

RESEARCH

Open Access



# Evolutionary dynamics of pseudoautosomal region 1 in humans and great apes

Juraj Bergman\*  and Mikkel Heide Schierup

\*Correspondence:  
jurajbergman@birc.au.dk

Bioinformatics Research  
Centre, Aarhus University,  
DK-8000 Aarhus C, Denmark

## Abstract

**Background:** The pseudoautosomal region 1 (PAR1) is a 2.7 Mb telomeric region of human sex chromosomes. PAR1 has a crucial role in ensuring proper segregation of sex chromosomes during male meiosis, exposing it to extreme recombination and mutation processes. We investigate PAR1 evolution using population genomic datasets of extant humans, eight populations of great apes, and two archaic human genome sequences.

**Results:** We find that PAR1 is fast evolving and closer to evolutionary nucleotide equilibrium than autosomal telomeres. We detect a difference between substitution patterns and extant diversity in PAR1, mainly driven by the conflict between strong mutation and recombination-associated fixation bias at CpG sites. We detect excess C-to-G mutations in PAR1 of all great apes, specific to the mutagenic effect of male recombination. Despite recent evidence for Y chromosome introgression from humans into Neanderthals, we find that the Neanderthal PAR1 retained similarity to the Denisovan sequence. We find differences between substitution spectra of these archaics suggesting rapid evolution of PAR1 in recent hominin history. Frequency analysis of alleles segregating in females and males provided no evidence for recent sexual antagonism in this region. We study repeat content and double-strand break hotspot regions in PAR1 and find that they may play roles in ensuring the obligate X-Y recombination event during male meiosis.

**Conclusions:** Our study provides an unprecedented quantification of population genetic forces governing PAR1 biology across extant and extinct hominids. PAR1 evolutionary dynamics are predominantly governed by recombination processes with a strong impact on mutation patterns across all species.

**Keywords:** Pseudoautosomal region, Comparative genomics, Nucleotide composition evolution, Meiotic recombination, Hominid evolution



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Introduction

The mammalian sex chromosomes, X and Y, originated from a pair of autosomal precursors around 180 million years ago [1]. At least four subsequent recombination suppression events and loss of sequence homology in the sex-determining region (SDR) have resulted in extreme divergence of sequence and function between the sex chromosomes [2–6]. Yet, regions with X-Y homology and genetic exchange, termed the pseudoautosomal regions (PARs), still persist across placental mammals. In great apes, the pseudoautosomal region 1 (PAR1) is just 2.7 Mb in length but has the important role of ensuring proper meiotic segregation of X and Y. PAR1 is also one of the most evolutionarily dynamic regions of the genome. Obligate male crossovers are restricted to this physically small region during male meiosis, so recombination rates per base pair are extremely high, and the region shows high nucleotide diversity. Recent progress in genome assembly of the sex chromosomes and the availability of population genomic datasets have now made it possible to study divergence and diversity processes of this important region in detail.

The importance of pseudoautosomal regions is evident in the association of PAR1-specific mutations with various phenotypic consequences in humans. Since homologous pairing and exchange of genetic material is crucial for successful gametogenesis (at least one cross-over per chromosome pair is necessary for proper segregation), PAR1 is especially important for male fertility [7, 8]. Additionally, polymorphisms in PAR1 are associated with various diseases, such as skeletal malformations due to variants in the *SHOX* gene [9, 10], schizophrenia [11], bipolar disorder [12], and hematological malignancies [13]. Large genomic rearrangements in PAR1 have also been reported—examples include an approximately 300 kb deletion associated with acute lymphoblastic leukemia, spanning several genes [14, 15], a 47.5 kb deletion in the enhancer region of the *SHOX* gene [16], and an extension of the PAR1 region on the Y chromosome through a translocation of a 110 kb region from the X chromosome [17].

Sequence evolution of the pseudoautosomal region after the split between the avian and mammalian lineages, and leading up to extant mammalian species, involved the formation of several evolutionary strata mediated by recombination suppression between the sex chromosomes [18–20]. Consequently, PAR1 of humans and great apes is a small genomic region evolving under a concentration of strong population genetic forces. Such a convergence of forces results in PAR1-specific patterns of mutations, nucleotide composition evolution and recombination that have been mostly studied for large human diversity datasets [21–23]. Here, we aim to provide a broader perspective on PAR1 evolution by including nucleotide divergence and diversity data from eight populations of great apes, as well as ancient hominin data. We quantify divergence rates, and deviations of sequence composition from nucleotide equilibrium, and infer substitution and diversity spectra. For comparison, we use autosomal telomeres, which are similar to PAR1 in having higher male recombination rates and GC content, typical of male-specific recombination processes [24, 25].

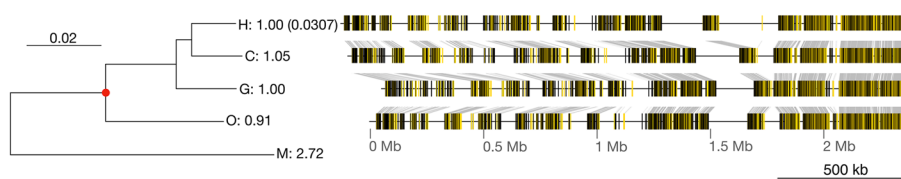
A central focus of our study is the relationship between recombination and nucleotide diversity in PAR1. Recombination in PAR1 has been subject to several studies, including sperm-typing, pedigree-based and sequence-based approaches. The general conclusion is that recombination in PAR1 during spermatogenesis occurs at a rate approximately

17-fold higher than the genome-wide average (1.2 cM/Mb), while the difference between male and female rates in the region is around 10-fold [6, 12, 21]. Recombination rate differences can affect sequence diversity in several ways. Firstly, recombination counters the effect of directional selection on linked diversity. A strong reduction of nucleotide diversity during selective sweeps is expected in low recombining regions [26, 27], and similarly, reduction in diversity due to background selection is diminished with increasing recombination rate [28]. Additionally, recombination directly affects diversity due to its mutagenicity [22, 29], which is likely an important contributor to PAR1 evolution [30–32]. The type of nucleotide diversity caused by the recombination process may also differ from the diversity introduced into the genome through DNA replication errors [22]. Recombination is also associated with GC-biased gene conversion, whereby purine-pyrimidine mismatches that arise during meiotic chromosomal pairing are most often resolved into GC, rather than AT pairs [33–35]. This effect on nucleotide diversity will be strongest in highly recombining genome regions, and is indistinguishable from directional selection favoring GC content [36].

Polymorphisms in the pseudoautosomal regions may also be maintained due to balancing and/or sexually antagonistic (SA) selection. Theory predicts that the recombination rate within the SDR will determine the conditions under which SA polymorphisms can persist [6, 37, 38]. In general, a sexually antagonistic variant that is tightly linked to the SDR can be maintained in a population under a broader range of selection coefficients compared to a freely recombining locus. Furthermore, as the recombination rate in PAR1 decreases with the distance from the telomere, PAR1 loci that are potentially under SA selection are therefore more likely to be found closer to the pseudoautosomal boundary. Additionally, a peak of diversity is expected around a PAR locus under balancing or SA selection [39]. Potential SA loci have been identified in various animal species using population genomic data to calculate  $F_{st}$  values between females and males [40–44], and sexual antagonism specific to the PAR has been suggested for the plant *Silene latifolia* [45, 46]. A recent study of sexual antagonism in human pseudoautosomal regions reached equivocal conclusions [23], and while SA in PAR1 of other great apes has not been directly investigated, a recent study on the location of the pseudoautosomal boundary argues for stronger SA acting in great apes compared to strepsirrhine primates [47].

We also investigate the evolutionary dynamics of PAR1 during recent hominin evolution. This topic is of special interest as hybridization events between ancient humans resulted in peculiar patterns of sex chromosome evolution. On the one hand, the X chromosome is devoid of ancient DNA introgression tracks in modern humans [48, 49], while the human Y chromosome invaded and replaced the Neanderthal Y lineage [50]. We therefore use high-coverage sequences of a Neanderthal and Denisovan individual [51, 52] to study PAR1 evolution during recent hominin speciation.

Lastly, we study the factors that ensure the occurrence of a recombination event in pseudoautosomal regions. Studies in mice have shown that repeat elements play a pivotal role in determining the physical structure of the PAR during recombination [53–55]. We therefore did a comparative analysis of repeat elements between PAR1 and autosomal telomeres. Another important determinant of PAR recombination is the recombination motif-recognition protein PRDM9, which is known to facilitate meiotic double-strand



**Fig. 1** PAR1 phylogenetic tree and sequence alignment between great ape species. The total alignment consists of 620,054 syntenic sites of which 239,799 are putatively neutral (highlighted in yellow). As indicated by letters in the tree diagram, the species are as follows, from top to bottom: human (H), chimpanzee (C), gorilla (G), orangutan (O) and macaque (M). The number associated with each branch label is the divergence rate ratio with respect to the human rate (presented in parentheses), since the branching of the great ape ancestor (red point)

**Table 1** Divergence estimates based on phyloFit estimation and divergence times from [62] for the PAR1 sequence and the concatenated sequence of autosomal telomeres

	PAR1		Autosomal telomeres	
	Total divergence	Per site and year divergence	Total divergence	Per site and year divergence
Human	0.0105	$0.9652 \times 10^{-9}$	0.0082	$0.7543 \times 10^{-9}$
Chimpanzee	0.0120	$1.1023 \times 10^{-9}$	0.0081	$0.7483 \times 10^{-9}$
HC ancestor	0.0036	$1.7496 \times 10^{-9}$	0.0027	$1.3179 \times 10^{-9}$
Gorilla	0.0143	$1.1007 \times 10^{-9}$	0.0108	$0.8300 \times 10^{-9}$
HCG ancestor	0.0166	$1.5248 \times 10^{-9}$	0.0120	$1.1047 \times 10^{-9}$
Orangutan	0.0279	$1.1696 \times 10^{-9}$	0.0210	$0.8819 \times 10^{-9}$

breaks and operate in human pseudoautosomal regions [56], as well as to evolve rapidly between primate lineages [57–59]. We therefore explored sequence evolution of double-strand break hotspots in PAR1 to gain insight into the role of PRDM9 in PAR1 sequence evolution across great apes.

## Results

### PAR1 divergence is exceptionally high across the great ape phylogeny

We used the Progressive Cactus aligner [60] to produce an alignment between the human, chimpanzee, gorilla, orangutan, and macaque PAR1 sequences. The macaque sequence was used as an outgroup to root the tree and for estimating divergence rates within the great ape phylogeny since the branching of the great ape ancestor. The estimated phylogenetic tree and alignment are shown in Fig. 1. For tree estimation, we excluded sites in coding and repetitive regions, as well as CpG islands and conserved sites across primates. This left us with an alignment of 239,799 putatively neutral sites across the great ape phylogeny. Since the branching of the great ape ancestor into extant species, PAR1 appears to have evolved at different rates (Fig. 1; Table 1). The divergence rate in the chimpanzee lineage is approximately 5% larger compared to the PAR1 sequence of humans and gorillas, while the orangutan PAR1 sequence is the least diverged with a 10–16% lower divergence compared to the other species. Since the human-chimpanzee split, the divergence along the chimpanzee lineage is 14% higher than the human rate. Nevertheless, PAR1 sequences evolve faster than autosomal ones; pairwise divergence estimates of autosomal sequences for the human-chimp (0.0137),

human-gorilla (0.0175), and human-orangutan species pairs (0.034) [61] are on average 40% lower than our PAR1 estimates.

Based on great ape divergence times [62], we estimate that the PAR1 divergence rate is within the range of  $0.97\text{--}1.17 \times 10^{-9}$  substitutions per site and year for the terminal branches of the phylogeny (Table 1). These rates are 1.51–1.83 fold higher than the yearly mutation rate for non-human great apes, estimated using trio sequencing ( $0.64 \times 10^{-9}$  mutations per site and year [62]). While it has been suggested that this mutation rate estimate is adequate for explaining autosomal divergence rates, the magnitude of PAR1 divergence implies additional mutational processes acting in this region. For comparison, we estimated the substitution rate for autosomal telomeric sequences (defined as 3 Mb syntenic regions that are present at the tips of autosomes in great ape species and the macaque outgroup). We concatenated sequences from all autosomal telomeres, in order to obtain a single estimate for each branch of the phylogeny. These sequences have a 1.16–1.38 fold higher divergence rate than the trio-based mutation rate, indicating that higher divergence is a general feature of telomeres, likely due to a combination of higher substitution rates and larger ancestral polymorphism caused by a higher local effective population size of telomeres due to a smaller effect of linked selection. However, even compared to autosomal telomeres, PAR1 has a 1.28–1.47 fold higher divergence rate across all branches of the phylogeny (Table 1), implying an exceptionally high rate compared to other fast-evolving regions.

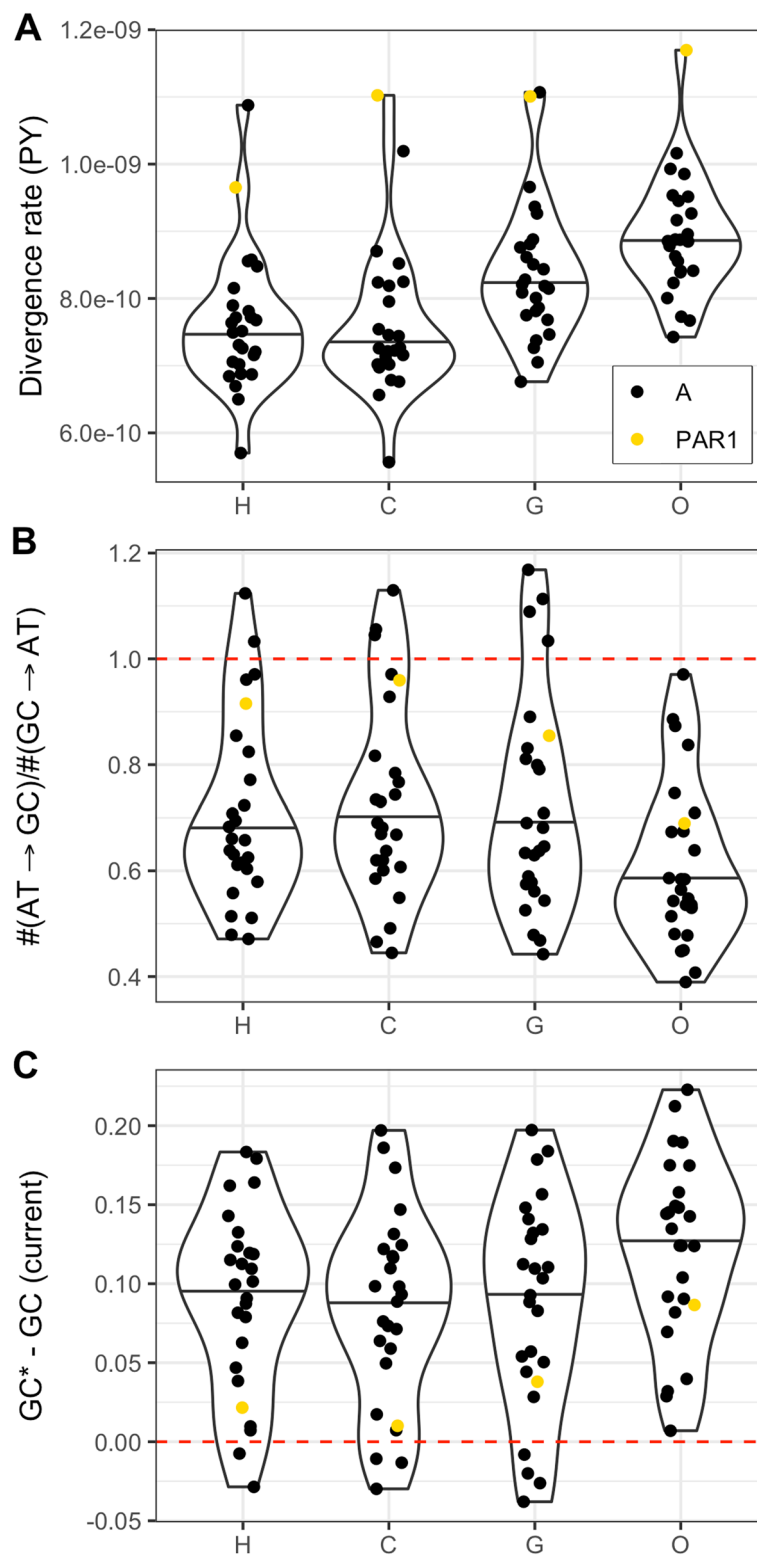
#### Nucleotide composition in PAR1 is closer to equilibrium than for other telomeres

We next consider PAR1 divergence as a function of derived nucleotide state and substitution type. For comparison, and to obtain enough data, we computed divergence rates for each telomere separately. Specifically, we studied GC content evolution by comparing counts of AT→GC and GC→AT substitutions, and rates inferred from the posterior mean substitution counts between the corresponding nucleotides. These counts were inferred using the phyloFit program [63].

In line with the results in Table 1, the divergence rates of PAR1 in the four branches studied are on average higher than those for autosomal sequences, even when inferring divergence rates for each telomere separately (Fig. 2A). Although divergence rates of some autosomal telomeres overlap with PAR1 rates, the PAR1 rates are significantly higher (Wilcoxon  $W = 9$ ,  $p = 0.0013$ ). Generally, transitions occur more frequently than transversions, with the average ts:tv ratio of approximately 2.4:1 across all telomeres and species, which is somewhat higher than the autosomal estimate of 2.1:1 [64, 65]. For PAR1 alone, the ts:tv is among the lowest of all telomeres (average of 1.92:1), ranging from 1.75:1 in the chimpanzee to 2.04:1 in the orangutan (Additional file 1: Table S1).

(See figure on next page.)

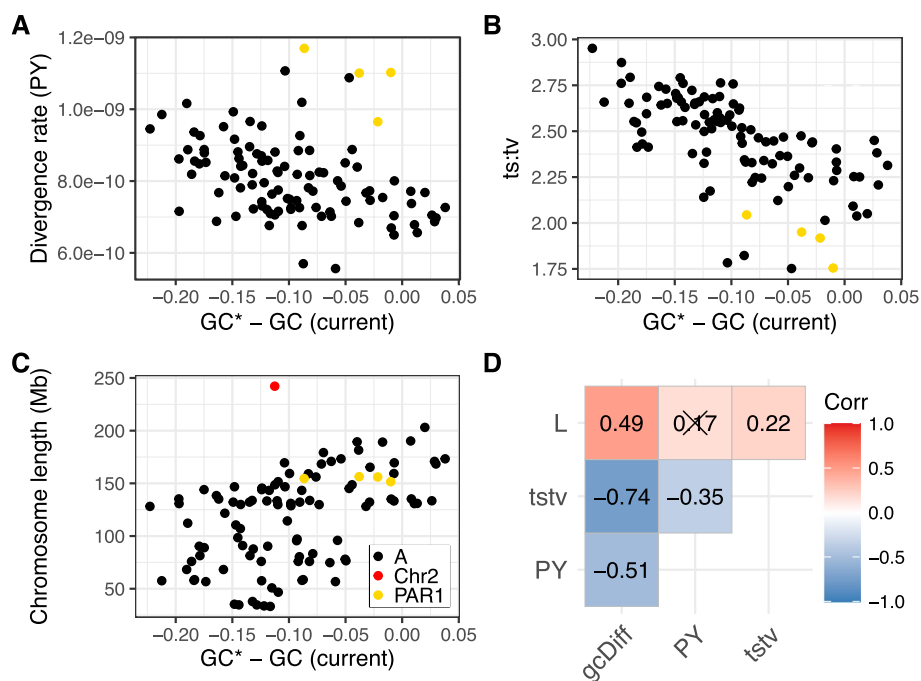
**Fig. 2** **A** Distribution of per site and year (PY) divergence rates for autosomal ( $A$ ;  $N = 25$ ) telomeres and PAR1 in the four external branches of the great ape phylogeny; human (H), chimpanzee (C), gorilla (G), and orangutan (O). Telomere 8L in both the human and gorilla lineage exceeds the PAR1 rate. **B** Distribution of AT→GC/GC→AT substitution count ratios across telomeres. **C** Distribution of the difference between the equilibrium GC proportion ( $GC^*$ ), expected given the inferred AT→GC and GC→AT substitution rates, and current telomeric GC proportion. The red dashed lines indicate the expectation under nucleotide composition equilibrium



**Fig. 2** (See legend on previous page.)

To assess whether telomeric GC content is at equilibrium, we plotted the distribution of the ratio of AT→GC versus GC→AT substitution counts (Fig. 2B). We expect this ratio to be close to 1 if the nucleotide composition of the sequence is close to equilibrium. However, it is below 1 for most telomeres, indicating that telomeric sequences are evolving towards lower GC content, as observed previously for autosomal sequences [66]. However, we observe that average AT→GC/GC→AT count ratios are greater in PAR1 compared to autosomal telomeres (Wilcoxon  $W = 317, p = 0.0245$ ). We next calculate the equilibrium GC content of telomeres ( $GC^*$ ). The  $GC^*$  value was calculated as the per site divergence rate for AT→GC mutations, divided by the sum of AT→GC and GC→AT rates, and it represents the stationary expectation for the proportion of GC nucleotides in a region evolving under the inferred divergence rates [66]. Figure 2C shows that the current PAR1 GC proportion is closer to its  $GC^*$  value compared to the mean of the autosomal telomeres. Furthermore, a bootstrap analysis shows that GC proportions across telomeres are distinct, likely reflecting telomere-specific evolutionary dynamics (Additional file 2: Fig. S1). PAR1 is likely closer to equilibrium both because its GC content has decreased more than for the other telomeres, and because its expected  $GC^*$  is larger.

We next correlate the difference between  $GC^*$  and current GC (*i.e.*, a measure of deviation from nucleotide composition equilibrium) with divergence rate (Fig. 3A), the



**Fig. 3** Correlation of the difference between equilibrium GC proportion ( $GC^*$ ) and current GC proportion with **A** per site and year (PY) divergence rates, **B** the ts:tv ratios, and **C** chromosome lengths (a telomere of the human chromosome 2 is highlighted in red). **D** Spearman's partial correlation coefficients for all pairwise combinations of the three parameters: difference between equilibrium  $GC^*$  proportion and current GC (gcDiff), per site and year divergence rate (PY), the ts:tv ratio (tstv) and chromosome length (L). Crossed-out coefficients are non-significant ( $p > 0.05$ ). For each telomere, we consider four values corresponding to the four external branches of the great ape phylogeny. In total, we consider  $N = 100$  estimates for autosomes (across 25 alignable telomeres) and  $N = 4$  estimates for PAR1

ts:tv ratio (Fig. 3B) and length of the chromosome upon which the telomere is located (Fig. 3C). Partial correlation coefficients between these parameters are presented in Fig. 3D. We observe that telomeres closer to equilibrium generally have lower divergence rates, with PAR1 as an obvious outlier. Similarly, the ts:tv ratio is negatively correlated to GC difference, with PAR1 having some of the lowest ts:tv values. On the other hand, chromosome length is positively correlated with GC difference. This is due to a larger current GC for small chromosomes and not differences in GC\*, as current GC declines with increased chromosome length (Spearman's  $\rho = -0.4483$ ,  $p < 0.0001$ ), but the equilibrium GC\* is not related to chromosome length (Spearman's  $\rho = 0.1436$ ,  $p = 0.1457$ ). These results indicate that smaller chromosomes have been subject to stronger GC accumulation in the past, perhaps due to higher telomeric recombination rate on small chromosomes causing stronger GC-biased gene conversion. The telomere of the human chromosome 2 (red point in Fig. 3C) is an outlier to this trend, as this chromosome is a human-specific fusion and its telomeres likely still retain a stronger deviation from composition equilibrium characteristic to their pre-fusion status as parts of the shorter 2A and 2B chromosomes (still present in other great apes). We conducted a linear regression analysis with GC difference as the response variable and the other three parameters as explanatory variables (Table 2). Every explanatory variable had a significant effect, and together, they explain ~68% of the variance in the deviation of telomeric nucleotide composition from equilibrium.

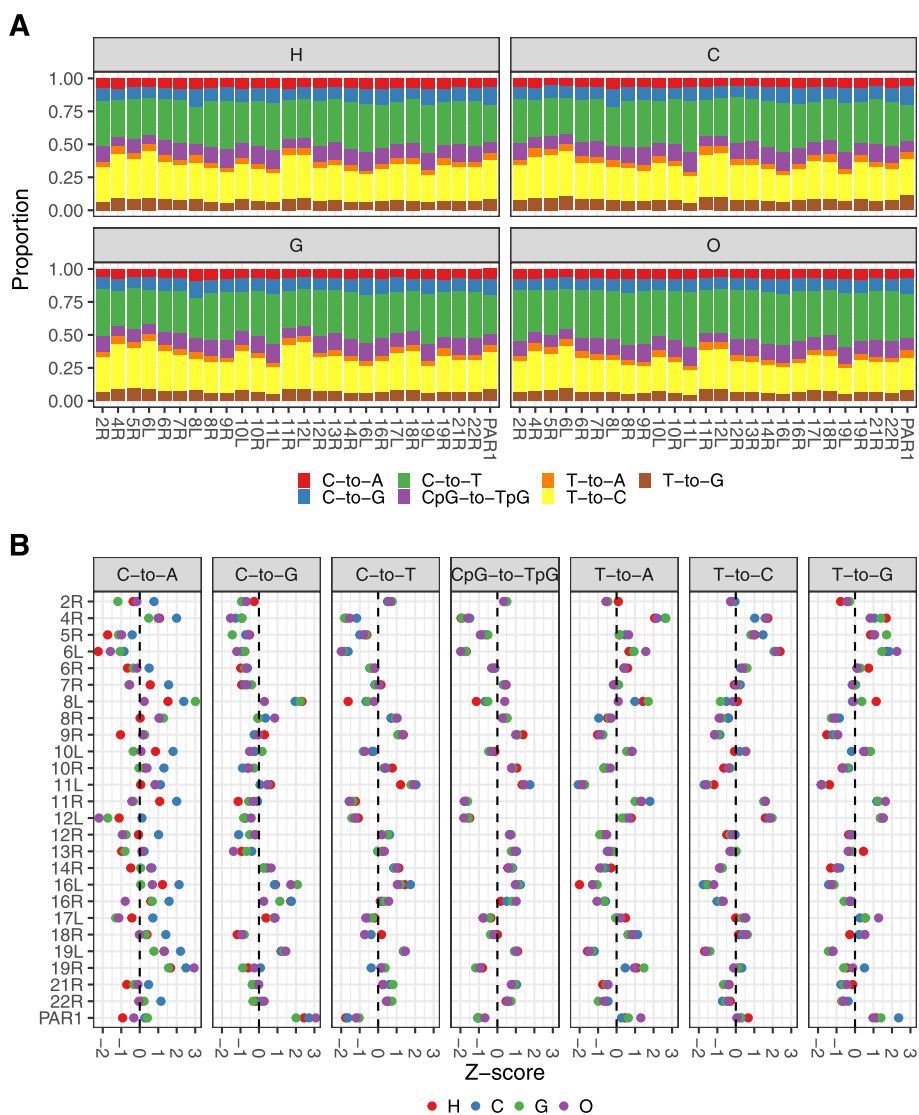
### Substitution spectra of telomeres are similar across great apes

We next divided substitutions into seven classes, including a CpG→TpG class, and calculated the count proportion of each class (*i.e.*, substitution spectrum) for every telomere and great ape species (Fig. 4A). We calculated the difference between substitution proportions and the corresponding mean proportion (specific to a species and substitution class), normalized by the standard deviation (*Z*-score), for all telomeres across all species and substitution classes (Fig. 4B). Interestingly, PAR1 and telomere 8L have a similar pattern of proportion differences, characterized most strongly by an excess of C→G transversions. This similarity between the two telomeres likely stems from two distinct sources, as the C→G mutagenic signature has been associated with male meiotic double-strand breaks on the X chromosome in humans [22], while telomere 8L has been shown to be a female de novo mutation (DNM) hotspot, with a maternal C→G

**Table 2** Linear regression model with the response variable as the deviation from nucleotide composition equilibrium (GC\*-GC) and explanatory variables as the per site and year divergence rate, the transition to transversion ratio and chromosome length (GC\*-GC ~ div. rate + ts:tv + length). For each telomere, we consider four values corresponding to the four external branches of the great ape phylogeny. In total, we consider  $N = 100$  estimates for autosomes (across 25 alignable telomeres) and  $N = 4$  estimates for PAR1

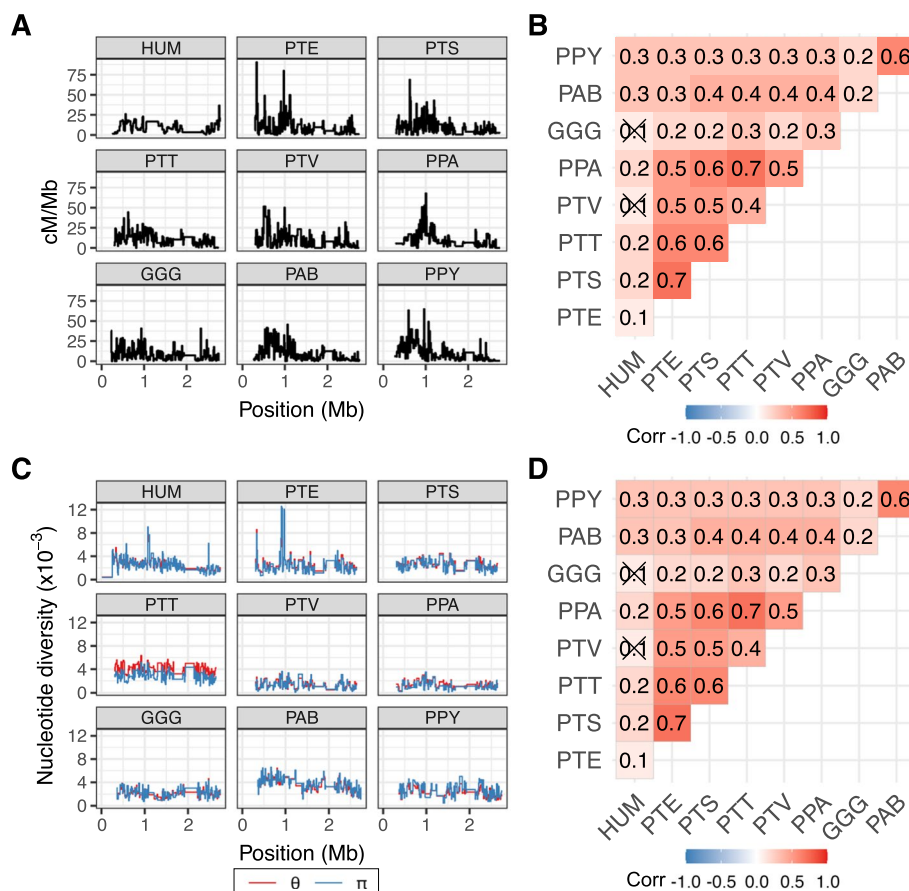
	Estimate	Standard error	P-value
(Intercept)	0.4309	0.0506	$1.7317 \times 10^{-13}$
div. rate	-223.2882	30.6623	$7.6641 \times 10^{-11}$
ts:tv	-0.1611	0.0145	$4.1528 \times 10^{-19}$
length (Mb)	0.0004	$8.0120 \times 10^{-5}$	$1.9489 \times 10^{-6}$
$R^2 = 0.6774$ ( $p < 2.2 \times 10^{-16}$ )			





**Fig. 4** **A** Substitution spectra for human (H), chimpanzee (C), gorilla (G), and orangutan (O) across 25 autosomal telomeres and PAR1. **B** Difference between relative substitution proportions and mean relative proportions normalized by the standard deviation (Z-score) for each substitution class, telomere and species. For each telomere and substitution class, we plot four values corresponding to the four external branches of the great ape phylogeny. In total, we plot  $N = 700$  estimates for autosomes (across 25 alignable telomeres) and  $N = 28$  estimates for PAR1

mutation rate that is 50-fold greater than the genome average [67]. Furthermore, non-human great ape species also exhibit the C→G substitution excess on 8L and PAR1, indicating that these mutation hotspots are conserved across the great ape phylogeny. In addition to an excess of C→G transversions, 8L and PAR1 have a general decrease in transition and increase in transversion proportions, resulting in the lowest ts:tv ratios among all telomeres (Fig. 3B; Additional file 1: Table S1). Telomeres 16L and 16R have also been identified as DNM hotspots and, while they do show an excess of C→G transversions, their ts:tv ratios are similar to those of other autosomal telomeres. The higher similarity between 8L and PAR1 may stem from the dependence of telomere substitution dynamics on divergence rates, the ts:tv ratio and chromosome length (Fig. 3;



**Fig. 5** **A** Sex-averaged recombination rates for 10 kb windows of PAR1 in humans (HUM) and eight subspecies of great apes (PTE = *P. troglodytes ellioti*; PTS = *P. troglodytes schweinfurthii*; PTT = *P. troglodytes troglodytes*; PTV = *P. troglodytes verus*; PPA = *P. paniscus*; GGG = *G. gorilla gorilla*; PAB = *P. abelii*; PPY = *P. pygmaeus*). Numbers of polymorphic sites used to infer recombination maps are as follows: HUM ( $N = 220$ ), PTE ( $N = 4039$ ), PTS ( $N = 5191$ ), PTT ( $N = 6765$ ), PTV ( $N = 2438$ ), PPA ( $N = 2064$ ), GGG ( $N = 4323$ ), PAB ( $N = 6957$ ) and PPY ( $N = 4371$ ). **B** Spearman's coefficients for correlations between recombination maps at the 10 kb scale. **C** Nucleotide diversity for 10 kb windows of PAR1 in humans and great apes, measured as  $\pi$  and Watterson's  $\theta$ . Numbers of polymorphic sites used to infer diversity statistics as follows: HUM ( $N = 29,172$ ), PTE ( $N = 5192$ ), PTS ( $N = 6161$ ), PTT ( $N = 9379$ ), PTV ( $N = 3181$ ), PPA ( $N = 2687$ ), GGG ( $N = 4876$ ), PAB ( $N = 7606$ ) and PPY ( $N = 4547$ ). **D** Spearman's coefficients for correlations between nucleotide diversity  $\pi$  at the 10 kb scale. Crossed-out coefficients are non-significant ( $p > 0.05$ ). We only consider 10 kb regions with more than 2500 callable sites

Table 2)—indeed, chromosome 8 and X are very similar with respect to all three parameters (Additional file 1: Table S1).

### Recombination and diversity in PAR1

In the absence of family data, we used the LDhat program to generate PAR1 recombination rates ( $\rho = 4N_e r$ ; where  $r$  is the rate in Morgans) for eight subspecies of great apes, to compare with the published human PAR1 map, inferred from directly observed crossovers in human pedigrees [21] (Fig. 5A). The LDhat recombination rate estimates represent population-wide sex-averages and were converted into cM/Mb by assuming a sex-averaged recombination rate of 9.01 cM/Mb, as inferred in the human PAR1 [21].

**Table 3** Estimates of PAR1 population genetic parameters for human (HUM) and great ape populations

	$\rho$	$\pi$	$\theta_{\text{PAR}} (\theta_{\text{A}})$	Tajima's $D$	$N_e$	$\mu$ (PG)	$\mu$ (PY)	$B$
HUM	-	0.0025	0.0025 (0.0009)	-0.0541	-	-	-	0.6745
PTE	12,922	0.0024	0.0026 (0.0013)	-0.2500	14,138	$4.31 \times 10^{-8}$	$1.72 \times 10^{-9}$	0.6801
PTS	16,141	0.0023	0.0026 (0.0016)	-0.3273	17,661	$3.31 \times 10^{-8}$	$1.32 \times 10^{-9}$	0.6064
PTT	32,855	0.0028	0.0039 (0.0023)	-1.0783	35,948	$1.97 \times 10^{-8}$	$0.79 \times 10^{-9}$	0.8125
PTV	9,136	0.0013	0.0014 (0.0008)	-0.2132	9,996	$3.33 \times 10^{-8}$	$1.33 \times 10^{-9}$	0.7302
PPA	15,293	0.0011	0.0014 (0.0005)	-0.7477	16,732	$1.70 \times 10^{-8}$	$0.68 \times 10^{-9}$	0.7704
GGG	12,584	0.0023	0.0022 (0.0016)	0.1314	13,769	$4.10 \times 10^{-8}$	$2.16 \times 10^{-9}$	0.7640
PAB	11,901	0.0038	0.0037 (0.0019)	0.1374	13,021	$7.26 \times 10^{-8}$	$2.79 \times 10^{-9}$	0.6531
PPY	13,728	0.0027	0.0023 (0.0013)	0.6627	15,020	$4.48 \times 10^{-8}$	$1.72 \times 10^{-9}$	0.4962

$\rho = 4N_e r$  of the whole PAR1 region, as measured by LDhat

$\pi = 4N_e \mu$ ; average pairwise diversity estimator of PAR1

$\theta = 4N_e \mu$ ; Watterson's diversity estimator ( $\theta_{\text{PAR}} = \text{PAR1 } \theta$ ;  $\theta_{\text{A}} = \text{autosomal } \theta$ )

$N_e$  = effective size of the population

$\mu$  = per site and per generation (PG), or per year (PY) mutation rate

$B = 4N_e b$ ; strength of GC-biased gene conversion in PAR1

PTE *P. troglodytes ellioti*, PTS *P. troglodytes schweinfurthii*, PTT *P. troglodytes troglodytes*, PTV *P. troglodytes verus*, PPA *P. paniscus*, GGG *G. gorilla gorilla*, PAB *P. abelii*, PPY *P. pygmaeus*

Notably, across the subspecies studied, the recombination rate per physical distance is consistently high towards the telomeric end of PAR1, though humans have a significant uptick close to the pseudoautosomal boundary. The majority of recombination events occur between positions 0.25–1.25 Mb of PAR1 with the strongest peak of recombination inferred for *P. troglodytes ellioti*, located between base pairs 330,001–340,000. The peak has an extremely high recombination rate of 90.42 cM/Mb and overlaps the gene *PPP2R3B*, known to be a recombination hotspot in humans [68], and highly expressed during spermatogenesis [69]. We next lifted-over genomic coordinates of non-human species to the human genomic reference and estimated the between-species correlation of recombination rates for 10 kb genomic windows (Fig. 5B). Generally, correlation coefficients between maps of the different species are positive. The correlations are strongest between species of the same genus; among subspecies of the *Pan* genus, the overall strongest inferred correlation is between the *P. troglodytes ellioti* and *P. troglodytes schweinfurthii* maps (Spearman's  $\rho = 0.6739$ ,  $p < 0.0001$ ), and the *P. abelii* and *P. pygmaeus* map correlation is stronger than the correlations with non-orangutan species (Spearman's  $\rho = 0.5752$ ,  $p < 0.0001$ ). Taken together, these results imply that, while recombination in PAR1 is broadly similar across great apes, genus-specific patterns of PAR1 recombination are also observable at the studied phylogenetic scale.

Nucleotide diversity in PAR1 is on average  $2.3\times$  higher than for autosomal estimates (Table 3), and follows the recombination pattern, with elevated rates close to the telomere end (Fig. 5C), and a strong within-genus correlation (Fig. 5D). The correlation between recombination and average pairwise diversity  $\pi$  is positive for all species (Additional file 2: Fig. S2), consistent with recombination-associated mutagenesis [22, 30, 31].

Given LDhat estimates of  $\rho$  ( $4N_e r$ ) for PAR1, where  $r = \text{cM}/100$ , and assuming that the sex-averaged genetic length of PAR1 in great apes is equal to the human estimate

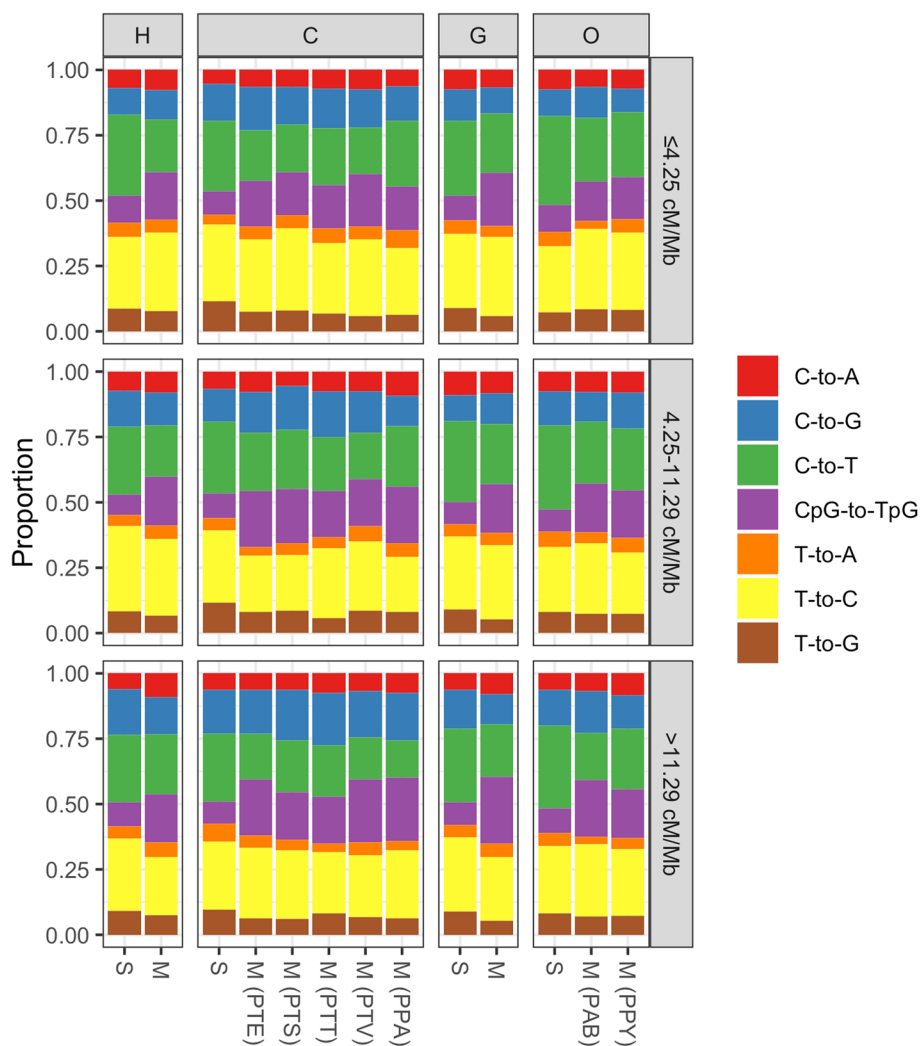
of 22.85 cM [21], we can calculate the effective size ( $N_e$ ) for each population (Table 3). For five out of eight great ape species,  $N_e$  estimates fall within the previously inferred ranges [70]. For *P. paniscus*,  $N_e$  is higher than the previously inferred upper limit (16,732 vs. 10,800 individuals), while for *G. gorilla gorilla* and *P. abelii*,  $N_e$  is lower than the previously inferred lower limit (13,769 vs. 20,132 and 13,021 vs. 16,731 individuals, respectively). We can similarly calculate the PAR1-specific per generation mutation rate  $\mu$  by taking the ratio  $\pi/\rho = \mu/r$  (Table 3). We also report  $\mu$  as a per year rate by assuming generation times of 25 years for chimpanzees, 19 years for the gorilla and 26 years for the orangutan. Compared to the per year mutation rate estimated from great ape trios ( $0.64 \times 10^{-9}$  [62]), PAR1 yearly  $\mu$  is on average 2.44-fold larger, ranging from a 1.06-fold increase in *P. paniscus* to a 4.36-fold increase in *P. abelii*. While this mutation range overlaps the per year divergence range of PAR1 (Table 1), mutations notably occur at a higher yearly rate compared to substitutions, indicating that extant nucleotide diversity levels do not fully reflect the long-term substitution dynamics of PAR1.

#### Differences between segregating polymorphisms and substitution spectra in PAR1

We next explore the differences between PAR1 diversity and substitution spectra, and their dependence on the recombination rate. We therefore divided the PAR1 sequence of each great ape population into 10 kb windows with low ( $\leq 4.25$  cM/Mb), medium ( $4.25 < \text{cM/Mb} \leq 11.29$ ), and high recombination rates ( $> 11.29$  cM/Mb), based on recombination rate terciles determined across all maps of the nine populations studied (Fig. 5A). Since the substitution spectrum estimates are based on the reference alignment, i.e., all sequences within the alignment are considered simultaneously, while the division of PAR1 regions into recombination bins is population-based, we applied the following protocol when inferring the substitution spectrum of a specific genus and recombination bin. Due to the strong within-genus correlation of recombination rates (Fig. 5B), we inferred recombination maps for each genus by calculating the mean recombination rate for each 10 kb region across all its sampled populations. For the human and gorilla datasets, with one population per genus, we simply used the inferred rates. Based on each genus-specific map, we categorized 10 kb regions into the three recombination bins to infer divergence rates and posterior substitution counts using the phyloFit program. In total, this procedure results in four different divisions of PAR1 (one for each

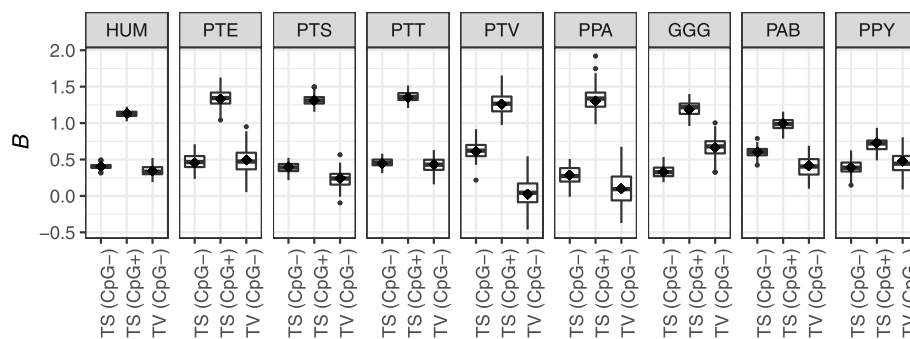
**Table 4** Divergence estimates based on phyloFit estimation and divergence times from [62] for the PAR1 sequence, across three recombination rate bins

Rec. rate (cM/Mb)	Total divergence			Per site and year divergence		
	$\leq 4.25$	4.25–11.29	$> 11.29$	$\leq 4.25$	4.25–11.29	$> 11.29$
Human	0.0092	0.0106	0.0119	$0.8411 \times 10^{-9}$	$0.9731 \times 10^{-9}$	$1.0927 \times 10^{-9}$
Chimpanzee	0.0109	0.0117	0.0139	$1.0042 \times 10^{-9}$	$1.0764 \times 10^{-9}$	$1.2797 \times 10^{-9}$
HC ancestor	0.0026	0.0039	0.0048	$1.2571 \times 10^{-9}$	$1.8603 \times 10^{-9}$	$2.3172 \times 10^{-9}$
Gorilla	0.0134	0.0137	0.0166	$1.0319 \times 10^{-9}$	$1.0600 \times 10^{-9}$	$1.2840 \times 10^{-9}$
HCG ancestor	0.0154	0.0177	0.0175	$1.4121 \times 10^{-9}$	$1.6236 \times 10^{-9}$	$1.6114 \times 10^{-9}$
Orangutan	0.0243	0.0286	0.0374	$1.0192 \times 10^{-9}$	$1.1984 \times 10^{-9}$	$1.5695 \times 10^{-9}$



**Fig. 6** Substitution and diversity spectra for the human (H), chimpanzee (C), gorilla (G), and orangutan (O) PAR1, across three recombination bins (rows). Substitution and diversity spectra are denoted by S and M, respectively. Multiple diversity spectra are plotted for chimpanzees and orangutans, each corresponding to a particular subspecies (PTE = *P. troglodytes ellioti*; PTS = *P. troglodytes schweinfurthii*; PTT = *P. troglodytes troglodytes*; PTV = *P. troglodytes verus*; PPA = *P. paniscus*; PAB = *P. abelii*; PPY = *P. pygmaeus*), while the diversity spectra for humans and gorillas correspond to the human YRI and *G. gorilla gorilla* populations, respectively

genus), each with three genus-specific recombination bins (i.e., 12 distinct combinations of PAR1 10 kb regions in total). Table 4 reports divergence rates for the different recombination bins. Notably, divergence rates increase with increasing recombination rates for all lineages studied. The increase in divergence between the low and high recombination bins for extant species ranges from 1.24-fold in gorillas to 1.54-fold in orangutans. In Fig. 6, we show the substitution and extant diversity spectra for PAR1 across the three recombination bins. The most notable difference between the two spectra is the decrease of CpG→TpG substitutions compared to extant segregating variation. This trend is observable across all three recombination bins. Furthermore, for all substitution and diversity spectra, the C→G proportion is higher in the highest, compared to the



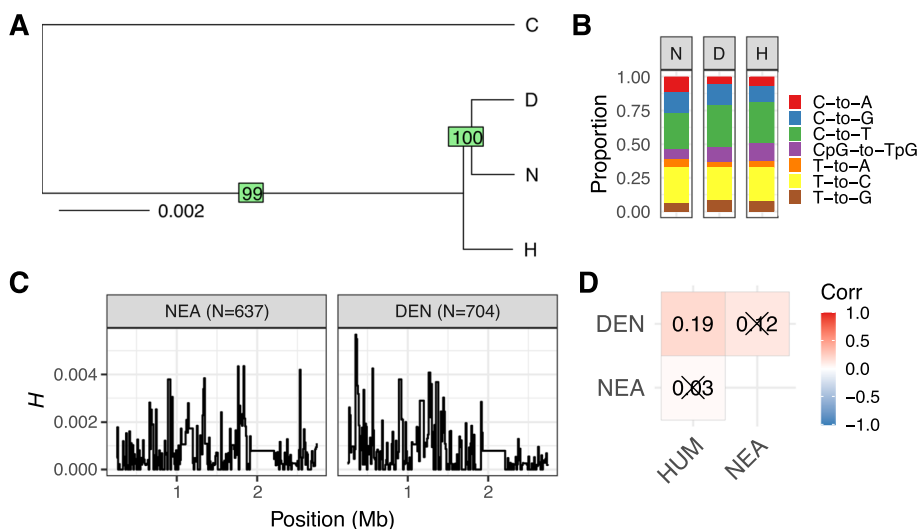
**Fig. 7** PAR1  $B$  estimates for non-CpG transitions (TS; CpG $-$ ), CpG transitions (TS; CpG $+$ ), and GC-changing transversions (TV; CpG $-$ ) for the human YRI population (HUM) and eight subspecies of great apes (PTE = *P. troglodytes ellioti*; PTS = *P. troglodytes schweinfurthii*; PTT = *P. troglodytes troglodytes*; PTV = *P. troglodytes verus*; PPA = *P. paniscus*; GGG = *G. gorilla gorilla*; PAB = *P. abelii*; PPY = *P. pygmaeus*). Distributions of  $B$  estimates were obtained from 100 bootstrapped allele frequency spectra for each mutation type and subspecies. The “diamond” points correspond to the maximum likelihood  $B$  estimate

lowest recombination bin, in line with the observation that this mutation type is associated with male meiotic double-strand breaks on the X [22].

Figure 6 indicates that CpG mutations are the main contributors to the discrepancy between long-term substitution rates and segregating polymorphism. Such an increase in extant diversity compared to the long-term neutral expectation (i.e., substitution spectrum) is expected when a fixation bias favoring GC nucleotides is opposed by AT-biased mutation [71]. In humans and great apes, the preferred fixation of GC nucleotides is mediated by recombination and is termed GC-biased gene conversion (gBGC [33–35]). To assess the strength of gBGC in PAR1, we constructed allele frequency spectra (AFS) based on GC frequency categories of segregating variants and inferred the bias parameter  $B = 4N_e b$ , where  $b$  is a bias coefficient [36]. For this analysis, we omitted sites segregating for CpG transversions, due to their idiosyncratic gBGC dynamics [72], and calculated  $B$  for the rest of the sites (Table 3), and separately for three mutation categories: non-CpG and CpG transitions and non-CpG transversions (Fig. 7). Generally, PAR1  $B$  estimates are approximately twice the autosomal estimates [72, 73], reflecting the high PAR1 recombination rate, with CpG sites experiencing the strongest gBGC dynamics. Therefore, the exceptionally strong gBGC at CpG sites likely contributes to the paucity of this mutation class in the substitution spectrum as CpG $\rightarrow$ TpG mutations are ultimately fixed back as GC nucleotides, while extant diversity at these sites is elevated due to the opposition between exceptionally strong GC $\rightarrow$ AT mutation and AT $\rightarrow$ GC fixation forces [71].

#### Evolution of PAR1 during recent hominin speciation

We next analyze archaic PAR1 sequences of a Neanderthal and Denisovan individual. In total, 251,571 putatively neutral base pairs were alignable between the archaic sequences and the human and chimpanzee reference sequences. Pairwise divergence between the human and archaic samples are 0.0021, 0.0022 and 0.0019 for human-Neanderthal, human-Denisovan, and Neanderthal-Denisovan pairs, respectively. Figure 8A shows a neighbor-joining tree of archaic individuals and the human reference, using the chimpanzee sequence as an outgroup. The Neanderthal and Denisovan individuals are



**Fig. 8** **A** PAR1 neighbor-joining tree for human (H), Neanderthal (N), Denisovan (D), and chimpanzee (C) sequences. Internal branch supports of 100 bootstrap samples are shown in green rectangles. **B** Substitution spectra for the Neanderthal, Denisovan and human PAR1. **C** Nucleotide heterozygosity for 10 kb windows of PAR1 in Neanderthal (NEA) and Denisovan (DEN) sequences, measured as per site heterozygosity  $H$ . Numbers of polymorphic sites used to infer  $H$  values are presented in the parentheses. **D** Spearman coefficients for correlations between diversity at the 10 kb scale, for archaic individuals and humans (HUM). Crossed-out coefficients are non-significant ( $p > 0.05$ ). We only consider 10 kb regions with more than 2500 callable sites

grouped into a separate clade, as previously observed for phylogenetic trees based on autosomal data [51, 74]. Therefore, while a recent study showed that the Y chromosome introgressed from humans into Neanderthals [50], our results indicate that the Neanderthal PAR1 retained the autosomal-like phylogenetic relationship (i.e., avoided partial replacement with the invading human Y-linked PAR1).

Interestingly, the Denisovan substitution spectrum resembles that of humans more closely than that of Neanderthals (Fig. 8B). We do, however, recover the pattern of relatively higher proportions of C→G transversions and a lack of CpG→TpG transitions in archaics compared to modern humans, as reported in [49]. The general reduction of C→T transitions is especially evident in the Neanderthal spectrum. Additionally, there is a notable excess of C→A transversions in the Neanderthal spectrum, but not in the Denisovan one. This mutation type is generally enriched in archaic sequences, but only significantly so given a specific nucleotide context [49].

We estimate average nucleotide heterozygosity to be  $7.34 \times 10^{-4}$  and  $7.84 \times 10^{-4}$  per base pair for the Neanderthal and Denisovan PAR1 sequences, respectively, approximately  $\sim 4\times$  higher than the genome-wide average [51, 52]. Figure 8C shows diversity values in 10 kb regions along PAR1. Interestingly, the Denisovan sequence shows high telomeric diversity, as in humans (Fig. 5C), while Neanderthal diversity is relatively low at the telomeric end. The correlation between diversity patterns is positive between human and Denisovan sequences, and non-significant for the other two comparisons (Fig. 8D), as might be expected due to higher similarity between the human and Denisovan substitution spectra (Fig. 8B). Importantly, despite the differences between the archaic sequences, support for an autosomal-like phylogeny of hominin PAR1 remains

strong (Fig. 8A), indicating that these discrepancies could be due to changes in mutation and/or recombination processes during recent hominin evolution [49].

## Discussion

PAR1 is a unique genomic region due to its linkage to the sex-determining region of the Y chromosome and the status as arguably the most important recombination hotspot in the genome. With this study, we show that PAR1 biology of great apes is mostly governed by recombination dynamics, in turn affecting mutation patterns of this region.

The nucleotide composition of the human genome is affected by recombination and is known to be evolving towards a lower GC content [34, 66]. On the other hand, the GC content of the PAR1 region is relatively close to the equilibrium expectation, compared to similar regions with high male-specific recombination rates (Fig. 2B, C). We also show that the deviation from the equilibrium nucleotide content depends on divergence rates, the ts:tv ratio and chromosome length (Fig. 3, Table 2). Greater deviations are generally associated with short chromosomes, which may have evolved under relatively stronger gBGC effects, due to higher effective sizes in ancestral populations compared to extant populations [62, 66].

As previously observed in humans, we detect a strong enrichment of C→G mutations, associated to male meiotic double-strand breaks on the X [22] in the substitution spectra of all species of great apes (Fig. 4). Additionally, the mutation signatures of the telomeric regions of chromosome arms 8L, 16L, and 16R [67] are well conserved, indicating a general stability of mutation hotspots and their effects on substitution spectra across the great ape phylogeny. Given the fact that these are not transient hotspots, their ubiquity across a broader set of species would be an interesting avenue of future research.

The analysis of recombination rate and nucleotide diversity revealed expected associations between the two parameters (Fig. 5 and S2). Evolution of recombination differences between the genera is evident in PAR1, in line with previous observations of differences between human and chimpanzee recombination maps [75, 76]. Additionally, an increased recombination rate close to the pseudoautosomal boundary detected in humans [21] is not found in other great apes. This may indicate a shift in human PAR1 recombination patterns, but it could also reflect the differences between the methods used to infer the maps (the human map is pedigree-based, whereas the great ape maps are based on LD analyses). An estimation of PAR1 recombination maps using LD-based methods [77, 78] for different human populations would be beneficial for a better understanding of this pattern.

The  $N_e$  estimates we obtained by using the recombination rate estimate  $\rho$  ( $4N_e r$ ) are fairly accurate for most populations (Table 3). For the *P. paniscus*, PAR1  $N_e$  is larger than previously inferred, while it is lower for the *G. gorilla gorilla* and *P. abelli* populations [70]. This could be due either to processes that cause PAR sequences to evolve differently from autosomal ones, and affect  $N_e$  in these populations, or simply due to a limitation of the LDhat method used to infer  $\rho$ . A more sophisticated method of recombination rate inference using demography-aware models [78] would be valuable in future studies.

Our results quantifying the PAR1-specific mutation and the divergence rates, and their association with recombination (Tables 3 and 4, Figs. 5 and 6 and S2), strongly suggest



that recombination-associated mutations play a more important role in PAR1 divergence, compared to other genomic regions. In a recent study on the mutagenic impact of recombination, [29] showed that the male de novo mutation rate is  $\sim 3.3 \times 10^{-8}$  per site and generation in the region  $\pm 20$  kb from male crossovers. If we assume that PAR1 experiences one such crossover per generation and an average background male mutation rate of  $\sim 0.96 \times 10^{-8}$  per site and generation [29], we estimate that a  $\sim 3.5\%$  increase of PAR1 mutations is due to this one mandatory male crossover event. However, this increase is too small to fully account for the high PAR1-specific  $\mu$  in great apes (Table 3). Therefore, additional mutation inputs, e.g., from non-crossover events [79] and double strand break (DSB) repair mechanisms [56], are likely to further contribute to a higher mutation rate of the PAR1 sequence.

We also showed that, as expected, the strength of GC-biased gene conversion (Table 3, Fig. 7) is higher than for autosomal sequences [72, 73, 80, 81]. By separately estimating the  $B$  parameters for different classes of mutations, we find that CpG sites are most strongly affected by gBGC dynamics. For gorillas and orangutans, the gBGC strength is somewhat reduced for CpG sites, as observed for autosomes [72]. This mutation class is also underrepresented in the substitution spectrum compared to its frequency in the diversity spectrum (Fig. 6). We hypothesize that the discrepancy between the CpG frequency in the diversity and substitution spectrum is due to a somewhat counterintuitive phenomenon [71] that occurs when mutation and fixation forces work in the opposite directions (i.e., CpG  $\rightarrow$  TpG mutations vs. gBGC at CpG-segregating sites). In this case, extant diversity, and therefore the proportion of CpG sites in the diversity spectrum, is elevated beyond the neutral expectation (i.e., the substitution spectrum). Importantly, this phenomenon may not be limited to CpG sites, and is probably occurring at all GC-changing sites evolving under the extreme PAR1-specific mutation and substitution dynamics. Therefore, the estimated range of PAR1-specific mutation rates  $\mu$  (Table 3) is also likely overestimated and should be interpreted with caution.

In a recent study, differences between frequencies of PAR1 alleles segregating in females and males have been reported for humans [23]. We have also conducted a similar analysis of frequency differences between the sexes but found no evidence for sexual antagonism (SA) in the PAR1 of great apes (Additional file 3). We see three possible reasons for this discrepancy. Firstly, for the majority of great ape populations, the sample size is simply too small to reach significance levels, even for sites fixed for different alleles between females and males. Secondly, in the study of human PAR sex-specific allele frequencies, to our knowledge, the authors of [23] did not apply any correction for multiple-testing when assessing frequency differences, making their results difficult to reconcile with our own analysis of the human dataset. Thirdly, while we consider only the African YRI population, the authors of [23] study frequency differences using either all individuals from the 1000 Genomes Project [82] or individuals sorted into superpopulations. Therefore, much of the population-specific structure is ignored, which may inflate the observed frequency differences. An adoption of a more sophisticated framework for detecting SA, such as the one presented in [40], would be more appropriate for future studies. On the other hand, the difference in PAR lengths between haplorrhine and strepsirrhine primates is suggestive of a historically stronger SA in great apes [47]. Additionally, shortening of the PAR to just  $\sim 1$  Mb in length has been recently detected

in the marmoset [83], indicative of further potential for sexually antagonistic evolution in the great ape PAR1. However, detection of ongoing SA processes is likely very difficult in highly derived PAR sequences of primates, compared to species with nascent sex chromosomes [84].

Another example of genus-specific PAR1 evolution comes from the analysis of ancient human sequences. Interestingly, we find that PAR1 retained the autosomal-like phylogeny patterns of Neanderthal and Denisovan sequences (Fig. 8A) [51, 74], despite the invading human Y chromosome [50]. Additionally, we recover population-specific mutation patterns of archaic PAR1 sequences (Fig. 8B) in line with previous observations [49]. These results indicate that the invading human Y chromosome largely managed to spread in the Neanderthal lineage without being linked with the human-specific PAR1 sequence. A possible explanation for this observation is the decoupling of human PAR1 from the rest of the Y chromosome during the introgression event (likely due to high PAR1 recombination), followed by loss from the Neanderthal lineage due to directional selection or genetic drift. Specifically, as sequence homology is crucial for successful chromosome pairing during meiosis, the production of gametes may have been impeded in reproductive cells that contained a pair of diverged PAR1 sequences, such as a Neanderthal and human PAR1. Additionally, as introgression of the human Y chromosome likely occurred at low introduction frequencies into the Neanderthal population, the probability of survival of the human PAR1 would also be low if no strong selective forces acted in favor of its maintenance. Our results therefore imply that the main selective force in favor of human Y introgression likely acted on the sex-determining region of the Y.

Lastly, we explored repeat content and evolution in PAR1-specific DSB regions (Additional file 4). The mo-2 minisatellite arrays in mice [53–55] are repetitive elements involved in chromosome axes elongation and sister chromatid separation, crucial for achieving a recombination event during male meiosis. Our analysis of human sequence also implies an important recombinogenic role of repetitive elements, especially in PAR1 (Additional file 4: Fig. S4). A more detailed analysis of PAR1, using sophisticated molecular biology and microscopy methods as in [55], would be of great value for our understanding of the assurance of human PAR recombination. When analyzing DSB regions, we find patterns indicative of higher recombination rates, such as high divergence rates (Additional file 4: Table S2) and recombination-specific substitution patterns (Additional file 4: Fig. S5). Even though these regions have been identified in human males, the observed patterns are also present in other great apes, indicating that DSB hotspots are likely conserved within the great ape lineage despite significant divergence of PRDM9-binding motifs in primates [57–59]. This observation could also indicate the existence of PRDM9-independent DSB hotspots in PAR1 across great apes. With these results in mind, we propose that in addition to recombination map estimation and PRDM9-binding studies, repeat content analysis and comparative approaches should be included into future analyses of recombination determinants.

## Conclusions

PAR1 is a unique genomic region that, despite its very small size, has a paradoxically large functional role as it ensures the proper inheritance of X and Y chromosomes. Due to its small size, PAR1 is the strongest hotspot of recombination in the genome,

which results in accelerated nucleotide evolution. Additionally, population genetics theory indicates that this region is especially prone to contain genes that preferentially benefit one of the two sexes. It is therefore evident that PAR1 is a stage for evolutionary conflicts that can lead to major shifts in evolution. Using the comparative framework of human and great ape species, we observe that while this region is fast evolving and has high between-species divergence, its nucleotide composition is close to equilibrium. Together with the observation that we detect no ongoing sexual antagonism within this region, the dominant forces shaping PAR1 evolution in great apes are mutation processes and male-specific recombination. To test the universality of these observations, future studies should focus on expanding the PAR1 analyses to other mammals, given the increasing availability of high quality sex chromosome assemblies in non-model species.

## Methods

### Data

Reference genomes used for alignment of the pseudoautosomal region 1 (PAR1) and autosomal telomeres were GRCh38 for human, Clint PTRv2 for chimpanzee (University of Washington; January 2018), gorGor4 for gorilla (Wellcome Trust Sanger Institute; October 2015), Susie PABv2 for orangutan (University of Washington, 2018), and Mmul\_10 for rhesus macaque (The Genome Institute at Washington University School of Medicine, 2019); <https://www.ncbi.nlm.nih.gov/>. We used the Progressive Cactus pipeline [60] to obtain a multiple-species alignment and ancestral sequence reconstruction. The aligned regions range from the beginning of the X chromosome to 100 kb upstream of the *XG* gene (which contains the pseudoautosomal boundary): 1-2,916,500 bp in human, 1-2,589,128 bp in chimpanzee, 1-2,486,644 in gorilla, and 1-2,498,610 in macaque reference. For the PABv2 assembly, we used the unlocalized X-linked scaffold (NW\_019937303.1) of length 3,293,409 which contains the orangutan PAR1 region. We only retained regions that were uniquely aligned between the sequences, which resulted in 1,066,683 base pairs of alignable sequence. We further curated the alignment by retaining only alignment blocks that are syntenic (non-inverted and non-overlapping) between the human and the three great ape species, resulting in an alignment of 620,054 sites. Finally, to obtain an alignment of putatively neutral sequences, we excluded coding regions, CpG islands, repetitive sequences, and conserved regions (UCSC tracks: ncbiRefSeq, cpgIslandExtUnmasked, rmsk and phastConsElements30way). After excluding sites in and upstream of the *XG* gene, we were left with 239,799 putatively neutral PAR1 sites defined across all species and ancestral nodes. For autosomal telomeres, we chose sequences that are conserved as telomeres (3 Mb regions at the tip of an autosome) in the macaque reference [85, 86]. This left us with 25 alignable autosomal telomeres with on average 872,720 putatively neutral bases for each telomere.

For human nucleotide diversity estimates we used the 30× high-coverage data of 107 individuals of the Yoruba (YRI) population from the 1000 Genomes Project [82]. The recombination map of human PAR1 was taken from [21]. The callable fraction for the human PAR1 was determined using the pilot accessibility mask of the 1000 Genomes Project, lifted-over [87] to GRCh38 coordinates.

The dataset of the eight subspecies of great apes used to estimate diversity and recombination maps was curated from three sequencing studies [70, 88, 89]. In total, we analyzed five subspecies of the *Pan* genus: *P. troglodytes ellioti* ( $N = 10$ ), *P. troglodytes schweinfurthii* ( $N = 19$ ), *P. troglodytes troglodytes* ( $N = 18$ ), and *P. troglodytes verus* ( $N = 11$ ) and *P. paniscus* ( $N = 13$ ); one subspecies of the *Gorilla* genus, *G. gorilla gorilla* ( $N = 23$ ); and two subspecies of the *Pongo* genus, *P. abelii* ( $N = 11$ ) and *P. pygmaeus* ( $N = 15$ ). We conducted a de novo read mapping and variant calling for PAR1 in all eight great ape populations as described in [72]. Recombination maps for great apes were inferred using the LDhat 2.2 program in interval mode [77], following the protocols in [75, 90]. The callable fraction of PAR1 for great apes was determined as the number of sites covered by at least 1.5 reads per haploid genome and no more than  $2 \times n \times cov$  reads, where  $n$  is the ploidy of the region and  $cov$  is the mean coverage per haploid genome. We lifted-over the great ape reference sequences to GRCh38 coordinates for estimating between-species correlations of recombination and diversity (Fig. 5B, D).

Estimates of autosomal nucleotide diversity ( $\theta_A$ ) for the nine studied populations in Table 3 were taken from [70, 88, 89, 91].

For the analysis of PAR1 evolution in archaic humans, we use data of the Neanderthal and Denisovan individuals sequenced to high coverage from [51, 52]. The vcf coordinates of these samples were lifted-over to the GRCh38 human reference prior to analysis. The callable fraction of the archaic sequences was determined from their vcf files, which included calls of invariant sites. Coordinates of double-strand break hotspot regions in PAR1 of human males were taken from [56] and lifted to GRCh38 coordinates. For both analyses we exclude coding regions, CpG islands, repetitive sequences and conserved regions.

### Statistical analysis

All statistical analyses and plotting were conducted using the R programming software [92]. The linear model presented in Table 2 was run using the R “lm” function.

### Divergence and substitution spectra estimation

We used the program phyloFit [63] to estimate divergence rates, using the expectation maximization (option -E) and the general unrestricted single nucleotide model (--subst-mod UNREST). Additionally, we ran phyloFit using the U2S substitution model (the general unrestricted dinucleotide model with strand symmetry), to infer posterior counts and rates of different substitution types, and construct substitution spectra. As input for phyloFit, we use the multiple-species Cactus alignment with reconstructed ancestral states. To ensure convergence of the expectation maximization algorithm, we ran phyloFit 10 times for each analysis with random parameter initialization (option -r) and random seed numbers (option -D). We report divergence estimates and substitution spectra for the runs with the highest likelihood values.

We use node-specific substitution rate matrices inferred by phyloFit to calculate the equilibrium expectation of GC content ( $GC^*$ ), as the ratio of the  $AT \rightarrow GC$  divergence rate to the full rate (i.e., the sum of  $AT \rightarrow GC$  and  $GC \rightarrow AT$  rates). To convert phyloFit

divergence rates into per year estimates, we use the times of the last common ancestors for all nodes within the great ape phylogeny that have been independently estimated in [62].

#### Estimation of GC-biased gene conversion

Inference of the GC-biased gene conversion (gBGC) parameter  $B$  ( $4N_e b$ , where  $N_e$  is the effective size of the population and  $b$  is the conversion parameter as defined by [36]) follows the framework in [72, 93]. In short, we constructed allele frequency spectra (AFSs) with categories corresponding to GC frequency bins. For great ape samples, these bins correspond to discrete values of GC counts at a segregating site, while for the larger human population, we defined 50 equally-sized GC frequency ranges for use as AFS categories. We used only segregating sites with complete data, i.e., with defined nucleotide states for all individuals within a population. Prior to  $B$  inference, we omitted the two most extreme categories of AFSs as these are most likely to be affected by mutation processes and potentially bias  $B$  inference.

#### Inference of the neighbor-joining tree of archaic sequences

We first constructed reference sequences for the Neanderthal and Denisovan individuals by converting their vcf files into fasta files using a custom python script. For polymorphic sites in the vcf, we randomly selected one of the two segregating nucleotides as the reference. To construct the neighbor-joining tree in Fig. 8A, we followed the protocol of [50] and used the R packages *ape* [94] and *phangorn* [95] to infer the tree and internal bootstrap support.

#### Inference of repeat content

Repeat content of PAR1 and autosomal telomeres was characterized using the Tandem Repeats Finder program [96]. For all telomeres, the program was run on the unfiltered sequence (i.e., including coding regions, CpG islands, repetitive sequences and conserved regions), using the recommended parameter settings: match weight of 2, mismatch and indel penalty of 5 and 7, respectively, match and indel probability of 80 and 10, respectively, minimum alignment score of 50, and maximum period size of 2000 bp.

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-022-02784-x>.

Additional file 1: Table S1. Telomere statistics and phyloFit substitution counts.  
Additional file 2. Supplementary figures and legends.  
Additional file 3. Analysis of sexual antagonism in PAR1 of humans and great apes.  
Additional file 4. Analysis of factors that ensure recombination in PAR1.  
Additional file 5. Review history.

#### Acknowledgements

We thank Deborah Charlesworth for critical reading of the manuscript. We thank Genome DK (<https://genome.au.dk/>) for bioinformatics support.

#### Peer review information

Kevin Pang was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Review history**

The review history is available as Additional file 5.

**Authors' contributions**

JB and MHS conceptualized the study design. JB conducted data curation and analysis. JB and MHS interpreted the results and jointly wrote the manuscript. The author(s) read and approved the final manuscript.

**Authors' Twitter handles**

Twitter handles: @juraj\_bergman (Juraj Bergman), @mikkelschierup (Mikkel Heide Schierup).

**Funding**

The study was supported by grant NNF18OC0031004 from the Novo Nordisk Foundation to MHS.

**Availability of data and materials**

The sequence data used in this study is openly available from GenBank accessions: PRJNA189439 [70], PRJEB15083 [88], PRJEB19688 [97], and [www.internationalgenome.org](http://www.internationalgenome.org) [82]. Data and scripts needed to reproduce the analyses are deposited in GitHub (<https://github.com/jbergman/par1EvoDynamics>) and zenodo (<https://zenodo.org/badge/latestdoi/497624380>) [98]. The scripts provided in these repositories are free to use under the GNU General Public License v3 (or later), as published by the Free Software Foundation.

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

Received: 30 November 2021 Accepted: 30 September 2022

Published online: 17 October 2022

**References**

- Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, et al. Origins and functional evolution of Y chromosomes across mammals. *Nature*. 2014;508:488–93.
- Nei M. Linkage modifications and sex difference in recombination. *Genetics*. 1969;63:681–99.
- Bull JJ. Evolution of sex determining mechanisms. San Francisco: Benjamin-Cummings Publishing Company; 1983.
- Charlesworth B. The evolution of sex chromosomes. *Science*. 1991;1030–3. <https://doi.org/10.1126/science.1998119>.
- Charlesworth D, Charlesworth B, Marais G. Steps in the evolution of heteromorphic sex chromosomes. *Heredity*. 2005;118–28. <https://doi.org/10.1038/sj.hdy.6800697>.
- Otto SP, Pannell JR, Peichel CL, Ashman T-L, Charlesworth D, Chippindale AK, et al. About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet*. 2011;27:358–67.
- Gabriel-Robez O, Rumpler Y, Ratomponirina C, Petit C, Levilliers J, Croquette MF, et al. Deletion of the pseudoautosomal region and lack of sex-chromosome pairing at pachytene in two infertile men carrying an X;Y translocation. *Cytogenet Genome Res*. 1990;38–42. <https://doi.org/10.1159/000132951>.
- Mohandas TK, Speed RM, Passage MB, Yen PH, Chandley AC, Shapiro LJ. Role of the pseudoautosomal region in sex-chromosome pairing during male meiosis: meiotic studies in a man with a deletion of distal Xp. *Am J Hum Genet*. 1992;51:526–33.
- Blaschke RJ, Rappold G. The pseudoautosomal regions, SHOX and disease. *Curr Opin Genet Dev*. 2006;233–9. <https://doi.org/10.1016/j.gde.2006.04.004>.
- Binder G. Short stature due to SHOX deficiency: genotype, phenotype, and therapy. *Horm Res Paediatr*. 2011;81–9. <https://doi.org/10.1159/000324105>.
- Lencz T, Morgan TV, Athanasiou M, Dain B, Reed CR, Kane JM, et al. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry*. 2007;12:572–80.
- Flaquer A, Rappold GA, Wienker TF, Fischer C. The human pseudoautosomal regions: a review for genetic epidemiologists. *Eur J Hum Genet*. 2008;771–9. <https://doi.org/10.1038/ejhg.2008.63>.
- Weng S, Stoner SA, Zhang D-E. Sex chromosome loss and the pseudoautosomal region genes in hematological malignancies. *Oncotarget*. 2016;7:72356–72.
- Russell LJ, Capasso M, Vater I, Akasaka T, Bernard OA, Calasanz MJ, et al. Deregulated expression of cytokine receptor gene, CRLF2, is involved in lymphoid transformation in B-cell precursor acute lymphoblastic leukemia. *Blood*. 2009;114:2688–98.
- Mullighan CG, Collins-Underwood JR, Phillips LAA, Loudin MG, Liu W, Zhang J, et al. Rearrangement of CRLF2 in B-progenitor- and Down syndrome-associated acute lymphoblastic leukemia. *Nat Genet*. 2009;41:1243–6.
- Benito-Sanz S, Royo JL, Barroso E, Paumard-Hernández B, Barreda-Bonis AC, Liu P, et al. Identification of the first recurrent PAR1 deletion in Léri-Weill dyschondrosteosis and idiopathic short stature reveals the presence of a novel SHOX enhancer. *J Med Genet*. 2012;49:442–50.

17. Mensah MA, Hestand MS, Larmuseau MHD, Isrie M, Vanderheyden N, Declercq M, et al. Pseudoautosomal region 1 length polymorphism in the human population. *PLoS Genet.* 2014:e1004578. <https://doi.org/10.1371/journal.pgen.1004578>.
18. Lahn BT. Four evolutionary strata on the human X chromosome. *Science.* 1999;964–7. <https://doi.org/10.1126/science.286.5441.964>.
19. Pandey RS, Wilson Sayres MA, Azad RK. Detecting evolutionary strata on the human x chromosome in the absence of gametologous y-linked sequences. *Genome Biol Evol.* 2013;5:1863–71.
20. Kelkar A, Thakur V, Ramaswamy R, Deobagkar D. Characterisation of inactivation domains and evolutionary strata in human X chromosome through Markov segmentation. *PLoS One.* 2009;4:e7885.
21. Hinch AG, Altemose N, Noor N, Donnelly P, Myers SR. Recombination in the human Pseudoautosomal region PAR1. *PLoS Genet.* 2014;10:e1004503.
22. Agarwal I, Przeworski M. Signatures of replication timing, recombination, and sex in the spectrum of rare variants on the human X chromosome and autosomes. *Proc Natl Acad Sci U S A.* 2019;116:17916–24.
23. Monteiro B, Arenas M, Prata MJ, Amorim A. Evolutionary dynamics of the human pseudoautosomal regions. *PLoS Genet.* 2021;17:e1009532.
24. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet.* 1998;861–9. <https://doi.org/10.1086/302011>.
25. Kong A, Gudbjartsson DF, Sainz J, Jonsson GM, Gudjonsson SA, Richardsson B, et al. A high-resolution recombination map of the human genome. *Nat Genet.* 2002;241–7. <https://doi.org/10.1038/ng917>.
26. Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. *Genet Res.* 1974;23–35. <https://doi.org/10.1017/s0016672300014634>.
27. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. *Genet Res.* 2007;89:311–36.
28. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. *Genetics.* 1993;134:1289–303.
29. Halldorsson BV, Palsson G, Stefansson OA, Jonsson H, Hardarson MT, Eggertsson HP, et al. Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science.* 2019;363. <https://doi.org/10.1126/science.aau1043>.
30. Filatov DA, Gerrard DT. High mutation rates in human and ape pseudoautosomal genes. *Gene.* 2003;67–77. [https://doi.org/10.1016/s0378-1119\(03\)00697-8](https://doi.org/10.1016/s0378-1119(03)00697-8).
31. Filatov DA. A gradient of silent substitution rate in the human pseudoautosomal region. *Mol Biol Evol.* 2004;21:410–7.
32. Bussell JJ, Pearson NM, Kanda R, Filatov DA, Lahn BT. Human polymorphism and human–chimpanzee divergence in pseudoautosomal region correlate with local recombination rate. *Gene.* 2006;94–100. <https://doi.org/10.1016/j.gene.2005.10.020>.
33. Galtier N, Piganeau G, Mouchiroud D, Duret L. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics.* 2001;159:907–11.
34. Meunier J, Duret L. Recombination drives the evolution of GC-content in the human genome. *Mol Biol Evol.* 2004;21:984–90.
35. Duret L, Galtier N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet.* 2009;285–311. <https://doi.org/10.1146/annurev-genom-082908-150001>.
36. Nagyaki T. Evolution of a large population under gene conversion. *Proc Natl Acad Sci U S A.* 1983;80:5941–5.
37. Jordan CY, Charlesworth D. The potential for sexually antagonistic polymorphism in different genome regions. *Evolution.* 2012;66:505–16.
38. Charlesworth B, Jordan CY, Charlesworth D. The evolutionary dynamics of sexually antagonistic mutations in pseudoautosomal regions of sex chromosomes. *Evolution.* 2014;68:1339–50.
39. Kirkpatrick M, Guerrero RF. Signatures of sex-antagonistic selection on recombining sex chromosomes. *Genetics.* 2014;531–41. <https://doi.org/10.1534/genetics.113.156026>.
40. Lucotte EA, Laurent R, Heyer E, Ségurel L, Toupan B. Detection of allelic frequency differences between the sexes in humans: a signature of sexually antagonistic selection. *Genome Biol Evol.* 2016;8:1489–500.
41. Cheng C, Kirkpatrick M. Sex-specific selection and sex-biased gene expression in humans and flies. *PLoS Genet.* 2016:e1006170. <https://doi.org/10.1371/journal.pgen.1006170>.
42. Flanagan SP, Jones AG. Genome-wide selection components analysis in a fish with male pregnancy. *Evolution.* 2017;1096–105. <https://doi.org/10.1111/evo.13173>.
43. Dutoit L, Mugal CF, Bolívar P, Wang M, Nadachowska-Brzyska K, Smeds L, et al. Sex-biased gene expression, sexual antagonism and levels of genetic diversity in the collared flycatcher (*Ficedula albicollis*) genome. *Mol Ecol.* 2018;3572–81. <https://doi.org/10.1111/mec.14789>.
44. Wright AE, Fumagalli M, Cooney CR, Bloch NI, Vieira FG, Buechel SD, et al. Male-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture. *Evol Lett.* 2018;2:52–61.
45. Qiu S, Bergero R, Charlesworth D. Testing for the footprint of sexually antagonistic polymorphisms in the pseudoautosomal region of a plant sex chromosome pair. *Genetics.* 2013;663–72. <https://doi.org/10.1534/genetics.113.152397>.
46. Guirao-Rico S, Sánchez-Gracia A, Charlesworth D. Sequence diversity patterns suggesting balancing selection in partially sex-linked genes of the plant *Silene latifolia* are not generated by demographic history or gene flow. *Mol Ecol.* 2017;26:1357–70.
47. Shearn R, Wright AE, Mousset S, Régis C, Penel S, Lemaître J-F, et al. Evolutionary stasis of the pseudoautosomal boundary in strepsirrhine primates. *Elife.* 2020;9. <https://doi.org/10.7554/eLife.63650>.
48. Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature.* 2014;354–7. <https://doi.org/10.1038/nature12961>.
49. Skov L, Coll Macià M, Sveinbjörnsson G, Mafessoni F, Lucotte EA, Einarsdóttir MS, et al. The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature.* 2020;582:78–83.
50. Petr M, Hajdinjak M, Fu Q, Essel E, Rougier H, Crevecoeur I, et al. The evolutionary history of Neanderthal and Denisovan Y chromosomes. *Science.* 2020;369:1653–6.

51. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. 2014;505:43–9.
52. Prüfer K, de Filippo C, Grote S, Mafessoni F, Korlević P, Hajdinjak M, et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*. 2017;358:655–8.
53. Harbers K, Francke U, Soriano P, Jaenisch R, Müller U. Structure and chromosomal mapping of a highly polymorphic repetitive DNA sequence from the pseudoautosomal region of the mouse sex chromosomes. *Cytogenet Cell Genet*. 1990;53:129–33.
54. Takahashi Y, Mitani K, Kuwabara K, Hayashi T, Niwa M, Miyashita N, et al. Methylation imprinting was observed of mouse mo-2 macrosatellite on the pseudoautosomal region but not on chromosome 9. *Chromosoma*. 1994;450–8. <https://doi.org/10.1007/s004120050054>.
55. Acquaviva L, Boekhout M, Karasu ME, Brick K, Pratto F, Li T, et al. Ensuring meiotic DNA break formation in the mouse pseudoautosomal region. *Nature*. 2020;582:426–31.
56. Pratto F, Brick K, Khil P, Smagulova F, Petukhova GV, Camerini-Otero RD. DNA recombination. Recombination initiation maps of individual human genomes. *Science*. 2014;346:1256442.
57. Schwartz JJ, Roach DJ, Thomas JH, Shendure J. Primate evolution of the recombination regulator PRDM9. *Nat Commun*. 2014;5:4370.
58. Lesecque Y, Glémin S, Lartillot N, Mouchiroud D, Duret L. The red queen model of recombination hotspots evolution in the light of archaic and modern human genomes. *PLoS Genet*. 2014;10:e1004790.
59. Altemose N, Noor N, Bitoun E, Tumian A, Imbeault M, Chapman JR, et al. A map of human PRDM9 binding provides evidence for novel behaviors of PRDM9 and other zinc-finger proteins in meiosis. *Elife*. 2017;6. <https://doi.org/10.7554/eLife.28383>.
60. Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature*. 2020;587:246–51.
61. Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, et al. Insights into hominid evolution from the gorilla genome sequence. *Nature*. 2012;483:169–75.
62. Besenbacher S, Hvilsum C, Marques-Bonet T, Mailund T, Schierup MH. Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nat Ecol Evol*. 2019;3:286–92.
63. Siepel A, Haussler D. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol Biol Evol*. 2004;21:468–88.
64. Ebersberger I, Metzler D, Schwarz C, Pääbo S. Genomewide comparison of DNA sequences between humans and chimpanzees. *Am J Hum Genet*. 2002;70:1490–7.
65. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491–8.
66. Duret L, Arndt PF. The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet*. 2008;4:e1000071.
67. Jónsson H, Sulem P, Kehr B, Kristmundsdóttir S, Zink F, Hjartarson E, et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature*. 2017;549:519–22.
68. Webb AJ. Meiotic recombination and linkage disequilibrium in the human genome; 2006.
69. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347:1260419.
70. Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, et al. Great ape genetic diversity and population history. *Nature*. 2013;499:471–5.
71. McVean GAT, Charlesworth B. A population genetic model for the evolution of synonymous codon usage: patterns and predictions. *Genet Res*. 1999;145–58. <https://doi.org/10.1017/s0016672399003912>.
72. Bergman J, Schierup MH. Population dynamics of GC-changing mutations in humans and great apes. *Genetics*. 2021. <https://doi.org/10.1093/genetics/iyab083>.
73. Glémin S, Arndt PF, Messer PW, Petrov D, Galtier N, Duret L. Quantification of GC-biased gene conversion in the human genome. *Genome Res*. 2015;25:1215–28.
74. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*. 2010;468:1053–60.
75. Auton A, Fledel-Alon A, Pfeifer S, Venn O, Ségurel L, Street T, et al. A fine-scale chimpanzee genetic map from population sequencing. *Science*. 2012;336:193–8.
76. Munch K, Mailund T, Dutheil JY, Schierup MH. A fine-scale recombination map of the human-chimpanzee ancestor reveals faster change in humans than in chimpanzees and a strong impact of GC-biased gene conversion. *Genome Res*. 2014;467–74. <https://doi.org/10.1101/gr.158469.113>.
77. McVean GAT, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. The fine-scale structure of recombination rate variation in the human genome. *Science*. 2004;304:581–4.
78. Spence JP, Song YS. Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Sci Adv*. 2019;5:eaaw9206.
79. Halldorsson BV, Hardarson MT, Kehr B, Styrkarsdóttir U, Gylfason A, Thorleifsson G, et al. The rate of meiotic gene conversion varies by sex and age. *Nat Genet*. 2016;48:1377–84.
80. Lachance J, Tishkoff SA. Biased gene conversion skews allele frequencies in human populations, increasing the disease burden of recessive alleles. *Am J Hum Genet*. 2014;95:408–20.
81. Borges R, Szöllösi GJ, Kosiol C. Quantifying GC-biased gene conversion in great ape genomes using polymorphism-aware models. *Genetics*. 2019;212:1321–36.
82. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
83. Yang C, Zhou Y, Marcus S, Formenti G, Bergeron LA, Song Z, et al. Evolutionary and biomedical insights from a marmoset diploid genome assembly. *Nature*. 2021:227–33. <https://doi.org/10.1038/s41586-021-03535-x>.
84. Zhou Q, Bachtrog D. Sex-specific adaptation drives early sex chromosome evolution in *Drosophila*. *Science*. 2012;337:341–5.



85. Rhesus Macaque Genome Sequencing and Analysis Consortium, Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, et al. Evolutionary and biomedical insights from the rhesus macaque genome. *Science*. 2007;316:222–34.
86. Kanthaswamy S, Ng J, Ross CT, Trask JS, Smith DG, Buffalo VS, et al. Identifying human-rhesus macaque gene orthologs using heterospecific SNP probes. *Genomics*. 2013;101:30–7.
87. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Brief Bioinform*. 2013;14:144–61.
88. de Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, et al. Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science*. 2016;354:477–81.
89. Nater N, Mattle-Greminger P, Nurcahyo N, Nowak G, Manuel D, Desai D, et al. Morphometric, behavioral, and genomic evidence for a new orangutan species (project): MorphoBank datasets; 2017. <https://doi.org/10.7934/p2591>.
90. Stevison LS, Woerner AE, Kidd JM, Kelley JL, Veeramah KR, McManus KF, et al. The time scale of recombination rate evolution in great apes. *Mol Biol Evol*. 2016;33:928–45.
91. Xue Y, Prado-Martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, et al. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science*. 2015;242–5. <https://doi.org/10.1126/science.aaa3952>.
92. Wirtschaftsuniversität Wien Department of Statistics and Mathematics. The R Project for statistical computing. 2008.
93. Vogl C, Bergman J. Inference of directional selection and mutation parameters assuming equilibrium. *Theor Popul Biol*. 2015;106:71–82.
94. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 2019;526–8. <https://doi.org/10.1093/bioinformatics/bty633>.
95. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics*. 2011:592–3. <https://doi.org/10.1093/bioinformatics/btq706>.
96. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27:573–80.
97. Nater A, Mattle-Greminger MP, Nurcahyo A, Nowak MG, de Manuel M, Desai T, et al. Morphometric, behavioral, and genomic evidence for a new orangutan species. *Curr Biol*. 2017;27:3487–98.e10.
98. Bergman J. jbergman/par1EvoDynamics: par1EvoDynamics. 2022; Available from: <https://zenodo.org/record/7116196>. Cited 2022 Sep 27.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

