



# Dynamic evolution of great ape Y chromosomes

Monika Cechova<sup>a,1,2</sup>, Rahulsham Vegesna<sup>a,b,2,3</sup>, Marta Tomaszewicz<sup>a</sup>, Robert S. Harris<sup>a</sup>, Di Chen<sup>a</sup>, Samarth Rangavittal<sup>a,b</sup>, Paul Medvedev<sup>c,d,4</sup>, and Kateryna D. Makova<sup>a,4</sup>

<sup>a</sup>Department of Biology, Penn State University, University Park, PA 16802; <sup>b</sup>Intercollege Graduate Program in Bioinformatics and Genomics, Penn State University, University Park, PA 16802; <sup>c</sup>Department of Computer Science and Engineering, Penn State University, University Park, PA 16802; and <sup>d</sup>Department of Biochemistry and Molecular Biology, Penn State University, University Park, PA 16802

Edited by Amanda M. Larracunte, University of Rochester, Rochester, NY, and accepted by Editorial Board Member Daniel L. Hartl September 3, 2020 (received for review January 30, 2020)

The mammalian male-specific Y chromosome plays a critical role in sex determination and male fertility. However, because of its repetitive and haploid nature, it is frequently absent from genome assemblies and remains enigmatic. The Y chromosomes of great apes represent a particular puzzle: their gene content is more similar between human and gorilla than between human and chimpanzee, even though human and chimpanzee share a more recent common ancestor. To solve this puzzle, here we constructed a dataset including Ys from all extant great ape genera. We generated assemblies of bonobo and orangutan Ys from short and long sequencing reads and aligned them with the publicly available human, chimpanzee, and gorilla Y assemblies. Analyzing this dataset, we found that the genus *Pan*, which includes chimpanzee and bonobo, experienced accelerated substitution rates. *Pan* also exhibited elevated gene death rates. These observations are consistent with high levels of sperm competition in *Pan*. Furthermore, we inferred that the great ape common ancestor already possessed multicopy sequences homologous to most human and chimpanzee palindromes. Nonetheless, each species also acquired distinct ampliconic sequences. We also detected increased chromatin contacts between and within palindromes (from Hi-C data), likely facilitating gene conversion and structural rearrangements. Our results highlight the dynamic mode of Y chromosome evolution and open avenues for studies of male-specific dispersal in endangered great ape species.

sex chromosomes | palindromes | gene content evolution

The mammalian male-specific sex chromosome—the Y—is vital for sex determination and male fertility and is a useful marker for population genetics studies. It carries *SRY*, which encodes the testis-determining factor that initiates male sex determination (1). The human Y also harbors azoospermia factor regions, deletions of which can cause infertility (2). Y chromosome sequences have been used to analyze male dispersal (3) and hybridization sex bias (4) in natural populations. Thus, the Y is important biologically and its sequences have critical practical implications. Moreover, study of the Y is needed to obtain a complete picture of mammalian genome evolution. Yet, due to its repetitive and haploid nature, the Y has been assembled for only a handful of mammalian species (5).

Among great apes, the Y has so far been assembled only in human (6), chimpanzee (7), and gorilla (8). A comparative study of these Y assemblies (8) uncovered some unexpected patterns which could not be explained with the data from three species alone. Despite a recent divergence of these species (~7 million years ago [MYA]) (9), their Y chromosomes differ enormously in size and gene content, in sharp contrast to the stability of the rest of the genome. For example, the chimpanzee Y is only half the size of the human Y, and the percentage of gene families shared by these two chromosomes (68%) that split ~6 MYA (9) is similar to that shared by human and chicken autosomes that split ~310 MYA (7). Puzzlingly, in terms of shared genes and overall architecture, the human Y is more similar to the gorilla Y than to the chimpanzee Y even though human and chimpanzee have a more recent common ancestor (8). Y chromosomes from additional great ape species should be sequenced to understand

whether high interspecific variability in gene content and architecture is characteristic of all great ape Ys.

All great ape Y chromosomes studied thus far include *pseudoautosomal regions (PARs)*, which recombine with the X chromosome, and *male-specific X-degenerate*, ampliconic, and heterochromatic regions, which evolve as a single linkage group (6–8). The X-degenerate regions are composed of segments with different levels of homology to the X chromosome, which are called strata, corresponding to stepwise losses of X-Y recombination. Because of lack of recombination (except for occasional X-Y gene conversion (10, 11), X-degenerate regions are expected to accumulate gene-disrupting mutations; however, this has not been examined in detail. The ampliconic regions consist of repetitive sequences that have >50% identity to each other and contain palindromes—inverted repeats (separated by a spacer) up to several megabases long the arms of which are >99.9% identical (6). Palindromes are thought to evolve to allow intrachromosomal (Y-Y) gene conversion (12) which rescues the otherwise non-recombining male-specific regions from deleterious mutations (13). We presently lack knowledge about how conserved palindrome sequences are across great apes. In general, X-degenerate

## Significance

The male-specific Y chromosome harbors genes important for sperm production. Because Y is repetitive, its DNA sequence was deciphered for only a few species, and its evolution remains elusive. Here we compared the Y chromosomes of great apes (human, chimpanzee, bonobo, gorilla, and orangutan) and found that many of their repetitive sequences and multicopy genes were likely already present in their common ancestor. Y repeats had increased intrachromosomal contacts, which might facilitate preservation of genes and gene regulatory elements. Chimpanzee and bonobo, experiencing high sperm competition, underwent many DNA changes and gene losses on the Y. Our research is significant for understanding the role of the Y chromosome in reproduction of nonhuman great apes, all of which are endangered.

Author contributions: M.C., R.V., M.T., P.M., and K.D.M. designed research; M.C., R.V., M.T., and R.S.H. performed research; M.C. and R.V. contributed new reagents/analytic tools; M.C., R.V., M.T., R.S.H., D.C., and S.R. analyzed data; and M.C., R.V., M.T., P.M., and K.D.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. A.M.L. is a guest editor invited by the Editorial Board.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>Present address: Central European Institute of Technology, Department of Genetics and Reproduction, Veterinary Research Institute, Brno, 62100 Czech Republic.

<sup>2</sup>M.C. and R.V. contributed equally to this work.

<sup>3</sup>Present address: Cancer Data Science Laboratory, National Cancer Institute, Bethesda, MD 20892.

<sup>4</sup>To whom correspondence may be addressed. Email: pzm11@psu.edu or kdm16@psu.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2001749117/-DCSupplemental>.

First published October 5, 2020.

regions are more conserved, whereas ampliconic regions are prone to rearrangements, and heterochromatic regions, which are rich in satellite repeats (14), evolve very rapidly among species (7, 8, 14). However, the evolution of great ape Y chromosomes outside of human, chimpanzee, and gorilla has not been explored.

Known Y chromosome protein-coding genes are located in either the X-degenerate or ampliconic regions. X-degenerate genes (16 on the human Y) are single-copy, ubiquitously expressed genes with housekeeping functions (15). Multicopy ampliconic genes (nine gene families on the human Y, eight of which—all but *TSPY*—are located in palindromes) are expressed only in testis and function during spermatogenesis (6). Some human Y genes are deleted or pseudogenized in other great apes and thus are not essential for all species (8, 16). To illuminate genes essential for male reproduction in nonhuman great apes, all of which are endangered species, a cross-species analysis of Y-gene content evolution is needed.

Here we compared the Y chromosomes in five species representing all four great ape genera: the human (*Homo*) lineage diverged from the chimpanzee (*Pan*), gorilla (*Gorilla*), and orangutan (*Pongo*) lineages ~6, ~7, and ~13 MYA, respectively (9), and the bonobo and chimpanzee lineages (belonging to the genus *Pan*), which diverged ~1 MYA (17). We produced draft assemblies of the bonobo and Sumatran orangutan Ys and combined them with the human, chimpanzee, and gorilla Y assemblies (6–8) to construct great ape Y multispecies alignments. This comprehensive dataset enabled us to answer several pivotal questions about the evolution of great ape Y chromosomes. First, we assessed lineage-specific substitution rates and identified species experiencing significant rate acceleration. Second, we determined interspecific gene content turnover. Third, we evaluated the conservation of palindromic sequences and examined chromatin interactions within ampliconic regions. Our results highlight the highly dynamic nature of great ape Y chromosome evolution.

## Results

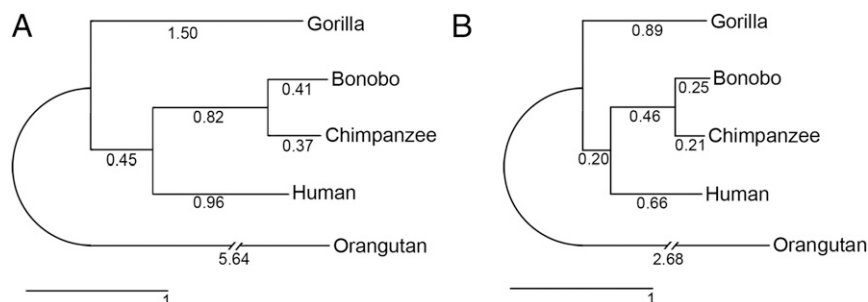
**Assemblies.** To obtain Y chromosome assemblies for all major great ape lineages, we augmented publicly available human and chimpanzee assemblies (6, 7) by producing draft bonobo and Sumatran orangutan (henceforth called “orangutan”) assemblies and by improving the gorilla assembly (8) of Y male-specific regions (*SI Appendix, Fig. S1*, and see *Materials and Methods* for details). The resulting assemblies (henceforth called “Y assemblies”) were of high quality, as evidenced by their high degree of homology to the human and chimpanzee Ys (*SI Appendix, Fig. S2*) and by the presence of sequences of the most expected homologs (16) of human Y genes (*SI Appendix, Fig. S3*). They also were of sufficient continuity (*SI Appendix, Table S1*), particularly when the highly repetitive structure of the Y is taken into account.

**Ampliconic and X-Degenerate Scaffolds.** To determine which scaffolds are ampliconic and which are X-degenerate in our bonobo,

gorilla, and orangutan Y assemblies [such annotations are already available for the human and chimpanzee Ys (6, 7)], we developed a classifier which combines the copy count in the assemblies with mapping read depth information from whole-genome sequencing of male individuals (*SI Appendix, Supplemental Note S1*). This approach was needed as ampliconic regions can be collapsed in assemblies based on next-generation sequencing data (5). Using this classifier, we identified 12.5, 10.0, and 14.5 Mb of X-degenerate scaffolds in bonobo, gorilla, and orangutan, respectively. The length of ampliconic regions was more variable: 10.8 Mb in bonobo, 4.0 Mb in gorilla, and 2.2 Mb in orangutan. Due to potential collapse of repeats, we might have underestimated the true lengths of ampliconic regions. However, their length estimates are expected to reflect their complexity: e.g., the complexity might be low in the orangutan Y, which is consistent with a high read depth in its Y ampliconic scaffolds (*SI Appendix, Fig. S4*) and with its long gene-harboring repetitive arrays previously found cytogenetically (18).

**Alignments.** We aligned the sequences of the Y chromosomes from five great ape species (see *Materials and Methods* for details). The resulting multispecies alignment allowed us to identify species-specific sequences, sequences shared by all species, and sequences shared by some but not all species (*SI Appendix, Fig. S5* and *Table S2A*). These results were confirmed by pairwise alignments (*SI Appendix, Table S2B*). For example, as was shown previously (8), the gorilla Y had the highest percentage of its sequence aligning to the human Y (75.5 and 89.6% from multispecies and pairwise alignments, respectively). In terms of sequence identity (*SI Appendix, Tables S2, C and D*), the chimpanzee and bonobo Ys were most similar to each other (99.1 to 99.2% and ~98% from multispecies and pairwise alignments, respectively), while the orangutan Y had the lowest identity to any other great ape Y chromosomes (~93 to 94% and ~92% from multispecies and pairwise alignments, respectively). From multispecies alignments (*SI Appendix, Table S2C*), the human Y was most similar in sequence to the chimpanzee or bonobo Ys (97.9 and 97.8%, respectively), less similar to the gorilla Y (97.2%), and the least similar to the orangutan Y (93.6%), in agreement with the accepted phylogeny of these species (9). The pairwise alignments confirmed this trend (*SI Appendix, Table S2D*). These results argue against incomplete lineage sorting at the male-specific Y chromosome locus in great apes.

**Substitution Rates on the Y.** We next asked whether the chimpanzee Y chromosome, the architecture and gene content of which differ drastically from the human and gorilla Ys (8), experienced an elevated substitution rate. Using our multispecies Y chromosome alignment, we estimated substitution rates along the branches of the great ape phylogenetic tree (Fig. 1A and see *Materials and Methods* for details). A similar analysis was performed using an alignment of autosomes (Fig. 1B). A higher substitution rate on the Y than on the autosomes, i.e., male mutation bias (19), was found for each branch of the phylogeny (Fig. 1 and *SI Appendix, Supplemental*



**Fig. 1.** Phylogenetic tree of nucleotide sequences. (A) Y chromosome. (B) Autosomes. Branch lengths (substitutions per 100 sites) were estimated from multispecies alignment blocks including all five species.

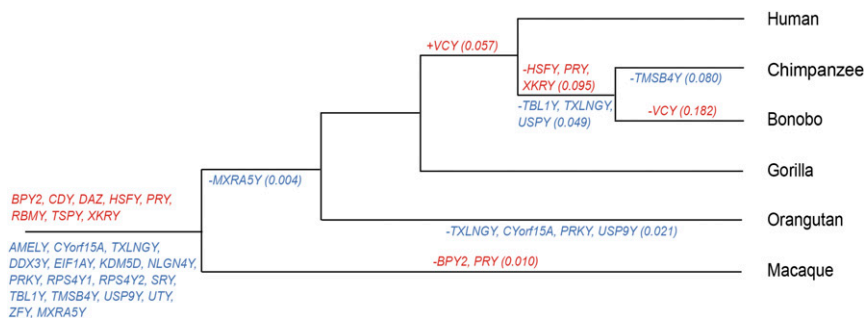
Note S2). Notably, the Y-to-autosomal substitution rate ratio was higher in the *Pan* lineage, including the chimpanzee (1.76) and bonobo (1.64) lineages and the lineage of their common ancestor (1.78), than in the human lineage (1.45). These trends did not change after correcting for ancestral polymorphism (*SI Appendix, Supplemental Note S2*). We subsequently used a test akin to the relative rate test (20) and addressed whether the *Pan* lineage experienced more substitutions than the human lineage (*SI Appendix, Table S3*). Using gorilla as an outgroup, we observed a significantly higher number of substitutions that occurred between chimpanzee and gorilla than between human and gorilla. For autosomes, this number was 0.6% higher in the chimpanzee–gorilla than in the human–gorilla comparison, whereas for the Y, it was 7.9% higher ( $P < 1 \times 10^{-5}$  in both cases,  $\chi^2$ -test on the contingency table). Similarly, we observed a higher number of substitutions that occurred between bonobo and gorilla than between human and gorilla. This number was 2.9% higher in the bonobo–gorilla than in the human–gorilla comparison for autosomes and as much as 9.6% higher for the Y ( $P < 1 \times 10^{-5}$  in both cases,  $\chi^2$ -test on the contingency table). Thus, while the *Pan* lineage experienced an elevated substitution rate at both autosomes and the Y, this elevation was particularly strong on the Y.

**Gene Content Evolution.** Utilizing sequence assemblies and testis expression data (21), we evaluated gene content and the rates of gene birth and death on the Y chromosomes of five great ape species. First, we examined the presence/absence of homologs of human Y chromosome genes (16 X-degenerate genes + 9 ampliconic gene families = 25 gene families; for multicopy ampliconic gene families, we were not studying copy number variation, but only presence/absence of a family in a species; *SI Appendix, Fig. S6*). Such data were previously available for the chimpanzee Y, in which 7 of 25 human Y gene families became pseudogenized or deleted (7), and for the gorilla Y, in which only one gene family (*VCY*) of 25 is absent (8). Here, we compiled the data for bonobo and orangutan. From the 25 gene families present on the human Y, the bonobo Y lacked 7 (*HSFY*, *PRY*, *TBL1Y*, *TXLNGY*, *USP9Y*, *VCY*, and *XKRY*) and the orangutan Y lacked 5 (*TXLNGY*, *CYorf15A*, *PRKY*, *USP9Y*, and *VCY*). Second, our gene annotation pipeline did not identify novel genes in the bonobo and orangutan Y assemblies (*SI Appendix, Supplemental Note S3*), similar to previous results for the chimpanzee (7) and gorilla (8) Ys. Thus, we obtained the complete information about gene family content on the Y chromosome in five great ape species.

Using this information and utilizing the macaque Y chromosome (22) as an outgroup, we reconstructed gene content at ancestral nodes and studied the rates of gene birth and death (23)

across the great ape phylogeny. Because X-degenerate and ampliconic genes might exhibit different trends, we analyzed them separately (Fig. 2 and *SI Appendix, Table S4*). Considering gene births, none can occur for X-degenerate genes because they were present on the proto-sex chromosomes. Only one gene birth (*VCY*, in the human–chimpanzee–bonobo common ancestor) was observed for ampliconic genes, leading to overall low gene birth rates. Considering gene deaths, three ampliconic gene families and three X-degenerate genes were lost by the chimpanzee–bonobo common ancestor, leading to death rates of 0.095 and 0.049 events/MY, respectively. Bonobo lost an additional ampliconic gene, whereas chimpanzee lost an additional X-degenerate gene, leading to death rates of 0.182 and 0.080 events/MY, respectively. In contrast, no deaths of either ampliconic or X-degenerate genes were observed in human and gorilla. Orangutan did not experience any deaths of ampliconic genes, but lost four X-degenerate genes. Its X-degenerate gene death rate (0.021 events/MY) was still lower than that in the chimpanzee lineage (0.080 events/MY) or in the bonobo–chimpanzee common ancestor (0.049 events/MY). To summarize, the *Pan* genus exhibited the highest death rates for both X-degenerate and ampliconic genes across great apes. Additionally, we observed significantly higher nonsynonymous-to-synonymous rate ratios for four X-degenerate genes (*DDX3Y*, *EIF1AY*, *PRKY*, and *ZFY*) and one ampliconic gene (*CDY*) in bonobo, chimpanzee, and/or their common ancestor (*SI Appendix, Supplemental Note S4*). However, none of these ratios was significantly greater than one, providing no evidence for positive selection.

**Conservation of Human and Chimpanzee Palindrome Sequences.** Did the palindromes (human palindromes are labeled with P and chimpanzee palindromes are labeled with C) now present on the human Y (P1 to P8) and chimpanzee Y (C1 to C19) evolve before or after the great ape lineages split? To answer this question, we identified the proportions of human and chimpanzee palindrome sequences that aligned to bonobo, orangutan, and gorilla Ys in our multispecies alignments (Fig. 3A and *SI Appendix, Tables S5 and S6*). Among human palindromes, P5 and P6 were the most conserved (covered by 86 to 96% of other great ape Y assemblies), whereas the majority of P3 sequences were human specific (covered by only 31 to 37% of other great ape Y assemblies). Nevertheless, the common ancestor of great apes likely already had substantial lengths of sequences homologous to P1, P2, and P4 to P8, and some sequences of P3 (Fig. 3B). Chimpanzee palindromes C17, C18, and C19 are homologous to human palindromes P8, P7, and P6, respectively (7). Therefore, we focused on the other chimpanzee palindromes



**Fig. 2.** Evolution of Y chromosome gene content in great apes. The reconstructed history of gene birth and death for X-degenerate (blue) and ampliconic (red) genes was overlaid on the great ape phylogenetic tree (not drawn to scale), using macaque as an outgroup. The rates of gene birth and death (in events per million years) are shown in parentheses (for complete data, see *SI Appendix, Fig. S6 and Table S4*). The list at the root includes the genes that were present in the common ancestor of great apes and macaque. In addition to most of the genes on the human Y, the macaque Y harbors the X-degenerate *MXRASY* gene, which we found to be deleted in orangutan and pseudogenized in bonobo, chimpanzee, gorilla, and human. We currently cannot find a full-length copy of the *VCY* gene in bonobo (*SI Appendix, Supplemental Note S7*). *TXLNGY* and *DDX3Y* are also known as *CYorf15B* and *DBY*, respectively.

and, following ref. 7, divided them into five homologous groups: C1 (C1+C6+C8+C10+C14+C16), C2 (C2+C11+C15), C3 (C3+C12), C4 (C4+C13), and C5 (C5+C7+C9) (*SI Appendix, Table S6*). The palindromes in the C3, C4, and C5 groups had substantial proportions (47 to 95%) of their sequences covered by alignments with other great ape Ys (Fig. 3A). In contrast, most C2 sequences (85%) were shared only with bonobo, and a substantial proportion of C1 sequences was chimpanzee specific. Nonetheless, the common ancestor of great apes likely already had large amounts of sequences homologous to group C3, C4, and C5 palindromes and also some sequences homologous to group C1 and C2 palindromes (Fig. 3B).

To determine whether the bonobo, orangutan, and gorilla sequences homologous to human or chimpanzee palindromes were multicopy (i.e., present in more than one copy), and thus could have been arranged as palindromes in the common ancestor of great apes, we obtained their read depths from whole-genome sequencing of their respective males (Fig. 3A and *Materials and Methods*). This approach was used because we expected that some palindromes were collapsed in our Y assemblies. We also used the data on the homology between human and chimpanzee palindromes summarized from the literature (6–8) (*SI Appendix, Table S7*). Using maximum parsimony reconstruction, we concluded (*SI Appendix, Supplemental Note S5*) that sequences homologous to P4, P5, P8, and C4 and partial sequences homologous to P1, P2, and C2 were multicopy in the common ancestor of great apes (Fig. 3B). Sequences homologous to P3, P6, and C1 were multicopy in the human–gorilla common ancestor, and those homologous to P7 and C5 were multicopy in the human–chimpanzee common ancestor (Fig. 3B and *SI Appendix, Supplemental Note S5*).

#### Species-Specific Multicopy Sequences in Bonobo, Gorilla, and Orangutan.

In addition to finding sequences homologous to human and/or chimpanzee palindromes, we detected 9.36, 1.73, and 3.35 Mb of species-specific sequences in our bonobo, gorilla, and orangutan Y assemblies (*SI Appendix, Table S8, Fig. S5, and Supplemental Note S6*). By mapping male whole-genome sequencing reads to these sequences (*Materials and Methods*), we found that 81, 44, and 30% of them in bonobo, gorilla, and orangutan had a copy number of 2 or above (*SI Appendix, Table S8*). Thus, large portions of Y species-specific sequences are multicopy and might harbor species-specific palindromes.

#### Frequent Chromatin Interactions between and within Palindromes.

Because Y ampliconic regions undergo Y-Y gene conversion and Non-Allelic Homologous Recombination (NAHR) (13), we hypothesized that these processes are facilitated by increased chromatin interactions. To evaluate this, we studied chromatin interactions on the Y utilizing a statistical approach specifically developed for handling Hi-C data originating from repetitive sequences (24). We used publicly available Hi-C data generated for human and chimpanzee induced pluripotent stem cells (iPSCs) (25) and for human umbilical vein endothelial cells (26). We found prominent chromatin contacts both between and within palindromes located inside ampliconic regions on the Y (Fig. 4A and B). In fact, the contacts in the human palindromic regions were significantly overrepresented when compared with the expectation based on the proportion of the Y occupied by palindromes ( $P < 0.001$ , permutation test with palindromic/nonpalindromic group categories; *SI Appendix, Table S9 and Fig. S8*), suggesting biological importance. Notably, we observed similar patterns for two different human cell types, as well as for both human and chimpanzee iPSCs (Fig. 4A and B and *SI Appendix, Fig. S9*).

We also hypothesized that arms of the same palindrome interact with each other via chromatin contacts. Our analysis of human Hi-C data from iPSCs (25) suggests that palindrome arms are indeed colocalized—a pattern particularly prominent for the large palindromes P1 and P5 (Fig. 4C). These results suggest

that, in addition to the enrichment in the local interactions expected to be present in the Hi-C data (27), homologous regions of the two arms of a palindrome interact with each other with high frequency.

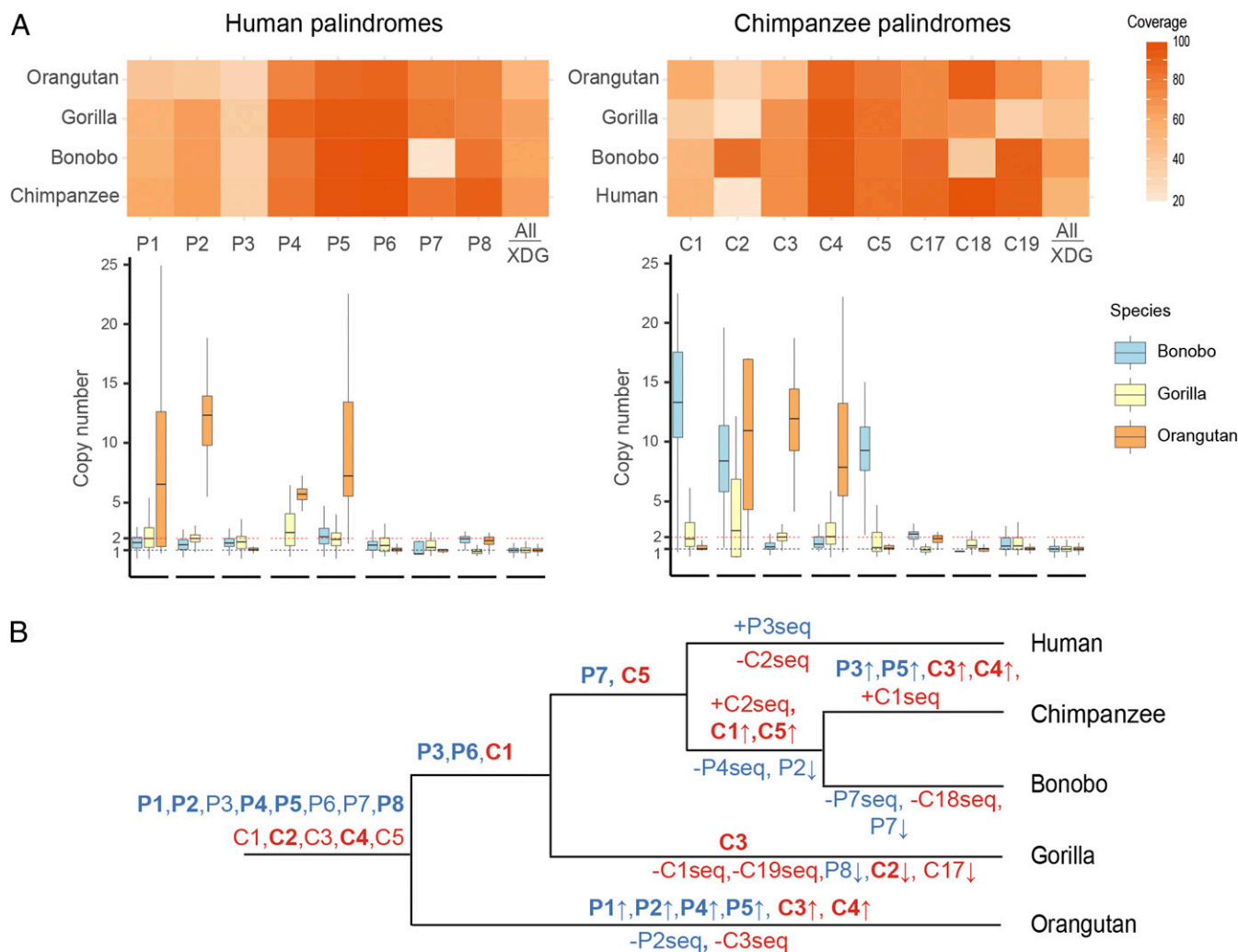
#### Discussion

**Substitution Rates.** Higher substitution rates on the Y than on the autosomes, which we found across the great ape phylogeny, confirm another study (28) and are consistent with male mutation bias likely caused by a higher number of cell divisions in the male than in the female germline (19). Higher autosomal substitution rates that we detected in the *Pan* than *Homo* lineage corroborate yet another study (29) and can be explained by a shorter generation time in *Pan*. A higher Y-to-autosomal substitution ratio (i.e., stronger male mutation bias) in the *Pan* than in the *Homo* lineage, as observed by us here, could be due to several reasons. First, species with sperm competition produce more sperm and thus undergo a greater number of replication rounds, generating more mutations on the Y and potentially leading to stronger male mutation bias than species without sperm competition (19). Consistent with this expectation, chimpanzee and bonobo experience sperm competition and exhibit strong male mutation bias, as compared with no sperm competition (30) and weak male mutation bias in human and gorilla (*SI Appendix, Supplemental Note S2*). Contradicting this expectation, orangutans have limited sperm competition (30), but exhibit strong male mutation bias (*SI Appendix, Supplemental Note S2*). Second, a shorter spermatogenic cycle can increase the number of replication rounds per time unit and can elevate Y substitution rates, leading to stronger male mutation bias. In agreement with this explanation, the spermatogenic cycle is shorter in chimpanzee than in human (31, 32); the data are limited for other great apes. Third, a stronger male mutation bias would be expected in *Pan* than in *Homo* if the ratio of male-to-female generation times was respectively higher (33). However, the opposite is true: this ratio is higher in *Homo* than in *Pan* (33).

Phylogenetic studies produce estimates of male mutation bias that might be affected by ancient genetic polymorphism in closely related species (28). Even though we corrected for this effect (*SI Appendix, Supplemental Note S2*), our results should be taken with caution because of incomplete data on the sizes of ancestral great ape populations (34). Pedigree studies inferring male mutation bias are unaffected by ancient genetic polymorphism. One such study detected significantly higher male mutation bias in chimpanzee than in human (35), in agreement with our results, while another study found no significant differences in male mutation bias among great apes (36). These two studies analyzed only a handful of trios per species, and thus their conclusions should be reevaluated in larger studies.

**Ampliconic Sequences.** We found that substantial portions of most human palindromes, and of most chimpanzee palindrome groups, were likely multicopy (and thus potentially palindromic) in the common ancestor of great apes, suggesting conservation over >13 MY. Moreover, two of the three rhesus macaque palindromes are conserved with human palindromes P4 and P5 (22), indicating conservation over >25 MY. Our study also found species-specific amplification or loss of ampliconic sequences, indicating that their evolution is rapid. Thus, repetitive sequences constitute a biologically significant component of great ape Y chromosomes, and their multicopy state might be selected for.

Ampliconic sequences are thought to have evolved multiple times in diverse species to enable Y-Y NAHR including intrachromosomal gene conversion and nonallelic crossing-over (reviewed in ref. 37). Y-Y NAHR can compensate for degeneration in the absence of interchromosomal recombination on the Y by removing deleterious mutations (38, 39), can decrease the drift-driven loss of less mutated alleles, can lead to concerted



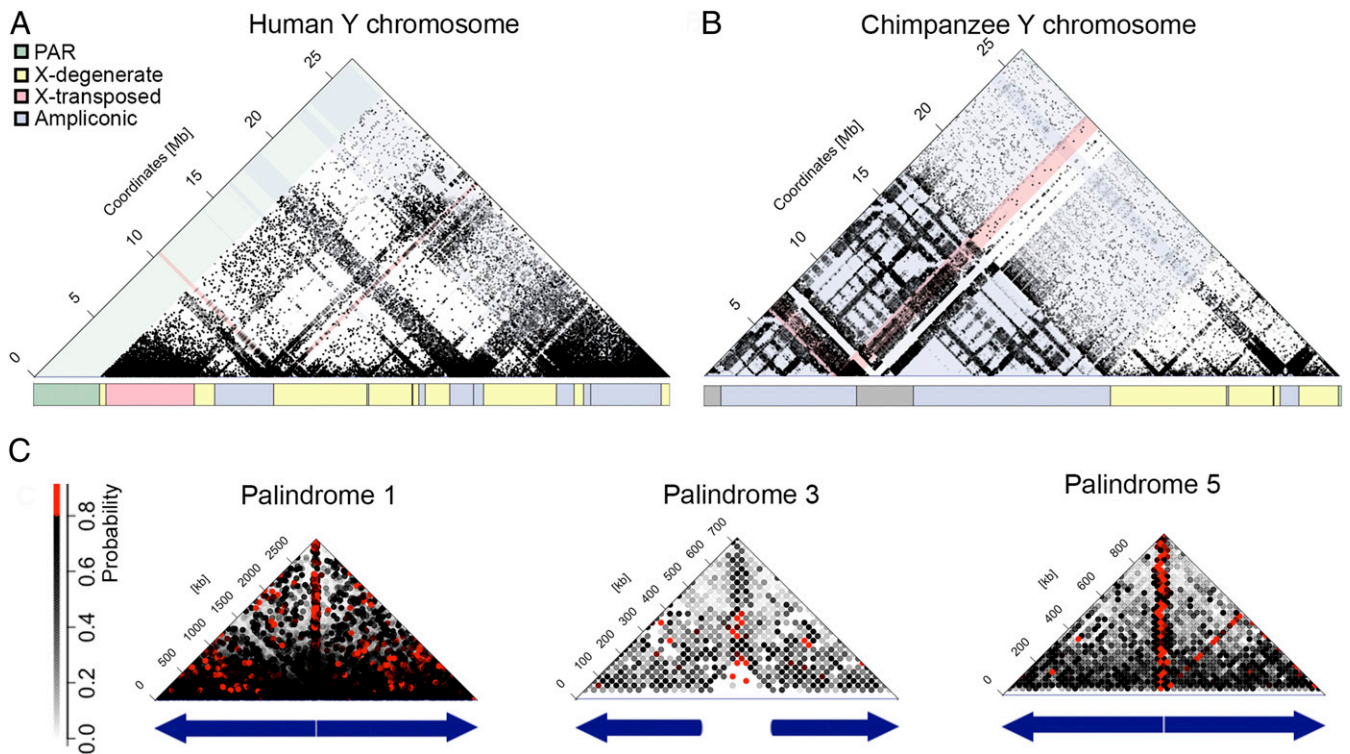
**Fig. 3.** Evolution of sequences homologous to human and chimpanzee palindromes. (A) Heatmaps showing coverage for each palindrome in each species in the multispecies alignment (the last column includes all palindromes) and box plots representing copy number (natural log) of 1-kb windows which have homology with human or chimpanzee palindromes (the last box plot is for X-degenerate genes; the data for human and chimpanzee can be found in *SI Appendix, Table S7*). (B) The great ape phylogenetic tree (not drawn to scale) with evolution of human (shown in blue) and chimpanzee (red) palindromic sequences overlaid on it. Palindrome names in boldface indicate that their sequences were present in two or more copies. Negative (-) and positive (+) signs indicate loss and gain of palindrome sequence (possibly only partial), respectively. Arrows represent gain (↑) or loss (↓) of palindrome copy number. If several equally parsimonious scenarios were possible, we conservatively assumed a later date of acquisition of the multicopy state for a palindrome (*SI Appendix, Supplemental Note S5*, for additional details; the data are shown in *SI Appendix, Table S7*).

evolution of repeats (13), and can increase the fixation rate of beneficial mutations (37). Yet, despite its critical importance for the Y, how Y-Y NAHR occurs mechanistically is not well understood. Our analysis of Hi-C data suggested that ampliconic sequences and palindrome arms colocalize on the Y in both human and chimpanzee, potentially facilitating Y-Y NAHR. The latter process is frequently used to explain rapid evolution of the ampliconic gene families' copy number (40), as well as structural rearrangements (41), some of which lead to spermatogenic failure, sex reversal, and Turner syndrome (42).

Previous studies (e.g., reviewed in refs. 12, 13, 37) focused on the role of Y-Y recombination in preserving Y ampliconic gene families, which are critical for spermatogenesis and fertility (6), and suggested that this phenomenon explains the major adaptive role of palindromic sequences. However, two human palindromes, P6 and P7, do not harbor any known protein-coding genes (6) and are multicopy in most great ape species that we examined (Fig. 3A and *SI Appendix, Table S7*). We hypothesize that conservation of these palindromes is driven not by

spermatogenesis-related genes, but by elements regulating gene expression (*SI Appendix, Fig. S7*). Indeed, by analyzing ENCODE (43) datasets (*SI Appendix, Supplemental Methods*), we found candidate open-chromatin and protein-binding sites in P6 and P7 (*SI Appendix, Fig. S7*). Interestingly, these sites were found in tissues other than testis, suggesting that they regulate expression of genes outside of the Y chromosome and echoing findings in *Drosophila* and mouse Y chromosomes (44, 45). Note that our observations should be considered preliminary because of the limitations (e.g., low read mappability) of studying regulatory elements in repetitive (in this case palindromic) regions and should be confirmed in future studies.

**Gene Content Evolution.** We inferred that the gene content in the common ancestor of great apes likely was the same as is currently found in gorilla and included eight ampliconic and 16 X-degenerate genes (Fig. 2). Analyzing the data on the evolution of ampliconic gene content (Fig. 2), palindrome sequence (Fig. 3B), and ampliconic gene copy number (fig. 2 in ref. 21) jointly, we can infer which



**Fig. 4.** Chromatin contacts on the human and chimpanzee Y chromosomes, as evaluated from iPSCs. (A) Human Y chromosome contacts with palindromes (highlighted in light blue), pseudoautosomal regions (green), and centromere (red). The schematic representation of the sequencing classes on the Y chromosome is adapted from ref. 6. (B) Chimpanzee Y chromosome contacts with palindromes (highlighted in light blue). The schematic representation of the sequencing classes on the Y chromosome is adapted from ref. 7. (C) Chromatin interactions for the three largest palindromes on the human Y (the next three largest palindromes are displayed in [SI Appendix, Fig. S10](#)). To resolve ambiguity due to multimapping reads, each interaction was assigned a probability based on the fraction of reads supporting it (see [SI Appendix, Supplemental Methods](#), for details). Palindrome arms are shown as blue arrows, and the spacer is shown as white space between them.

ampliconic genes were present in the multicopy state in the great ape common ancestor. Our results suggest that such an ancestor had multicopy sequences homologous to P1, P2, P4, P5, and P8 (Fig. 3B), which carry *DAZ*, *BPY2*, *CDY*, *HSFY*, *XKRY*, and *VCY* on the human Y (5). Except for *VCY*, which was likely acquired by the human–chimpanzee common ancestor ([SI Appendix, Supplemental Note S7](#)), the remaining five genes were presumably present as multicopy gene families in the common ancestor of great apes, because three of them—*DAZ*, *BPY2*, and *CDY*—are present as multicopy in all great ape species (21), and the other two—*HSFY* and *XKRY*—are present in all great ape species but chimpanzee and bonobo (21), in which they were lost (Fig. 2). The macaque Y ampliconic region has the *HSFY* and *CDY* gene families located in palindromes that are homologous to human P4 and P5, respectively (22), providing further evidence of their ancient origins. Additionally, Bhowmick and colleagues (46) argued that major expansions of *CDY*, *HSFY*, *TSPY*, and *XKRY* families had already occurred in the common ancestor of Old World monkeys and apes. The *RBMV* gene family was likely present in a single-copy state in the common ancestor of great apes on palindrome P3 (the palindrome P3 sequence is present in a single copy in orangutan) (Fig. 3A). However, in human, some *RBMV* copies are located outside of P3 in inverted repeat 2 (IR2) (6), which implies that this gene family was expanding in part independently of P3. Bhowmick and colleagues (46) suggested that the divergence between *RBMV* copies located in P3 and IR2 occurred in the common ancestor of Old World monkeys and apes.

We discovered that there is only one gene family that was born across the entire great ape Y phylogeny: *VCY* was acquired by the common ancestor of human and chimpanzee ([SI Appendix, Supplemental Note S7](#)). As a result, except for this branch, we

found uniformly low rates of gene birth. A low rate of ampliconic gene birth contradicts predictions of high birth rate made in previous studies for such genes (37), but suggests that great ape radiation does not provide sufficient time for gene acquisition by ampliconic regions. Ampliconic regions on the Y chromosomes of several other mammals acquired such genes (47, 48); however, the timing of such acquisitions is unknown.

We expected to observe a high death rate for X-degenerate genes, but a low death rate for ampliconic genes, because the former genes do not undergo Y–Y gene conversion and thus should accumulate deleterious mutations, whereas the latter genes are multicopy and can be rescued by Y–Y gene conversion. Unexpectedly, the rates of gene death were similar between ampliconic and X-degenerate genes. Indeed, across the great ape Y phylogenetic tree, ~44.4% of ampliconic gene families were either deleted or pseudogenized, as compared with ~43.8% of X-degenerate genes. While our data did not support our hypothesis, other findings suggest that death of ampliconic genes is a gradual process. Indeed, ampliconic gene families dead in some great ape species have reduced copy number in other species (21, 40), lowering the chances for Y–Y gene conversion.

The rates of gene death varied among great ape species. In particular, we observed high rates of death in the lineages of bonobo, chimpanzee, and their common ancestor. What could be the evolutionary forces driving such a high rate of gene death, likely operating in the *Pan* lineage continuously since its divergence from the human lineage? First, gene-disrupting or gene deletion mutations could be hitchhiking in haplotypes with beneficial mutations. Positive selection might be acting in the *Pan* lineage due to sperm competition (49). No gene deaths in the human and gorilla lineages, experiencing no sperm competition,

and low gene death rates in orangutan, experiencing limited sperm competition, are consistent with this explanation. Our tests for positive selection acting at protein-coding genes did not produce significant results, although they indicated significantly elevated nonsynonymous-to-synonymous rate ratios for five genes in bonobo, chimpanzee, and/or their common ancestor (*SI Appendix, Supplemental Note S4*). We might have limited power to detect positive selection from phylogenetic data collected for closely related species. Second, the *Pan* Y could have undergone stronger drift leading to fixation of variants lacking genes that were already in the process of becoming nonessential. At first sight, the existing data contradict this explanation, as nucleotide diversity for the chimpanzee and bonobo Y chromosomes was found to be high (50) and in fact higher than that for the gorilla and orangutan Ys (51). However, only a small number of orangutan and gorilla individuals were examined in the latter study (51), and thus this conclusion needs to be reevaluated in the future.

**Future Directions.** Future studies should include sequencing of the Y chromosome for a substantial number of individuals per species, and such data are expected to provide a resolution between the evolutionary scenarios driving high substitution and gene death rates in the *Pan* lineage. Future investigations should also focus on deciphering the sequences of different copies and isoforms of ampliconic genes (52), which should allow one to examine natural selection potentially operating on those genes in more detail. Because chromatin organization depends on the tissue of origin (26), the high prevalence of intra-ampliconic contacts that we found in two somatic tissues should be confirmed in testis and sperm. Additionally, comparing chromatin organization and evolution of palindromes in the Y vs. X chromosomes should aid in understanding the unique role that repetitive regions might play on the Y.

From a more applied perspective, the bonobo and orangutan Y assemblies presented here are useful for developing genetic markers to track male dispersal in these endangered species. This is of utmost importance because both species experience population decreases due to habitat loss. Therefore, our results are expected to be of great utility to conservation genetics efforts aimed at restoring these populations.

## Materials and Methods

See *SI Appendix, Supplemental Methods*, for details.

**Assemblies and Alignments.** For bonobo and Sumatran orangutan, we generated and assembled (53) deep-coverage short sequencing reads from male individuals and identified putative Y contigs by mapping them against the

corresponding female reference assemblies (54). These contigs were then scaffolded with mate-pair reads (55). The orangutan Y assembly was further improved by merging (56) with another high-quality assembly generated with 10× Genomics technology (57). The bonobo Y assembly was improved by additional scaffolding with long Y-enriched Pacific Biosciences reads (58, 59). We improved the continuity of the gorilla Y assembly by merging two previously published assemblies (8, 60). To remove PARs, we filtered each species-specific Y assembly against the corresponding female reference genome. Great ape Y assemblies were aligned with PROGRESSIVECTACTUS (61). Substitution rates were estimated for alignment blocks containing all five species with the GTR model (62) implemented in PHYLOFIT (63).

**Gene Content Analysis.** To retrieve the bonobo and orangutan genes, we aligned the scaffolds from their Y chromosome assemblies to the respective species-specific or closest-species-specific reference coding sequences using BWA-MEM (64). Novel gene predictions were evaluated with AUGUSTUS (65). The evolutionary history of Y-gene content and gene birth and death rates was reconstructed using procedures in ref. 23.

**Palindrome Analysis.** To analyze conservation of human and chimpanzee palindromes, we found all multispecies alignment blocks that overlap their coordinates and identified the percentage of nonrepetitive bases in such blocks per species. To evaluate the copy number of sequences homologous to human and chimpanzee palindromes in bonobo, gorilla, and orangutan, we mapped whole-genome male sequencing reads to the corresponding 1-kb windows which overlap intervals of the human and chimpanzee Y palindromes using BWA-MEM (64) and compared their read depth with that of single-copy X-degenerate genes. The copy number of species-specific ampliconic sequences in bonobo and orangutan were evaluated similarly. Regulatory factor-binding sites in human palindromes P6 and P7 were extracted from ENCODE (43). To analyze a potential enrichment of ampliconic interactions, Hi-C data (25, 26) were processed with mHi-C (24).

**Data Availability.** New sequencing data and assemblies are available under BioProject accession no. [PRJNA602326](https://doi.org/10.26207/9han-5s18) (66). Multiple sequence alignments of great ape Y chromosomes and of great ape X and Y chromosomes are available at ScholarSphere, <https://doi.org/10.26207/9han-5s18> (67). Code is available at GitHub, <https://github.com/makovalab-psu/great-ape-Y-evolution>.

**ACKNOWLEDGMENTS.** We thank L. Carrel, O. Ryder, M. Ferguson-Smith, S. Warris, K. Sahlin, F. Chiaromonte, A. Kenney, San Diego Zoo Institute for Conservation Research, and the Smithsonian Institution for their assistance. This research was supported by NIH Grant R01GM130691 (to K.D.M.) and NSF Grants DBI-1356529, IIS-1453527, and CCF-1439057 (to P.M.). Support was also provided by the Clinical and Translational Sciences Institute, the Institute of Computational and Data Sciences, the Huck Institutes of the Life Sciences, and the Eberly College of Science of Penn State University. Finally, this research was supported by the Computation, Bioinformatics, and Statistics (CBIOS) Predoctoral Training Program awarded to Penn State by the NIH (M.C. and R.V. are trainees). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

- P. Berta *et al.*, Genetic evidence equating SRY and the testis-determining factor. *Nature* **348**, 448–450 (1990).
- V. S. Vineeth, S. S. Malini, A journey on Y chromosomal genes and male infertility. *Int. J. Hum. Genet.* **11**, 203–215 (2011).
- M. I. Douadi *et al.*, Sex-biased dispersal in western lowland gorillas (Gorilla gorilla gorilla). *Mol. Ecol.* **16**, 2247–2259 (2007).
- C. Roos *et al.*, Nuclear versus mitochondrial DNA: Evidence for hybridization in colobine monkeys. *BMC Evol. Biol.* **11**, 77 (2011).
- M. Tomaszewicz, P. Medvedev, K. D. Makova, Y and W chromosome assemblies: Approaches and discoveries. *Trends Genet.* **33**, 266–282 (2017).
- H. Skalaetsky *et al.*, The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
- J. F. Hughes *et al.*, Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463**, 536–539 (2010).
- M. Tomaszewicz *et al.*, A time- and cost-effective strategy to sequence mammalian Y chromosomes: An application to the de novo assembly of gorilla Y. *Genome Res.* **26**, 530–540 (2016).
- G. V. Glazko, M. Nei, Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* **20**, 424–434 (2003).
- Z. H. Rosser, P. Balaresque, M. A. Jobling, Gene conversion between the X chromosome and the male-specific region of the Y chromosome at a translocation hotspot. *Am. J. Hum. Genet.* **85**, 130–134 (2009).
- B. Trombetta, E. D'Atanasio, F. Cruciani, Patterns of inter-chromosomal gene conversion on the male-specific region of the human Y chromosome. *Front. Genet.* **8**, 54 (2017).
- S. Rozen *et al.*, Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**, 873–876 (2003).
- E. Betrán, J. P. Demuth, A. Williford, Why chromosome palindromes? *Int. J. Evol. Biol.* **2012**, 207958 (2012).
- M. Cechova *et al.*, High satellite repeat turnover in great apes studied with short- and long-read technologies. *Mol. Biol. Evol.* **36**, 2415–2431 (2019).
- D. W. Bellott *et al.*, Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494–499 (2014).
- P. Hallast, M. A. Jobling, The Y chromosomes of the great apes. *Hum. Genet.* **136**, 511–528 (2017).
- J. Hey, The divergence of chimpanzee species and subspecies as revealed in multi-population isolation-with-migration analyses. *Mol. Biol. Evol.* **27**, 921–933 (2010).
- B. Gläser *et al.*, Simian Y chromosomes: Species-specific rearrangements of DAZ, RBM, and TSPY versus contiguity of PAR and SRY. *Mamm. Genome* **9**, 226–231 (1998).
- M. A. Wilson Sayres, K. D. Makova, Genome analyses substantiate male mutation bias in many species. *BioEssays* **33**, 938–945 (2011).
- P. Moorjani, C. E. G. Amorim, P. F. Arndt, M. Przeworski, Variation in the molecular clock of primates. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10607–10612 (2016).
- R. Vegesna *et al.*, Ampliconic genes on the great ape Y chromosomes: Rapid evolution of copy number but conservation of expression levels. *Genome Biol. Evol.* **12**, 842–859 (2020).
- J. F. Hughes *et al.*, Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* **483**, 82–86 (2012).
- W. Iwasaki, T. Takagi, Reconstruction of highly heterogeneous gene-content evolution across the three domains of life. *Bioinformatics* **23**, i230–i239 (2007).

24. Y. Zheng, F. Ay, S. Keles, Generative modeling of multi-mapping reads with mHi-C advances analysis of Hi-C studies. *eLife* **8**, e38070 (2019).
25. I. E. Eres, K. Luo, C. J. Hsiao, L. E. Blake, Y. Gilad, Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates. *PLoS Genet.* **15**, e1008278 (2019).
26. S. S. P. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
27. B. R. Lajoie, J. Dekker, N. Kaplan, The Hitchhiker's guide to Hi-C analysis: Practical guidelines. *Methods* **72**, 65–75 (2015).
28. K. D. Makova, W.-H. Li, Strong male-driven evolution of DNA sequences in humans and apes. *Nature* **416**, 624–626 (2002).
29. N. Elango, J. W. Thomas, S. V. Yi; NISC Comparative Sequencing Program, Variable molecular clocks in hominoids. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1370–1375 (2006).
30. M. J. Anderson, A. F. Dixon, Sperm competition: Motility and the midpiece in primates. *Nature* **416**, 496 (2002).
31. E. B. Smithwick, L. G. Young, K. G. Gould, Duration of spermatogenesis and relative frequency of each stage in the seminiferous epithelial cycle of the chimpanzee. *Tissue Cell* **28**, 357–366 (1996).
32. C. G. Heller, Y. Clermont, Spermatogenesis in man: An estimate of its duration. *Science* **140**, 184–186 (1963).
33. G. Amster, G. Sella, Life history effects on the molecular clock of autosomes and sex chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1588–1593 (2016).
34. T. Marques-Bonet, O. A. Ryder, E. E. Eichler, Sequencing primate genomes: What have we learned? *Annu. Rev. Genomics Hum. Genet.* **10**, 355–386 (2009).
35. O. Venn *et al.*, Nonhuman genetics. Strong male bias drives germline mutation in chimpanzees. *Science* **344**, 1272–1275 (2014).
36. S. Besenbacher, C. Hvilsom, T. Marques-Bonet, T. Mailund, M. H. Schierup, Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nat. Ecol. Evol.* **3**, 286–292 (2019).
37. B. Trombetta, F. Cruciani, Y chromosome palindromes and gene conversion. *Hum. Genet.* **136**, 605–619 (2017).
38. G. A. B. Marais, P. R. A. Campos, I. Gordo, Can intra-Y gene conversion oppose the degeneration of the human Y chromosome? A simulation study. *Genome Biol. Evol.* **2**, 347–357 (2010).
39. T. Connallon, A. G. Clark, Gene duplication, gene conversion and the evolution of the Y chromosome. *Genetics* **186**, 277–286 (2010).
40. R. Vegesna, M. Tomaszewicz, P. Medvedev, K. D. Makova, Dosage regulation, and variation in gene expression and copy number of human Y chromosome ampliconic genes. *PLoS Genet.* **15**, e1008369 (2019).
41. L. Skov, M. H. Schierup; Danish Pan Genome Consortium, Analysis of 62 hybrid assembled human Y chromosomes exposes rapid structural changes and high rates of gene conversion. *PLoS Genet.* **13**, e1006834 (2017).
42. J. Lange *et al.*, Isodicentric Y chromosomes and sex disorders as byproducts of homologous recombination that maintains palindromes. *Cell* **138**, 855–869 (2009).
43. C. A. Davis *et al.*, The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).
44. B. Lemos, A. T. Branco, D. L. Hartl, Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15826–15831 (2010).
45. S. Kaufmann *et al.*, Inter-chromosomal contact networks provide insights into mammalian chromatin organization. *PLoS One* **10**, e0126125 (2015).
46. B. K. Bhowmick, Y. Satta, N. Takahata, The origin and evolution of human ampliconic gene families and ampliconic structure. *Genome Res.* **17**, 441–450 (2007).
47. Y. Q. S. Soh *et al.*, Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell* **159**, 800–813 (2014).
48. T.-C. Chang, Y. Yang, E. F. Retzel, W.-S. Liu, Male-specific region of the bovine Y chromosome is gene rich with a high transcriptomic activity in testis development. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12373–12378 (2013).
49. J. F. Hughes *et al.*, Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* **437**, 100–103 (2005).
50. A. C. Stone, R. C. Griffiths, S. L. Zegura, M. F. Hammer, High levels of Y-chromosome nucleotide diversity in the genus Pan. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 43–48 (2002).
51. P. Hallast *et al.*, Great ape Y Chromosome and mitochondrial DNA phylogenies reflect subspecies structure and patterns of mating and dispersal. *Genome Res.* **26**, 427–439 (2016).
52. K. Sahlin, M. Tomaszewicz, K. D. Makova, P. Medvedev, Deciphering highly similar multigene family transcripts from Iso-Seq data with IsoCon. *Nat. Commun.* **9**, 4601 (2018).
53. N. I. Weisenfeld *et al.*, Comprehensive variation discovery in single human genomes. *Nat. Genet.* **46**, 1350–1355 (2014).
54. S. Rangavittal *et al.*, DiscoverY: A classifier for identifying Y chromosome sequences in male assemblies. *BMC Genomics* **20**, 641 (2019).
55. K. Sahlin, F. Vezzi, B. Nystedt, J. Lundeberg, L. Arvestad, BESST: Efficient scaffolding of large fragmented assemblies. *BMC Bioinformatics* **15**, 281 (2014).
56. A. H. Wences, M. C. Schatz, Metassembler: Merging and optimizing de novo genome assemblies. *Genome Biol.* **16**, 207 (2015).
57. N. I. Weisenfeld, V. Kumar, P. Shah, D. M. Church, D. B. Jaffe, Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).
58. M. Boetzer, W. Pirovano, SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* **15**, 211 (2014).
59. A. C. English *et al.*, Mind the gap: Upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* **7**, e47768 (2012).
60. S. Warris *et al.*, Correcting palindromes in long reads after whole-genome amplification. *BMC Genomics* **19**, 798 (2018).
61. B. Paten *et al.*, Cactus: Algorithms for genome multiple sequence alignment. *Genome Res.* **21**, 1512–1528 (2011).
62. S. Tavaré, Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* **17**, 57–86 (1986).
63. A. Siepel, D. Haussler, Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol. Biol. Evol.* **21**, 468–488 (2004).
64. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997v2* (26 May 2013).
65. M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19** (suppl. 2), ii215–ii225 (2003).
66. M. Cechova *et al.*, Next-generation sequencing and Y assemblies of bonobo, orangutan, and gorilla using whole-genome and Y flow-sorted reads. *Sequence Read Archive*. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA602326/>. Deposited 29 January 2020.
67. M. Cechova *et al.*, Multiple sequence alignments of great ape Y chromosomes and of great ape X and Y chromosomes. *ScholarSphere*. <https://doi.org/10.26207/9han-5a18>. Deposited 18 September 2020.