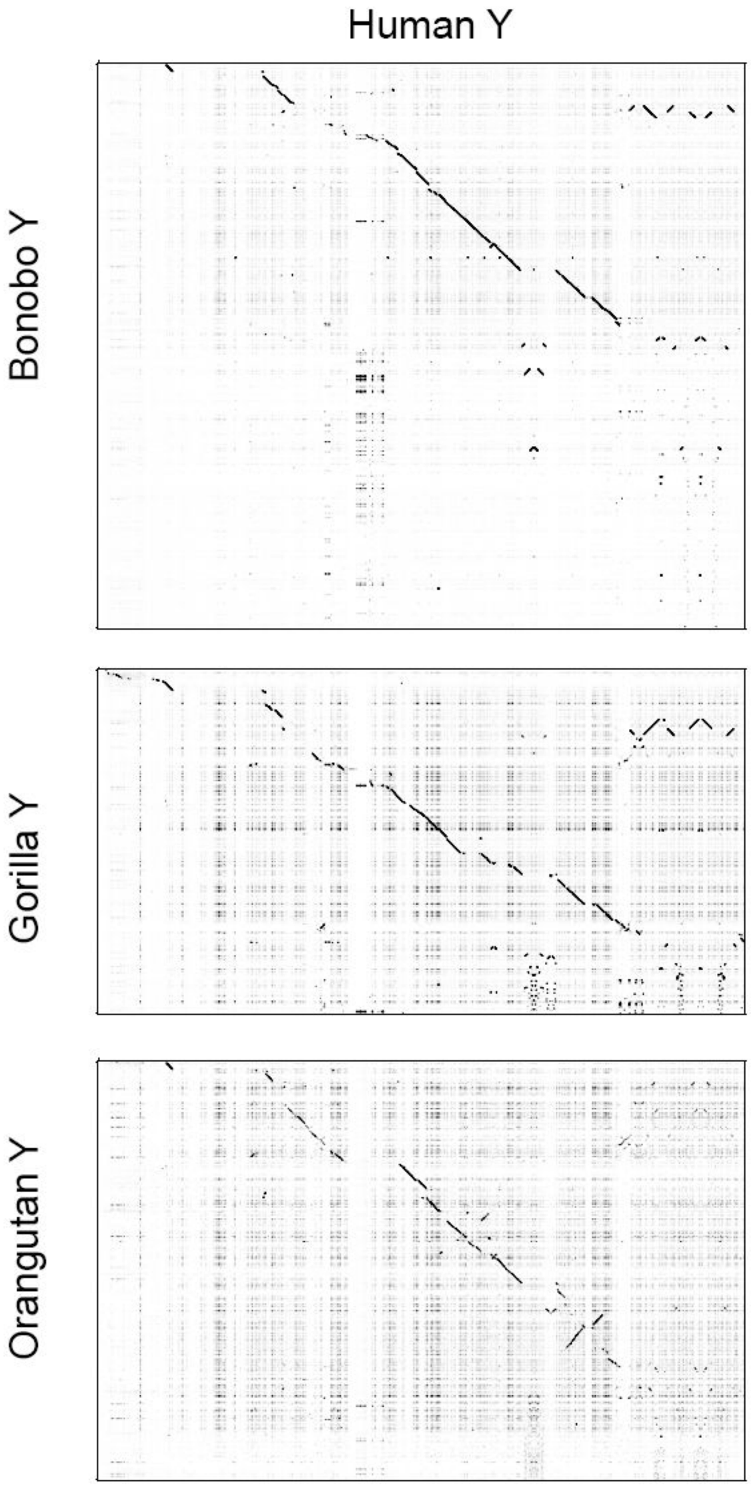


Figure S2B. Sequence comparisons between chimpanzee (panTro6) and bonobo, gorilla and Sumatran orangutan Y chromosome assemblies.

The order of scaffolds and/or contigs was modified in each assembly to match the chimpanzee Y chromosome using Mauve (32) v.2015-02-25. Dot plot was generated with Gepard version 1.40 (33) using word length 50.

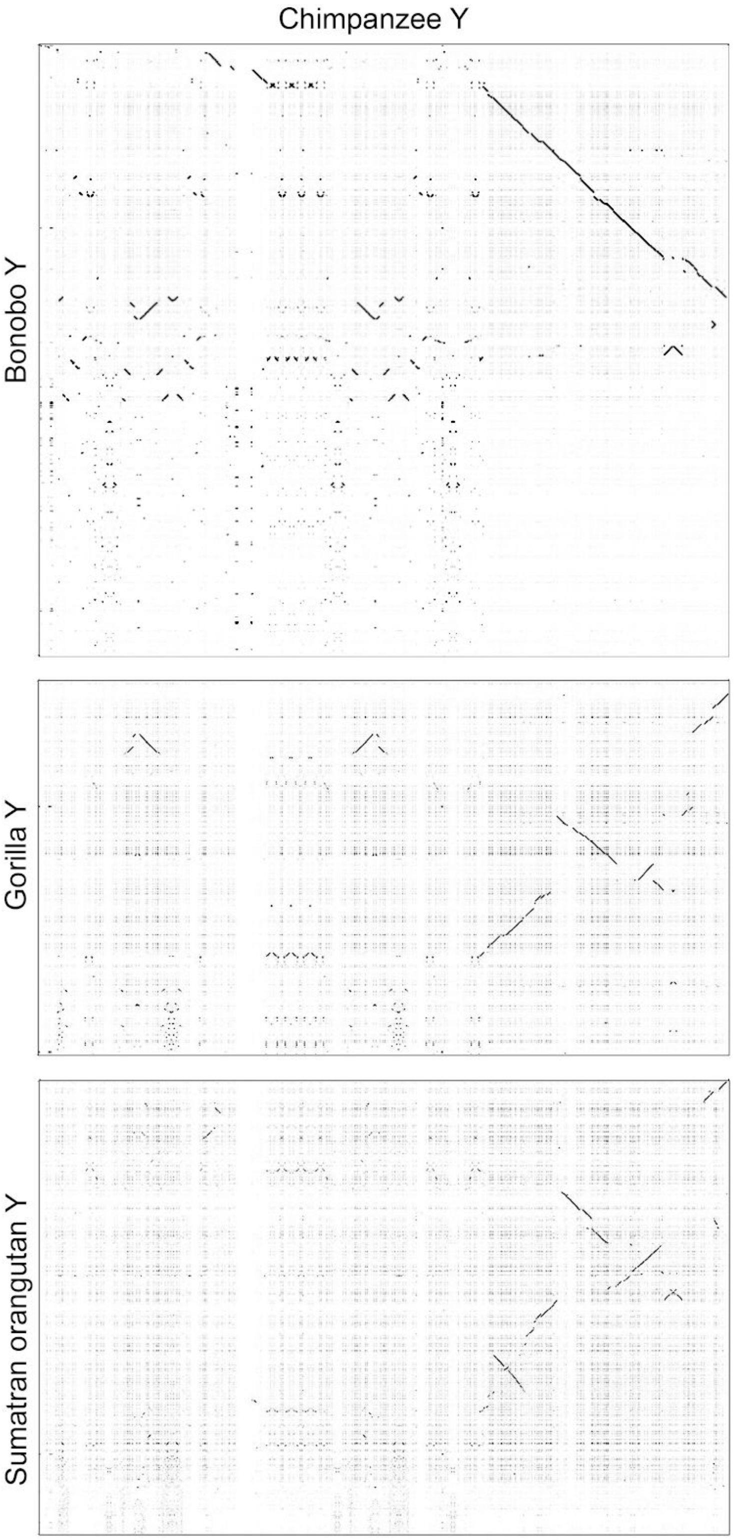
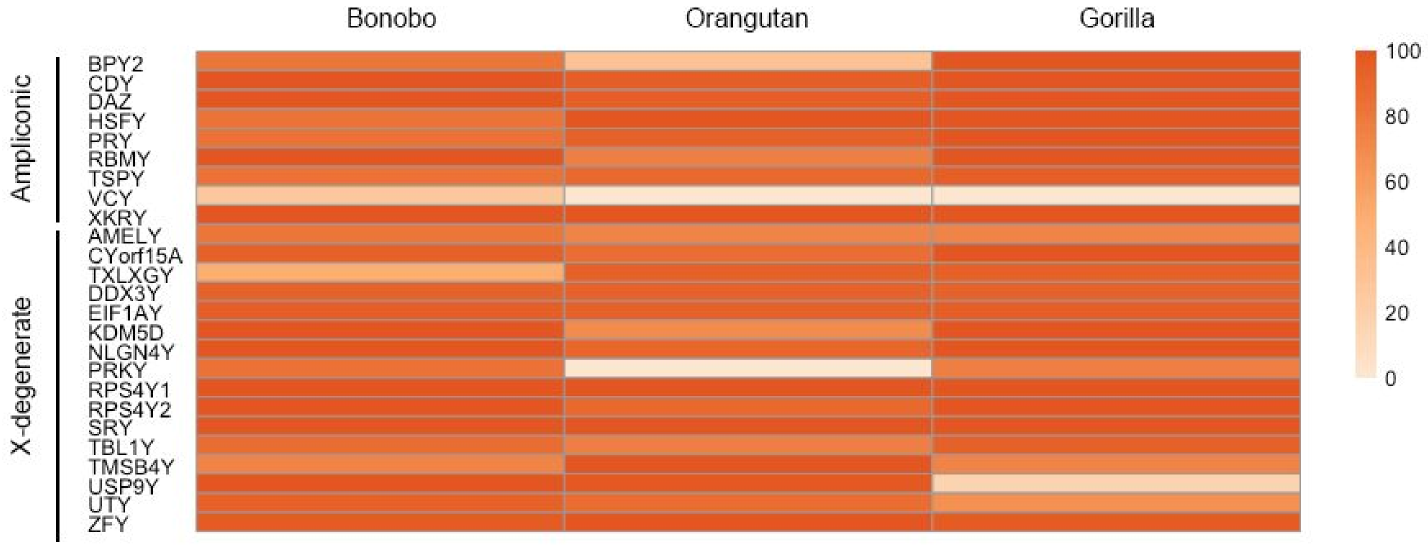


Figure S3. Protein-coding gene sequence retrieval in the new Y assemblies.
(A) For evaluation purposes, we aligned the scaffolds from each of bonobo, Sumatran orangutan and gorilla Y chromosome assemblies to species-specific or closest-species-specific reference coding sequences using BWA-MEM (v.0.7.10) (6). Next, we visualized the alignment results in Integrative Genomics Viewer (IGV) (v.2.3.72). Consensus sequences were retrieved to evaluate sequence coverage over the reference sequence (in percentage) using BLAST (34). The results were represented as heatmaps using pheatmap package in R.
(B) Table of accession numbers used as queries.

A



B

Coding sequence	Bonobo	Orangutan	Gorilla
BPY2	AY958084.1	KP141770.1	GATR01000016.1
CDY	AY958081.1	KP141772.1	GATR01000022.1
DAZ	AY958083.1	KP141773.1	GATR01000021.1
HSFY	CCDS35475.1	KP141774.1	GATR01000004.1
PRY	CCDS14799.1	KP141776.1	GATR01000017.1
RBMY	AH014838.2	KP141777.1	GATR01000007.1
TSPY	AY958082.1	KP141780.1	GATR01000012.1
VCY	XM_003318999.5	CCDS56617.1	CCDS56617.1
XKRY	NM_004677.2	NM_004677.2	NM_004677.2
AMELY	NM_001102459.1	ENST00000215479.10	FJ532255.1
CYorf15A	AY633113.1	NR_045128.1	FJ532256.1
TXLNGY	NR_045129.1	GATK01000021.1	FJ532257.1

DDX3Y	AY633112.1	NM_001131248.1	FJ532258.1
EIF1AY	AY633115.1	GATK01000002.1	FJ532259
KDM5D	AY736376.1	GATK01000003.1	FJ532260.1
NLGN4Y	AY728015.1	KP141775.1	FJ532261
PRKY	AY728014.1	ENST00000533551.5	FJ532262
RPS4Y1	AY633110.1	GATK01000004.1	FJ532263.1
RPS4Y2	AY633111.1	KP141778.1	FJ532264
SRY	AY679780.1	KP141779.1	X86382.1
TBL1Y	ENST00000383032.6	GATK01000018.1	FJ532265
TMSB4Y	ENST00000284856.4	GATK01000007.1	FJ532266
USP9Y	ENST00000338981.7	GATK01000005.1	FJ532267
UTY	AY679781.1	GATK01000020.1	FJ532268.1
ZFY	AY679779.1	GATK01000006.1	AY913764

Figure S4. (A, C, E) Thresholds used for classification of windows into X-degenerate versus ampliconic, and (B, D, F) average copy number for overlapping 5-kb windows.
X-degenerate scaffolds are shown in yellow, whereas ampliconic windows are shown in blue (or turquoise).

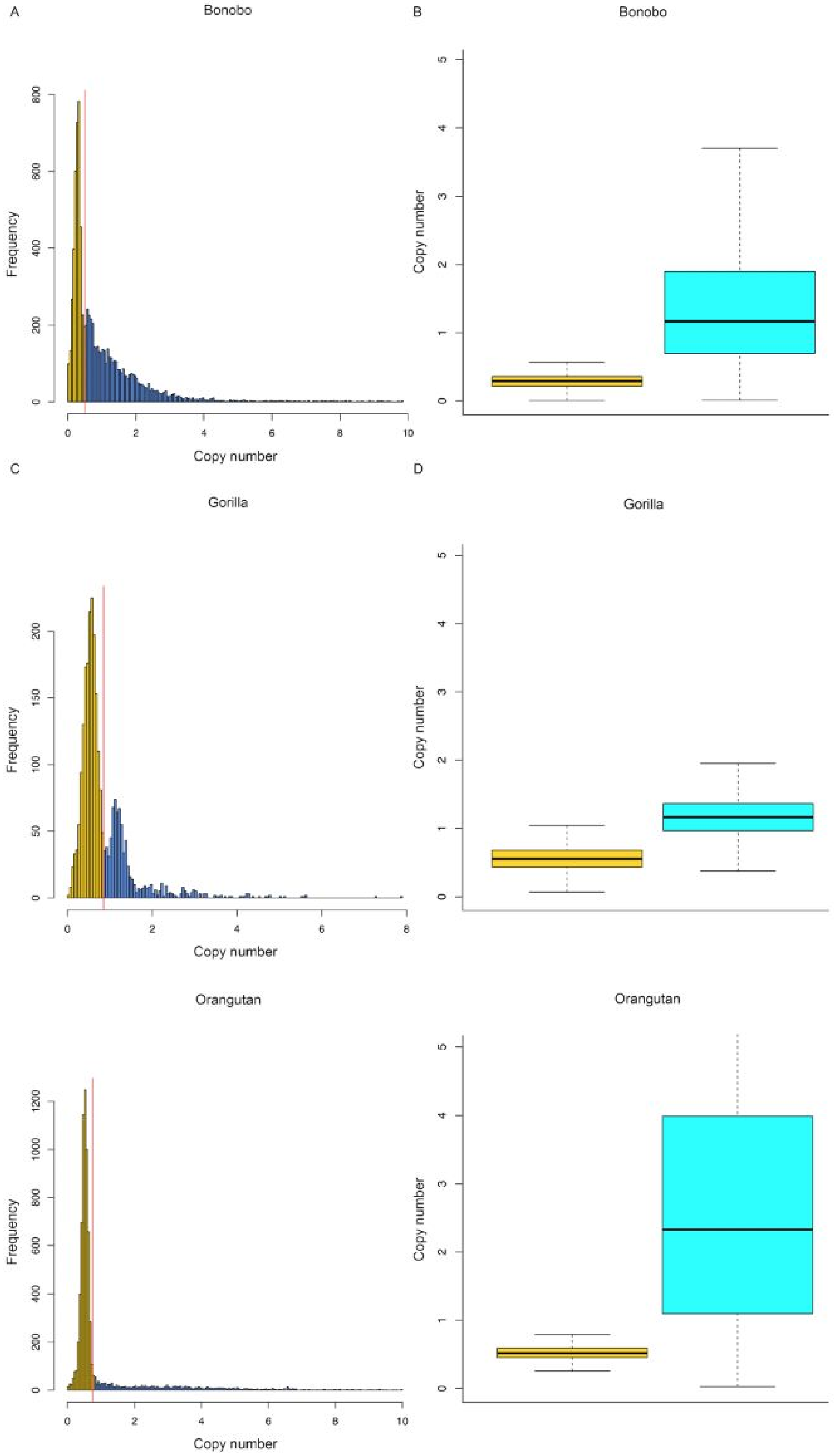


Figure S5. Shared and lineage-specific sequences in multi-species alignments.
Counts of aligned bases in each set of species. For example, the first five bars reflect alignments involving human, chimpanzee, bonobo, gorilla, and orangutan; the sixth through ninth bars reflect alignments involving human, chimpanzee, bonobo, and gorilla but not orangutan, etc.; the last five columns reflect species-specific sequences.

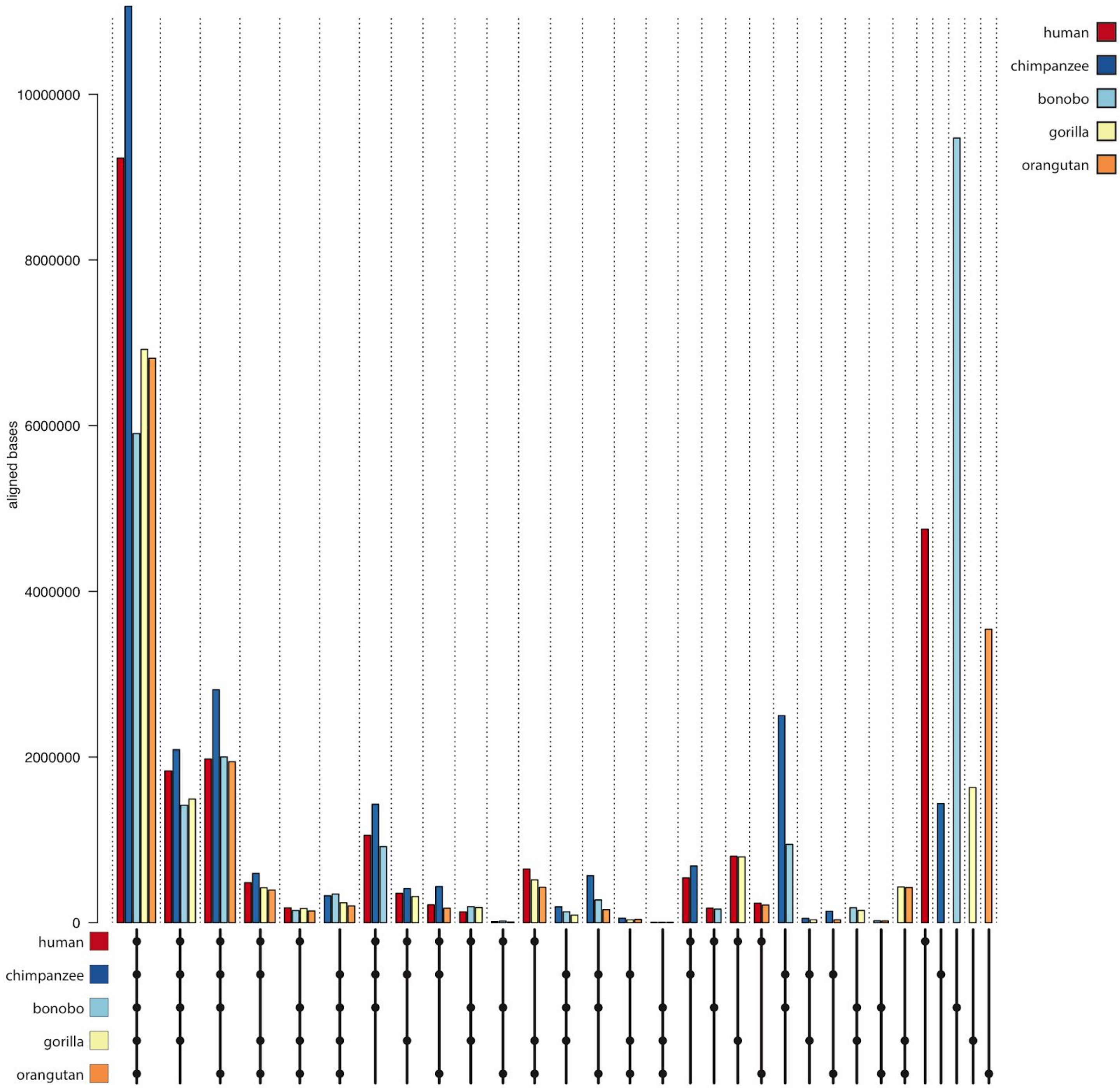


Figure S6. Reconstructed gene content of great apes.
The first six rows have information about the gene content of great apes and macaque, which were used as an input for the model of Iwasaki and Takagi (27). The other rows were reconstructed by the model. BC - common ancestor of bonobo and chimpanzee; BCH - common ancestor of bonobo, chimpanzee, and human; BCHG - common ancestor of bonobo, chimpanzee, human, and gorillas. GA - common ancestor of great apes. We defined the presence of *RPS4Y2* and *MXRA5Y* in bonobo and orangutan based on AUGUSTUS, and Y chromosome specific testis transcriptome assembly results (35). The presence of *RPS4Y2* gene was confirmed in bonobo through gene prediction (shares 100% identity with chimpanzee *RPS4Y2*) and assembled transcript sequences (shares 99.6% identity with chimpanzee *RPS4Y2*). *MXRA5Y*, which is psedogenized in human and chimpanzee (36), was missing in orangutan (no gene prediction or transcript found) and pseudogenized in bonobo (gene prediction annotated as X chromosome homolog *MXRA5* and missing the first three exons of *MXRA5* in its sequence). In the case of gorilla, we did not find its transcript in transcriptome assembly and a BLAT search of human *MXRA5Y* gene (NC_000024.10:11952465-11993293) resulted in a 12-kb long hit which is around 20% of the gene. We assumed the gene is lost in gorilla as well.

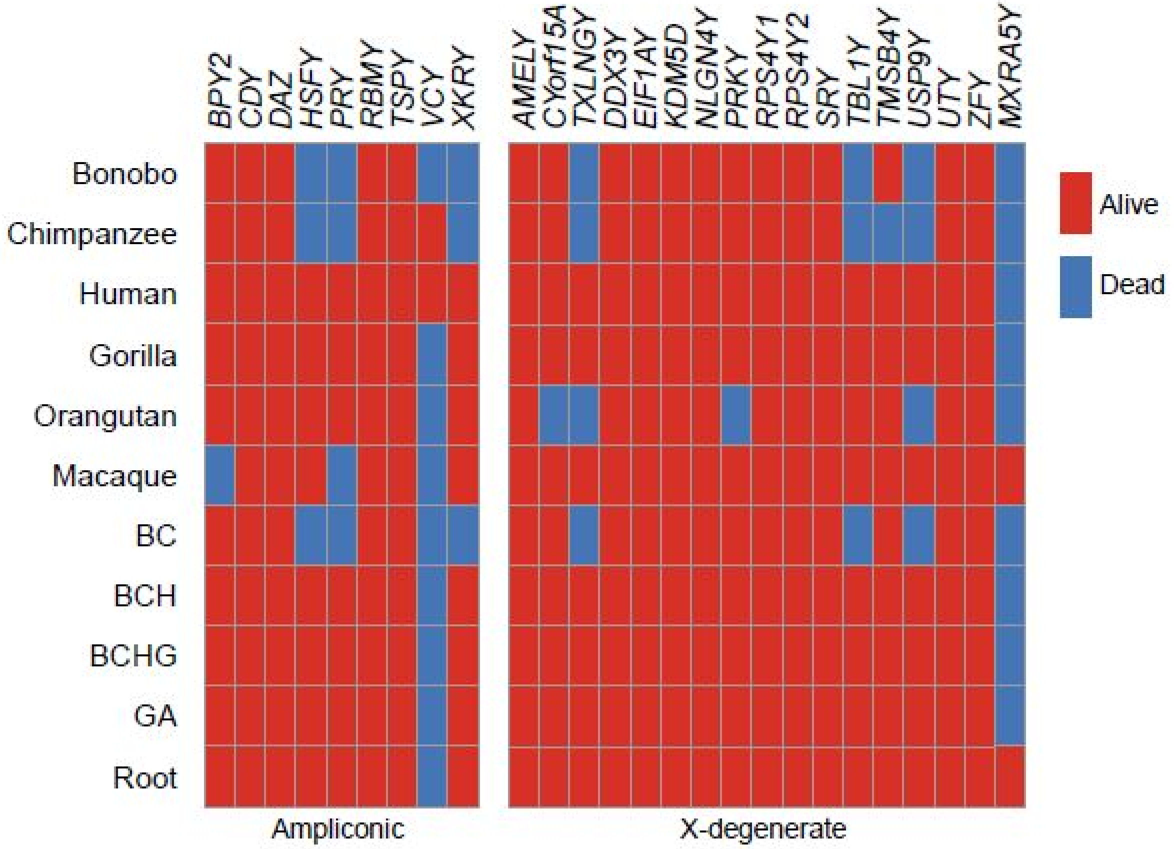


Figure S7. IGV screen shots of peaks of DNase-seq, H3K4me1 and H3K27ac marks on human palindrome P6, and of CREB1 on human palindrome P7.

A. Peaks on both arms of P6 are shown within the blue and grey boxes. **B**. Zoom-in view of peaks on the left arm of palindrome P6. The coverage track represents the depth of coverage and peaks track represents the peaks identified by macs2 (37). **C**. Zoom-in view of peaks on the right arm of palindrome P6. The coverage track (top) represents the depth of coverage, and the peaks track (bottom) represents the peaks identified by macs2 (37). **D.** Peaks on both arms of palindrome P7 are shown. The coverage track (top) represents the depth of coverage and the peaks track (bottom) represents the peaks identified by macs2 (37). See Table S12 for the coordinates of P7.

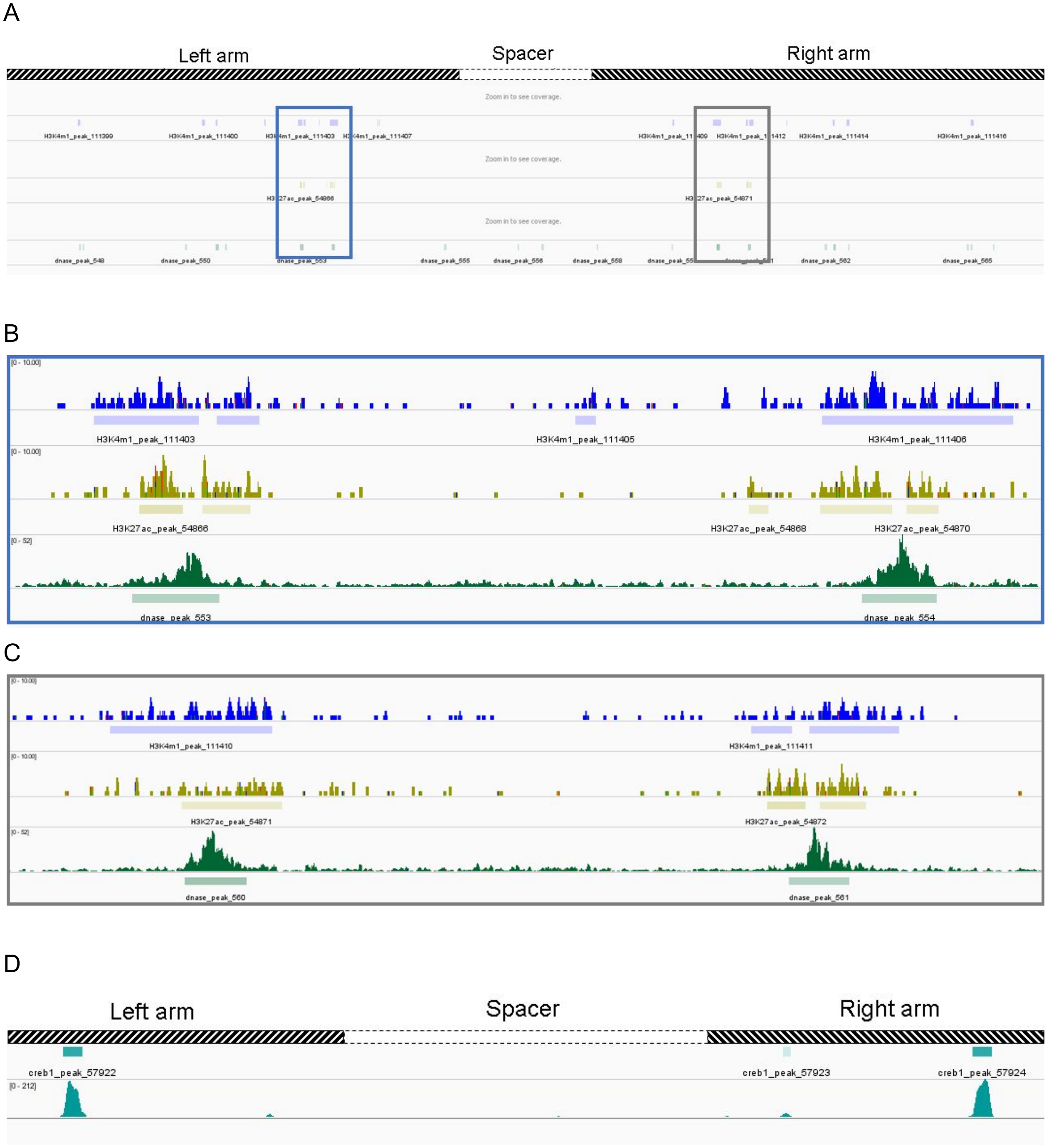


Figure S8. Chromatin interactions on the human Y chromosome.
A. Chromatin contacts split by groups: palindrome-palindrome contacts, palindrome-other (i.e. mixed) and other-other (chromatin interactions in which palindromes are not involved). **B**. The probability of interactions as estimated by (28); the probability is the highest for the palindromic group, in which both reads from a pair fall into human palindromes. The table is based on human iPSC data (29). See also Table S9.

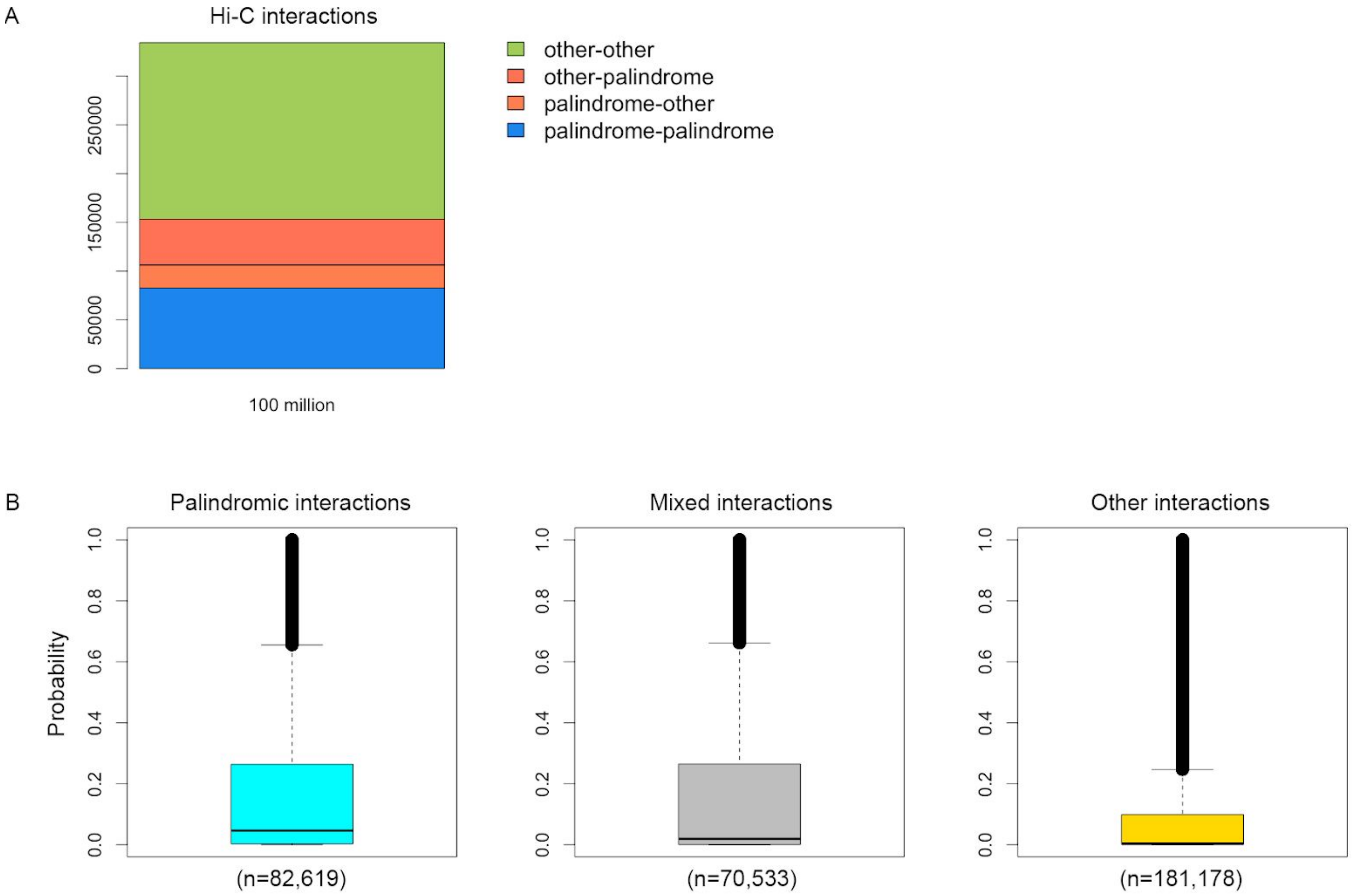
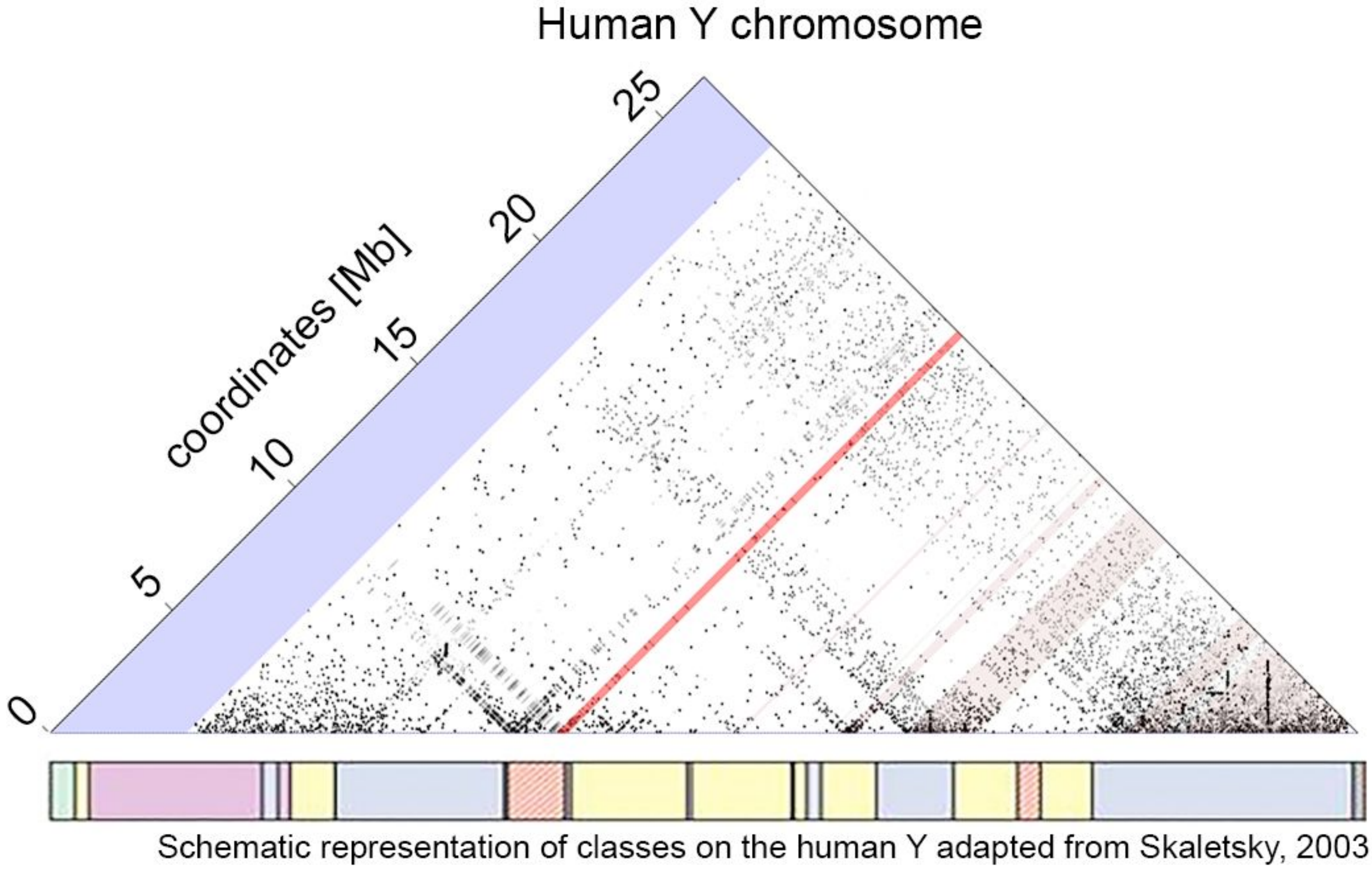


Figure S9. Hi-C contact map generated for Human Umbilical Vein Cells (HUVEC), using the data from (38)**.**



References

1. M. Tomaszkiewicz, P. Medvedev, K. D. Makova, Y and W Chromosome Assemblies: Approaches and Discoveries. *Trends Genet.* **33**, 266–282 (2017).

2. H. Skaletsky, *et al.*, The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).

3. J. F. Hughes, *et al.*, Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463**, 536–539 (2010).

4. R. Vegesna, M. Tomaszkiewicz, P. Medvedev, K. D. Makova, Dosage regulation, and variation in gene expression and copy number of human Y chromosome ampliconic genes. *PLoS Genet.* **15**, e1008369 (2019).

5. T. Derrien, *et al.*, Fast computation and applications of genome mappability. *PLoS One* **7**, e30377 (2012).

6. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013).

7. A. Zeileis, G. Grothendieck, zoo:S3Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software* **14** (2005).

8. T. Miyata, H. Hayashida, K. Kuma, K. Mitsuyasu, T. Yasunaga, Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 863–867 (1987).

9. K. D. Makova, W.-H. Li, Strong male-driven evolution of DNA sequences in humans and apes. *Nature* **416**, 624–626 (2002).

10. M. A. Wilson Sayres, Genetic Diversity on the Sex Chromosomes. *Genome Biol. Evol.* **10**, 1064–1078 (2018).

11. N. Yu, M. I. Jensen-Seaman, L. Chemnick, O. Ryder, W.-H. Li, Nucleotide Diversity in Gorillas. *Genetics* **166**, 1375–1383 (2004).

12. M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19 Suppl 2**, ii215–25 (2003).

13. J. T. Robinson, *et al.*, Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

14. R. S. Harris, Improved pairwise Alignmnet of genomic DNA (2007).

15. SMIT, F. A. A., Repeat-Masker Open-3.0. *http://www.repeatmasker.org* (2004) (October 23, 2018).

16. B. Paten, *et al.*, Cactus: Algorithms for genome multiple sequence alignment. *Genome Res.* **21**, 1512–1528 (2011).

17. G. Hickey, B. Paten, D. Earl, D. Zerbino, D. Haussler, HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**, 1341–1342 (2013).

18. J.-M. Chen, D. N. Cooper, N. Chuzhanova, C. Férec, G. P. Patrinos, Gene conversion: mechanisms, evolution and human disease. *Nat. Rev. Genet.* **8**, 762–775 (2007).

19. A. J. Jeffreys, C. A. May, Intense and highly localized gene conversion activity in human meiotic crossover

hot spots. *Nat. Genet.* **36**, 151–156 (2004).

20. S. A. Sawyer, GENECONV: a computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University in St. Louis. *St. Louis* (1999).

21. Z. Yang, PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).

22. K. Sahlin, M. Tomaszkiewicz, K. D. Makova, P. Medvedev, Deciphering highly similar multigene family transcripts from Iso-Seq data with IsoCon. *Nat. Commun.* **9**, 4601 (2018).

23. M. A. Larkin, *et al.*, Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).

24. N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).

25. S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

26. P. Moorjani, C. E. G. Amorim, P. F. Arndt, M. Przeworski, Variation in the molecular clock of primates. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10607–10612 (2016).

27. W. Iwasaki, T. Takagi, Reconstruction of highly heterogeneous gene-content evolution across the three domains of life. *Bioinformatics* **23**, i230–9 (2007).

28. Y. Zheng, F. Ay, S. Keles, Generative modeling of multi-mapping reads with mHi-C advances analysis of Hi-C studies. *eLife* **8** (2019).

29. I. E. Eres, K. Luo, C. J. Hsiao, L. E. Blake, Y. Gilad, Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates. *PLoS Genet.* **15**, e1008278 (2019).

30. E. Garrison, G. Marth, Haplotype-based variant detection from short-read sequencing. *arXiv [q-bio.GN]* (2012).

31. M. Tomaszkiewicz, *et al.*, A time- and cost-effective strategy to sequence mammalian Y Chromosomes: an application to the de novo assembly of gorilla Y. *Genome Res.* **26**, 530–540 (2016).

32. A. C. E. Darling, B. Mau, F. R. Blattner, N. T. Perna, Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403 (2004).

33. J. Krumsiek, R. Arnold, T. Rattei, Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* **23**, 1026–1028 (2007).

34. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

35. Rahulsimham Vegesna, Marta Tomaszkiewicz, Oliver A. Ryder, Rebeca Campos-Sánchez, Paul Medvedev, Michael DeGiorgio, and Kateryna D. Makova, Ampliconic genes on the great ape Y chromosomes: Rapid evolution of copy number but conservation of expression levels. *In Review* (2020).

36. J. F. Hughes, *et al.*, Strict evolutionary conservation followed rapid gene loss on human and rhesus y chromosomes. *Nature* **483**, 82–87 (2012).

37. J. M. Gaspar, Improved peak-calling with MACS2. *bioRxiv*, 496521 (2018).

38. S. S. P. Rao, *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).