OXFORD

Human Molecular Genetics, 2023, Vol. 32, 1, 65–78
https://doi.org/10.1093/hmg/ddac144
Advance access publication date 3 August 2022
Original Article

Disclosing complex mutational dynamics at a Y chromosome palindrome evolving through intra- and inter-chromosomal gene conversion

Maria Bonito[1,†], Francesco Ravasini[1,†], Andrea Novelletto[2], Eugenia D'Atanasio[3], Fulvio Cruciani[1,3,†] and Beniamino Trombetta[1,†,*]

[1]Department of Biology and Biotechnology 'Charles Darwin', Sapienza University of Rome, Laboratory affiliated to Istituto Pasteur Italia - Fondazione Cenci Bolognetti, Rome 00185, Italy
[2]Department of Biology, University of Rome Tor Vergata, Rome 00133, Italy
[3]Institute of Molecular Biology and Pathology (IBPM), CNR, Rome 00185, Italy
*To whom correspondence should be addressed at: Dipartimento di Biologia e Biotecnologie 'Charles Darwin', Sapienza Università di Roma, P.le Aldo Moro 5, Rome 00185, Italy. Tel: +39 0649912852; Email: beniamino.trombetta@uniroma1.it

[†]These authors contributed equally to this work.

Abstract

The human MSY ampliconic region is mainly composed of large duplicated sequences that are organized in eight palindromes (termed P1–P8), and may undergo arm-to-arm gene conversion. Although the importance of these elements is widely recognized, their evolutionary dynamics are still nuanced. Here, we focused on the P8 palindrome, which shows a complex evolutionary history, being involved in intra- and inter-chromosomal gene conversion. To disclose its evolutionary complexity, we performed a high-depth (50×) targeted next-generation sequencing of this element in 157 subjects belonging to the most divergent lineages of the Y chromosome tree. We found a total of 72 polymorphic paralogous sequence variants that have been exploited to identify 41 Y-Y gene conversion events that occurred during recent human history. Through our analysis, we were able to categorize P8 arms into three portions, whose molecular diversity was modelled by different evolutionary forces. Notably, the outer region of the palindrome is not involved in any gene conversion event and evolves exclusively through the action of mutational pressure. The inner region is affected by Y-Y gene conversion occurring at a rate of 1.52×10^{-5} conversions/base/year, with no bias towards the retention of the ancestral state of the sequence. In this portion, GC-biased gene conversion is counterbalanced by a mutational bias towards AT bases. Finally, the middle region of the arms, in addition to intra-chromosomal gene conversion, is involved in X-to-Y gene conversion (at a rate of 6.013×10^{-8} conversions/base/year) thus being a major force in the evolution of the *VCY/VCX* gene family.

Introduction

In many organisms, sex chromosomes are heteromorphic and participate in sex determination. It is widely recognized that mammalian sex-chromosomes evolved from a single pair of ancient autosomes which, because of the suppression of meiotic recombination, underwent morphological differentiation (1–3). As a consequence, the evolution of the heterogametic sex-specific chromosome has been characterized by rapid structural decay and loss of most ancestral genes (2,4).

The human Y chromosome is the most studied model of this evolutionary process (1,2). It is one of the shortest human chromosomes and can be structurally divided into two portions: 1) the Pseudoautosomal Regions (PAR1 and PAR2), where X-Y meiotic crossing-over is still active, and 2) the Male Specific region of the Y (MSY), where no meiotic recombination occurs. For this reason, the MSY has long been considered a recombinationally inert genetic element. This view radically changed with the discovery that the sequence landscape of this region can be modulated by inter- and intra-chromosomal gene conversion (5–13).

Gene conversion mainly affects the evolution of the ampliconic portion of the MSY (5). The main feature of this genomic portion is the presence of eight large palindromic sequences, termed P1-P8, each consisting of two nearly identical inverted repeats, the palindrome arms, separated by a short single-sequence spacer (14). Due to their structural organization, palindromes can be considered 'pseudo-diploid' (11,13); if a Paralogous Sequence Variant (PSV, i.e. a single nucleotide difference between the two palindrome arms) exists, a chromosome can be considered in a 'pseudo-heterozygous' state. The main effect of Y-Y gene conversion is to change the state of the genotype from 'pseudo-heterozygous' (e.g. A/C) to 'pseudo-homozygous' (A/A or C/C depending on the direction of the conversion event). Because of this strong homogenizing effect, palindrome arms share a sequence identity higher than 99.9%.

Received: April 14, 2022. **Revised:** June 21, 2022. **Accepted:** June 21, 2022
© The Author(s) 2022. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/hmg/article/32/1/65/6654770 by guest on 31 October 2024

Y-chromosome palindromes also exhibit a high number of multi-copy genes with a testis-specific expression, which are essential for sperm production and fertility (14–17). It has been initially suggested that arm-to-arm gene conversion may be necessary to protect the integrity of these genes (in absence of meiotic recombination) by restoring the ancestral state of mutations, thus preserving the sequence identity of Y palindromes (5,11,12). However, this conservative bias of gene conversion within palindromes has not been clearly confirmed yet. Indeed, it has been recently demonstrated that, in at least one singleton palindrome (P6), there is no bias towards the retention of the ancestral state at variable positions (13), pointing out that different palindromes can probably evolve through different molecular mechanisms and that, notwithstanding their importance, the study of the evolutionary dynamics of these particular elements is still in its infancy.

To deepen our knowledge about the evolutionary dynamics of palindromic sequences, we focused our attention on the P8 palindrome because: (i) it is the only singleton palindrome which contains genes (14); (ii) it evolves through the action of intra-chromosomal gene conversion (5) and (iii) it shares a high sequence identity with four gametologous regions on the X chromosome that may shape its genetic diversity through the action of X-to-Y gene conversion (7,10). How intra-chromosomal and X-to-Y gene conversion interact with each other and with mutational events in shaping the molecular evolution of P8 palindrome has yet to be clarified.

To this aim, we performed a high-depth (>50×) targeted next-generation sequencing (NGS) of palindrome P8 in 157 unrelated males whose phylogenetic relationships were previously determined through the analysis of non-duplicated MSY sequences (13). We used this phylogeny as an essential tool for investigating gene conversion in this portion of the genome as it allows fine mapping and timing of the mutation and gene conversion events (see Materials and methods).

Through this analysis we identified, within P8 arms, 72 PSVs, a figure higher than in previous studies (12), thus increasing our ability to understand the dynamics of the gene conversion (both Y-Y and X-Y) events during the recent human history. Within the P8 palindrome arms, we identified three discrete classes of sequences showing very different evolutionary paths. We highlighted and resolved the evolutionary complexity of P8 by demonstrating how different molecular mechanisms act differentially on different portions of the palindrome and how the modulation of these forces may influence the evolution of the palindromic structures of human MSY.

Results

By analysing the genetic diversity of about 3.3 Mb of the X-degenerate region of the human MSY, the phylogenetic relationships among 157 Y chromosomes (Supplementary Material, Table S1) have been previously reconstructed (13). The resulting phylogenetic tree (Supplementary Material, Fig. S1), based on 7240 variants (Supplementary Material, Table S3 in Bonito *et al.* (13)) was used in this study as a tool for mapping the mutational and gene conversion events revealed by high depth next-generation sequencing of the P8 palindrome in exactly the same samples.

Structure of the P8 palindrome in the reference sequence

In the human reference sequence, the P8 palindrome covers a genomic region of about 79 kb on the long arm of the Y chromosome (ChrY:16093532–16 172 355, assembly: GRCh37/hg19). More specifically, it is characterized by a length of 38 006 bp for the proximal arm (ChrY:16093532–16 131 537) and 37 404 bp for the distal one (ChrY:16134952–16 172 355), separated by a central spacer of 3414 bp (Fig. 1 and Supplementary Material, Table S2). From the alignment of the P8 arms, based on the reference sequence (GRCh37/hg19), it is possible to divide each arm into two separated portions with different Y-Y sequence similarities. At the outer boundaries of the palindrome, we identified two peculiar arm-segments, hereafter called Additional Flanking Regions (AFRs), covering ∼2.8 kb on the proximal arm and ∼2.2 kb on the distal one (Fig. 1 and Supplementary Material, Table S2). These portions exhibit a sequence similarity (90.8%) significantly lower ($P < 0.00001$, Fisher's exact test) than the rest of the palindrome arms, which show the highest Y-Y identity (99.997%) observed among all the human MSY palindromes (Supplementary Material, Fig. S2). The lower arm-to-arm similarity in the AFRs could be attributed to a limited activity of intra-chromosomal gene conversion in this part of the palindrome.

Two remarkable features of P8 are that it is the only singleton palindrome containing gene (*VCY*) and portions of each arm share an elevated sequence similarity with four gametologous portions of the X chromosome spanning from ∼7 to ∼15 kb (Supplementary Material, Table S3). Therefore, for subsequent analyses we divided the P8 palindrome arms into three discrete portions: the AFR, an XY gametologous region (g-XY, containing the VCY gene) and a non-gametologous one (Fig. 1 and Supplementary Material, Table S2).

Structural variation analysis within P8 palindrome

It is well recognized that palindromes may be involved in extensive structural rearrangements, including deletion or duplication of entire arms (17–20). By using specific primers (Supplementary Material, Table S4), we confirmed the presence of both proximal and distal arms in the whole sample set through the amplification of both the inner and outer boundaries of the arms in each sample.

Moreover, to detect deletion/duplication events within arms (or of entire arms), we performed an *in silico* depth analysis. By calculating the exponential moving average

Downloaded from https://academic.oup.com/hmg/article/32/1/65/6654770 by guest on 31 October 2024

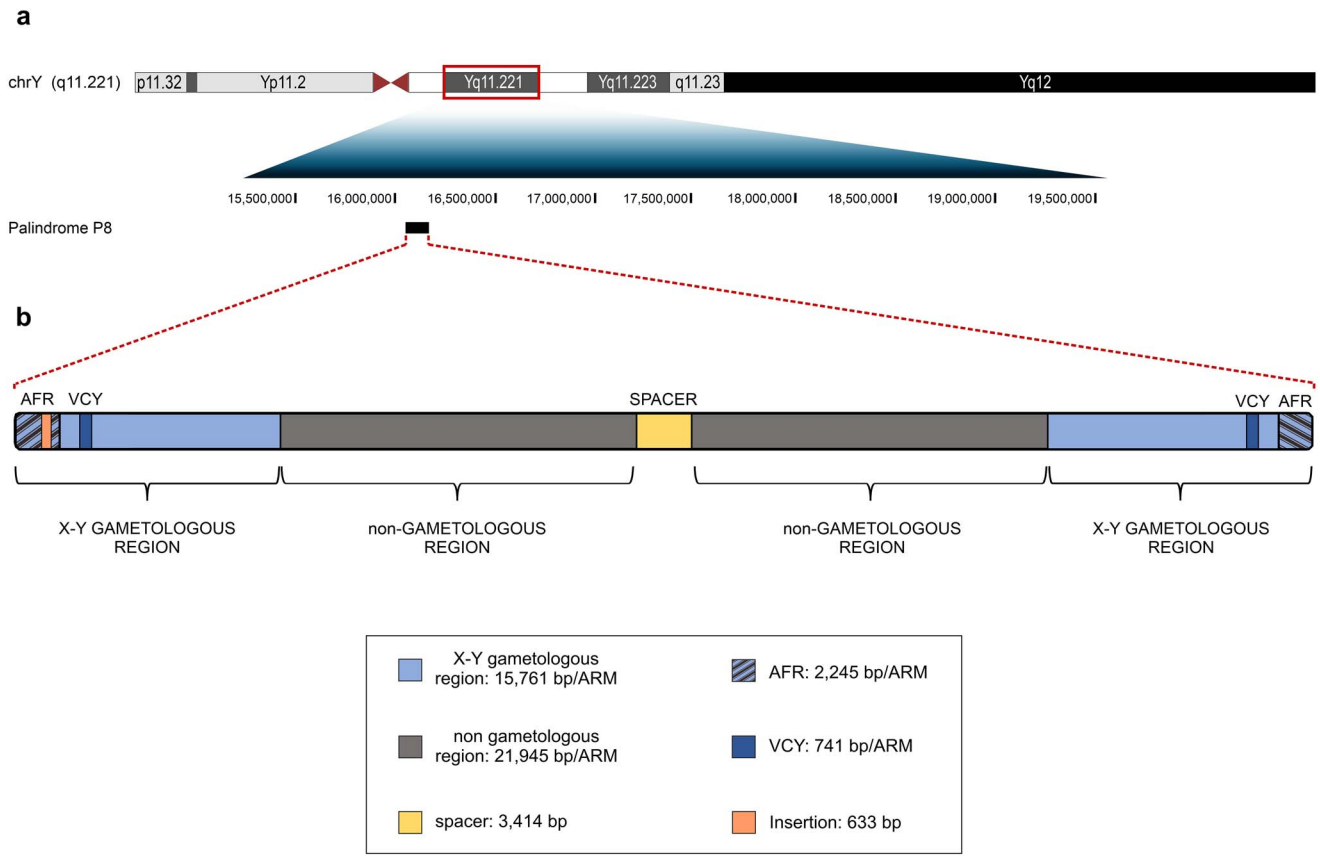

Figure 1. Structure of palindrome P8 (based on the reference sequence—GRCh37/hg19). (**A**) Ideogram of Y chromosome, showing positions of the P8 palindrome, with structure and coordinates. (**B**) Structure of the P8 palindrome. Each arm is made up of two portions: the non-gametologous region (grey), the XY gametologous region (g-XY, light blue). This region comprises the small *VCY* gene (dark blue) and a low Y-Y identity portion, named Additional Flanking Regions (AFRs, striped). The 633 bp insertion on the proximal arm is represented in orange.

(EMA) of the standardized sequencing depth values of the palindrome ((13); see Materials and methods), we inferred no duplications or deletions among our samples.

This result could seem to be at odds with the observation of eight copy number variants in P8 arms among 1216 samples (from the 1000 Genomes project) belonging to several branches of the Y chromosome tree, but these variations occurred in cell lines where somatic mutations cannot be excluded (17,19).

Through arm-to-arm alignment of P8 in the reference sequence, we observed a 633 bp sequence on the proximal arm, which is absent on the distal one (Supplementary Material, Fig. S2). Since we obtained sequencing data for this region from all the 157 samples analysed in this study, we know that this block is present in at least one arm of all our samples. However, it is virtually possible that this sequence is present in both arms, but it is not detectable because its sequencing reads will be wrongly mapped against the proximal arm of the reference, as a consequence of the short-read mapping issues previously described (13). For this reason, we tested the absence of the 633 bp sequence in the distal arm of all our samples through an arm-specific PCR (primers in Supplementary Material, Table S4) and we confirmed that the reference state (i.e. 633 bp insertion on the proximal arm and its absence on the distal one) pervades the Y chromosome phylogeny.

Genetic diversity of P8 palindrome

To study the genetic diversity of P8, we sequenced (at a depth higher than 50×) this palindrome in all the 157 samples previously analysed for P6 (13). After removing the interspersed repeated elements, we obtained, for each sample, sequencing data for a total of 34 677 bp from the arms (17 456 and 17 221 bp from proximal and distal arm, respectively) and 1886 bp from the haploid spacer (Supplementary Material, Table S5). Owing to the ambiguous read mapping recently described for palindromic sequences (13), we could not perfectly establish on which arm the mutation occurred for 'pseudo-heterozygous' samples, except for PSVs shared with the reference genome. However, this aspect did not affect the possibility to identify both the 'pseudo-heterozygous' genotypes and gene conversion events (see Materials and Methods).

Our deep-sequencing analysis revealed a total of 72 polymorphic PSVs across the phylogeny, six of which were already present in the reference sequence (Supplementary Material, Fig. S3, Supplementary Material, Table S6). Interestingly, V657 and V659 show a peculiar mutational pattern along the phylogeny, which is compatible

Downloaded from https://academic.oup.com/hmg/article/32/1/65/6654770 by guest on 31 October 2024

with a mutation that occurred on the stem lineage of the tree before the human Y chromosome radiation. Moreover, both sites show evidence of new mutational events occurred on the proximal arm in two different branches of the phylogeny (branch 52 for V657 and branch 75 for V659 as reported in Supplementary Material, Table S6). In addition, 7 PSVs (V119.1*, V675, V676, V109*, V683, V687 and V704) show evidence of recurrent mutations. Thus, the observed diversity of P8 arms can be explained by 81 mutational events in our phylogeny (Supplementary Material, Fig. S3, Supplementary Material, Table S6).

The mutational pattern of P8 palindrome arms is partially consistent with previous findings based on the analysis of the entire palindromic region (13,21). As expected, we found a higher number of transitions ($N = 56$) compared with transversions ($N = 28$), which resulted in a Ti/Tv ratio = 2.0, not significantly higher than the one observed in the X-degenerate region and similar to the value reported for the P6 palindrome arms (13). More generally, our value turned out to be similar to that of the whole ampliconic portion as reported by Helgason *et al.* (21).

Within the haploid spacer (about 2 kb sequenced) we detected a total of five variants across the 157 Y chromosomes-based phylogeny (Supplementary Material, Table S7).

Among the mutational events, we can differentiate the A or T (W) nucleotides changing in G or C (S) nucleotides, or vice versa. Recently, a sequence analysis of all the palindromes on a limited portion of the Y phylogeny revealed a mutational bias of the arms towards AT (12), but in the P6 palindrome arms it was revealed that this bias was artificial, because it was exclusively due to the hypermutability of the CpG sites (13). Interestingly, the P8 palindrome arms show two main differences in the mutational behaviour compared to P6. Firstly, we found a higher ($P = 0.0013$, Fisher's exact test) proportion of S-to-W substitutions (0.29) with respect to W-to-S mutations (0.13) (Table 1). This pattern can be interpreted as a real AT mutational bias of the P8 palindrome arms, because this excess remains after a correction for the hypermutable CpG sites (Table 1). Secondly, while in P6 it has been described an excess of S-to-W mutational events within the spacer compared to the arms (13), in P8 no mutational differences have been found between these two palindromic elements (Table 1).

Dynamics of Y-Y gene conversion in P8 palindrome

In our phylogeny we found that 16 out of the 72 identified PSVs (22%) showed footprints of gene conversion and that about half of them (7 PSVs) were affected by multiple events (Supplementary Material, Table S6, Supplementary Material, Fig. S3). By exploiting the MSY phylogeny (Supplementary Material, Fig. S1), we found a total of 41 Y-Y gene conversion events (Supplementary Material, Table S6, Supplementary Material, Fig. S3). Interestingly, in the AFRs we identified a total of 16 polymorphic PSVs, which showed no signals of gene conversion activity. The lack of arm-to-arm gene conversion in this portion of the palindrome could explain the increased Y-Y sequence diversity of ARFs as compared to the rest of the arms.

By mapping the events within the Y chromosome tree, we found 14 gene conversion events restoring the ancestral 'pseudo-homozygous' and 27 events fixing the derived 'pseudo-homozygous' genotype (Supplementary Material, Table S6 and Supplementary Material, Fig. S3). This observation is at odds with previous findings in which a significant excess of conversions restoring the ancestral state or no particular ancestral/derived bias have been observed in different palindromic elements (11–13). Differently, we observed a higher number of conversions generating the derived 'pseudo-homozygous' state. This weakly significant difference (14 vs 27, $P = 0.042$, Chi-square test) may suggest that Y-Y gene conversion in P8 palindrome is not involved in maintaining the ancestral state of sequences, but rather it is a molecular mechanism that would increase the evolutionary rate of the palindrome arms.

In this regard, we should note that Y-Y conversions towards the ancestral state are an underestimate of the actual number of events. This is because it is not possible to detect the to-ancestral events occurring exactly on the same branch where the mutation generating the PSV took place. Because of this, in order to investigate for a real ancestral/derived conversion bias in this palindrome, we performed a recalibration of the number of gene conversions by discarding the to-derived events which we would not have been observed if they had occurred towards the ancestral state (Supplementary Material, Table S8). By this approach, we discarded a total of 17 gene conversion events, resulting in a final number of 24 conversions, 14 of which towards the ancestral 'pseudo-homozygous' state. After this correction, the number of to-derived events considerably dropped and the difference between the two directions of gene conversion was not significant (14 vs 10, $P = 0.252$, Chi-square test). This result denies the former suggestion and is in line with what has been observed for another singleton palindrome of the human MSY (13), showing that no bias of gene conversion towards the ancestral or derived state is ongoing on at least two human MSY palindromes.

We also analysed the GC-biased gene conversion, i.e. the tendency towards the fixation of GC base pairs rather than AT in a gene conversion event. We found a total of 31 informative conversion events changing the GC content of the arms; 25 resulted in the fixation of GC whereas only 6 are involved in the conversion towards AT ($P = 6.4 \times 10^{-4}$, Chi-square test), suggesting a strong GC-biased gene conversion within this element. The existence of the GC-biased gene conversion raises the possibility that a tendency towards the retention of the ancestral state may actually exist but that it can be masked. It can happen when, for example, there is a greater number of events in which the derived base is represented by a G or a C. To test this hypothesis, we performed a

Downloaded from https://academic.oup.com/hmg/article/32/1/65/6654770 by guest on 31 October 2024

Table 1. Mutational behaviour of P8 palindrome

	S-to-W mut/GCnt (%)	W-to-S mut/ATnt (%)	ratio (S-to-W/W-to-S)[a]
With CpG			
Arms	38/13029 (0.29)	28/21648 (0.13)	2.23**
Spacer	3/717 (0.41)	0/1169 (0)	-
Spacer-arm ratio[a]	1.41	-	
Without CpG			
Arms	30/13021 (0.23)	-[b]	1.77*
Spacer	2/716 (0.28)	-[b]	-
Spacer-arm ratio[a]	1.22	-[b]	

Incidence and ratio of strong to weak (S-to-W) and weak to strong (W-to-S) mutations in P8 palindrome arms and spacer, considering (above) and not considering (below) CpG. [a]2×2 contingency table, Fisher Exact Test. *P-value < 0.05; **P-value < 0.01. [b]The correction for CpG sites does not affect the W-to-S mut/ATnt ratio, which exhibits the same values.

new ancestral/derived bias analysis by using 18 events towards GC bases, discarding from the 25 GC conversions all the derived events that we would not have been observed if they had occurred towards the ancestral ($N = 7$) (Supplementary Material, Table S9). Among these GC-biased events, the number of to-ancestral conversions ($N = 9$) was not different from the to-derived ones ($N = 9$), confirming the absence of an ancestral/derived gene conversion bias in P8. This result makes the GC fixation bias the unique driving force of the Y-Y recombination in this element, as also reported for the P6 palindrome (13).

Since we found no signals of gene conversion within the AFRs, we excluded this portion for the estimate of the arm-to-arm conversion rate.

By exploiting the mutation rate of the Y tree (13), we obtained an average Y-Y gene conversion rate of 1.52×10^{-5} conversions per base per year, ranging between a minimum value of 1.44×10^{-5} and a maximum value of 1.60×10^{-5} events per base per year. This value is significantly higher than the rate estimated for P6 palindrome (6.01×10^{-6}, $P < 0.0001$, test of comparison of two rates). Considering a 25-year human generation, this corresponds to a rate of 3.8×10^{-4} conversions per base per generation. Thus, in the transmission from father to son, we expect to have an average of 13 bases affected by gene conversion within the 35 kb of the P8 palindrome arm.

P8 palindrome mutation rate

The mutation rate calculated for palindrome arms (based on the observed number of mutational events) is probably an underestimate of the actual mutation rate, because it does not consider the mutations which generate new PSVs immediately converted to the ancestral state through gene conversion (13).

Through the approach described in Bonito *et al.* (13) (see Materials and methods), we calculated an arm mutation rate of 8.81×10^{-10} (SD $= 0.45 \times 10^{-10}$) mutations per base per year, which was significantly higher ($P = 0.0096$, test of comparison of two rates) than the one calculated for the arms of P6 (6.18×10^{-10}) (13). Interestingly, the mutational events seem to be unevenly distributed with an increased number of mutations in the portion of the palindrome arms that shares a sequence identity with the four gametologous regions on the X chromosome.

Thus, we recalculated a new mutation rate in the g-XY region corresponding to 11.12×10^{-10} (SD $= 0.57 \times 10^{-10}$) mutations per base per year, and a significantly lower rate of 6.25×10^{-10} (SD $= 0.32 \times 10^{-10}$) mutations per base per year in the non-gametologous one ($P = 0.0125$, test of comparison of two rates). It is worth noting that the mutation rate of the non-gametologous region is consistent with the mutation rate previously calculated for palindrome P6 ($P = 0.9560$, test of comparison of two rates), suggesting that this P8 portion evolves through the same mechanisms as P6. On the contrary, the higher mutation rate in the g-XY portion may suggest that the X-to-Y gene conversion may have a mutagenic effect on this part of the palindrome (see next paragraph).

Finally, we used the five variants identified in the P8 spacer to estimate a mutation rate for this region and we found an average of 9.0×10^{-10} mutations/base/year (SD $= 0.47 \times 10^{-10}$), which was consistent with the P6 spacer mutation rate (9.16×10^{-10}, $P = 0.9709$, test of comparison of two rates).

Dynamics of X-to-Y gene conversion in P8 palindrome

Although it is widely recognized that X-to-Y gene conversion may shape the genetic diversity of the *VCY* genes (7,10), the pervasiveness of this molecular process in the evolution of the entire palindrome has yet to be exhaustively analysed.

It is possible to find an X-to-Y gene conversion event exclusively if it acts on a region in which a Gametologous Sequence Variant (GSV, i.e. a single nucleotide difference between gametologous sequences) is present (see materials and methods). Interestingly, the main effect of this molecular mechanism will be to increase the similarity among gametologous regions (by erasing GSVs) and to decrease the arm-to-arm similarity by the introduction of new PSVs.

Within the g-XY regions we identified a total of 15 PSVs that have been introduced through the action of X-to-Y gene conversion. Following the criteria described

Downloaded from https://academic.oup.com/hmg/article/32/1/65/6654770 by guest on 31 October 2024

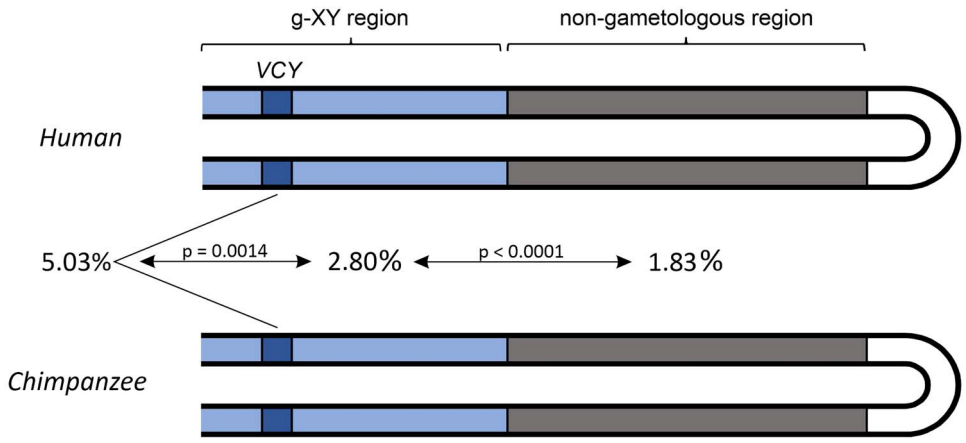

Figure 2. Human–chimpanzee sequence comparison in palindrome P8. Overview of sequence divergence between human and chimpanzee for different portions of the palindrome arms. The P8 palindromes (human and chimpanzee) are folded about the centre of the spacer. For each portion human-chimp divergence is reported (as a percentage). The significance between different divergence values is indicated by the arrows.

in Trombetta *et al.* (7), we identified a minimum of 21 independent X-to-Y gene conversion events, 3 of which involved multiple GSVs and with some GSVs that are interested in more than one event (Supplementary Material, Table S10, Supplementary Material, Fig. S3). These events occurred in the g-XY region excluding the AFR portion, which seems to evolve exclusively through the action of mutational pressure.

The observed minimum gene-conversion tract, measured as the nucleotide segment including the outermost converted GSVs, ranged from 1 to 14 bp (Supplementary Material, Table S10), whereas the maximum gene-conversion tract, measured as the distance between the two nearest non-converted GSVs flanking the converted site/s, ranged from 21 to 170 bp (Supplementary Material, Table S10).

It should be noted that the PSVs introduced by inter-chromosomal gene conversion are unevenly distributed, being significantly ($P < 0.00001$, Fisher's exact test) accumulated in the *VCY* gene (7/41—converted GSVs/Total GSVs—17% of the GSVs converted) compared to the remaining part of the g-XY region (13/849—converted GSVs/Total GSVs −1.6% of the GSVs converted). This result suggests that although X-to-Y gene conversion is active on the entire g-XY region its effect is significantly stronger on the *VCY* gene with respect to the rest of the palindrome.

Interestingly, by aligning P8 arms between human and chimpanzee, we observed a considerably higher human-chimpanzee sequence divergence in the g-XY region compared to the non-gametologous one (2.80% vs 1.83%, $P < 0.0001$, Fisher's Exact test) (Fig. 2) with the significantly higher inter-species divergence observed within the *VCY* genes (5.03%, $P = 0.0014$). These results suggest that X-to-Y gene conversion may be interpreted as an evolutionary force able to increase the evolutionary rate of this genomic portion.

Accordingly, by using the method described in Cruciani *et al.* (22), we obtained for the whole g-XY region an average X-to-Y gene conversion rate of 1.01×10^{-8} ($SD = 5.22 \times 10^{-10}$) events per base per year, that represents the probability per year that a site is involved in an X-Y gene conversion event (22). This is clearly an underestimate of the true value because gene conversion involving sites that are identical between X and Y would be undetectable.

Similarly, we selectively calculated the X-to-Y gene conversion rate for the *VCY* genes, resulting in a value of 6.013×10^{-8} ($SD = 3.12 \times 10^{-9}$) events per base per year. This value turned out to be significantly different ($P < 0.0001$, test of comparison of two rates) and one order of magnitude higher than the rate calculated in the remaining g-XY region (excluding *VCY*) ($5.57 \times 10^{-9} \pm SD = 2.88 \times 10^{-10}$). Thus, although the X-to-Y gene conversion is ongoing over the entire g-XY region, it is strongly active in the evolution of the *VCY*/*VCX* gene family.

Discussion

The euchromatic portion of the human MSY contains a high proportion of intra-chromosomal segmental duplications mainly organized into eight large palindromic structures named P1-P8, consisting of two repeated and inverted sequences (the palindrome arms) separated by a non-duplicated spacer (5,14). Interestingly, palindromes are not a peculiarity of the human Y chromosome, but they independently arose in the sex-specific haploid chromosome of several taxa and are overrepresented on the human X chromosome too (23–33). The presence of these pseudo-diploid elements within the haploid portions of the nuclear genome of several species suggests a fundamental biological significance for palindromic structures. Nevertheless, although some hypotheses have been proposed (29,34), the evolution and the functional roles of Y palindromes are not completely understood.

In general, Y palindromes exhibit an excess of multi-copy genes (with a tissue-specific expression in testes) that are essential for sperm production and fertility (5,14), so it has been proposed that the duplication and the following establishment of an arm-to-arm gene

Downloaded from https://academic.oup.com/hmg/article/32/1/65/6654770 by guest on 31 October 2024

conversion activity may have evolved to protect these fundamental genes against the genetic erosion that has characterized the evolution of the mammalian Y chromosome, owing to the lack of meiotic recombination (1,5,35,36). More precisely, it has been hypothesized that gene conversion could be a mechanism evolved to counteract the emergence of new mutations in important genes by conserving the ancestral state of gene sequences (1,5). According to this hypothesis, a *de novo* mutation on a paralog will be preferentially back-mutated to the ancestral state rather than being transmitted to the other arm through a gene conversion event. Some studies confirmed this hypothesis showing weak evidence that Y-Y gene conversion may be apparently biased towards the retention of the ancestral state of the variants (11,12). On the other hand, it has been demonstrated the lack of gene conversion bias in the maintenance of the ancestral state for at least the singleton palindrome P6 (13). Thus, if a gene conversion bias towards the ancestral state of the mutation exists in other MSY palindromes remains to be elucidated.

In order to clarify this issue, we aimed to deeply investigate the evolutionary dynamics of the only MSY singleton palindrome containing genes: P8.

Interestingly, through the analysis of the reference sequences, we were able to divide P8 arms in three different portions: a region showing a low Y-Y sequence identity (AFRs), a region homologous to four different portions of the X chromosome (g-XY) and a portion showing only Y-Y similarity (non-gametologous region) (Fig. 1). Given these peculiarities, we focused our attention on the action of three main evolutionary forces acting on P8 palindrome, i.e. mutation, intra- and inter-chromosomal gene conversion, and their interplay in the evolution of the whole palindromic portion.

To this aim, we used a robust phylogeny of 157 Y chromosomes and a high-depth sequencing to carry out an unbiased study on the evolution of the P8 palindrome, allowing us to compare it with P6 palindrome which has been already analysed for the same samples (13).

It is known that palindromic sequences can undergo copy number variations (17,20,37,38) and the presence of a different arm number with respect to the ancestral one (two arms per palindrome) may introduce distortions into the analysis of the evolution of these elements by mutation and gene conversion. Therefore, we firstly confirmed the presence of two arms of the palindrome in our samples by arm-specific PCRs and depth analysis.

Although previous studies that exploited the 1000 Genomes dataset showed that P8 arm copies may vary among different haplogroups (17,19), this observation is not in contrast with our findings. Indeed, among the 1216 samples of the 1000 Genomes Project, a duplication (or a deletion) of an entire P8 arm occurred exclusively eight times along the phylogeny (17). Moreover, it should be noted that, since the 1000 Genomes samples consist of cell lines, most of the copy number variations observed in the terminal branches of the Y chromosome phylogenetic tree (7 out of 8) may have originated as a consequence of somatic mutations (17). Thus, given that the probability to spot such events is low, it would have been difficult to observe arm copy variants in our smaller dataset. From these observations, we concluded that, in our phylogeny, palindrome P8 is as stable as P6.

The fine structural analysis of P8 revealed a 633 bp sequence specific to the proximal arm. After confirming the presence of this element only on the proximal arm of all our samples, we analysed its evolutionary conservation between human and chimpanzee. By performing a sequence alignment, we observed that chimpanzee palindrome P8 lacks the 633 bp block, suggesting that it probably appeared in the stem lineage of the human Y chromosome diversity. A BLAT analysis against the human reference genome (GRCh37/hg19) revealed that this element shares high similarity (94.7%) with one of the four gametologous sequences of the X chromosome (the one comprising *VCX3B* gene—chrX:8428059–8 438 764). Given that the evolution of this portion of the palindrome may be driven by intra-chromosomal gene conversion, it is tempting to speculate that a gene conversion event from the X to the Y chromosome may have possibly led to the inclusion of this fragment in the proximal arm of P8.

From the analysis of the genetic variability of palindrome P8, we identified 72 variable PSVs and 41 Y-Y gene conversion events involving about 22% of the PSVs. For P8 palindrome, we did not observe a significant difference between the number of the to-ancestral and to-derived gene conversions, suggesting the absence of the hypothesized gene conversion bias towards the ancestral state (5,11,12).

Interestingly, a preferential trend of Y-Y recombination emerged from the analysis of a bias towards the fixation of specific nucleotides. A gene conversion bias is expected when one paralog, bearing a particular variant state of the PSV, is more likely to act as a donor (or acceptor) sequence. In particular, the GC-biased gene conversion tends to favour the paralog bearing the G or C variant as a donor rather than the paralog with the A or T variant, which will act as an acceptor sequence (39–43). We observed this bias in P8 palindrome by detecting a significant excess of conversions fixing GC bases over AT. These results are in line with what has been recently observed for the human singleton palindrome P6 (13), confirming the absence of Y-Y gene conversion bias towards the ancestral state and that the unique bias of conversion in these palindromic sequences is the fixation of GC bases.

We then analysed the possible influence on the base content of the arms compared to the spacer in P8 (this study) and in P6 (13) owing to the GC-conversion bias we observed. As reported by Hallast *et al.* (11), there is a difference between P8 and P6 in the GC content: P6 has a significant excess of GC bases in the arms (38.78%) compared to the spacer (36.98%) ($P < 10^{-10}$ chi-square test), presumably due to the GC gene conversion bias. On

Downloaded from https://academic.oup.com/hmg/article/32/1/65/6654770 by guest on 31 October 2024

the contrary, this difference has not been observed in P8 (GC content 40.93% in arms vs 40.66% in spacer, $P = 0.78$ chi-square test).

The GC content depends on several factors, including mutation pressure. Probably, in P8 there is no increased GC content in the arms compared to spacer because there is a counterbalance effect due to the arm mutational bias towards AT (Table 1), which is completely absent in P6 (13). The AT-mutational bias can also explain the lack of differences in the mutational pattern between arms and spacer of P8.

By precisely knowing the evolutionary time of each branch of the Y tree (13) and the distribution of the PSVs within the phylogeny, we estimated an observed Y-Y gene conversion rate of 1.52×10^{-5} conversions per base per year, which turned to be about one order of magnitude higher than the Y-Y conversion rate (obtained with the same method and analysing the same samples) of P6 palindrome.

One major difference between P6 and P8 palindromes is that the latter includes the *VCY* gene, whereas P6 is a gene-free palindrome. The higher rate of Y-Y conversion in P8 could depend on the gene content: the presence of a gene in P8 may require a higher recombination rate for its structural and functional maintenance, whereas in P6 there could be a lower selection pressure in maintaining a high conversion rate. Whether, as a rule, gene-rich palindromes have a higher conversion rate compared to gene-free elements, remains to be elucidated.

Intra-chromosomal conversion is not equally distributed along the entire length of the arms. The AFRs are interested only by mutational pressure, are devoid of any recombination signature and evolve similarly to the palindrome spacer. The culling of recombination in this region may be due to several factors, but it should be mentioned that the AFRs are ~2.8 Kb long and they have on one side (externally) the unique sequence of MSY and on the other side (internally) a large structural difference between the arms (an insertion of 633 bp on the proximal arm). Probably, the presence of these two portions with no Y-Y similarity has influenced the inhibition of intra-chromosomal recombination in this portion of the palindrome.

Our estimated mutation rate for the P8 palindrome arms turned out to be significantly higher than the one for P6 (Fig. 3). To investigate whether the higher mutation rate of P8 is due to a higher mutational pressure caused by X-to-Y gene conversion, we separately evaluated the mutation rate of the g-XY region and the non-gametologous one, and found the former significantly higher than the latter (Fig. 3). Interestingly, the mutation rate of the non-gametologous region was statistically indistinguishable from the one of palindrome P6, suggesting similar mutational dynamics for these two structurally similar elements (Fig. 3) and underlying the mutagenic effect of inter-chromosomal gene conversion.

It is known that mutation introduces new differences between palindrome arms, whereas intra-chromosomal gene conversion dilutes this diversity (5,11,13), so it is possible to test if a mutation/conversion balance (which should maintain constant the average Y-Y diversity over time) has been established in P8 palindrome. By using the method described in Bonito *et al.* (13), it is possible to calculate the expected gene conversion rate assuming the existence of such a balance. Thus, we used a π average of 1.77×10^{-4} and a mutation rate of 8.25×10^{-10} ($SD = 0.43 \times 10^{-10}$) mutations per base per year, both specifically calculated for P8 arms (excluding the AFRs), and we obtained an expected gene conversion rate of 9.30×10^{-6} ($8.82–9.78 \times 10^{-6}$) events per duplicated nucleotide per year. Unlike P6 (13), this value is not consistent with the observed gene conversion rate calculated independently, suggesting that the equilibrium between mutation and gene conversion is not reached in palindrome P8. The lack of such an equilibrium is probably due to the introduction of new PSVs through the action of inter-chromosomal gene conversion.

We also investigated the dynamics of X-to-Y gene conversion along the P8 arms. The presence of an XY Gene Conversion Hotspot (GCH) in P8 (located within the *VCY* gene) has already been described (7,10), but the extent of this mechanism on the entire region has never been exhaustively investigated. We found 21 independent gene conversion events from the X chromosome that increased the Y-Y divergence through the introduction of PSVs. The gene-conversion-tract lengths here observed (mean of the maximum tract lengths across sites: 74 bp) are comparable with those previously obtained for other X-to-Y gene conversion hotspots: 118 bp at HSA (22), 47 bp at CERs (8) and 64 bp in the ARSDP pseudogene (7). The diversity of the entire g-XY region is shaped by the X chromosome acting as a donor sequence, but the conversion hotspot is located in the *VCY* gene which shows a gene conversion rate about 11 times higher than that reported for the rest of the g-XY portion (Fig. 3). Compared to other GCHs (10) located within the MSY, the *VCY* gene has the highest inter-chromosomal gene conversion rate. This does not necessarily reflect a more intense gene conversion activity, but can be the consequence of multiple *VCX* sequences acting as donors. The four sequences involved in *VCX*-to-*VCY* gene conversion could be a continuous source of X-Y GSVs. Conversely, since all the other MSY-GCHs can have only one donor sequence, conversion events may only decrease the number of GSVs.

Our estimates of X-to-Y gene conversion rate are considerably lower than the one here observed for Y-Y gene conversion, but similar or even higher than our estimates of P8 arms mutation rate (Fig. 3). Thus, it is clear that X-to-Y gene conversion can be highly effective in increasing the level of diversity in P8 palindrome arms. Although the higher *VCY* rate compared to the MSY-GCH may be due to the presence of multiple donor sequences, the difference between *VCY* and the rest of the g-XY region may suggest a possible functional role in preserving high similarity between X-Y gene copies in humans in a form

Downloaded from https://academic.oup.com/hmg/article/32/1/65/6654770 by guest on 31 October 2024

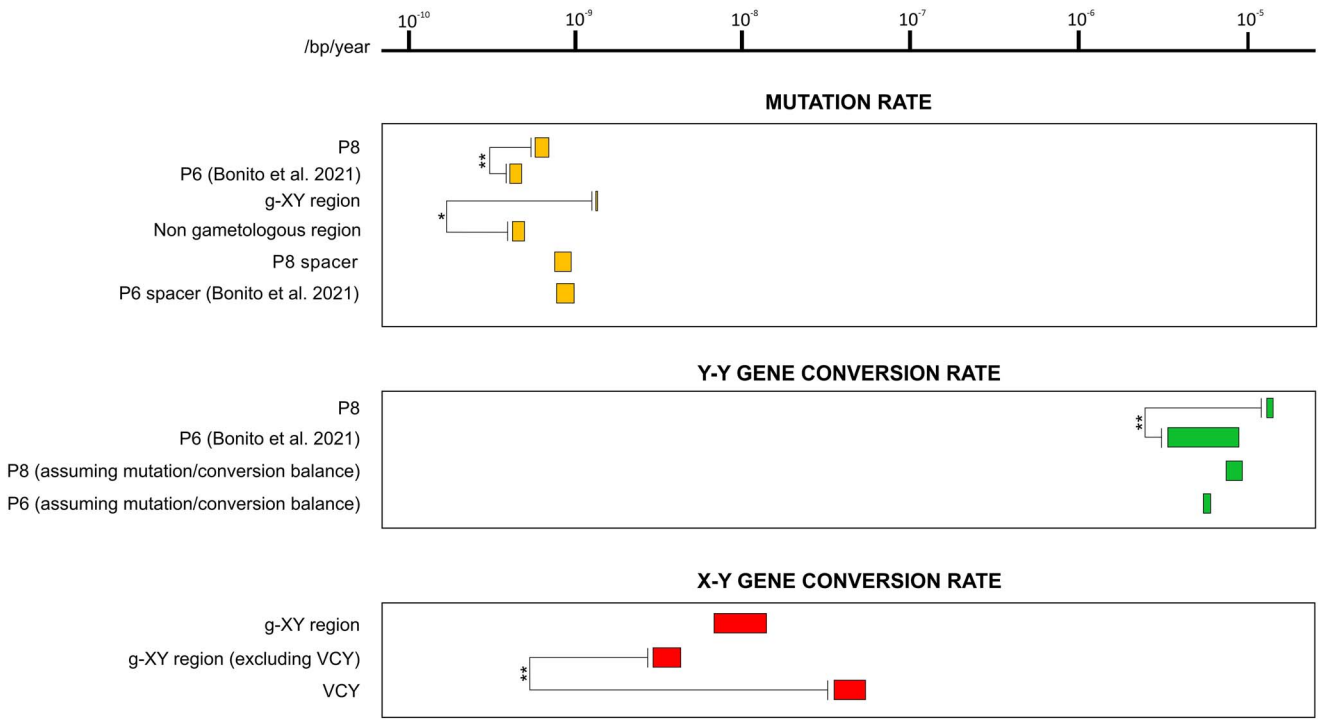

Figure 3. Comparison of the evolutionary forces acting on the P8 palindrome. Mutation rates (yellow), Y-Y gene conversion rates (green) and X-to-Y gene conversion rates (red) of the different portions of palindrome P8 and comparison with palindrome P6. Each value ranges from the minimum to the maximum, as reported in the text. For the XY conversion rate the bar spans three SDs around the mean value. The significance is shown between different rates. * indicates a *P*-value < 0.05, ** indicates a *P*-value < 0.01.

of concerted evolution. This is consistent with previous hypotheses predicting that members of the *VCX/VCY* family may work together by complementing each other in functions involved in spermatogenesis (44,45).

Recently it has been suggested that the *VCY* genes originated in the human-chimpanzee-bonobo common ancestor with the subsequent loss in the bonobo lineage (30). However, it seems that these genes preserve an essential function only in the human species, as they may be not functional in chimpanzees (19,46). Moreover, the sequence divergence that we observed between human and chimpanzee *VCY* (5.03%) (Fig. 2) is higher than the mean divergence observed among other orthologous Y chromosome ampliconic genes (47), supporting the hypothesis that these genes may have different evolutionary and functional histories in the two species.

The P8 palindrome may be considered a key model for studying the differences between intra- and inter-chromosomal gene conversion occurring in the same portion of MSY. Indeed, we can recognize two main differences between these molecular mechanisms. First, the gene conversion rate is much higher for Y-Y conversion than for the inter-chromosomal one (Fig. 3). Probably, this is observed because the activity of non-allelic gene conversion negatively correlates with the distance between interacting sequences; indeed, although not for the same portion of the genome, a higher frequency of intra-chromosomal as opposed to inter-chromosomal conversion has been already observed (9,48). Second, the length of the conversion tracts is considerably different. Probably due to the higher sequence divergence between gametologous regions with respect to paralogous ones, the X-to-Y conversion tracts seem to be extremely shorter than the Y-Y ones, which on average exceed one kilobase (11). A further possibility is that the two non-allelic conversions involve different molecular mechanisms that may lead to different tract lengths.

In summary, the P8 palindrome has a complex evolutionary history, being divided into several portions with different molecular forces acting on them. From an evolutionary point of view, AFRs and the spacer are the simplest portions because they evolve exclusively through the action of mutational pressure. On the contrary, the genetic diversity of the g-XY region is influenced by all three evolutionary forces acting on the palindrome: mutation, intra- and inter-chromosomal gene conversion. The observed diversity results from a balance between mutation and gene conversion, with X-to-Y events introducing differences between the arms (and also increasing genetic diversity) and Y-Y conversion reducing both intra-chromosomal and allelic diversity. The final effect is an increase in the evolutionary rate of this genomic element. Finally, the non-gametologous portion is the only region of P8 that behaves like P6: with the exception of mutational bias, the evolutionary dynamics acting on the two palindromes are very similar.

Materials and Methods

The sample

In this study we analysed the same 157 samples (Supplementary Material, Table S1), as in Bonito *et al.* (13).

Downloaded from https://academic.oup.com/hmg/article/32/1/65/6654770 by guest on 31 October 2024

Samples were chosen from our laboratory collection to maximize the haplogroup differentiation of the Y phylogeny. They were obtained from peripheral blood or buccal swab and DNA was extracted using appropriate procedures. For the same samples, we used the phylogenetic information already used by Bonito *et al.* (13) to map mutations and gene conversion events within palindrome P8. This study was approved by the 'Sapienza Università di Roma' ethical committee (protocol number 1158/13 and 496/13) and by the 'University of Tor Vergata' (protocol number 164/14) who considered the list of collaborators, anonymity of samples and the compliance with consent regulations. All the procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Phylogenetic tree

The maximum parsimony tree was reconstructed by following the criteria described in Bonito *et al.* (13). In summary, after generating a.meg input file we obtained the tree by using the MEGA software (49). Since we could not univocally define how many mutations were private of the A00 chromosome or occurred at A0-T branch (both branches indicated as branch 1 in Supplementary Material, Fig. S1), the root of the tree was positioned at midpoint by default. The Network software (50) was used to produce a median joining network of the samples, submitting a.rdf file as input, and to obtain the list of mutations for each branch and the positions of recurrent ones.

Analysis of the reference sequence

We retrieved the reference sequence of P8 (arms and spacer) from the UCSC genome browser (assembly GRCh37/hg19) using the 'segmental duplications' function of the browser. We then performed an arm-to-arm alignment by using Vista Lagan (51). We identified the gametologous sequences of the arms either by using the 'segmental duplications' function or by performing a BLAT analysis, both using the UCSC genome browser.

DNA quality control

To perform targeted NGS of P8 palindrome, we used a quantity ≥3 μg of DNA and we checked the quality parameters for each sample. We assessed the low amount of degradation by means of an electrophoretic run on a 1% agarose gel. A concentration higher than ≥37.5 ng/μl and a purity of A260/280 = 1.8–2.0 were monitored using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific).

Selection of palindromic regions to be sequenced.

The total number of bases selected was ∼34.6 kb/sample (∼32.8 kb of the arms and ∼1.8 kb of the spacer), after discarding the interspersed repeated elements (Supplementary Material, Table S5). For these selection steps, we used the 'Table browser' tool of the UCSC Genome browser, considering the aligned annotation tracks for the February 2009 (GRCh37/hg19) assembly.

Targeted NGS

Library preparation, targeting, sequencing and alignment steps were performed by BGI Tech (Hong Kong). The targeted P8 portions were enriched using a Roche Nimblegen capture array, composed of 200 bp probes overlapping the selected regions. The captured regions were loaded onto an Illumina Hi-Seq 2500 platform to produce 100 bp paired-end reads and a ≥50× mean depth sequences per sample. The raw output was refined discarding low-quality reads and contaminations with adapters. The sequences of each subject were aligned to the human reference genome (Human Feb. 2009—GRCh37/hg19 assembly) with BWA (Burrows-Wheeler Aligner) software (52). In the present study a target enrichment of haploid regions was performed, for which a sequencing depth (DP) equal to N is expected (13). However, due to the duplicated nature of palindrome arms, each read maps at the two different paralogous positions of P8 palindrome, resulting in a DP = 2 N, whereas a precise mapping for the sequenced reads of the spacer has been obtained (DP = N). The data underlying this article are incorporated into the online supplementary material. The alignment.bam files of palindrome P8 for the 157 Y chromosomes analysed here have been deposited in the European Nucleotide Archive (https://www.ebi.ac.uk/ena/) under the study accession number PRJEB52142.

Analysis of P8 structural variation.

To assess possible structural rearrangements involving an entire arm of P8, we performed a PCR for each boundary by using in-house primer pairs (Supplementary Material, Table S4).

To identify deletions/duplications within P8, we extracted the depth values from each sequenced position and we performed the standardized Exponential Moving Average (EMA) analysis as described in Bonito *et al.* (13). In summary, we extracted DP values from each sequenced position by means of SAMtools platform (53,54). We calculated the standardized DP value for each sample by using the average depth of the ∼3.3 Mb of the MSY non-repetitive regions. Then, we calculated the EMA in P8 with the 'TTR' package in R software, setting 100 bp windows sliding by 1 bp. To detect possible deletions and duplication, we specifically selected sequences showing EMA values lower than 1.5 and higher than 2.5, respectively. This is because within 'pseudo-diploid' arms we expect to observe standardized EMA values ∼2. From the arm-to-arm alignment of P8 in the reference sequence, we observed a 633 bp difference on the proximal AFR that is absent on the distal one. We tested the absence of this sequence in the distal arm of our samples through a distal arm-specific PCR (Supplementary Material, Fig. S4). More in detail, we used an in-house primer pair (Supplementary Material, Table S4), which amplifies a 200 bp sequence only if the 633 bp sequence is absent on the distal arm.

Downloaded from https://academic.oup.com/hmg/article/32/1/65/6654770 by guest on 31 October 2024

Variant calling and filtering

For the variant calling within P8 palindrome we used the mpileup command in SAMtools (53,54). We obtained a VCF (Variant Call Format) file for each sample, from which we removed the indels. Within the duplicated arms, to discard false-positive variants and to assess the genotype of true variants, we applied the criteria listed in Supplementary Material, Table S11 in Bonito *et al.* (13), set based on the 'pseudo-diploid' features of palindromic regions. These parameters took into account the total number of reads covering each position (DP), the number of reads calling the alternative base (DP_{ALT}) and the number of reads showing the reference base (DP_{REF}). We excluded all the variants showing $DP \geq 2$ and $DP_{ALT} \leq 2$, because of the 'pseudo-diploid' state of palindrome arms. Then, we discarded all the variants with $DP_{ALT}/DP_{REF} < 0.1$. Moreover, we defined the PD parameter as $[PD = (DP_{ALT})/(DP_{REF} + DP_{ALT})]$ to refine the list of real variants performing the subsequent filtering. We discarded positions with $PD < 0.1$ because they likely derive from false-positive calls. Positions with PD value ≥ 0.9 were considered as alternative 'pseudo-homozygous' genotype, whereas we assigned a 'pseudo-heterozygous' genotype to the positions with $0.4 \leq PD \leq 0.6$, because for these variants we have half of the calls as 'alternative' and half as 'reference'. The positions with a PD value out of these ranges were considered as variants that needed experimental validation.

We used the software IGVtools (55) to manually check (in the bam files) all the variants that were retained after the filtering steps. Finally, we considered different criteria to determine whether to keep or discard a variant. In particular, we analysed the phylogenetic context, the depth and the quality of the region where the variant is located and if two variants were closely spaced (because it may indicate a common origin through the same event, such as gene conversion). For the variants called in the haploid spacer, we exploited the filtering criteria used for the Y chromosome X-degenerate region as reported in D'Atanasio *et al.* (56).

Validation of variants

We validated the variant positions showing ambiguous genetic status using PCRs and Sanger sequencing. All markers have been amplified following a standard protocol of touchdown PCR. The amplification reaction was performed starting from 50/100 ng of genomic DNA. The 20-mer primers selected for both amplification and sequencing have been designed to specifically amplify the Y chromosome referring to the GRCh37/hg19 human genome sequence and using Primer3 v. 0.4.0. Software. The purification of the PCR products and the sequencing reaction were carried out at Eurofins srl in Milan (http://www.eurofins.it) or at Bio-Fab Research srl in Rome (http://www.biofabresearch.it). Fluorescent sequencing reactions were performed and run on an automatic Applied Biosystems 3730xl DNA Analyzer using 20-mer internal oligonucleotides as sequencing primers. The sequences obtained were aligned and compared with Sequencher v. 4.8 (Gene Codes Corporation) to establish the allelic variants. The primer list for sequencing and amplification is available upon request.

Detection of PSVs and Y-Y gene conversion events

In palindrome arms, a 'pseudo-heterozygous' state is modified into a 'pseudo-homozygous' one by a gene conversion event. Therefore, the possibility to detect these events is greatly influenced by the observation of the 'pseudo-heterozygous' state, i.e. a PSV within the examined sequences. However, it is important to note that the identification of a gene conversion event does not depend on which arm the mutation that generated the PSV occurred in. In this study we used a maximum parsimony approach to find the minimum number of mutation and gene conversion events, despite the fact that we recognize that several scenarios are possible to explain the observed genetic diversity.

The minimum number of mutations (generating new PSVs) and gene conversion events is given by mapping each event within the phylogeny, according to the criteria described in Supplementary Material, Fig. S3 in Bonito *et al.* (13).

In summary:

- We considered a single chromosome showing a PSV as the result of a single mutational event occurring on a palindrome arm of that chromosome. A phylogenetic clade of chromosomes showing the same PSV is indicative of a mutational event occurring at the branch joining all the interested chromosomes. On the contrary, a PSV shared between two or more phylogenetically unrelated chromosomes has been considered as generated by different mutational events, in this case we designed the PSVs occurring on different branches using a progressive number after the PSV name (Supplementary Material, Table S6). We inferred the ancestral/derived state of PSVs according to their phylogenetic context. For PSVs generated by mutations occurring on the basal branches of the phylogeny the ancestral state has been determined by the observation of the orthologous base on the chimpanzee (Clint_PTRv2/panTro6).
- The observation of 'pseudo-homozygous' chromosomes descending from the branch where the PSV arose is indicative that one or more gene conversion events occurred. To investigate the direction of these events (ancestral to derived or vice versa), we used the ancestral/derived state information of the PSV.
- The observation in the phylogeny of a site showing exclusively a derived 'pseudo-homozygous' state suggests that a mutational event generating a PSV and a subsequent gene conversion towards the derived state has occurred on the same branch of the phylogeny in a close time frame.
- It is important to note that some PSVs in palindrome P8 have been already identified in (7). These PSVs

Downloaded from https://academic.oup.com/hmg/article/32/1/65/6654770 by guest on 31 October 2024

have been reported in this study with their name in (7) followed by a '*' (Supplementary Material, Table S6 and Supplementary Material, Fig. S3).

Estimation of the Y-Y gene conversion rate

We used the method described in Bonito *et al.* (13) to estimate the Y-Y gene conversion rate of palindrome arms. We estimated the lifetime of each branch of the tree by multiplying the number of mutations associated with that specific branch by the average time in which a mutation event can occur (406.6 year/mutation, calculated in Bonito *et al.* (13)).

We calculated a P8-specific gene conversion rate (*c*), according to the following equation:

$$c = \frac{\sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} t_i}$$

where *C* is the number of the independent gene conversion events observed along the phylogeny which occurred within the *i*th PSV and *n* is the total number of PSVs identified within P8. The time of persistence of a single PSV within the phylogeny (*t*) is calculated as the sum of the times of all the branches (internal and terminal ones) in which the PSV is present and it is an estimate of the time in which a gene conversion event could be observed for each PSV. With these parameters, we estimated a minimum and a maximum time which resulted in a maximum and a minimum gene conversion rate, respectively. Our calculation is based on the reasoning reported in Supplementary Material, Fig. S4 in Bonito *et al.* (13). In summary, to calculate the maximum time (and a minimum rate of conversion), we included the branch(es) carrying the PSV and the branch(es) where the gene conversion event(s) occurred. For the estimate of the minimum time (and a maximum rate of conversion), we excluded the exact branch(es) where the PSV arose and where the gene conversion event(s) occurred.

The expected conversion rate (*c*) assuming the mutation/conversion steady-state balance has been calculated using the method of Rozen *et al.* (5), as follows:

$$c = \frac{2\mu}{d}$$

where μ is the specific mutation rate estimated for P8 arms and *d* is the observed divergence between palindrome arms calculated as the average arm-to-arm nucleotide diversity of the 157 sequenced chromosomes.

P8 mutation rate

We calculated the mutation rate of the different portions of the P8 palindrome using the following formula:

$$\mu = \frac{N}{t_{tot} \times bp}$$

where *N* is the total number of mutational events, t_{tot} the time that encompasses the entire phylogeny (calculated as the total number of mutations of the tree times the average elapsed time for a single mutation) and *bp* is the length of the sequenced region.

Identification of X-to-Y gene conversion events

Gene conversion between gametologs can be individuated exclusively if it involves a GSV (7,8,10). This is because the gametologous base on the donor sequence will change the base on the acceptor chromosome (6,7) and a new Y chromosome SNP will be observed in the population. Moreover, a new YY-PSVs will be generated. When gene conversion between sex chromosomes is ongoing, we will expect to find an excess of SNPs at X–Y GSV sites, where the Y-linked derived allele is the same as the gametologous sequence on the X chromosome. Critical points for the success of this analysis regard the correct inference about the direction of the mutation and the identification of an actual ancestral state of the acceptor sequence. The direction of the mutation for each Y-linked polymorphism was unambiguously determined by placing it in the context of the Y chromosome phylogenetic information obtained in this study (Supplementary Material, Fig. S1).

Comparing the ancestral Y sequences with the X chromosome, a site was considered to be a GSV whenever a difference was found between them. We arbitrarily chose not to consider sequences of more than five non-aligning contiguous bases as an X-Y GSV. For the entire g-XY region number of mutations falling in GSVs was calculated.

X-to-Y gene conversion rate estimate

To estimate the rate of the X-to-Y gene conversion within P8, we used a slightly modified version of the method described in Cruciani *et al.* (22):

$$C_{x-y} = \frac{1}{Lt} \sum_{i=1}^{n} l_i$$

where $C_{x\text{-}y}$ is the estimated rate of gene conversion per base per year, *n* is the number of observed gene conversion events, l_i the length in bp of the *i*th gene conversion event, *L* is the length in bp of the region under study, and *t* is the time that encompasses the entire phylogeny (calculated as the total number of mutations of the tree times the average elapsed time for a single mutation). We considered the length of the minimum and maximum converted tracts (Supplementary Material, Table S10) and we divided it by the whole time of the phylogeny (13).

Human–chimpanzee comparison

We calculated human–chimpanzee divergence between different orthologous sequences of palindrome P8 (g-XY region, non-gametologous region and the spacer) by

Downloaded from https://academic.oup.com/hmg/article/32/1/65/6654770 by guest on 31 October 2024

performing pairwise alignments with LAGAN (51) and estimating the proportion of different sites.

Supplementary Material

Supplementary Material are available at *HMGJ* online.

Acknowledgements

The authors are grateful to all the anonymous donors for providing DNA samples and to the people that contributed to the sample collection.

Conflict of Interest statement. The authors declare that they have no competing and financial interests.

Consent to Participate

An informed consent was obtained from all individual participants included in the study.

Ethics Approval

This study was approved by the 'Sapienza Università di Roma' ethical committee (protocol numbers 1158/13 and 496/13) and by 'University of Rome Tor Vergata' (protocol number 164/14) who considered the list of collaborators, anonymity of samples and the compliance with consent regulations. All the procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Funding

Istituto Pasteur-Fondazione Cenci Bolognetti, Programmi di Ricerca 2018–2020 to F.C. (grant number 60); and Sapienza Università di Roma, Progetti per Avvio alla Ricerca—Tipo 1 to M.B. (grant number AR11816430EA 868F).

References

1. Charlesworth, B. (2003) The organization and evolution of the human Y chromosome. *Genome Biol.*, **4**, 226–226.
2. Bachtrog, D. (2013) Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.*, **14**, 113–124.
3. Larson, E.L., Kopania, E. and Good, J.M. (2018) Spermatogenesis and the evolution of mammalian sex chromosomes. *Trends Genet.*, **34**, 722–732.
4. Ellegren, H. (2011) Emergence of male-biased genes on the chicken Z-chromosome: sex-chromosome contrasts between male and female heterogametic systems. *Genome Res.*, **21**, 2082–2086.
5. Rozen, S., Skaletsky, H., Marszalek, J.D., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K. and Page, D.C. (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature*, **423**, 873–876.
6. Rosser, Z.H., Balaresque, P. and Jobling, M.A. (2009) 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.
7. Trombetta, B., Cruciani, F., Underhill, P.A., Sellitto, D. and Scozzari, R. (2010) Footprints of X-to-Y gene conversion in recent human evolution. *Mol. Biol. Evol.*, **27**, 714–725.
8. Trombetta, B., Sellitto, D., Scozzari, R. and Cruciani, F. (2014) Inter- and intraspecies phylogenetic analyses reveal extensive X-Y gene conversion in the evolution of gametologous sequences of human sex chromosomes. *Mol. Biol. Evol.*, **31**, 208–223.
9. Trombetta, B., Fantini, G., D'Atanasio, E., Sellitto, D. and Cruciani, F. (2016) Evidence of extensive non-allelic gene conversion among LTR elements in the human genome. *Sci. Rep.*, **6**, 28710–28710.
10. Trombetta, B., D'Atanasio, E. and Cruciani, F. (2017) Patterns of inter-chromosomal gene conversion on the male-specific region of the human Y chromosome. *Front. Genet.*, **8**, 54.
11. Hallast, P., Balaresque, P., Bowden, G.R., Ballereau, S. and Jobling, M.A. (2013) Recombination dynamics of a human Y-chromosomal palindrome: rapid GC-biased gene conversion, multi-kilobase conversion tracts, and rare inversions. *PLoS Genet.*, **9**, e1003666.
12. Skov, L., The Danish Pan Genome Consortium and Schierup, M.H. (2017) Analysis of 62 hybrid assembled human Y chromosomes exposes rapid structural changes and high rates of gene conversion. *PLoS Genet.*, **13**, e1006834.
13. Bonito, M., D'Atanasio, E., Ravasini, F., Cariati, S., Finocchio, A., Novelletto, A., Trombetta, B. and Cruciani, F. (2021) New insights into the evolution of human Y chromosome palindromes through mutation and gene conversion. *Hum. Mol. Genet.*, **30**, 2272–2285.
14. Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T. *et al.* (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, **423**, 825–837.
15. Kuroda-Kawaguchi, T., Skaletsky, H., Brown, L.G., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Silber, S., Oates, R., Rozen, S. *et al.* (2001) The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat. Genet.*, **29**, 279–286.
16. Krausz, C. and Casamonti, E. (2017) Spermatogenic failure and the Y chromosome. *Hum. Genet.*, **136**, 637–655.
17. Teitz, L.S., Pyntikova, T., Skaletsky, H. and Page, D.C. (2018) Selection has countered high mutability to preserve the ancestral copy number of Y chromosome amplicons in diverse human lineages. *Am. J. Hum. Genet.*, **103**, 261–275.
18. Massaia, A. and Xue, Y. (2017) Human Y chromosome copy number variation in the next generation sequencing era and beyond. *Hum. Genet.*, **136**, 591–603.
19. Shi, W., Massaia, A., Louzada, S., Handsaker, J., Chow, W., McCarthy, S., Collins, J., Hallast, P., Howe, K., Church, DM., *et al.* (2019). Birth, expansion, and death of VCY-containing palindromes on the human Y chromosome. *Genome Biol.* **20**, 207.
20. Ravasini, F., D'Atanasio, E., Bonito, M., Bonucci, B., Della Rocca, C., Berti, A., Trombetta, B. and Cruciani, F. (2021) Sequence read depth analysis of a monophyletic cluster of Y chromosomes characterized by structural rearrangements in the AZFc region resulting in DYS448 deletion and DYF387S1 duplication. *Front. Genet.*, **12**, 669405.
21. Helgason, A., Einarsson, A.W., Guðmundsdóttir, V.B., Sigurðsson, Á., Gunnarsdóttir, E.D., Jagadeesan, A., Ebenesersdóttir, S.S., Kong, A. and Stefánsson, K. (2015) The Y-chromosome point mutation rate in humans. *Nat. Genet.*, **47**, 453–457.
22. Cruciani, F., Trombetta, B., Macaulay, V. and Scozzari, R. (2010) About the X-to-Y gene conversion rate. *Am. J. Hum. Genet.*, **86**, 495–498.

Downloaded from https://academic.oup.com/hmg/article/32/1/65/6654770 by guest on 31 October 2024

23. Ross, M.T., Grafham, D.V., Coffey, A.J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G.R., Burrows, C., Bird, C.P. *et al.* (2005) The DNA sequence of the human X chromosome. *Nature*, **434**, 325–337.
24. Davis, J.K., Thomas, P.J., Comparative Sequencing Program, N.I.S.C. and Thomas, J.W. (2010) A W-lined palindrome and gene conversion in new world sparrows and blackbirds. *Chromosom. Res.*, **18**, 543–553.
25. Méndez-Lago, M., Bergman, C.M., de Pablos, B., Tracey, A., Whitehead, S.L. and Villasante, A. (2011) A large palindrome with interchromosomal gene duplications in the pericentromeric region of the D. melanogaster Y chromosome. *Mol. Biol. Evol.*, **28**, 1967–1971.
26. Soh, Y.Q., Alföldi, J., Pyntikova, T., Brown, L.G., Graves, T., Minx, P.J., Fulton, R.S., Kremitzki, C., Koutseva, N., Mueller, J.L. *et al.* (2014) Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell*, **159**, 800–813.
27. Skinner, B.M., Sargent, C.A., Churcher, C., Hunt, T., Herrero, J., Loveland, J.E., Dunn, M., Louzada, S., Fu, B., Chow, W. *et al.* (2016) The pig X and Y chromosomes: structure, sequence, and evolution. *Genome Res.*, **26**, 130–139.
28. Tomaszkiewicz, M., Rangavittal, S., Cechova, M., Campos Sanchez, R., Fescemyer, H.W., Harris, R., Ye, D., O'Brien, P.C., Chikhi, R., Ryder, O.A. *et al.* (2016) 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.
29. Trombetta, B. and Cruciani, F. (2017) Y chromosome palindromes and gene conversion. *Hum. Genet.*, **136**, 605–619.
30. Cechova, M., Vegesna, R., Tomaszkiewicz, M., Harris, R.S., Chen, D., Rangavittal, S., Medvedev, P. and Makova, K.D. (2020) Dynamic evolution of great ape Y chromosomes. *Proc. Natl. Acad. Sci. U. S. A.*, **117**, 26273–26280.
31. Swanepoel, C.M., Gerlinger, E.R. and Mueller, J.L. (2020) Large X-linked palindromes undergo arm-to-arm gene conversion across mus lineages. *Mol. Biol. Evol.*, **37**, 1979–1985.
32. Zhou, R., Macaya-Sanz, D., Carlson, C.H., Schmutz, J., Jenkins, J.W., Kudrna, D., Sharma, A., Sandor, L., Shu, S., Barry, K. *et al.* (2020) A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biol.*, **21**, 38.
33. Jackson, E.K., Bellott, D.W., Cho, T.J., Skaletsky, H., Hughes, J.F., Pyntikova, T. and Page, D.C. (2021) Large palindromes on the primate X chromosome are preserved by natural selection. *Genome Res.*, **31**, 1337–1352.
34. Betrán, E., Demuth, J.P. and Williford, A. (2012) Why chromosome palindromes? *Int. J. Evol. Biol.*, **2012**, 1–14.
35. Connallon, T. and Clark, A.G. (2010) Gene duplication, gene conversion and the evolution of the Y chromosome. *Genetics*, **186**, 277–286.
36. Marais, G.A.B., Campos, P.R.A. and Gordo, I. (2010) Can intra-Y gene conversion oppose the degeneration of the human Y chromosome? A simulation study. *Genome Biol. Evol.*, **2**, 347–357.
37. Repping, S., Skaletsky, H., Brown, L., van Daalen, S.K., Korver, C.M., Pyntikova, T., Kuroda-Kawaguchi, T., de Vries, J.W., Oates, R.D., Silber, S. *et al.* (2003) Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nat. Genet.*, **35**, 247–251.
38. Hughes, J.F. and Rozen, S. (2012) Genomics and genetics of human and primate y chromosomes. *Annu. Rev. Genomics Hum. Genet.*, **13**, 83–108.
39. Galtier, N. (2003) Gene conversion drives GC content evolution in mammalian histones. *Trends Genet.*, **19**, 65–68.
40. Kudla, G., Helwak, A. and Lipinski, L. (2004) Gene conversion and GC-content evolution in mammalian Hsp70. *Mol. Biol. Evol.*, **21**, 1438–1444.
41. Duret, L. and Galtier, N. (2009) Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genomics Hum. Genet.*, **10**, 285–311.
42. Dutta, R., Saha-Mandal, A., Cheng, X., Qiu, S., Serpen, J., Fedorova, L. and Fedorov, A. (2018) 1000 human genomes carry widespread signatures of GC biased gene conversion. *BMC Genomics*, **19**, 256.
43. Jackson, E., Bellott, D.W., Skaletsky, H. and Page, D.C. (2021) GC-biased gene conversion in X-chromosome palindromes conserved in human, chimpanzee, and rhesus macaque. *G3 (Bethesda)*, **11**, jkab224.
44. Lahn, B.T. and Page, D.C. (2000) A human sex-chromosomal gene family expressed in male germ cells and encoding variably charged proteins. *Hum. Mol. Genet.*, **9**, 311–319.
45. Van Esch, H., Hollanders, K., Badisco, L., Melotte, C., Van Hummelen, P., Vermeesch, J.R., Devriendt, K., Fryns, J.P., Marynen, P. and Froyen, G. (2005) Deletion of VCX-A due to NAHR plays a major role in the occurrence of mental retardation in patients with X-linked ichthyosis. *Hum. Mol. Genet.*, **14**, 1795–1803.
46. Kuroki, Y., Toyoda, A., Noguchi, H., Taylor, T.D., Itoh, T., Kim, D.S., Kim, D.W., Choi, S.H., Kim, I.C., Choi, H.H. *et al.* (2006) Comparative analysis of chimpanzee and human Y chromosomes unveils complex evolutionary pathway. *Nat. Genet.*, **38**, 158–167.
47. Hughes, J.F., Skaletsky, H., Pyntikova, T., Graves, T.A., van Daalen, S.K.M., Minx, P.J., Fulton, R.S., McGrath, S.D., Locke, D.P., Friedman, C. *et al.* (2010) Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature*, **463**, 536–539.
48. Ezawa, K. and OOta, S., Saitou, N. (2006) Genome-wide search of gene conversions in duplicated genes of mouse and rat. *Mol. Biol. Evol.*, **23**, 927–940.
49. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, **28**, 2731–2739.
50. Bandelt, H.J., Forster, P. and Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.*, **16**, 37–48.
51. Brudno, M., Do, C.B., Cooper, G.M., Kim, M.F., Davydov, E., Comparative Sequencing Program, N.I.S.C., Green, E.D., Sidow, A. and Batzoglou, S. (2003) LAGAN and multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.*, **13**, 721–731.
52. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows–wheeler transform. *Bioinformatics*, **25**, 1754–1760.
53. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
54. Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**, 2987–2993.
55. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.
56. D'Atanasio, E., Trombetta, B., Bonito, M., Finocchio, A., Di Vito, G., Seghizzi, M., Romano, R., Russo, G., Paganotti, G.M., Watson, E. *et al.* (2018) The peopling of the last Green Sahara revealed by high-coverage resequencing of trans-Saharan patrilineages. *Genome Biol.*, **19**, 20.

Downloaded from https://academic.oup.com/hmg/article/32/1/65/6654770 by guest on 31 October 2024