



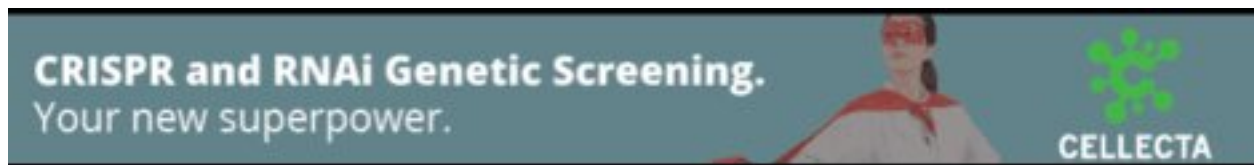
Epigenetic and evolutionary features of ape subterminal heterochromatin

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](https://doi.org/10.1101/gr.280987.125)

P<P	Published online October 22, 2025 in advance of the print journal.
Accepted Manuscript	Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.
Open Access	Freely available online through the <i>Genome Research</i> Open Access option.
Creative Commons License	This manuscript is Open Access. This article, published in <i>Genome Research</i> , is available under a Creative Commons License (Attribution 4.0 International license), as described at http://creativecommons.org/licenses/by/4.0/ .
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here .



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Published by Cold Spring Harbor Laboratory Press

1 **Epigenetic and evolutionary features of ape subterminal**
2 **heterochromatin**

3

4 DongAhn Yoo¹, Katherine M. Munson¹, Evan E. Eichler^{1,2}

5 **Affiliations**

6 1. Department of Genome Sciences, University of Washington School of Medicine, Seattle,
7 WA 98195, USA

8 2. Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

9

10 *Correspondence to ee3@uw.edu

11

12

13 **Corresponding author**

14 Evan E. Eichler, Ph.D.

15 Department of Genome Sciences

16 University of Washington School of Medicine

17 3720 15th Ave NE, S413C

18 Box 355065

19 Seattle, WA 98195-5065

20 Email: ee3@uw.edu

21 Tel: 1-206-543-9526

22

23 **Keywords:** Genomics, Subterminal heterochromatin, Evolution, Great ape, Comparative
24 genomics

25

26 **ABSTRACT**

27 Many African great ape chromosomes possess large subterminal heterochromatic caps at their
28 telomeres that are conspicuously absent from the human lineage. Leveraging the complete
29 sequences of great ape genomes, we characterize the organization of subterminal caps and
30 reconstruct the evolutionary history of these regions in chimpanzees and gorillas. Detailed
31 analyses of the composition of the associated terminal 32 bp satellite array from chimpanzee
32 (termed pCht) and intervening segmental duplication (SD) spacers confirm two independent
33 origins in the *Pan* and gorilla lineages. In chimpanzee and bonobo, we estimate these structures
34 emerged ~7.7 million years ago (MYA) in contrast to gorilla where they expanded more
35 recently, ~5.0 MYA, and now make up 8.5% of the total gorilla genome. In both lineages, the
36 SD spacers punctuating the pCht heterochromatic satellite arrays correspond to pockets of
37 decreased methylation, although in gorilla such regions are significantly less methylated ($p < 2.2 \times 10^{-16}$)
38 than chimpanzee or bonobo. Allelic pairs of subterminal caps show a higher degree of
39 sequence divergence than euchromatic sequences, with bonobo showing less divergent
40 haplotypes and less differentially methylated spacers. In contrast, we identify virtually identical
41 subterminal caps mapping to nonhomologous chromosomes within a species, suggesting ectopic
42 recombination potentially mediated by SD spacers. We find that the transition regions from
43 heterochromatic subterminal caps to euchromatin are enriched for structural variant insertions
44 and lineage-specific duplicated genes. Our findings suggest independent evolution of
45 subterminal caps converging on a common genetic and epigenetic structure that promoted
46 ectopic exchange as well as the emergence of novel genes at transition regions between
47 euchromatin and heterochromatin.

48 INTRODUCTION

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

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

71 2012). Jiang and colleagues further refined the Chromosome 2 fusion breakpoint and subterminal
72 SDs in humans (Jiang et al. 2024). In the most recent study, the heterochromatic cap sequences
73 were fully resolved using a hybrid long-read assembly approach (Oxford Nanopore Technologies
74 [ONT] and PacBio high-fidelity [HiFi]) and assembly phasing (HiC and parent–child trios) (Yoo
75 et al. 2025).

76 While these subterminal satellites are absent in other apes like orangutans (Haaf and Schmid
77 1987; Ijdo et al. 1991; Ventura et al. 2011), such structures are not unique to African great apes.
78 A similar, albeit larger, subterminal heterochromatic cap structure composed of alpha-satellite
79 DNA was described for siamang genomes from the gibbon lineage (Koga et al. 2012). We
80 recently characterized the heterochromatic cap sequence and structure of the siamang genome
81 (Yoo et al. 2025). While it is similar in form to that of the great apes (i.e., blocks of satellite
82 interdigitated with a spacer that shows epigenetically reduced methylation), the component
83 sequences completely differ. In siamang, there is no pCht satellite but instead alpha-satellite
84 sequence and the spacer is made up of a 50 kbp SD of different origin than either the *Pan* or
85 gorilla spacer. The independent emergence of subterminal satellite structures in multiple primate
86 lineages (gorilla, chimpanzee and siamang) so similar in structure and epigenetic form suggests
87 convergent evolution of potential functional significance. In this study, we leverage the fully
88 resolved sequence of telomere-to-telomere ape genomes to more systematically investigate the
89 evolution, structure, and epigenetic properties. The availability of two fully resolved haplotype
90 assemblies from each ape species allowed for investigations into patterns of allelic variation and
91 between-species comparisons of subterminal organization and differences in methylation
92 patterns. We examined the telomeric transition regions from heterochromatin to euchromatin

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

95 **RESULTS**

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

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

117 As previously reported (Yoo et al. 2025), analysis of the subterminal caps revealed a higher-
118 order organization of pCht satellite arrays of variable length (ranging from <10 kbp to hundreds
119 of kbp with a mean of 341-398 kbp) interdigitated with SD spacers (**Fig. 1B-D; Supplemental**
120 **Fig. S1**). We find that pCht satellites are generally more than 80% identical with the ones in
121 closer proximity showing higher identity within a chromosomal arm (**Fig. 1B**). While different
122 subtypes of pCht can be distinguished, we derived a consensus for each depending on species
123 (**Fig. 1C & Supplemental Fig. S2**). Despite the sequence variation, we find that the gorilla
124 satellite consensus is almost identical to that of the *Pan* lineage (chimpanzee and bonobo), with
125 the exception of the thymine in the 15th position, which shows less variability in the *Pan* lineage.
126 In terms of epigenetic status, we investigated methylation patterns of both spacers and satellites.
127 While we have data from both PacBio and ONT, we chose to focus on methylation calls from
128 ONT for two reasons. First, previous studies have shown (Liu et al. 2022; Sigurpalsdottir et al.
129 2024) greater sensitivity in detection, and second, ultra-long reads are necessary to uniquely
130 anchor reads in long stretches of DNA. Our study finds these satellite sequences show high
131 levels of 5mC methylation, with gorilla showing average methylation of 83.3% compared to
132 78.4% in bonobo and 73.6% in chimpanzee (**Supplemental Figs. S3 & S4A**).

133 The pCht satellite arrays are interrupted irregularly by pockets of less methylated subterminal SD
134 spacers (**Supplemental Figs. S3 & S4**). These SDs originate from euchromatic sequence in the
135 ancestral ape lineage (**Fig. 1C**). For example, the ancestral subterminal SD spacer in gorilla maps
136 to the intronic region of the gene *MALDI* of Chr8 (hsa10, p12.31) and subsequently expanded to
137 ~700 copies per gorilla haplotype. In contrast, the spacer of the *Pan* lineage originates from a

138 subsequence of *PGM5*, Chr11 (hsa9, q21.11) (**Supplemental Fig. S5**), and expanded to ~413
139 and ~430 copies per haplotype for chimpanzee and bonobo, respectively (**Supplemental Table**
140 **S3**). The gorilla spacers contained 1.8-fold greater LINE content and *Pan* spacers 2.4-fold
141 greater DNA transposon content compared to euchromatic regions; however, we found no
142 statistical significance from this comparison (one-sided permutation test $p=0.15$ and 0.07 ,
143 respectively). The average length of gorilla and *Pan* spacers are 34 and 32 kbp, respectively (**Fig.**
144 **1C-D**). Of note, the length of both spacers and pCht arrays are significantly longer in gorilla
145 compared to chimpanzee and bonobo (two-sided Wilcox test $p<2.2e-16$ and $p=6e-10$,
146 respectively) consistent with their independent origin and expansion in each lineage (Ventura et
147 al. 2012).

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

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

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

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

185 Intraspecies sequence comparisons of the subterminal satellites show that some nonhomologous
186 chromosomes are more similar in composition than their allelic counterparts—i.e., homologous
187 subterminal caps do not always pair (e.g., GGO.h2.3p vs. GGO.h1.Xp; **Fig. 2A**). Under the
188 assumption that nonhomologous subterminal caps are subject to ectopic recombination (Ventura
189 et al. 2012; Hirai et al. 2019), we sought to classify the major pCht subtypes using a two-step *k*-
190 means clustering approach (**Methods, Supplemental Fig. S6**). We identified eight higher-order
191 satellite block types across the 167 subterminal caps (**Fig. 2B**) from the three ape species.

192 Consistent with the hierarchical clustering tree, the gorilla and *Pan* lineages are largely distinct.
193 Gorilla subterminal caps consist of k2, k4, and k5 block types while the *Pan* lineage caps are
194 composed of k3, k6, k7, and k8 types. Once again, gorilla Chromosome 5 (hsa6) is an exception
195 with a largely uniform composition of the k1 block type, which is shared with the X
196 Chromosome p-arms of bonobo and chimpanzee and to a lesser degree with a fraction of other
197 *Pan* autosomes. The k1 block type, thus, is distinct in being the only block type shared among
198 the gorilla and *Pan* lineages, potentially identifying the ancestral pCht. We also note that this
199 block type is composed of the largest portions of common pChts shared among *Pan* and gorilla
200 (**Supplemental Fig. S7**).

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

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

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

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

234 ***Patterns of allelic and nonallelic variation.*** A striking feature of the subterminal
235 heterochromatic caps is the high degree of allelic sequence diversity observed between two
236 haplotypes (**Fig. 4A-C; Supplemental Data**). Only 12% (10/82) of haplotype comparisons show
237 >99% sequence identity with >50% alignment coverage. Most haplotype comparisons show
238 greater than 10% sequence divergence (**Supplemental Figs. S10-S13**). A comparison with other
239 genomic regions, including centromere, acrocentric and remaining regions, reveals that
240 subterminal caps were the regions with the greatest degree of allelic divergence (**Supplemental**
241 **Fig. S14**). Notably, the degree of allelic divergence differs among the species. Gorilla and
242 chimpanzee show on average greater sequence divergence (10.8 and 11%, respectively) when
243 compared to bonobo (8.7%) (**Fig. 4D**). When comparing nonallelic versus allelic sequence
244 identity, gorilla and chimpanzee distributions are largely indistinguishable, in contrast to bonobo
245 where allelic alignments now show significantly greater allelic sequence identity when compared
246 to nonallelic alignments (**Fig. 4E**). This feature is reflected in a matrix percent identity plot
247 where each subterminal haplotype is aligned to every other within the individual (**Fig. 4B,C**). In
248 contrast to bonobo where more than a third of allelic comparisons show the highest degree of
249 sequence identity (as indicated by the diagonal in **Fig. 4C**), gorilla shows no such pattern.
250 Instead, allelic and nonallelic patterns are largely indistinguishable perhaps providing evidence
251 of more rapid, ongoing ectopic recombination occurring in this lineage.

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

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

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

279 ***Subterminal heterochromatic-euchromatic transition regions and the evolution of new genes.***

280 Finally, we examined the euchromatin boundaries and putative evolutionary consequences and
281 epigenetic features of the subterminal caps (**Fig. 6**). Specifically, we compared ape species with
282 (gorilla/chimpanzee) and without (Bornean and Sumatran orangutans) subterminal caps to
283 determine if the rate of structural variation, gene density, and methylation differ in association
284 with the evolution of these large terminal heterochromatic structures (Yoo et al. 2025). Using the
285 human genome as a reference, we find ~24–72 Mbp of novel insertions mapping within 2 Mbp
286 from subterminal caps (African nonhuman apes) or telomeres (Asian great apes without
287 subterminal caps) (**Fig. 6A,B**). The insertions in the African apes are notably more abundant
288 (31.8–72.2 Mbp vs. 23.8–23.9 Mbp in orangutans) and larger (>10–500 kbp in size) when
289 compared to the Asian great apes (**Fig. 6C & Supplemental Fig. S27**). This is despite the fact
290 that divergence time between humans to African great apes are close to half the divergence to
291 Asian great apes. We estimate that 36%–42% of the inserted sequences in African apes
292 originated from nonhomologous chromosomes (**Supplemental Fig. S28**)—a potential
293 consequence of the ectopic exchanges extending beyond the subterminal heterochromatic cap.

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

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

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

320

321 **DISCUSSION**

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

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

359 Nevertheless, under this model, gorilla would be predicted to be the most active subterminal
360 heterochromatic cap perhaps explaining why it is much more prevalent in this genome adding
361 almost three chromosomes worth of DNA (584 Mbp).

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

367 prophase somewhat akin to the nuclear organizing structures associated with the repeat-rich short
368 arms of mammalian acrocentric chromosome (Nurk et al. 2022). Hirai and colleagues further
369 concluded that such subterminal associations may be selectively beneficial by providing greater
370 stability or buffering capacity to subterminal regions of ape chromosomes—areas known to be
371 genetically and evolutionarily unstable (Knight and Flint 2000; Mefford and Trask 2002). In our
372 study, we provide further support for their involvement in ectopic recombination by identifying
373 11 potential examples of interchromosomal translocations (8 in gorilla and 3 in the common
374 chimpanzee; **Fig. 5C**). Both our breakpoint analysis of these events and our phylogenetic
375 analysis indicate that the SD spacer regions are preferential drivers of this ectopic recombination
376 between nonhomologous chromosome arms. It is possible that both the high degree of sequence
377 identity of these 30 kbp segments and their reduced methylation status provide an ideal open-
378 chromatin substrate for nonallelic homologous recombination, although the mechanism still
379 remains to be determined. However, unlike Hirai’s prediction of conferring genomic stability,
380 our evolutionary analysis suggests the complete opposite effect. We find far stronger enrichment
381 of insertions, especially interchromosomal SDs, near the euchromatin boundaries of chimpanzee,
382 bonobo, and gorilla genomes when compared to Asian great apes without subterminal satellites
383 (**Fig. 6**). More importantly, this enrichment has led to the emergence of 204 species-specific
384 genes (Yoo et al. 2025; **Supplemental Tables S6-S7**) recently identified in chimpanzee, bonobo,
385 and gorilla. These contain distinct splice junctions and transcript models with open reading
386 frames and thus are candidates for the emergence of novel functional proteins in each species.
387 Therefore, subterminal heterochromatic chromatin [similar to the pericentromeric and SDs at the
388 boundaries of inversions (Linardopoulou et al. 2005; Porubsky et al. 2020; Yoo et al. 2025)] may
389 function as an incubator for the birth of new genes in ape genomes. This implicit genomic

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

392

393 **METHODS**

394 **Identification of subterminal satellites and SDs**

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

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

411 variants in that subterminal arm) across each subterminal arm. The higher-order structure of
412 subterminal satellites was further investigated based on two-step *k*-means clustering (Hartigan
413 and Wong 1979). The initial *k*-means clustering was performed under the assumption that some
414 combinations of pCht variants are more frequently found together than others. This was done by
415 computing the fraction of each pCht variant (n=19,012) across each of the subterminal arms
416 (Chr1p, Chr1q, etc.). The optimal number of clusters was inferred by selecting the optimal
417 Silhouette score (Rousseeuw 1987), varying *k* from at least 5 to 20 groups. The second *k*-means
418 clustering was performed using the cluster of pChts (n=6) across subterminal caps divided into
419 20 kbp nonoverlapping blocks. Again, the optimal number was determined by computing the
420 Silhouette score for *k* ranging from 5 to 20. The second clustering identified a total of eight
421 higher-order blocks of subterminal satellites that contain similar composition of pChts.

422 **Phylogeny of SD spacers**

423 Based on the mutations accumulated in 32 and 34 kbp SD spacers, phylogeny was inferred using
424 the maximum likelihood approach implemented by IQtree2 (Minh et al. 2020) (v2.3.6). All SD
425 spacers greater than 90% of original lengths were used. To estimate the position of the root, we
426 identified ortholog copies of 32 and 34 kbp spacers in the remaining great ape lineages, using
427 siamang as the outgroup (Yoo et al. 2025); the ortholog copies were obtained with minimap2 (Li
428 2018) (v2.26). The MSA of the SD spacers was generated using MAFFT (Katoh and Standley
429 2013) (v7). Based on the MSA, two phylogenetic trees of SD spacers were constructed with -B
430 1,000, and with GTR+F+I+R6 and TVM+F+I+R8 models, which were chosen as the best fit
431 model via Bayesian information criterion, for gorilla and *Pan* spacer trees, respectively. The
432 trees were time-scaled with the syntenic original copies corresponding to the single-copy
433 ortholog of orangutan and siamang. The previously reported divergence between human to

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

439 **Analysis of ectopic recombination**

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

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

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

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

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

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

472 **Methylation**

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

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

482 **Analysis of transcriptome and novel genes**

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

491

492 **ACKNOWLEDGMENTS**

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

498 This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have
499 previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to
500 HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of
501 this article can be made freely available under a CC BY 4.0 license immediately upon
502 publication.

503

504 **CONFLICTS OF INTERESTS**

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

507

508 **AUTHOR CONTRIBUTIONS**

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

512

513 **FIGURE LEGENDS**

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

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

534 **Figure 3. Evolution of subterminal SD spacers.** (A) Maximum likelihood phylogenetic tree of gorilla
 535 (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

561 >99.8%) with annotation tracks for higher-order pCht blocks, satellites, and SDs are shown on the right.
562 (B) Enlarged view of the two exchange breakpoints (black arrows) mapping within the 3rd and 25th SD
563 spacers. (C) Overview of potential ectopic exchange events between the subterminal caps (left-StCap1 vs
564 right-StCap2 tracks) identified in African great ape genomes (eight and three events in gorilla and
565 chimpanzee, respectively, including hsa3p vs. hsaXp in red). Assembly quality at the breakpoints was
566 validated by a read-depth analysis in **Supplemental Figs. S20-S26**. (D) Simulation test of exchange event
567 break points, suggesting that the exchange breaks (>99.8% identity alignment blocks; dotted line) are
568 more likely to be located within or close to subterminal SD spacers than a random distribution (red
569 histogram). (E,F) Comparison of maximum likelihood phylogeny to test recombination of SD spacer. A
570 phylogenetic tree of the complete spacer sequence (left) is compared to that of a subset of the sequence
571 (right). The tip nodes of the left tree are linked to the corresponding nodes of the right tree to visualize
572 phylogenetic shifts in the topology. Robinson-Foulds distance is used to measure the extent of
573 phylogenetic shift (first or last half sequences, in red and blue dotted lines and is compared to a null
574 distribution generated by computing distances of consensus tree vs. bootstrap replicate trees).

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

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

592

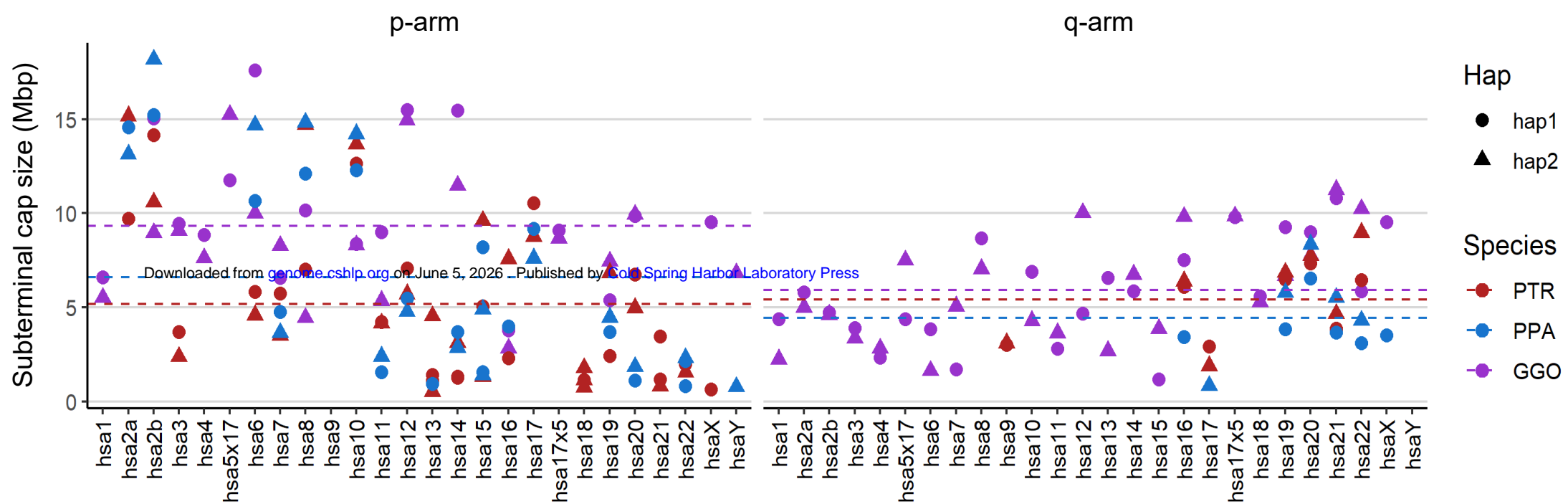
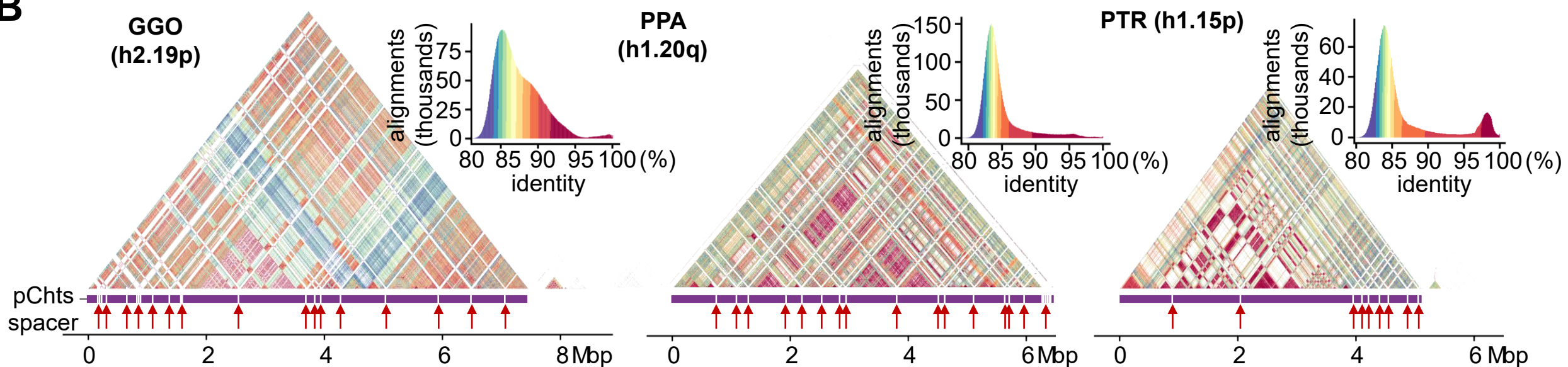
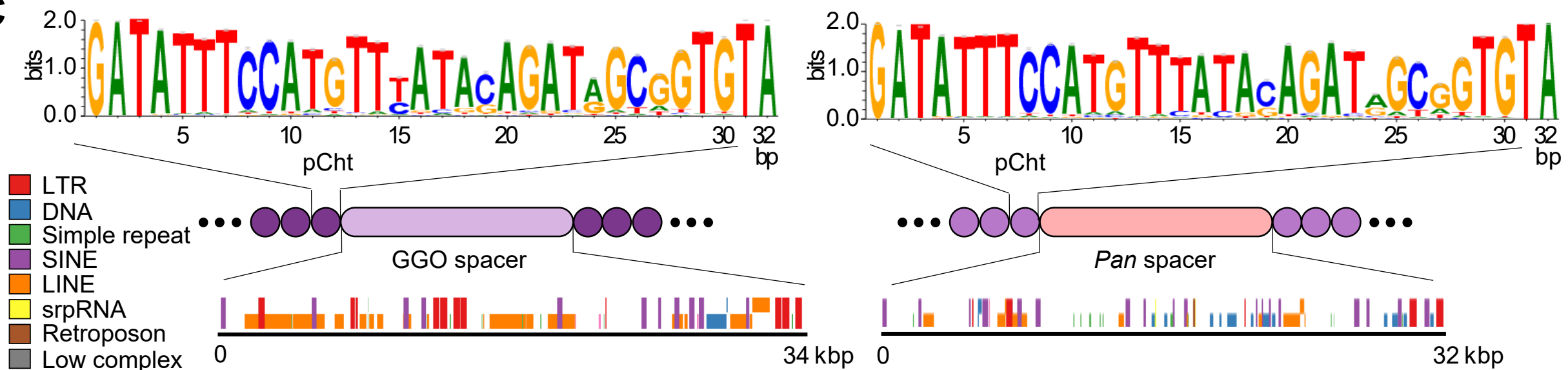
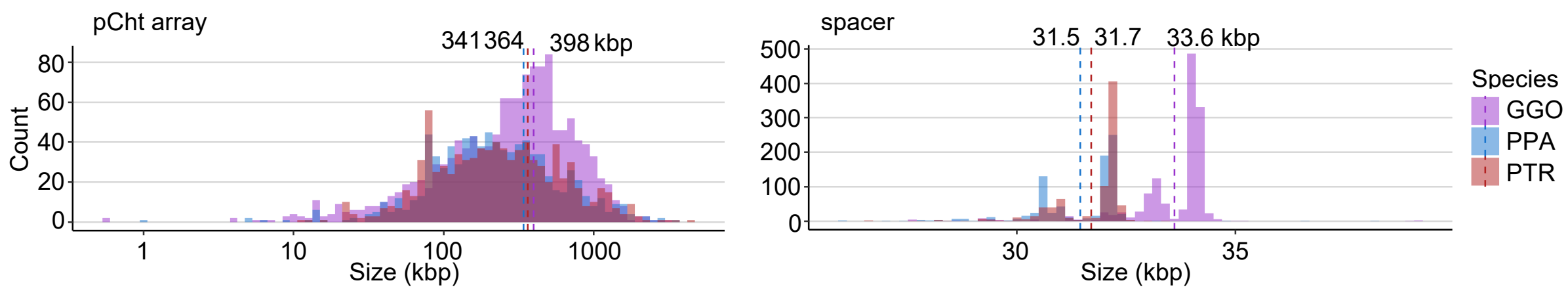
593 REFERENCES

- 594 Chen Y, Ye W, Zhang Y, Xu Y. 2015. High speed BLASTN: an accelerated MegaBLAST search tool.
595 *Nucleic acids research* **43**: 7762-7768.
- 596 Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, Sulovari A, Ebler J, Zhou W,
597 Serra Mari R. 2021. Haplotype-resolved diverse human genomes and integrated analysis of
598 structural variation. *Science* **372**: eabf7117.
- 599 Fontserè C., Kuhlwil M., Morcillo-Suarez C., Alvarez-Estape M., Lester J.D., Gratton P., Schmidt J.M.,
600 Dieguez P., Aebischer T., Álvarez-Varona P. and Agbor A., 2022. Population dynamics and
601 genetic connectivity in recent chimpanzee history. *Cell Genomics*, 2(6). Haaf T, Schmid M. 1987.
602 Chromosome heteromorphisms in the gorilla karyotype: Analyses with distamycin A/DAPI,
603 quinacrine and 5-azacytidine. *Journal of Heredity* **78**: 287-292.
- 604 Hartigan JA, Wong MA. 1979. Algorithm AS 136: A k -means clustering algorithm. *Journal of the royal*
605 *statistical society series c (applied statistics)* **28**: 100-108.
- 606 Hirai H, Hirai Y, Udono T, Matsubayashi K, Tosi AJ, Koga A. 2019. Structural variations of subterminal
607 satellite blocks and their source mechanisms as inferred from the meiotic configurations of
608 chimpanzee chromosome termini. *Chromosome Research* **27**: 321-332.

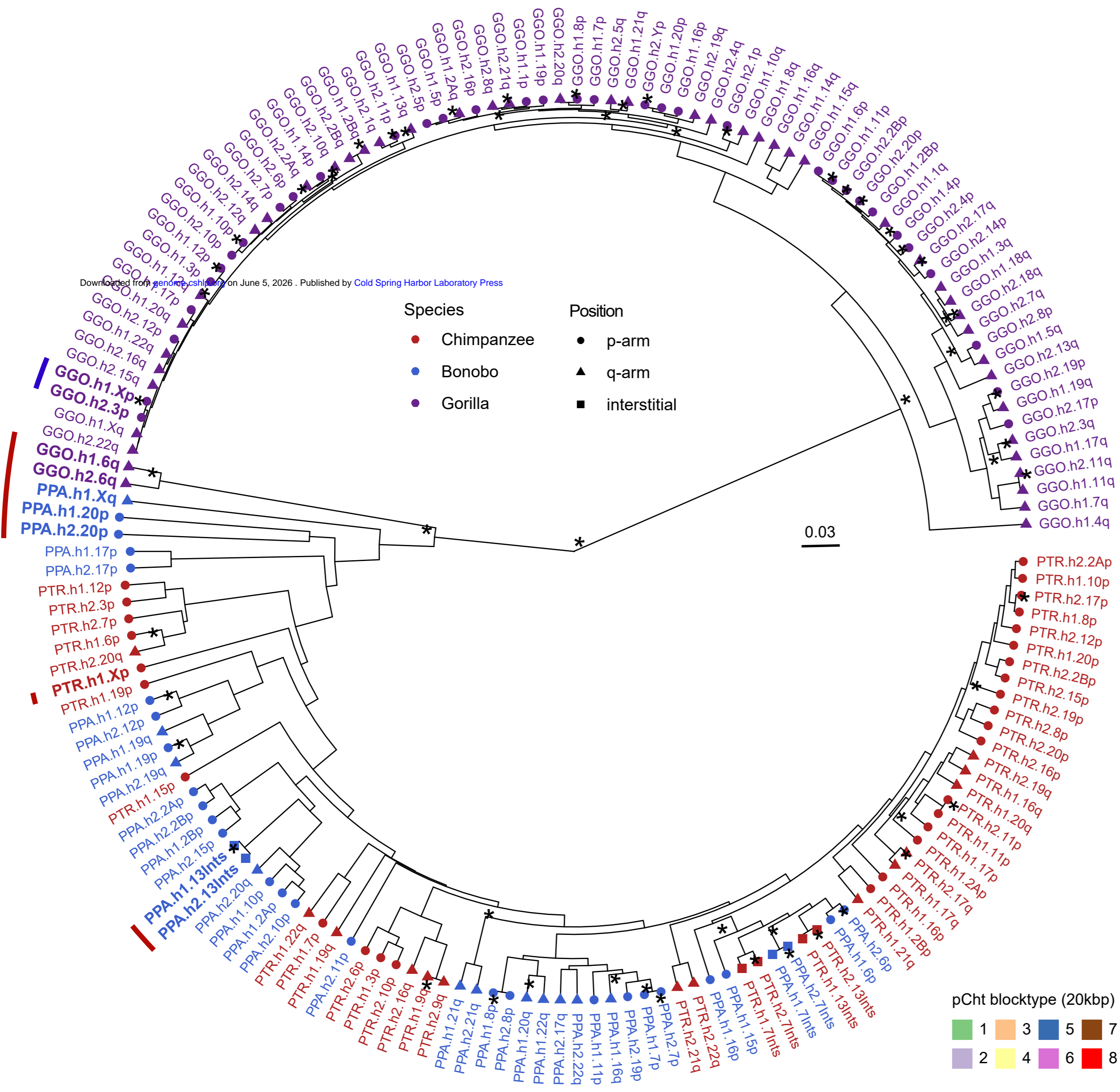
- 609 Ijdo JW, Baldini A, Ward D, Reeders S, Wells R. 1991. Origin of human chromosome 2: an ancestral
610 telomere-telomere fusion. *Proceedings of the National Academy of Sciences* **88**: 9051-9055.
- 611 Jiang X, Zhang L, Yang Z, Yang X, Ma K, Yoo D, Lu Y, Zhang S, Chen J, Nie Y. 2024. Incomplete
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, Kasprowitz 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.

- 635 Makova KD, Pickett BD, Harris RS, Hartley GA, Cechova M, Pal K, Nurk S, Yoo D, Li Q, Hebbar P.
636 2024. The complete sequence and comparative analysis of ape sex chromosomes. *Nature*: 1-11.
- 637 Mao Y, Harvey WT, Porubsky D, Munson KM, Hoekzema K, Lewis AP, Audano PA, Rozanski A, Yang
638 X, Zhang S et al. 2024. Structurally divergent and recurrently mutated regions of primate
639 genomes. *Cell* **187**: 1547-1562.e1513.
- 640 McConkey E. 2004. Orthologous numbering of great ape and human chromosomes is essential for
641 comparative genomics. *Cytogenetic & Genome Research* **105**.
- 642 Mefford HC, Trask BJ. 2002. The complex structure and dynamic evolution of human subtelomeres.
643 *Nature Reviews Genetics* **3**: 91-102.
- 644 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020.
645 IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.
646 *Mol Biol Evol* **37**: 1530-1534.
- 647 Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X and Thomas PD. 2019. Protocol Update for
648 large-scale genome and gene function analysis with the PANTHER classification system (v.
649 14.0). *Nature protocols* 14(3): 703-721.
- 650 Novo C, Arnoult N, Bordes WY, Castro-Vega L, Gibaud A, Dutrillaux B, Bacchetti S, Londoño-Vallejo
651 A. 2013. The heterochromatic chromosome caps in great apes impact telomere metabolism.
652 *Nucleic Acids Res* **41**: 4792-4801.
- 653 Nurk S, Koren S, Rhie A, Rautiainen M, Bizikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky
654 L, Gershman A. 2022. The complete sequence of a human genome. *Science* **376**: 44-53.
- 655 Porubsky D, Guitart X, Yoo D, Dishuck PC, Harvey WT, Eichler EE. 2024. SVbyEye: A visual tool to
656 characterize structural variation among whole-genome assemblies. *Bioinformatics*, 41(6):
657 btaf332.
- 658 Porubsky D, Sanders AD, Höps W, Hsieh P, Sulovari A, Li R, Mercuri L, Sorensen M, Murali SC,
659 Gordon D. 2020. Recurrent inversion toggling and great ape genome evolution. *Nature genetics*
660 **52**: 849-858.

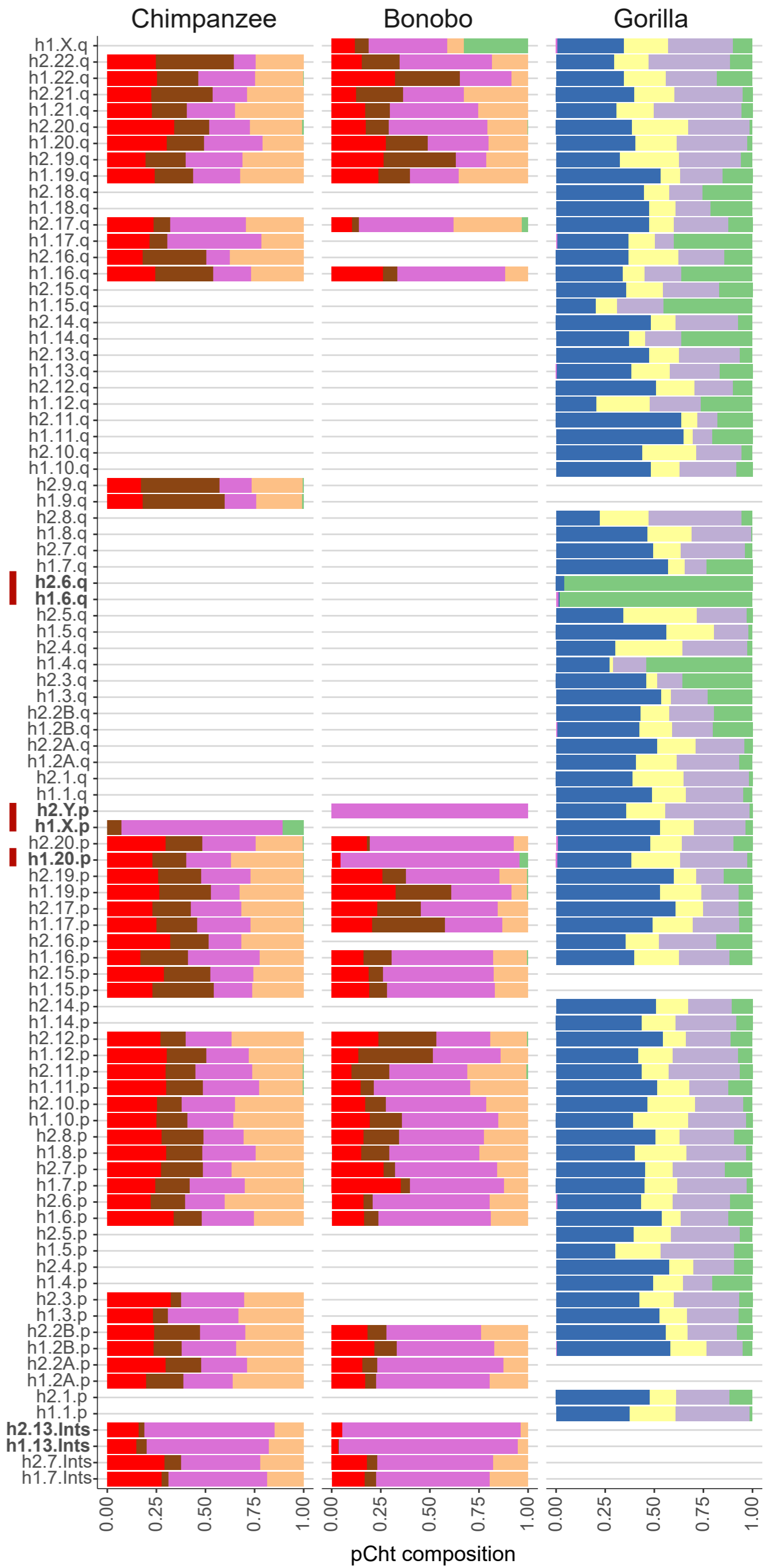
- 661 Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT and Salzberg SL. 2015. StringTie enables
662 improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology* 33(3):
663 290-295.
- 664 Rousseeuw PJ. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis.
665 *Journal of computational and applied mathematics* 20: 53-65.
- 666 Royle NJ, Baird DM, Jeffreys AJ. 1994. A subterminal satellite located adjacent to telomeres in
667 chimpanzees is absent from the human genome. *Nature genetics* 6: 52-56.
- 668 Sigurpalsdottir BD, Stefansson OA, Holley G, Beyter D, Zink F, Hardarson MP, Sverrisson SP,
669 Kristinsdottir N, Magnusdottir DN, Magnusson OP. and Gudbjartsson DF. 2024. A comparison
670 of methods for detecting DNA methylation from long-read sequencing of human genomes.
671 *Genome Biology* 25(1): 69.
- 672 Suzuki R, Shimodaira H. 2006. Pvcust: an R package for assessing the uncertainty in hierarchical
673 clustering. *Bioinformatics* 22: 1540-1542.
- 674 Ventura M, Catacchio CR, Alkan C, Marques-Bonet T, Sajjadian S, Graves TA, Hormozdiari F, Navarro
675 A, Malig M, Baker C. 2011. Gorilla genome structural variation reveals evolutionary parallelisms
676 with chimpanzee. *Genome research* 21: 1640-1649.
- 677 Ventura M, Catacchio CR, Sajjadian S, Vives L, Sudmant PH, Marques-Bonet T, Graves TA, Wilson RK,
678 Eichler EE. 2012. The evolution of African great ape subtelomeric heterochromatin and the
679 fusion of human chromosome 2. *Genome research* 22: 1036-1049.
- 680 Vollger MR, Kerpedjiev P, Phillippy AM, Eichler EE. 2022. StainedGlass: interactive visualization of
681 massive tandem repeat structures with identity heatmaps. *Bioinformatics* 38: 2049-2051.
- 682 Yoo D, Rhie A, Hebbar P, Antonacci F, Logsdon GA, Solar SJ, Antipov D, Pickett BD, Safonova Y,
683 Montinaro F. 2025. Complete sequencing of ape genomes. *Nature*: 1-18.
- 684 Yunis JJ, Prakash O. 1982. The origin of man: a chromosomal pictorial legacy. *Science* 215: 1525-1530.
- 685

A**B****C****D**

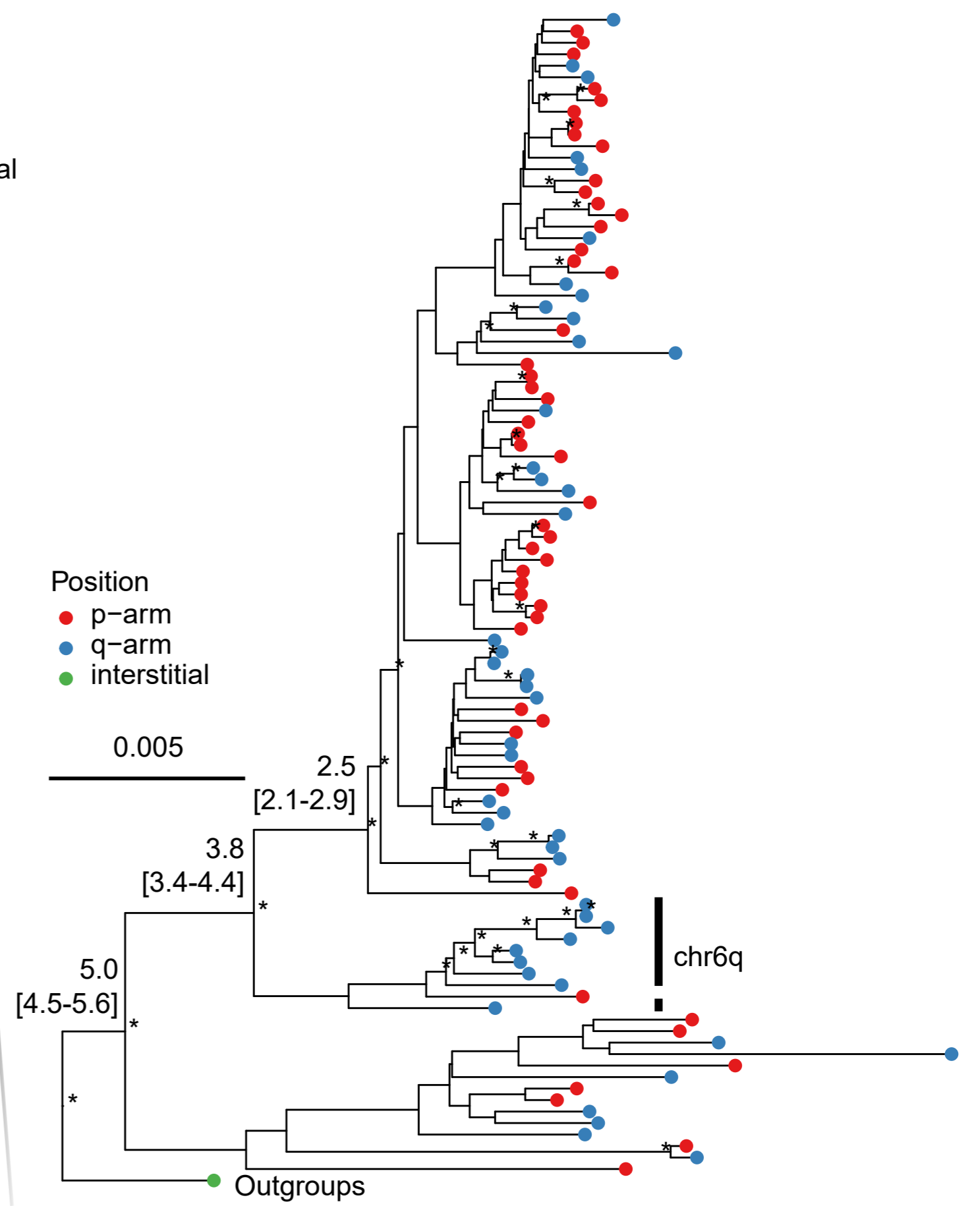
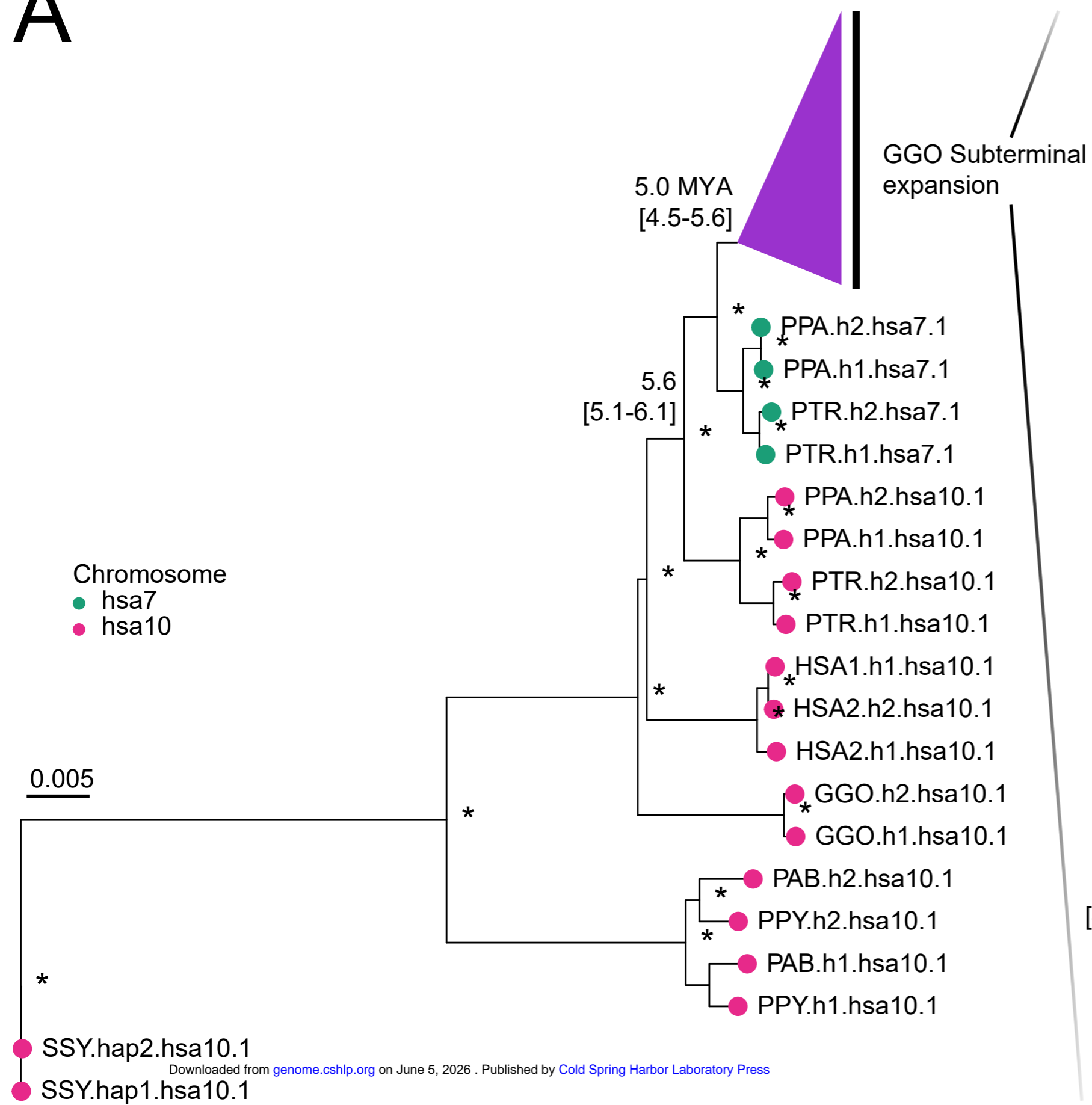
A



B



A



B

