# Epigenetic and evolutionary features of ape subterminal heterochromatin

DongAhn Yoo[1], Katherine M. Munson[1], Evan E. Eichler[1,2]

**Affiliations**

1. Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA
2. Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

*Correspondence to ee3@uw.edu

**Corresponding author**
Evan E. Eichler, Ph.D.
Department of Genome Sciences
University of Washington School of Medicine
3720 15th Ave NE, S413C
Box 355065
Seattle, WA 98195-5065
Email: ee3@uw.edu
Tel: 1-206-543-9526

Keywords: Genomics, Subterminal heterochromatin, Evolution, Great ape, Comparative genomics

## ABSTRACT

Many African great ape chromosomes possess large subterminal heterochromatic caps at their telomeres that are conspicuously absent from the human lineage. Leveraging the complete sequences of great ape genomes, we characterize the organization of subterminal caps and reconstruct the evolutionary history of these regions in chimpanzees and gorillas. Detailed analyses of the composition of the associated terminal 32 bp satellite array from chimpanzee (termed pCht) and intervening segmental duplication (SD) spacers confirm two independent origins in the *Pan* and gorilla lineages. In chimpanzee and bonobo, we estimate these structures emerged ~7.7 million years ago (MYA) in contrast to gorilla where they expanded more recently, ~5.0 MYA, and now make up 8.5% of the total gorilla genome. In both lineages, the SD spacers punctuating the pCht heterochromatic satellite arrays correspond to pockets of decreased methylation, although in gorilla such regions are significantly less methylated ($p$<2.2e-16) than chimpanzee or bonobo. Allelic pairs of subterminal caps show a higher degree of sequence divergence than euchromatic sequences, with bonobo showing less divergent haplotypes and less differentially methylated spacers. In contrast, we identify virtually identical subterminal caps mapping to nonhomologous chromosomes within a species, suggesting ectopic recombination potentially mediated by SD spacers. We find that the transition regions from heterochromatic subterminal caps to euchromatin are enriched for structural variant insertions and lineage-specific duplicated genes. Our findings suggest independent evolution of subterminal caps converging on a common genetic and epigenetic structure that promoted ectopic exchange as well as the emergence of novel genes at transition regions between euchromatin and heterochromatin.

## INTRODUCTION

With respect to humans, chromosome karyotypes of nonhuman African great apes (chimpanzee, bonobo and gorilla) differ by the presence of subterminal heterochromatic caps, which were recognized cytogenetically more than 40 years ago (Yunis and Prakash 1982). Among chimpanzee, gorilla and bonobo, the subterminal caps are differentially distributed among chromosomes. These caps are composed of hundreds of kilobase pairs (kbp) of long satellite arrays where the basic repeat unit, known as pCht satellite (also called as StSat or subterminal satellite), is 32 bp in length (Royle et al. 1994; Koga et al. 2011; Ventura et al. 2012). Previous cytogenetic studies in chimpanzees have shown that subterminal caps can form unique terminal associations in 32% of spermatocytes during late meiotic prophase I, so called post-bouquet structures. These cytogenetic structures are thought to promote ectopic recombination through persistent interactions between subtelomeric sequences among homologous and nonhomologous chromosomes (Hirai et al. 2019). The function of the subterminal caps is not known though have been proposed to help stabilize African great ape genomes by preventing or buffering against interchromosomal exchanges of more proximal subtelomeric sequences (Hirai et al. 2019) or contributing to the replication biology of telomeres (Novo et al. 2013).

Subsequent investigations into the organization of these regions discovered segmental duplication (SD) “spacers” interdigitated between the large arrays of pCht satellite DNA (Ventura et al. 2012; Yoo et al. 2025). SD spacers of different phylogenetic origin were identified in *Pan* and gorilla lineages suggesting that the subterminal cap structures arose independently. Previous studies proposed an evolutionary model, where a human–chimpanzee pericentric inversion followed by Chromosome 2 fusion contributed to the predisposition or loss of subterminal caps in nonhuman apes and human genomes (Royle et al. 1994; Ventura et al.

2012). Jiang and colleagues further refined the Chromosome 2 fusion breakpoint and subterminal SDs in humans (Jiang et al. 2024). In the most recent study, the heterochromatic cap sequences were fully resolved using a hybrid long-read assembly approach (Oxford Nanopore Technologies [ONT] and PacBio high-fidelity [HiFi]) and assembly phasing (HiC and parent–child trios) (Yoo et al. 2025).

While these subterminal satellites are absent in other apes like orangutans (Haaf and Schmid 1987; Ijdo et al. 1991; Ventura et al. 2011), such structures are not unique to African great apes. A similar, albeit larger, subterminal heterochromatic cap structure composed of alpha-satellite DNA was described for siamang genomes from the gibbon lineage (Koga et al. 2012). We recently characterized the heterochromatic cap sequence and structure of the siamang genome (Yoo et al. 2025). While it is similar in form to that of the great apes (i.e., blocks of satellite interdigitated with a spacer that shows epigenetically reduced methylation), the component sequences completely differ. In siamang, there is no pCht satellite but instead alpha-satellite sequence and the spacer is made up of a 50 kbp SD of different origin than either the *Pan* or gorilla spacer. The independent emergence of subterminal satellite structures in multiple primate lineages (gorilla, chimpanzee and siamang) so similar in structure and epigenetic form suggests convergent evolution of potential functional significance. In this study, we leverage the fully resolved sequence of telomere-to-telomere ape genomes to more systematically investigate the evolution, structure, and epigenetic properties. The availability of two fully resolved haplotype assemblies from each ape species allowed for investigations into patterns of allelic variation and between-species comparisons of subterminal organization and differences in methylation patterns. We examined the telomeric transition regions from heterochromatin to euchromatin

leading to potentially new insights into the genic content and stability of these regions across ape genomes.

## RESULTS

***African great ape subterminal heterochromatic cap structure and methylation.*** Using the recently published ape telomere-to-telomere (T2T) genomes (Yoo et al. 2025) (**Supplemental Table S1**), we estimated the size and content of each chromosomal cap in chimpanzee, bonobo, and gorilla (**Fig. 1 & Supplemental Table S2**) defining chromosomes based on their synteny with human (McConkey 2004). Overall, gorilla chromosomes have larger and greater number of subterminal caps where they account for 584.1 Mbp (8.46%) of the diploid genome, while that of chimpanzee and bonobo genomes represent 337.9 and 314.7 Mbp (5.4% and 5.0%), respectively. We identified subterminal caps on both homologs (allelic pairs) in both chimpanzee and gorilla, while bonobo showed patterns consistent with heteromorphic variation (present on only one of two homologous chromosomes). For example, the bonobo p-arms of Chr14 (hsa13) and Chr19 (hsa17), and both p- and q-arms of Chr18 (hsa16) possessed subterminal caps on only one of the two haplotypes for this individual. This heteromorphism and the fact that bonobo possesses the lowest number of subterminal caps (n=46 compared to 57 in chimpanzees and 79 in gorillas) suggest that the caps are less prominent or less stable in this species, although we caution that additional samples need to be examined. Notably, the short (p-) arms of chimpanzee and bonobo more consistently harbor a greater number of subterminal caps (39 and 31) when compared to q-arms (14 and 11). In contrast, the gorilla genome shows a more uniform distribution with more q-arm (45) subterminal caps than p-arms (34). The lengths of the subterminal caps were highly variable ranging from <1 Mbp in chimpanzee Chr14 (hsa13), to a 35-fold longer length of 18.9

Mbp in bonobo Chr6 (hsa7) (**Fig. 1A**). Unlike gorilla, two interstitial satellite arrays are observed in chimpanzee and bonobo mapping to Chr6 (hsa7) and 14 (hsa13) (Royle et al. 1994).

As previously reported (Yoo et al. 2025), analysis of the subterminal caps revealed a higher-order organization of pCht satellite arrays of variable length (ranging from <10 kbp to hundreds of kbp with a mean of 341-398 kbp) interdigitated with SD spacers (**Fig. 1B-D; Supplemental Fig. S1**). We find that pCht satellites are generally more than 80% identical with the ones in closer proximity showing higher identity within a chromosomal arm (**Fig. 1B**). While different subtypes of pCht can be distinguished, we derived a consensus for each depending on species (**Fig. 1C & Supplemental Fig. S2**). Despite the sequence variation, we find that the gorilla satellite consensus is almost identical to that of the *Pan* lineage (chimpanzee and bonobo), with the exception of the thymine in the 15$^{th}$ position, which shows less variability in the *Pan* lineage. In terms of epigenetic status, we investigated methylation patterns of both spacers and satellites. While we have data from both PacBio and ONT, we chose to focus on methylation calls from ONT for two reasons. First, previous studies have shown (Liu et al. 2022; Sigurpalsdottir et al. 2024) greater sensitivity in detection, and second, ultra-long reads are necessary to uniquely anchor reads in long stretches of DNA. Our study finds these satellite sequences show high levels of 5mC methylation, with gorilla showing average methylation of 83.3% compared to 78.4% in bonobo and 73.6% in chimpanzee (**Supplemental Figs. S3 & S4A**).

The pCht satellite arrays are interrupted irregularly by pockets of less methylated subterminal SD spacers (**Supplemental Figs. S3 & S4**). These SDs originate from euchromatic sequence in the ancestral ape lineage (**Fig. 1C**). For example, the ancestral subterminal SD spacer in gorilla maps to the intronic region of the gene *MALD1* of Chr8 (hsa10, p12.31) and subsequently expanded to ~700 copies per gorilla haplotype. In contrast, the spacer of the *Pan* lineage originates from a

subsequence of *PGM5*, Chr11 (hsa9, q21.11) (**Supplemental Fig. S5**), and expanded to ~413 and ~430 copies per haplotype for chimpanzee and bonobo, respectively (**Supplemental Table S3**). The gorilla spacers contained 1.8-fold greater LINE content and *Pan* spacers 2.4-fold greater DNA transposon content compared to euchromatic regions; however, we found no statistical significance from this comparison (one-sided permutation test $p$=0.15 and 0.07, respectively). The average length of gorilla and *Pan* spacers are 34 and 32 kbp, respectively (**Fig. 1C-D**). Of note, the length of both spacers and pCht arrays are significantly longer in gorilla compared to chimpanzee and bonobo (two-sided Wilcox test $p$<2.2e-16 and $p$=6e-10, respectively) consistent with their independent origin and expansion in each lineage (Ventura et al. 2012).

Examining the epigenetic status of SD spacers, we find a significantly lower rate of 5mC methylation in gorilla (average of 38.8%) relative to chimpanzee (40.2%) and bonobo (49.9%), with $p$<2.2e-16, two-sided Wilcox test (**Supplemental Figs. S3 & S4B**). Using SD spacers mapping outside of pCht satellite arrays as a comparative control, we find that 99.8% and 98.5% of gorilla and chimpanzee spacers are less methylated while only 17.8% of bonobo spacers are less methylated than their interstitial paralogs (**Supplemental Fig. S4B**). In chimpanzee and bonobo, spacer length positively correlates with CpG methylation although no such pattern is observed for gorilla where reduced methylation of spacer is much more uniform (**Supplemental Fig. S4C**). Using previously generated Iso-Seq data (Makova et al. 2024; Yoo et al. 2025), we were unable to identify spliced transcripts originating from the pCht satellite arrays or SD spacers (**Supplemental Tables S4-S5**) although note that Iso-Seq has relatively low sensitivity of detection (Novo et al., 2013).

***Evolution of the spacer and satellite sequence.*** In order to reconstruct the evolutionary history of the subterminal satellites, we performed an independent analysis of the two classes of repeats which define its composition: the pCht (~32 bp) satellite that exists in millions of copies throughout the ape genomes and the *Pan* and gorilla SD spacers (32-34 kbp) that represent independent SD expansions (400-700 copies per genome) in association with the pCht satellites. While the SD spacer is far less abundant than the pChts, its length is thousands of times longer, providing a robust signal for multiple sequence alignment (MSA), phylogenetic reconstruction, and timing estimates.

For the pCht satellites, we first extracted all pCht units from chimpanzee (n=6,881,426), bonobo (n=6,616,396), and gorilla (n=14,248,529) subterminal caps. From this, we selected the most abundant 19,012 pCht variants (85% of the total pChts) that occur more than 100 times in these genomes. Under the assumption that more closely related subterminal arms would likely share a similar pCht composition, we performed an all-to-all pairwise comparison between subterminal heterochromatic caps that contain at least 2,000 pChts to identify those that are more closely related to one another. We applied unsupervised hierarchical clustering (pvclust) (Suzuki and Shimodaira 2006) across the three species and constructed a tree based on normalized counts of each pCht variant between subterminal heterochromatic caps (**Fig. 2A, Methods**). The composition of pCht variant types clearly distinguishes the gorilla and *Pan* lineages with the exception of Chr5 (hsa6) q-arms of gorilla, which groups with the subterminal caps (ChrX) from the *Pan* lineage (**Fig. 2A**). This suggests a possible common origin, but the majority of the subterminal caps have formed independently as a result of recurrent expansions in each lineage. Among the subterminal caps from chimpanzee and bonobo, pCht variant compositions do not distinguish between the two species. The similarity in composition may arise from the pCht

sequences sharing a common origin or subsequent gene flow between the two lineages after speciation (Fontsere et al. 2022).

Intraspecies sequence comparisons of the subterminal satellites show that some nonhomologous chromosomes are more similar in composition than their allelic counterparts—i.e., homologous subterminal caps do not always pair (e.g., GGO.h2.3p vs. GGO.h1.Xp; **Fig. 2A**). Under the assumption that nonhomologous subterminal caps are subject to ectopic recombination (Ventura et al. 2012; Hirai et al. 2019), we sought to classify the major pCht subtypes using a two-step *k*-means clustering approach (**Methods, Supplemental Fig. S6**). We identified eight higher-order satellite block types across the 167 subterminal caps (**Fig. 2B**) from the three ape species. Consistent with the hierarchical clustering tree, the gorilla and *Pan* lineages are largely distinct. Gorilla subterminal caps consist of k2, k4, and k5 block types while the *Pan* lineage caps are composed of k3, k6, k7, and k8 types. Once again, gorilla Chromosome 5 (hsa6) is an exception with a largely uniform composition of the k1 block type, which is shared with the X Chromosome p-arms of bonobo and chimpanzee and to a lesser degree with a fraction of other *Pan* autosomes. The k1 block type, thus, is distinct in being the only block type shared among the gorilla and *Pan* lineages, potentially identifying the ancestral pCht. We also note that this block type is composed of the largest portions of common pChts shared among *Pan* and gorilla (**Supplemental Fig. S7**).

As a second approach, we used the SD spacer sequences of the gorilla and *Pan* lineages as a proxy to trace the evolutionary history of the subterminal heterochromatic caps. In this approach, we extracted a shared segment (>20 kbp in length and >90% identity) from the spacer for gorilla and the spacer for the *Pan* lineage and generated an MSA to construct a maximum likelihood phylogenetic tree for each lineage (**Methods & Fig. 3**). Using Asian apes as an outgroup

(**Methods**), the gorilla spacer phylogenetic tree predicts that the ape ancestral locus (hsa10) began to duplicate 5.6 (5.1-6.1) million years ago (MYA) creating an interstitial copy mapping to *Pan* Chr6 (hsa7) that is located 2 kbp away from an interstitial pCht satellite array. This interstitial spacer in the *Pan* lineage is ancestral to all other gorilla subterminal cap spacers beginning to duplicate soon thereafter (5.0 [4.5-5.6] MYA) (**Fig. 3A**), although it is no longer identified at the syntenic location in gorilla or human. The topology of the gorilla phylogenetic tree suggests a series of initial stepwise duplications, e.g., the gorilla Chr5 (hsa6) q-arm is estimated to have occurred 3.8 (3.4-4.4) MYA. After these initial duplications, a subsequent burst of gorilla spacers began to occur 2.5 (2.1-2.9) MYA giving rise to the majority of SD spacers in the gorilla genome. We also note that the phylogenetic tree of the duplicated spacers shows signs of incomplete lineage sorting, which does not follow the species tree (**Fig. 3**), consistent with additional trees (**Supplemental Fig. S8**).

Although the SD spacer in the *Pan* lineage differs with respect to its ancestral origin (hsa9), similar to the gorilla it is predicted to begin its duplication in the common ancestor of humans, chimpanzees, and gorillas creating multiple interstitial copies (hsa4, 5, 8, and X) as well as copies located subterminal of hsa2A/2B—the fusion point (Jiang et al. 2024) for the formation of human Chromosome 2 (**Fig. 3B**). Overall, branch lengths or sequence identity of the chimpanzee phylogeny are significantly longer ($p$<2.2e-16 by two-sided Wilcox test; **Supplemental Fig. S9**) than gorilla consistent with a series of more ancient duplications in the *Pan* lineage. We estimate the oldest divergence time among the *Pan* lineage SD spacers to be 7.7 (7.2-8.5) MYA, which is close to the divergence time of human and gorilla ancestral orthologs (8.6 [8.3-9.3] MYA; Kumar et al. 2022). The chimpanzee SD spacers associated with the ChrX p-arm subterminal caps, which also share ancient higher-order pCht blocks, represent some of the most ancient

spacer duplications (**Fig. 3B**). Similar to gorilla, the topology of the chimpanzee spacer tree predicts a burst of duplications, albeit much older (1.3 MYA—average of terminal branch length) when compared to gorilla (average of 0.4 MYA; **Supplemental Fig. S9B**). It should be noted, however, that the spacers from homologous chromosome arms do not always cluster, potentially as a result of ectopic recombination.

***Patterns of allelic and nonallelic variation.*** A striking feature of the subterminal heterochromatic caps is the high degree of allelic sequence diversity observed between two haplotypes (**Fig. 4A-C; Supplemental Data**). Only 12% (10/82) of haplotype comparisons show >99% sequence identity with >50% alignment coverage. Most haplotype comparisons show greater than 10% sequence divergence (**Supplemental Figs. S10-S13**). A comparison with other genomic regions, including centromere, acrocentric and remaining regions, reveals that subterminal caps were the regions with the greatest degree of allelic divergence (**Supplemental Fig. S14**). Notably, the degree of allelic divergence differs among the species. Gorilla and chimpanzee show on average greater sequence divergence (10.8 and 11%, respectively) when compared to bonobo (8.7%) (**Fig. 4D**). When comparing nonallelic versus allelic sequence identity, gorilla and chimpanzee distributions are largely indistinguishable, in contrast to bonobo where allelic alignments now show significantly greater allelic sequence identity when compared to nonallelic alignments (**Fig. 4E**). This feature is reflected in a matrix percent identity plot where each subterminal haplotype is aligned to every other within the individual (**Fig. 4B,C**). In contrast to bonobo where more than a third of allelic comparisons show the highest degree of sequence identity (as indicated by the diagonal in **Fig. 4C**), gorilla shows no such pattern. Instead, allelic and nonallelic patterns are largely indistinguishable perhaps providing evidence of more rapid, ongoing ectopic recombination occurring in this lineage.

As part of our all-by-all alignment analysis of subterminal caps, we searched for specific examples where there was evidence of a recent ectopic exchange (>99.5% identity) between nonhomologous chromosomes. We identified 11 candidates for ectopic exchange (8 in gorilla, 3 in chimpanzee, and none in bonobo). For example, we identified a potential 5 Mbp ectopic exchange between GGO.3p (Chr2/hsa3) and Chromosome Xp (**Fig. 5A**). We refined the breakpoints of the sequence exchange and mapped both to different SD-spacer regions—corresponding to the 3$^{rd}$ and 25$^{th}$ SD spacer positions on the short arm of the gorilla X Chromosome (**Fig. 5B**). Analyzing the remaining candidates (**Fig. 5C; Supplemental Figs. S15-S17**), we found 3/8 examples where the breakpoint mapped precisely within the spacer and two additional examples where one of the breakpoints mapped within 10 kbp of a spacer. Given that the SD spacers represent a small fraction of the subterminal heterochromatic cap, we performed a simulation to test the proximity of breakpoints to SD spaces. We find that candidate breakpoints of recent ectopic exchange (>99.8 identity) are more likely ($p$=0.035, one-sided Wilcox ranked sum test; **Supplemental Fig. S18**) to map in close proximity to SD spacers than expected based on a null subterminal distribution (**Fig. 5D**).

As a final test of the potential role of SD spacers in mediating ectopic exchange, we developed a phylogenetic approach to test for this association. We reasoned that if spacers were preferential sites of ectopic recombination during evolution, dividing the ~30 kbp spacer sequence into two segments (first and second half, located at telomere and centromere directions, respectively) would capture two distinct phylogenetic histories when comparing to each other (**Supplemental Fig. S19**) or to the topology of the full SD spacer (**Fig. 5E**). To test for significance, we calculated the distribution of Robinson-Foulds metric between individual bootstrap tree topologies (each replicate with n=1000) to their consensus topology to generate a null

distribution of the expected statistic and contrasted this with the partial sequence trees (**Fig. 5F**). We find a significant shift in the topology of the SD spacer subset trees ($p$<0.001) for all three species (gorilla, chimpanzee and bonobo), suggesting that historically the SD spacers may have been the preferential point of ectopic exchange between nonhomologous chromosome arms.

***Subterminal heterochromatic-euchromatic transition regions and the evolution of new genes.*** Finally, we examined the euchromatin boundaries and putative evolutionary consequences and epigenetic features of the subterminal caps (**Fig. 6**). Specifically, we compared ape species with (gorilla/chimpanzee) and without (Bornean and Sumatran orangutans) subterminal caps to determine if the rate of structural variation, gene density, and methylation differ in association with the evolution of these large terminal heterochromatic structures (Yoo et al. 2025). Using the human genome as a reference, we find ~24–72 Mbp of novel insertions mapping within 2 Mbp from subterminal caps (African nonhuman apes) or telomeres (Asian great apes without subterminal caps) (**Fig. 6A,B**). The insertions in the African apes are notably more abundant (31.8–72.2 Mbp vs. 23.8–23.9 Mbp in orangutans) and larger (>10–500 kbp in size) when compared to the Asian great apes (**Fig. 6C** & **Supplemental Fig. S27**). This is despite the fact that divergence time between humans to African great apes are close to half the divergence to Asian great apes. We estimate that 36%–42% of the inserted sequences in African apes originated from nonhomologous chromosomes (**Supplemental Fig. S28**)—a potential consequence of the ectopic exchanges extending beyond the subterminal heterochromatic cap.

Leveraging CpG methylation signals associated with the long-read sequence data used to generate these ape assemblies (Yoo et al. 2025), we investigated methylation at the euchromatin-heterochromatin boundary defined here as the last proximal subterminal pCht satellite unit. Within 200 kbp of this transition, we observe an immediate and sharp drop (~54%) in

methylation among nonhuman African great apes (**Fig. 6D & Supplemental Fig. S29**; Asian great apes shown in **Supplemental Fig. S30**), which quickly resets to a moderate level of hypermethylation (71%) when extended further euchromatically. Correlating with the transition to the lower methylation, we also observe the highest density of SDs and species-specific annotated duplicated genes (Yoo et al. 2025), which also reduces when extended more proximally into the euchromatin of gorilla (**Fig. 6E**) and chimpanzee (**Supplemental Fig. S31**). In Asian great apes, the signature of SDs and duplicated genes is less variable (**Fig. 6E** & **Supplemental Fig. S31**).

In total we identify 204 novel duplicated genes at these subterminal euchromatin boundaries in gorilla (n=85), chimpanzee (n=57), and bonobo (n=62) (**Fig. 6F; Supplemental Tables S6-S7**). The enrichment of lineage-specific duplicated genes at this heterochromatin-euchromatin boundary is most extreme for gorilla (14.6-fold difference) followed by relatively high levels in the *Pan* lineage (4.3- to 4.7-fold) dropping to lower levels in the orangutan (1.6- to 2.8-fold) (**Fig. 6G & Supplemental Fig. S32**). Compared to genome-wide average of each species, we find that the species with subterminal caps (nonhuman African great apes) have a greater enrichment of SDs (3.9- to 7.1-fold difference) than humans (3.0) or Asian great apes (2.3-2.5) at these boundaries (**Fig. 6G & Supplemental Figs. S31 & S33**). In addition, we also find that the pairwise identity between the paralogous SD pairs at the euchromatin boundaries is the highest in gorilla when compared to other species (**Supplemental Fig. S34**). In total, the result suggests an accelerated rate of evolutionary duplication and gene innovation among ape lineages possessing subterminal caps, potentially a consequence of both ectopic exchange and reduced methylation at the boundaries.

## DISCUSSION

The complete sequencing and analysis of the subterminal caps in both haplotypes of a single individual of chimpanzee, bonobo, and gorilla have provided novel insights into both the evolution and functional properties of these complex genomic regions. First, we provide supporting evidence for the largely independent expansion of the subterminal caps in both chimpanzee and gorilla as originally proposed by Ventura et al. (Ventura et al. 2012). Our phylogenetic analyses of both the satellite and the SD spacer sequences generally support this model with some important differences. The detailed pCht satellite higher-order structure analysis, for example, clearly defines a block type (k1) that likely existed in the common ancestor of both gorilla and chimpanzee lineages, and we hypothesize that k1 was lost to the human lineage as a result of incomplete lineage sorting (ILS) or the Chromosome 2 fusion (Royle et al. 1994; Jiang et al. 2024). Subsequent hyperexpansion of pCht in the gorilla and *Pan* lineages created lineage-specific higher-order block types that now account for the bulk of the subterminal heterochromatic caps of each species. Third, using the SD spacer sequences as a phylogenetic marker also supports early duplication and ILS. However, the phylogenetic analysis distinguishes two very different evolutionary trajectories. In the *Pan* lineage, both the formation and expansion (~7.7 [7.2-8.5] MYA) of the subterminal heterochromatin is much more ancient, far predating that of the gorilla SDs and the speciation of bonobo and the common chimpanzee. In contrast, gorilla subterminal heterochromatin began to emerge 5.0 (4.5-5.6) MYA—and hyperexpanded millions of years later (2.5-3.8 MYA). We predict the shared higher-order blocks (k1) of subterminal satellites identified in this study (**Fig. 2B**), as well as multi-copy orthologs of *Pan* lineage SDs in gorilla and human are the remnants of the ancient African great ape subterminal heterochromatin in these genomes.

While the general organization of the subterminal heterochromatic caps is similar among the nonhuman African apes, we document important epigenetic and genetic differences within each species. Owing to its more recent origin, gorilla subterminal heterochromatic caps are more homogenous among the nonhomologous chromosomes, and gorilla SD spacers show the most extensive reduction in CpG methylation. In contrast, bonobo subterminal heterochromatic caps are composed of the oldest or most divergent subterminal SDs and showed a higher degree of heteromorphism (present on only one of the two allelic homologues) when compared to the other apes analyzed in this study, although this observation is based on only one individual per species. Combined with a demonstrable weaker methylation change for the SD spacers (**Supplemental Fig. S4**), much higher allelic identity (**Fig. 4**), and the fact that we find no evidence of recent ectopic sequence exchange in bonobo genomes, this again distinguishes the bonobo lineage, although we caution that additional bonobo genomes derived from primary tissues as opposed to cell lines will need to be assayed to establish this as a species-specific property. While this may indicate a slower expansion or increased rate of decay, it may also be the result of some species-specific differences in repair or transfer of material between chromosomal homologs. Nevertheless, under this model, gorilla would be predicted to be the most active subterminal heterochromatic cap perhaps explaining why it is much more prevalent in this genome adding almost three chromosomes worth of DNA (584 Mbp).

Since the first characterization of African great ape subterminal satellites, there has been limited investigation into the functional significance of subterminal heterochromatic caps. Based on FISH experiments in chimpanzee meiotic sperm cells, a previous study by Hirai et al. 2019(Hirai et al. 2019) suggested that the subterminal heterochromatin was driving ectopic recombination by helping to form bouquet structures among nonhomologous chromosomes during late meiotic

prophase somewhat akin to the nuclear organizing structures associated with the repeat-rich short arms of mammalian acrocentric chromosome (Nurk et al. 2022). Hirai and colleagues further concluded that such subterminal associations may be selectively beneficial by providing greater stability or buffering capacity to subterminal regions of ape chromosomes—areas known to be genetically and evolutionarily unstable (Knight and Flint 2000; Mefford and Trask 2002). In our study, we provide further support for their involvement in ectopic recombination by identifying 11 potential examples of interchromosomal translocations (8 in gorilla and 3 in the common chimpanzee; **Fig. 5C**). Both our breakpoint analysis of these events and our phylogenetic analysis indicate that the SD spacer regions are preferential drivers of this ectopic recombination between nonhomologous chromosome arms. It is possible that both the high degree of sequence identity of these 30 kbp segments and their reduced methylation status provide an ideal open-chromatin substrate for nonallelic homologous recombination, although the mechanism still remains to be determined. However, unlike Hirai's prediction of conferring genomic stability, our evolutionary analysis suggests the complete opposite effect. We find far stronger enrichment of insertions, especially interchromosomal SDs, near the euchromatin boundaries of chimpanzee, bonobo, and gorilla genomes when compared to Asian great apes without subterminal satellites (**Fig. 6**). More importantly, this enrichment has led to the emergence of 204 species-specific genes (Yoo et al. 2025; **Supplemental Tables S6-S7**) recently identified in chimpanzee, bonobo, and gorilla. These contain distinct splice junctions and transcript models with open reading frames and thus are candidates for the emergence of novel functional proteins in each species. Therefore, subterminal heterochromatic chromatin [similar to the pericentromeric and SDs at the boundaries of inversions (Linardopoulou et al. 2005; Porubsky et al. 2020; Yoo et al. 2025)] may function as an incubator for the birth of new genes in ape genomes. This implicit genomic

instability may confer a long-term selective advantage if the genes evolve function important for the survival of the species.

## METHODS

### Identification of subterminal satellites and SDs

Subterminal satellites or pCht repeats of African great ape genomes, identified by a previous study were used (Yoo et al. 2025). Briefly, African great ape genomes (chimpanzee, bonobo and gorilla) were screened for pCht satellites via BLASTN (v2.12.0) (Chen et al. 2015) with the consensus sequence (len = 32 bp): "gatatttccatgtttatacagatagcggtgta". BLAST hits with longer than 90% of the consensus (>28 bp) were recovered. The individual pCht unit was classified into different types based on the variants (small INS, DEL or substitution). The SD spacers interrupting the satellite arrays were identified using BEDTools "subtract" option, subtracting the subterminal satellite arrays from the entire subterminal satellite regions, and 32 kbp of highly conserved sequence was identified in the *Pan* lineage and 34 kbp independently in gorilla. For extracting more precise locations of the individual SD spacers, minimap2 (Li 2018) (v2.26) was used ("*-x asm20*").

The relationship between subterminal arms shown in Fig. 2A was predicted by performing hierarchical clustering implemented by pvclust v2.2 (Suzuki and Shimodaira 2006) analysis limiting to the most abundant pCht (n>100) (method.dist = "correlation", and method.hclust = "average", nboot = 100). The input matrix fed into pvclust was built by computing normalized counts of 19,012 pChts (count of one pCht variant in a subterminal arm/count of total pCht

variants in that subterminal arm) across each subterminal arm. The higher-order structure of subterminal satellites was further investigated based on two-step *k*-means clustering (Hartigan and Wong 1979). The initial *k*-means clustering was performed under the assumption that some combinations of pCht variants are more frequently found together than others. This was done by computing the fraction of each pCht variant (n=19,012) across each of the subterminal arms (Chr1p, Chr1q, etc.). The optimal number of clusters was inferred by selecting the optimal Silhouette score (Rousseeuw 1987), varying k from at least 5 to 20 groups. The second *k*-means clustering was performed using the cluster of pChts (n=6) across subterminal caps divided into 20 kbp nonoverlapping blocks. Again, the optimal number was determined by computing the Silhouette score for k ranging from 5 to 20 The second clustering identified a total of eight higher-order blocks of subterminal satellites that contain similar composition of pChts.

**Phylogeny of SD spacers**

Based on the mutations accumulated in 32 and 34 kbp SD spacers, phylogeny was inferred using the maximum likelihood approach implemented by IQtree2 (Minh et al. 2020) (v2.3.6). All SD spacers greater than 90% of original lengths were used. To estimate the position of the root, we identified ortholog copies of 32 and 34 kbp spacers in the remaining great ape lineages, using siamang as the outgroup (Yoo et al. 2025); the ortholog copies were obtained with minimap2 (Li 2018) (v2.26). The MSA of the SD spacers was generated using MAFFT (Katoh and Standley 2013) (v7). Based on the MSA, two phylogenetic trees of SD spacers were constructed with -B 1,000, and with GTR+F+I+R6 and TVM+F+I+R8 models, which were chosen as the best fit model via Bayesian information criterion, for gorilla and *Pan* spacer trees, respectively. The trees were time-scaled with the syntenic original copies corresponding to the single-copy ortholog of orangutan and siamang. The previously reported divergence between human to

chimpanzee, gorilla, orangutan, and siamang of 6.4, 8.6, 15.2, and 19.5 MYA, respectively, were used for the time recalibration (Kumar et al. 2022). We used 100 replicates to compute confidence intervals using IQtree2 parameters, "--date-ci 100". To check for consistency of maximum likelihood tree, we additionally constructed neighbor-joining tree and maximum parsimony tree using MEGA12 (Kumar et al. 2024).

**Analysis of ectopic recombination**

Under the assumption that ectopic sequence exchange takes place at subterminal arms, we investigated for evidence of such events via two different approaches, through pairwise comparisons of subterminal arms 1) restricted to SD spacers and 2) the full sequence. The alignment was performed independently in each species using minimap2 (Li 2018) (v2.26), and the average identity as well as breadth of coverage were computed. To locate the representative event, we narrowed down the search space, restricting to the alignment of identity >99.5% to obtain relatively recent events, and for the scale of sequences larger than 1 Mbp.

To investigate locational bias of ectopic sequence exchange breakpoints, we tested their relative distance to the nearest SD spacers. This was done through a simulation test, shuffling the breakpoints 1,000 times to generate the null distribution of average distances; the *p*-value was computed by quantifying the fraction of more extreme null values to the observed averages.

We further tested whether the breakpoints of ectopic sequence exchanges fall within SD spacers by computing phylogenetic trees. We divided the MSA of SD spacers into two equally sized regions and constructed the phylogenetic trees of the first and second half of sequences, which were then compared to the original tree computed by the full SD sequences. If the ectopic exchanges take place frequently at the SD spacers, the inferred phylogeny based on the first or

second half of the sequence would shift dramatically. To test significance, we generated null distribution by computing the Robinson-Foulds distance between bootstrap trees to that of the consensus trees. This was then compared to the tree built by half of the SD sequences.

**Quantification of structural variations at the euchromatic boundaries**

The subterminal euchromatic boundaries were defined as the 2 Mbp regions downstream or upstream of subterminal satellite arrays of p-arms or q-arms, respectively. In the case of species with no subterminal satellites, the boundaries were defined as the 2 Mbp tips. The structural variation calls from PAV (Ebert et al. 2021) (v2.3.2)—taking humans (T2T-CHM13 v2.0) as the reference and SDs identified from the previous study (Yoo et al. 2025)—were used. For the large insertions, we separately inferred the additional sequences of apes relative to humans by using syntenic alignment. The alignment between homologous chromosomes between human to nonhuman apes (i.e., Chr1 vs. hsa1) generated by minimap2 (v2.26) were broken using RustyBam (v0.1.29) with a minimum structural event size of 50 bp. The unaligned regions of apes that are missing corresponding human homologs were considered as insertions. We also performed alignment of individual chromosomes to the whole orangutan genome to investigate origins of the inserted sequences.

**Methylation**

Methylation status of the subterminal cap sequences was determined based on ONT long-read sequencing data as described in the previous study (Yoo et al. 2025; https://github.com/marbl/T2T-Browser/tree/main/src/epi). Briefly, the CpG methylation calls were generated from Guppy v6.3.7 using the model "dna_r9.4.1_450bps_modbases_5hmc_5mc_cg_sup_prom.cfg". Reads were aligned to the

genome with minimap2 v2.26, and counts of modified bases at each cytosine position were obtained using modbam2bed v0.10.0. Because CpG methylation is symmetrical, counts from both cytosines in a CpG dinucleotide were merged to represent a single CpG site. Fractional methylation levels were then calculated for CpG sites supported by at least five reads.

**Analysis of transcriptome and novel genes**

Gene annotations and transcriptome alignment analyzed by the previous study were used (Yoo et al. 2025). Briefly, the transcriptome data were aligned onto respective genome assemblies using minimap2 (v2.26) after generating index via "*minimap2 -ax splice -f 1000 --secondary=no --eqx -K 100M*" and aligning using "*minimap2 -ax splice -f 1000 --secondary=no --eqx*". Stringtie (v3) was used for assessment and quantification of the transcripts (Pertea et al. 2015). Gene ontology (GO) terms, including biological process, molecular function and cellular component, of the novel genes located at the subterminal cap boundaries were analyzed using Panther database released on 2024/02/26 (Mi et al. 2019).

## ACKNOWLEDGMENTS

We would like to thank Tonia Brown, Soojin V. Yi, and Dongmin R. Son for helpful comments in the preparation of this manuscript. This work was supported, in part, by funding from the National Institutes of Health (NIH grants R01HG002385 and R01HG010169) and Weill Neurohub Family Foundation support to E.E.E. E.E.E. is an investigator of the Howard Hughes Medical Institute.



## CONFLICTS OF INTERESTS

E.E.E. is a scientific advisory board (SAB) member of Variant Bio, Inc. All other authors declare no competing interests.

## AUTHOR CONTRIBUTIONS

D.Y. performed evolutionary and methylation analyses. D.Y. and E.E.E. wrote the manuscript. K.M.M. provided sequencing support. E.E.E. and K.M.M. reviewed the manuscript. E.E.E. supervised and funded the project.

## FIGURE LEGENDS

**Figure 1. Overview of African great ape subterminal cap organization.** *(A)* Size of p- or q-arm subterminal caps in three species: chimpanzee (PTR), bonobo (PPA) and gorilla (GGO), indicated by red, blue and purple, respectively. Circles and triangles indicate haplotypes 1 and 2, or maternal and paternal for bonobo and gorilla, respectively. *(B)* StainedGlass (Vollger et al. 2022) self-alignment plot of subterminal caps including pCht satellites indicated by the purple track below the triangular heatmap, as well as subterminal SD spacers interrupting satellite arrays, indicated by the red arrows below. *(C)* The structure of subterminal caps. From the top, the logo base profile of pCht satellite is shown, followed by the higher-order structure of subterminal cap sequences, and the sequence composition of the subterminal SDs at the bottom. *(D)* Size distribution of uninterrupted pCht array and subterminal SDs. The dotted line indicates the mean. The x-axis is in a $\log_{10}$ scale for the pCht array due to variation in its size.

**Figure 2. pCht satellite higher-order structure and phylogeny.** *(A)* Clustering of the subterminal arms based on relative abundance of pCht variants. Each terminal node indicates species (chimpanzee - PTR, bonobo – PPA, and gorilla - GGO), human chromosome number, and whether the position of the subterminal cap is located in p-arms, q-arms, or interstitial (Ints). The top 5% of subterminal caps with the largest proportion of shared pCht types (among GGO and *Pan*) are highlighted (bold/red line) (**Supplemental Fig. S7C**). GGO.h2.3p vs. GGO.h1.Xp (blue line) is an example where the nonallelic subterminal cap pair is more closely related than the allelic pair. Internal nodes with bootstrap score higher than 95 are indicated by '*'. *(B)* Identification of eight higher-order block types (20 kbp) of pCht satellites across the subterminal caps of three African great apes. The highest proportion of shared pCht types among chimpanzee and gorilla are indicated (red line).

**Figure 3. Evolution of subterminal SD spacers.** *(A)* Maximum likelihood phylogenetic tree of gorilla (34 kbp) and *(B) Pan* subterminal SDs (32 kbp) constructed from total spacers (showing subset of n=100

536 copies). In each, the left panel shows the overall topology with a time-scale tree calibrated to orthologous
537 ancestral primate copies. Chromosome numbers (based on human chromosome synteny) are color-coded
538 and labelled by haplotype (h1 or h2) and species acronym: human (HSA; HSA1-CHM13 and HSA2-
539 HG002), chimpanzee (PTR), bonobo (PPA), gorilla (GGO), Sumatran orangutan (PAB), Bornean
540 orangutan (PPY), and siamang (SSY). A colored triangle denotes the subterminal spacer expansion. The
541 right panel zooms into the subterminal spacer topology and time estimates in million years ago (MYA)
542 with 95% confidence interval in the bracket. Gorilla (top) and chimpanzee (bottom) spacers are classified
543 as p-arm (red), q-arm (blue), or interstitial (green). In addition, for the bottom panel, species are indicated
544 as outgroup (rectangle), chimpanzee (circle), or bonobo (triangle) in origin. Internal nodes with bootstrap
545 score higher than 95 are indicated by '*'.

546 **Figure 4. Subterminal allelic vs. nonallelic sequence identity.** *(A)* Example of low-identity (hsa20q)
547 and high-identity (hsa8p) allelic subterminal caps in bonobo. Annotations include: percent identity of
548 alignment (scaled from blue <95% to red >99.8%); higher-order block type of pCht satellite arrays (top),
549 satellites (middle) and SD content (bottom). *(B,C)* Sequence identity matrix between paternal (h2) and
550 maternal (h1) haplotypes of *(B)* gorilla and *(C)* bonobo comparing allelic versus nonallelic subterminal
551 loci. The color intensity of blue to red indicates sequence identity (<95-100%). *(D)* Distribution of allelic
552 sequence identity of 50 kbp nonoverlapping windows of three African great ape genomes. Allelic
553 sequence identity of subterminal caps (purple) and the euchromatic regions (gray; excluding centromere,
554 acrocentric and subterminal regions) are compared with mean (straight) and median (dashed line)
555 indicated. *(E)* Distribution of average sequence identity between allelic subterminal caps compared to
556 nonallelic subterminal caps. Each dot represents the pairwise comparison. Two-sided permutation test
557 significance is indicated on top; **: $p$<0.0001 and *: $p$<0.05.

558 **Figure 5. Evidence of nonallelic sequence exchange at SD spacers.** *(A)* An alignment view (SVbyEye)
559 (Porubsky et al. 2024) depicting a candidate ectopic exchange between subterminal satellite chromosomes
560 hsa3 and hsaX p-arms of gorilla. The percent identity of alignment (scaled from blue <95% to red

>99.8%) with annotation tracks for higher-order pCht blocks, satellites, and SDs are shown on the right. *(B)* Enlarged view of the two exchange breakpoints (black arrows) mapping within the 3rd and 25th SD spacers. *(C)* Overview of potential ectopic exchange events between the subterminal caps (left-StCap1 vs right-StCap2 tracks) identified in African great ape genomes (eight and three events in gorilla and chimpanzee, respectively, including hsa3p vs. hsaXp in red). Assembly quality at the breakpoints was validated by a read-depth analysis in **Supplemental Figs. S20-S26**. *(D)* Simulation test of exchange event break points, suggesting that the exchange breaks (>99.8% identity alignment blocks; dotted line) are more likely to be located within or close to subterminal SD spacers than a random distribution (red histogram). *(E,F)* Comparison of maximum likelihood phylogeny to test recombination of SD spacer. A phylogenetic tree of the complete spacer sequence (left) is compared to that of a subset of the sequence (right). The tip nodes of the left tree are linked to the corresponding nodes of the right tree to visualize phylogenetic shifts in the topology. Robinson-Foulds distance is used to measure the extent of phylogenetic shift (first or last half sequences, in red and blue dotted lines and is compared to a null distribution generated by computing distances of consensus tree vs. bootstrap replicate trees).

**Figure 6. Duplicated genes and epigenetic features of ape genomes with and without subterminal caps.** *(A)* An ape comparative analysis of the organization of subterminal sequences for the long arm of chromosome hsa16. The stacked alignment plot (SVbyEye) contrasts syntenic euchromatic regions (light blue) with heterochromatic subterminal satellite regions (purple) and with the location of intra (dark blue) and interchromosomal (red) SDs along with the origin of the duplicated sequences (human chromosome designation). *(B)* The total number of new sequences (insertions [INS] or SDs), not present in humans, at subterminal boundaries within 2 Mbp of the euchromatic-heterochromatic transition zone. *(C)* Size distribution of the total inserted sequences with respect to human genome, color-coded by species. *(D)* CpG methylation profile within 2 Mbp of the euchromatin-heterochromatin transition for chromosomes with subterminal caps in gorilla. The average % CpG methylation (blue line) with standard deviation (blue shaded area) is shown. *(E)* The densities of SDs and novel genes at the euchromatic

boundaries, highly enriched in gorilla, with subterminal caps, as opposed to orangutan. The transparent area indicates 95% interval of the observed density. Dotted horizontal line indicates the mean density. *(F)* Duplicated genes and their copy number located at the boundary of subterminal caps, including eight genes validated by read depth of 120 primates (Mao et al. 2024). *(G)* Simulations to test enrichment of novel genes (Yoo et al. 2025) and SDs at the boundary of subterminal caps (2 Mbp) in gray and blue, respectively. The *p* indicates empirical *p*-value of the one-sided permutation test.

## REFERENCES

Chen Y, Ye W, Zhang Y, Xu Y. 2015. High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic acids research* **43**: 7762-7768.

Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, Sulovari A, Ebler J, Zhou W, Serra Mari R. 2021. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* **372**: eabf7117.

Fontsere C., Kuhlwilm M., Morcillo-Suarez C., Alvarez-Estape M., Lester J.D., Gratton P., Schmidt J.M., Dieguez P., Aebischer T., Álvarez-Varona P. and Agbor A., 2022. Population dynamics and genetic connectivity in recent chimpanzee history. *Cell Genomics*, 2(6).Haaf T, Schmid M. 1987. Chromosome heteromorphisms in the gorilla karyotype: Analyses with distamycin A/DAPI, quinacrine and 5-azacytidine. *Journal of Heredity* **78**: 287-292.

Hartigan JA, Wong MA. 1979. Algorithm AS 136: A *k*-means clustering algorithm. *Journal of the royal statistical society series c (applied statistics)* **28**: 100-108.

Hirai H, Hirai Y, Udono T, Matsubayashi K, Tosi AJ, Koga A. 2019. Structural variations of subterminal satellite blocks and their source mechanisms as inferred from the meiotic configurations of chimpanzee chromosome termini. *Chromosome Research* **27**: 321-332.

609 Ijdo JW, Baldini A, Ward D, Reeders S, Wells R. 1991. Origin of human chromosome 2: an ancestral
610 telomere-telomere fusion. *Proceedings of the National Academy of Sciences* **88**: 9051-9055.
611 Jiang X, Zhang L, Yang Z, Yang X, Ma K, Yoo D, Lu Y, Zhang S, Chen J, Nie Y. 2024. Incomplete
612 lineage sorting of segmental duplications defines the human chromosome 2 fusion site early
613 during African great ape speciation. *bioRxiv*: 2024.2012. 2012.628057.
614 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in
615 performance and usability. *Mol Biol Evol* **30**: 772-780.
616 Knight SJ, Flint J. 2000. Perfect endings: a review of subtelomeric probes and their use in clinical
617 diagnosis. *Journal of medical genetics* **37**: 401-409.
618 Koga A, Hirai Y, Hara T, Hirai H. 2012. Repetitive sequences originating from the centromere constitute
619 large-scale heterochromatin in the telomere region in the siamang, a small ape. *Heredity* **109**:
620 180-187.
621 Koga A, Notohara M, Hirai H. 2011. Evolution of subterminal satellite (StSat) repeats in hominids.
622 *Genetica* **139**: 167-175.
623 Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022.
624 TimeTree 5: an expanded resource for species divergence times. *Molecular biology and evolution*
625 **39**: msac174.
626 Kumar S, Stecher G, Suleski M, Sanderford M, Sharma S and Tamura K. 2024. MEGA12: Molecular
627 Evolutionary Genetic Analysis version 12 for adaptive and green computing. *Molecular Biology*
628 *and Evolution*, 41(12): msae263.
629 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094-3100.
630 Liu Y, Liu Z, Jiang T, Zang T and Wang Y, 2022. Comparison of the Nanopore and PacBio sequencing
631 technologies for DNA 5-methylcytosine detection. *IEEE International Conference on*
632 *Bioinformatics and Biomedicine (BIBM)* 220-225.
633 Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, Trask BJ. 2005. Human subtelomeres
634 are hot spots of interchromosomal recombination and segmental duplication. *Nature* **437**: 94-100.

Makova KD, Pickett BD, Harris RS, Hartley GA, Cechova M, Pal K, Nurk S, Yoo D, Li Q, Hebbar P. 2024. The complete sequence and comparative analysis of ape sex chromosomes. *Nature*: 1-11.

Mao Y, Harvey WT, Porubsky D, Munson KM, Hoekzema K, Lewis AP, Audano PA, Rozanski A, Yang X, Zhang S et al. 2024. Structurally divergent and recurrently mutated regions of primate genomes. *Cell* **187**: 1547-1562.e1513.

McConkey E. 2004. Orthologous numbering of great ape and human chromosomes is essential for comparative genomics. *Cytogenetic & Genome Research* **105**.

Mefford HC, Trask BJ. 2002. The complex structure and dynamic evolution of human subtelomeres. *Nature Reviews Genetics* **3**: 91-102.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**: 1530-1534.

Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X and Thomas PD. 2019. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v. 14.0). *Nature protocols* 14(3): 703-721.

Novo C, Arnoult N, Bordes WY, Castro-Vega L, Gibaud A, Dutrillaux B, Bacchetti S, Londoño-Vallejo A. 2013. The heterochromatic chromosome caps in great apes impact telomere metabolism. *Nucleic Acids Res* **41**: 4792-4801.

Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A. 2022. The complete sequence of a human genome. *Science* **376**: 44-53.

Porubsky D, Guitart X, Yoo D, Dishuck PC, Harvey WT, Eichler EE. 2024. SVbyEye: A visual tool to characterize structural variation among whole-genome assemblies. *Bioinformatics*, 41(6): btaf332.

Porubsky D, Sanders AD, Höps W, Hsieh P, Sulovari A, Li R, Mercuri L, Sorensen M, Murali SC, Gordon D. 2020. Recurrent inversion toggling and great ape genome evolution. *Nature genetics* **52**: 849-858.

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT and Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology* 33(3): 290-295.

Rousseeuw PJ. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* **20**: 53-65.

Royle NJ, Baird DM, Jeffreys AJ. 1994. A subterminal satellite located adjacent to telomeres in chimpanzees is absent from the human genome. *Nature genetics* **6**: 52-56.

Sigurpalsdottir BD, Stefansson OA, Holley G, Beyter D, Zink F, Hardarson MÞ, Sverrisson SÞ, Kristinsdottir N, Magnusdottir DN, Magnusson OÞ. and Gudbjartsson DF. 2024. A comparison of methods for detecting DNA methylation from long-read sequencing of human genomes. *Genome Biology* 25(1): 69.

Suzuki R, Shimodaira H. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**: 1540-1542.

Ventura M, Catacchio CR, Alkan C, Marques-Bonet T, Sajjadian S, Graves TA, Hormozdiari F, Navarro A, Malig M, Baker C. 2011. Gorilla genome structural variation reveals evolutionary parallelisms with chimpanzee. *Genome research* **21**: 1640-1649.

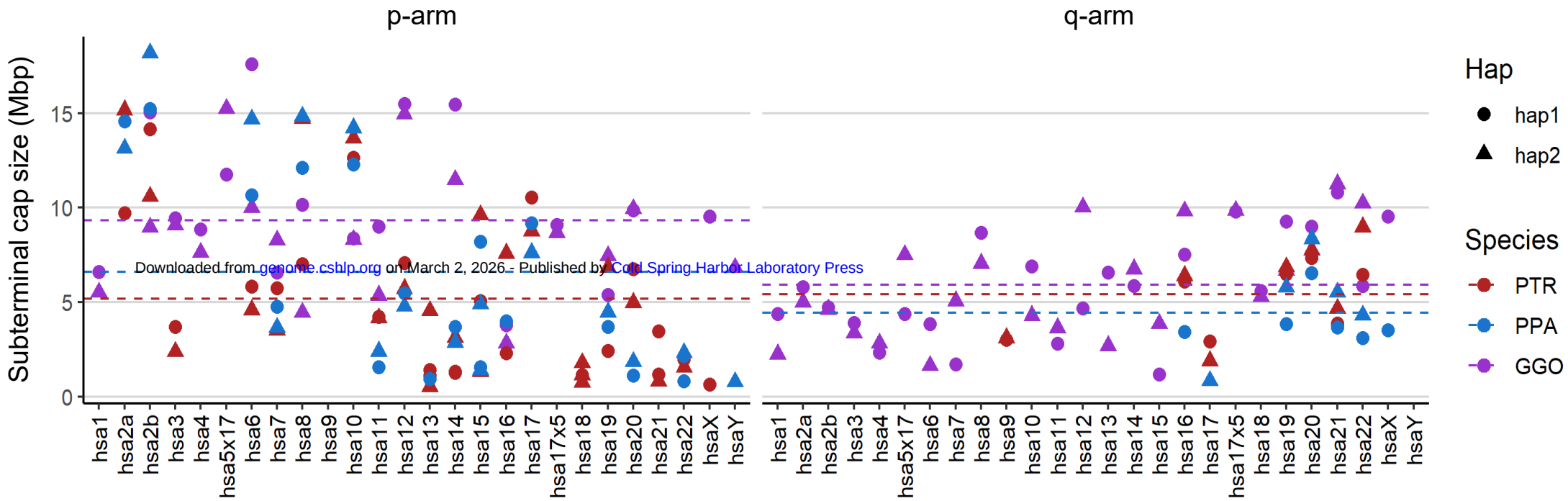
Ventura M, Catacchio CR, Sajjadian S, Vives L, Sudmant PH, Marques-Bonet T, Graves TA, Wilson RK, Eichler EE. 2012. The evolution of African great ape subtelomeric heterochromatin and the fusion of human chromosome 2. *Genome research* **22**: 1036-1049.

Vollger MR, Kerpedjiev P, Phillippy AM, Eichler EE. 2022. StainedGlass: interactive visualization of massive tandem repeat structures with identity heatmaps. *Bioinformatics* **38**: 2049-2051.

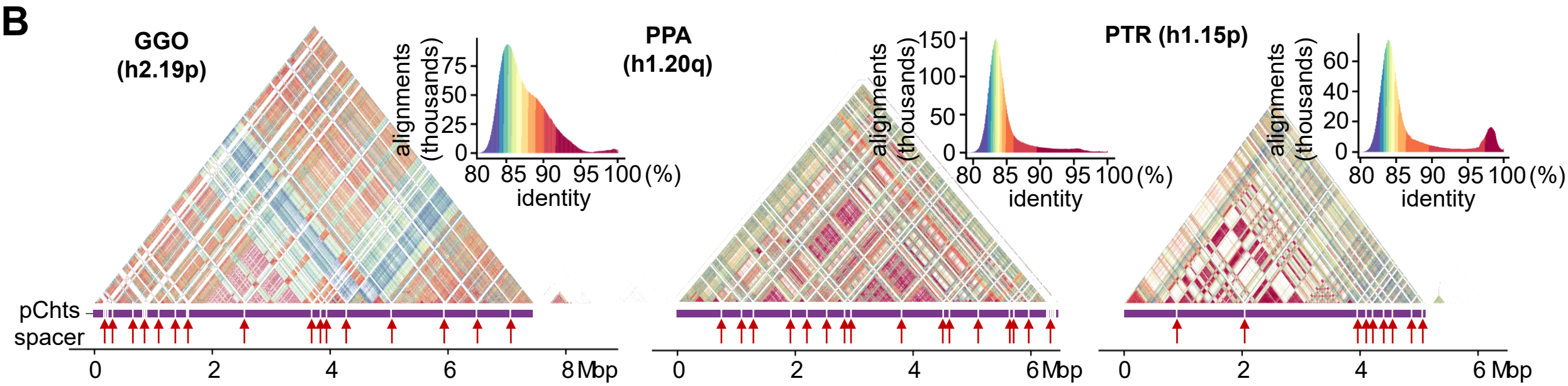
Yoo D, Rhie A, Hebbar P, Antonacci F, Logsdon GA, Solar SJ, Antipov D, Pickett BD, Safonova Y, Montinaro F. 2025. Complete sequencing of ape genomes. *Nature*: 1-18.

Yunis JJ, Prakash O. 1982. The origin of man: a chromosomal pictorial legacy. *Science* **215**: 1525-1530.

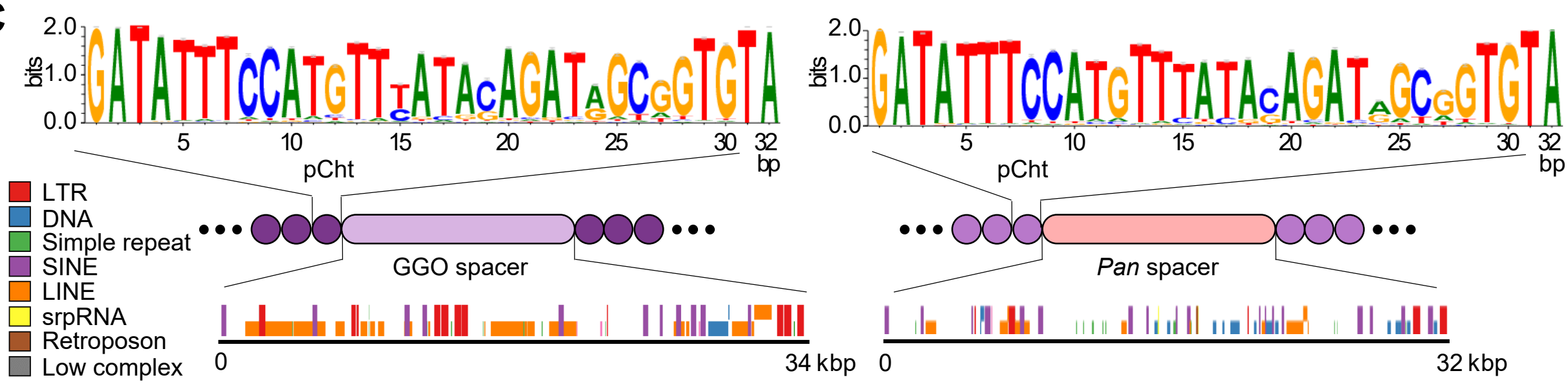
A



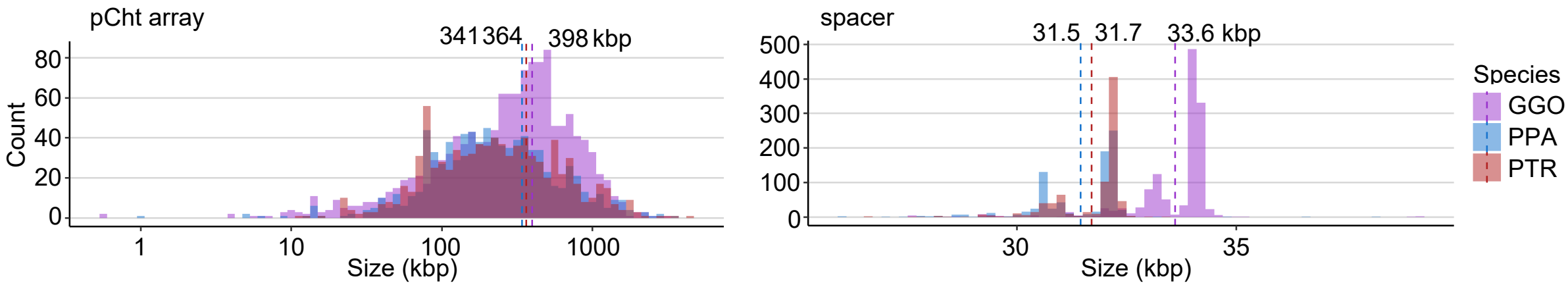

B




C




D



A



Species
- Chimpanzee
- Bonobo
- Gorilla

Position
- p-arm
- q-arm
- interstitial

0.03

pCht blocktype (20kbp)
1 2 3 4 5 6 7 8

B

Chimpanzee Bonobo Gorilla

pCht composition

A

5.0 MYA
[4.5-5.6]

GGO Subterminal
expansion

PPA.h2.hsa7.1
PPA.h1.hsa7.1
PTR.h2.hsa7.1
PTR.h1.hsa7.1
PPA.h2.hsa10.1
PPA.h1.hsa10.1
PTR.h2.hsa10.1
PTR.h1.hsa10.1
HSA1.h1.hsa10.1
HSA2.h2.hsa10.1
HSA2.h1.hsa10.1
GGO.h2.hsa10.1
GGO.h1.hsa10.1
PAB.h2.hsa10.1
PPY.h2.hsa10.1
PAB.h1.hsa10.1
PPY.h1.hsa10.1
SSY.hap2.hsa10.1
SSY.hap1.hsa10.1

5.6
[5.1-6.1]

Chromosome
hsa7
hsa10

0.005

Position
p−arm
q−arm
interstitial

0.005

2.5
[2.1-2.9]

3.8
[3.4-4.4]

5.0
[4.5-5.6]

chr6q

Outgroups



B

7.7 MYA
[7.2-8.5]

*Pan* lineage
expansion

GGO.h2.hsa2B.1
GGO.h1.hsa2B.1
GGO.h1.hsaX.1
GGO.h2.hsa4.1
GGO.h1.hsa4.1
GGO.h2.hsa2A.1
GGO.h1.hsa8.1
GGO.h2.hsa8.1
GGO.h2.hsa5.1
GGO.h1.hsa5.1
HSA1.h1.hsa9.1
HSA2.h1.hsa9.1
HSA2.h2.hsa9.1
HSA2.h2.hsa2.1
HSA2.h1.hsa2.1
HSA1.h1.hsa2.1
HSA1.h1.hsa9.3
HSA2.h2.hsa9.3
HSA2.h1.hsa9.3
HSA1.h1.hsa9.2
HSA2.h1.hsa9.2
HSA2.h2.hsa9.2
GGO.h2.hsa9.1
GGO.h1.hsa9.1
PPA.h2.hsa9.3
PPA.h1.hsa9.3
PTR.h2.hsa9.2
PTR.h1.hsa9.2
HSA1.h1.hsa9.4
HSA2.h2.hsa9.4
HSA2.h1.hsa9.4
PPY.h2.hsa9.1
PPY.h1.hsa9.1
PAB.h1.hsa9.1
PAB.h2.hsa9.1
SSY.h2.hsa9.1
SSY.h1.hsa9.1

8.1
[7.7-8.6]

Chromosome
hsa2/hsa2A
hsa2B
hsa4
hsa5
hsa8
hsa9
hsaX

0.005

Position
p−arm
q−arm
interstitial

Species
PTR
PPA
Outgroup

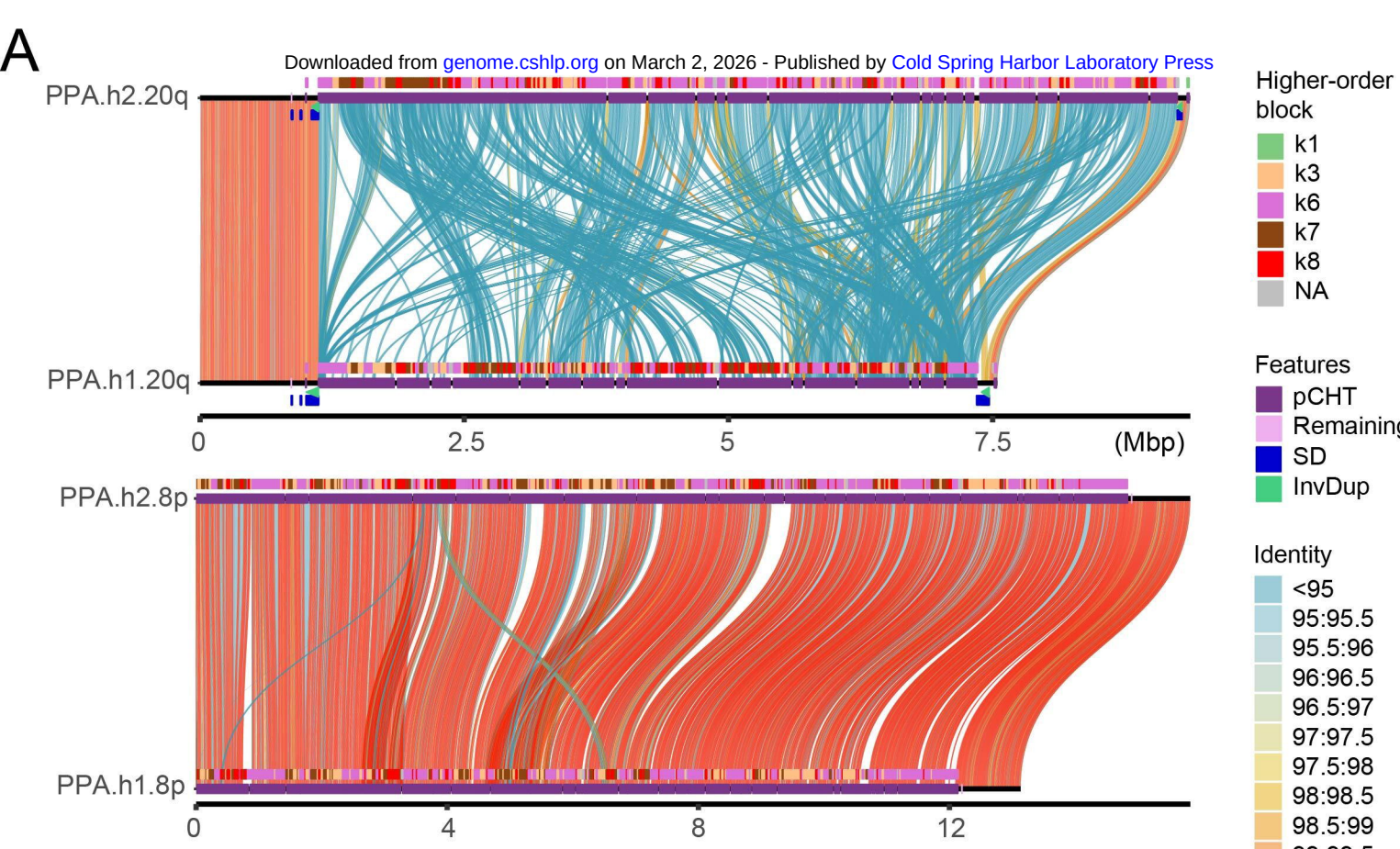
0.005

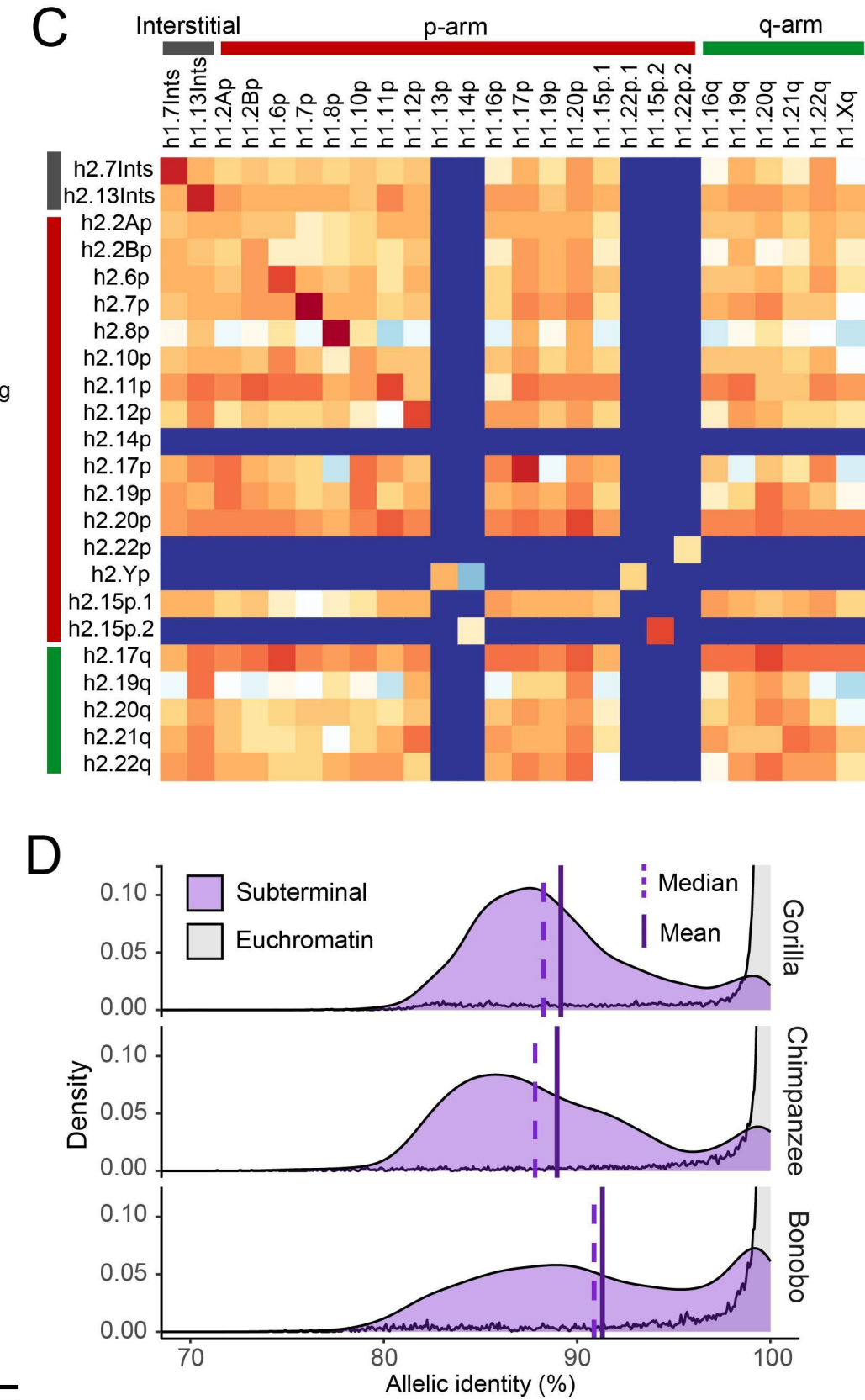
PTR.Xp

7.7
[7.2-8.5]

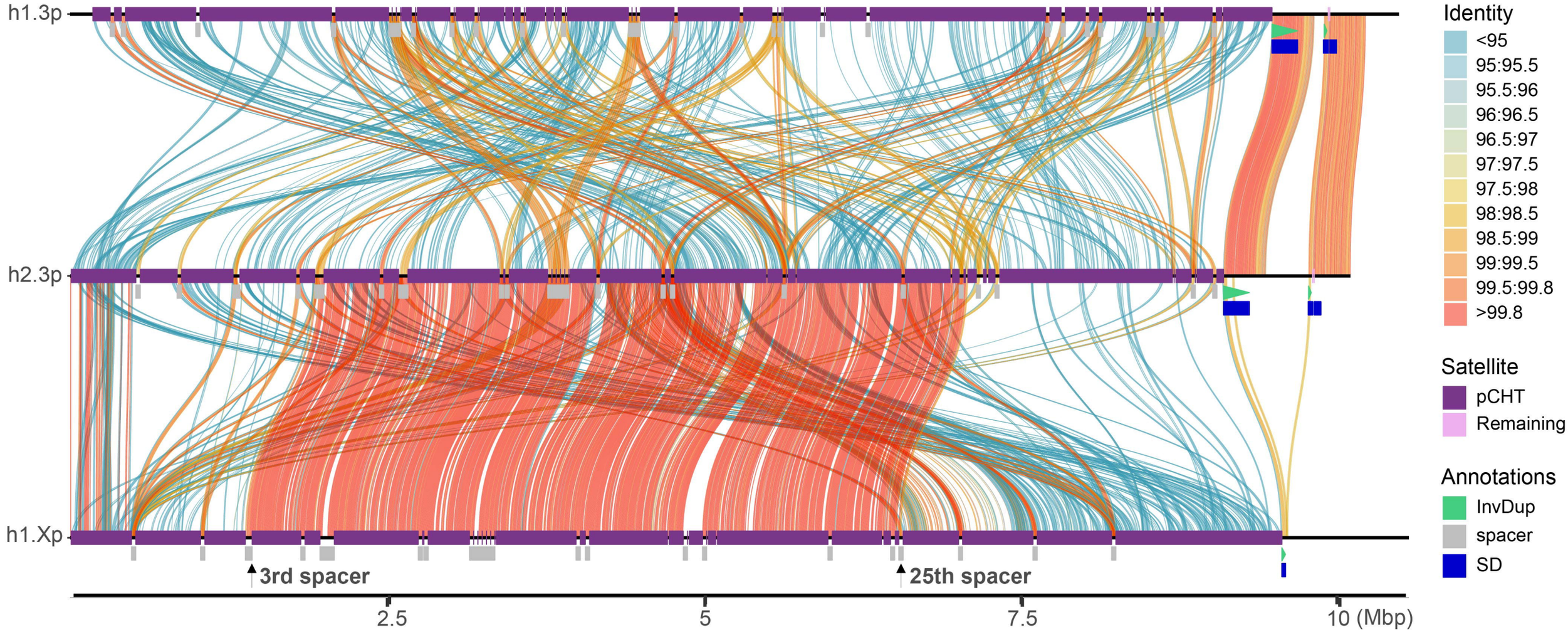
PPA.Xq

Outgroups

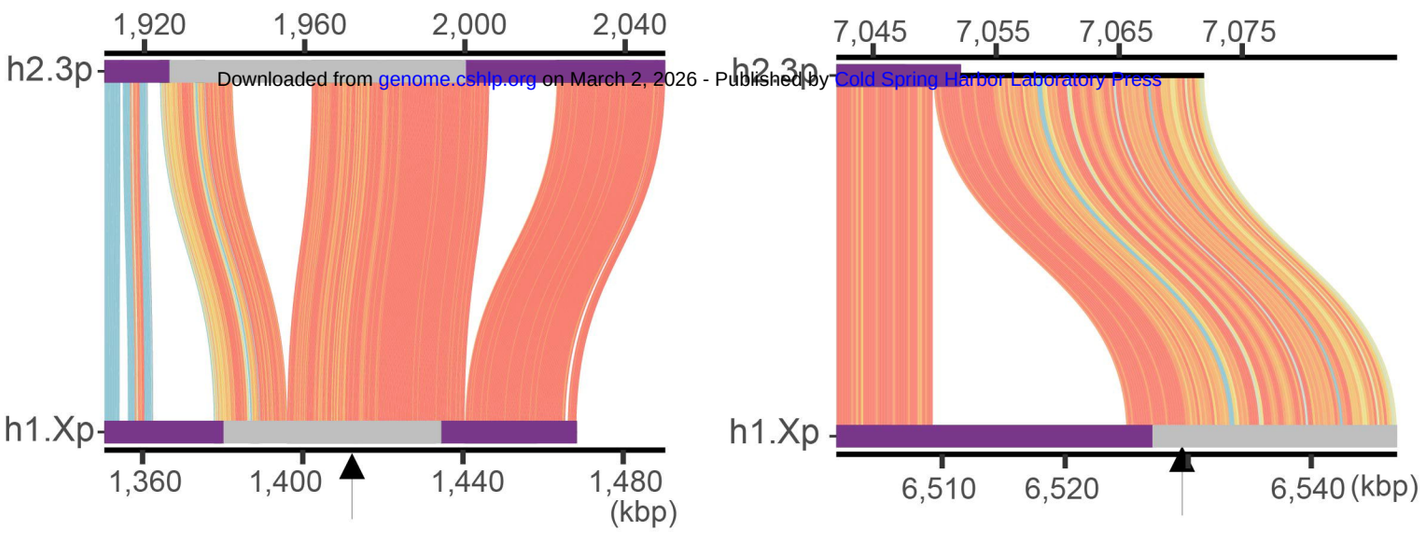


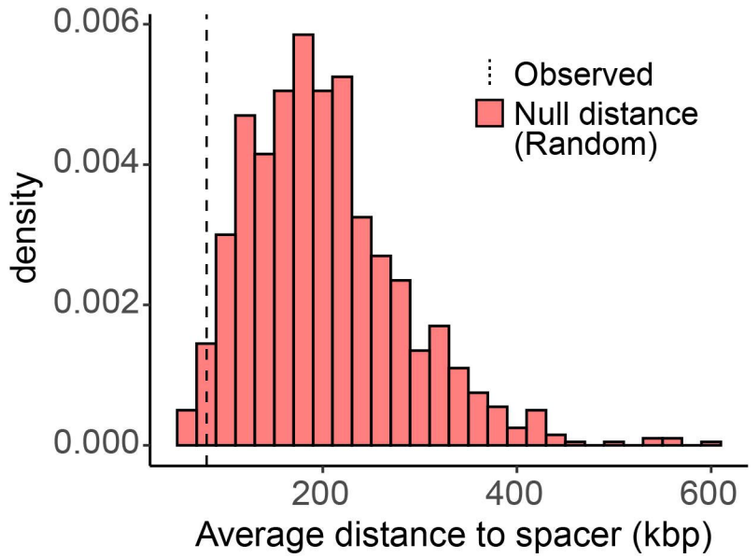
A



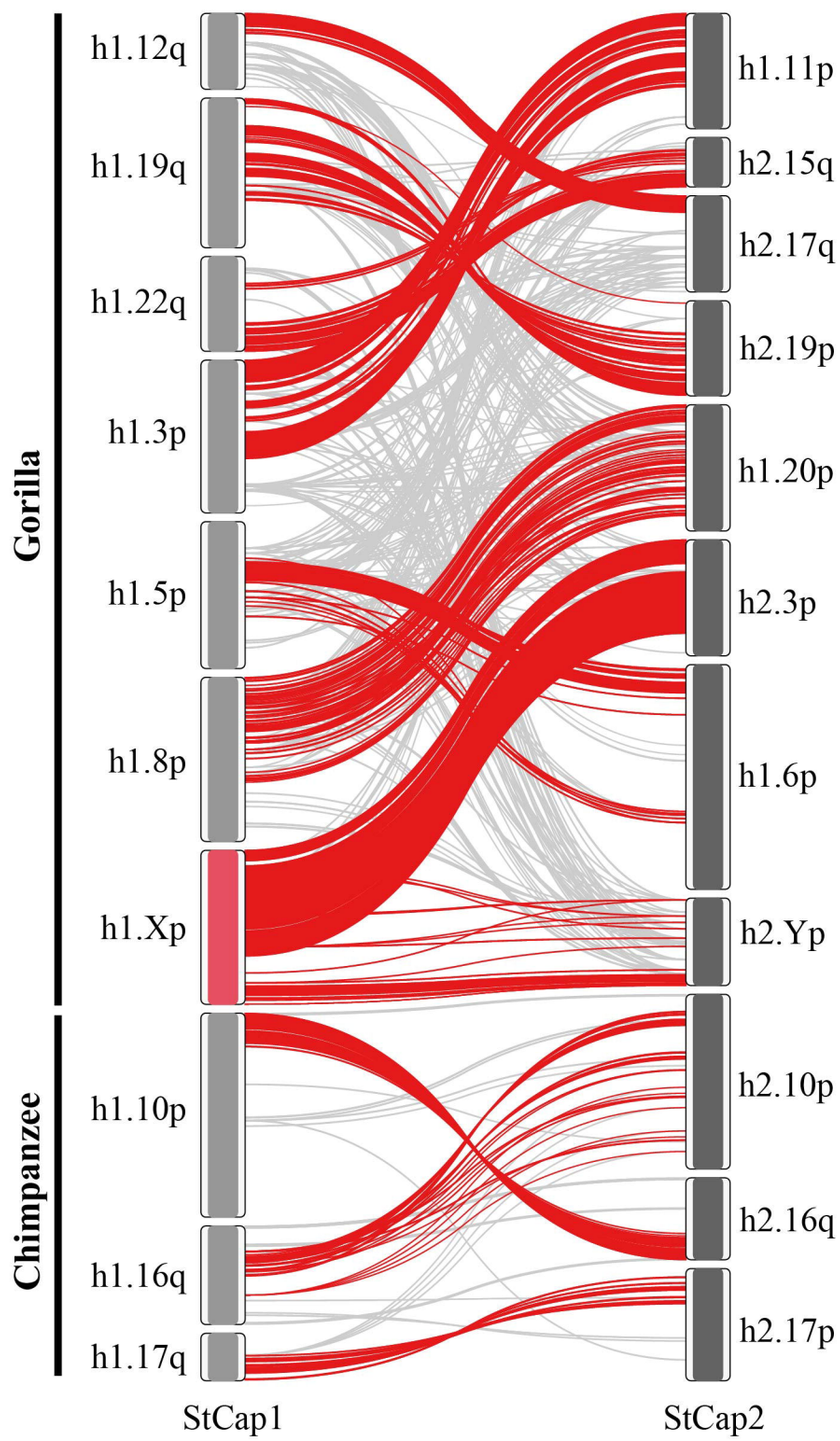

B



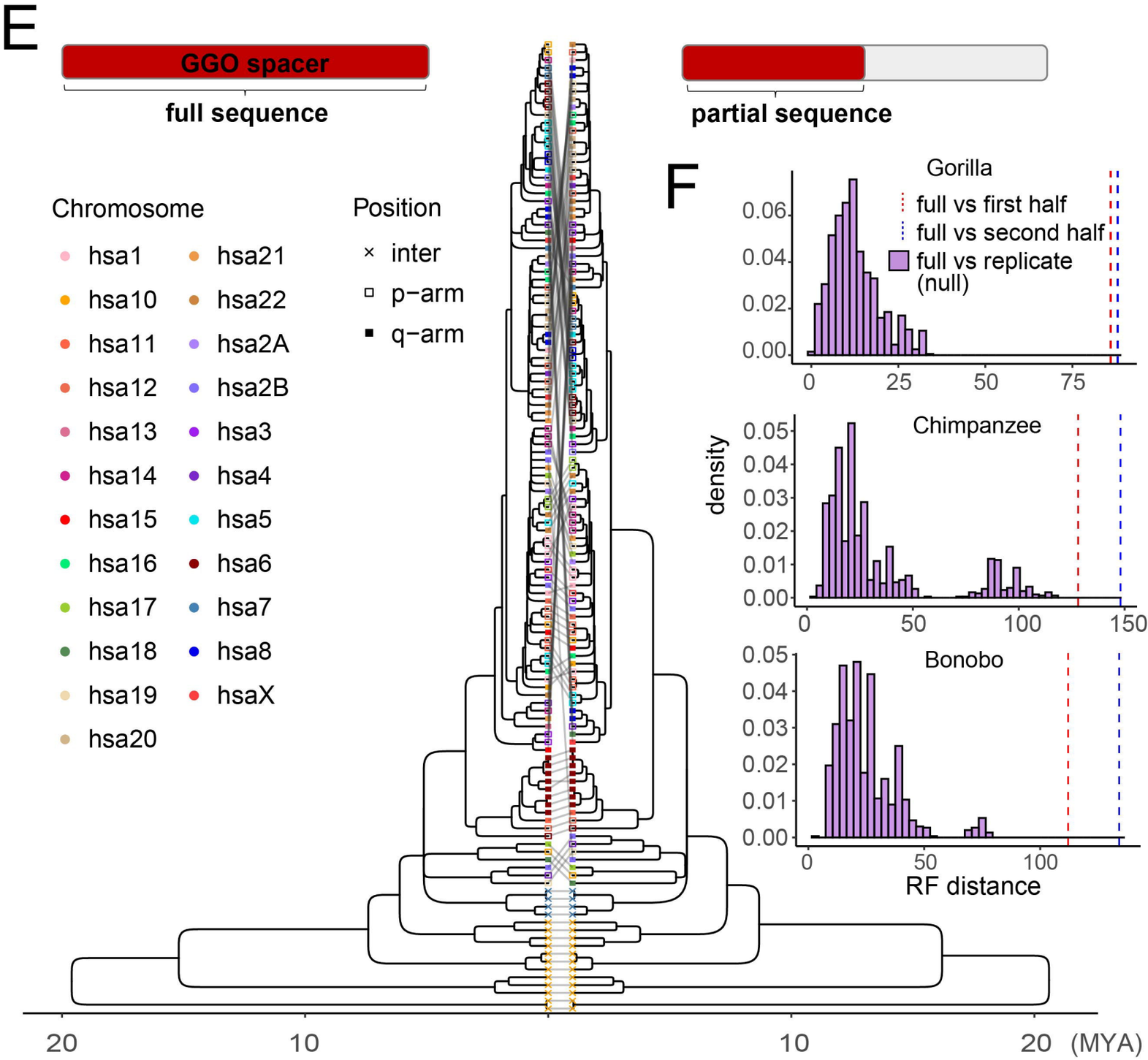

D



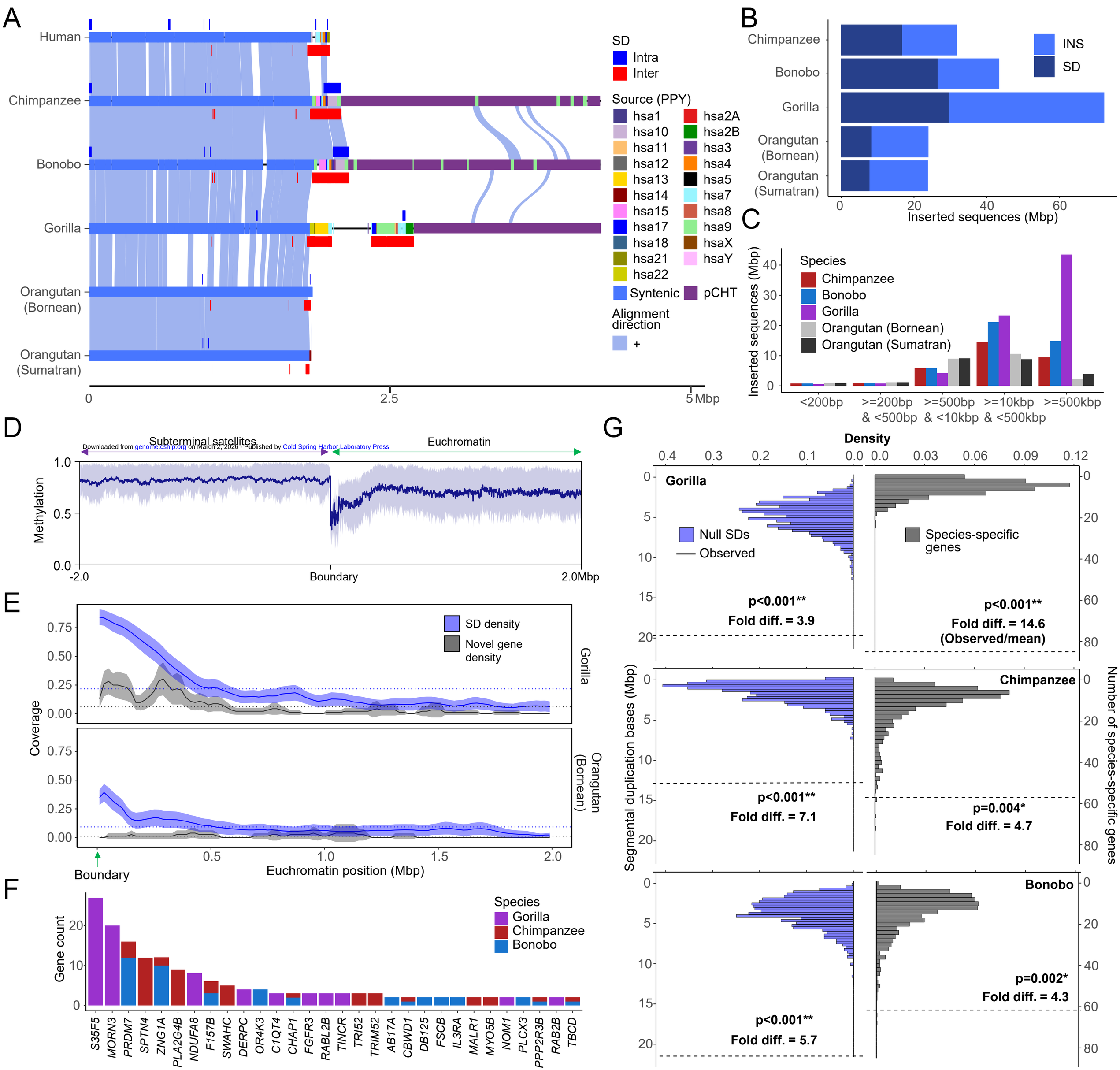

C




E




F


A
Human
Chimpanzee
Bonobo
Gorilla
Orangutan (Bornean)
Orangutan (Sumatran)
0
2.5
5 Mbp
SD
Intra
Inter
Source (PPY)
hsa1
hsa10
hsa11
hsa12
hsa13
hsa14
hsa15
hsa17
hsa18
hsa21
hsa22
hsa2A
hsa2B
hsa3
hsa4
hsa5
hsa7
hsa8
hsa9
hsaX
hsaY
Syntenic
pCHT
Alignment direction
+
B
Chimpanzee
Bonobo
Gorilla
Orangutan (Bornean)
Orangutan (Sumatran)
INS
SD
0
20
40
60
Inserted sequences (Mbp)
C
Inserted sequences (Mbp)
0
10
20
30
40
Species
Chimpanzee
Bonobo
Gorilla
Orangutan (Bornean)
Orangutan (Sumatran)
<200bp
>=200bp & <500bp
>=500bp & <10kbp
>=10kbp & <500kbp
>=500kbp
D
Subterminal satellites
Euchromatin
Downloaded from genome.cshlp.org on March 2, 2026 - Published by Cold Spring Harbor Laboratory Press
Methylation
1.0
0.5
0.0
-2.0
Boundary
2.0Mbp
E
SD density
Novel gene density
Coverage
0.75
0.50
0.25
0.00
Gorilla
Orangutan (Bornean)
0.5
1.0
1.5
2.0
Boundary
Euchromatin position (Mbp)
F
Gene count
0
10
20
Species
Gorilla
Chimpanzee
Bonobo
S35F5
MORN3
PRDM7
SPTN4
ZNG1A
PLA2G4B
NDUFA8
F157B
SWAHC
DERPC
OR4K3
C1QT4
CHAP1
FGFR3
RABL2B
TINCR
TRI52
TRIM52
AB17A
CBWD1
DB125
FSCB
IL3RA
MALR1
MYO5B
NOM1
PLCX3
PPP2R3B
RAB2B
TBCD
G
Density
0.4
0.3
0.2
0.1
0.0
0.0
0.03
0.06
0.09
0.12
Gorilla
Null SDs
Observed
p<0.001**
Fold diff. = 3.9
Species-specific genes
p<0.001**
Fold diff. = 14.6
(Observed/mean)
Chimpanzee
p<0.001**
Fold diff. = 7.1
p=0.004*
Fold diff. = 4.7
Bonobo
p<0.001**
Fold diff. = 5.7
p=0.002*
Fold diff. = 4.3
Segmental duplication bases (Mbp)
Number of species-specific genes
0
5
10
15
20
40
60
80

# Epigenetic and evolutionary features of ape subterminal heterochromatin

DongAhn Yoo, Katherine M Munson and Evan E Eichler

*Genome Res.* published online October 22, 2025
Access the most recent version at doi:10.1101/gr.280987.125