Can Intra-Y Gene Conversion Oppose the Degeneration of the Human Y Chromosome? A Simulation Study

Gabriel A. B. Marais*[1], Paulo R. A. Campos[2], and Isabel Gordo[3]

[1]Université de Lyon; Université Lyon 1; Centre National de la Recherche Scientifique; UMR5558, Laboratoire de Biométrie et Biologie évolutive; Villeurbanne, France

[2]Departamento de Física, Universidade Federal Rural de Pernambuco, Dois Irmãos, Recife-PE, Brazil

[3]Instituto Gulbenkian de Ciência, Oeiras, Portugal

*Corresponding author: E-mail: marais@biomserv.univ-lyon1.fr.

Abstract

The human Y is a genetically degenerate chromosome, which has lost about 97% of the genes originally present. Most of the remaining human Y genes are in large duplicated segments (ampliconic regions) undergoing intense Y–Y gene conversion. It has been suggested that Y–Y gene conversion may help these genes getting rid of deleterious mutations that would inactivate them otherwise. Here, we tested this idea by simulating the evolution of degenerating Y chromosomes with or without gene conversion using the most up-to-date population genetics parameters for humans. We followed the fate of a variant with Y–Y gene conversion in a population of Y chromosomes where Y–Y gene conversion is originally absent. We found that this variant gets fixed more frequently than the neutral expectation, which supports the idea that gene conversion is beneficial for a degenerating Y chromosome. Interestingly, a very high rate of gene conversion is needed for an effect of gene conversion to be observed. This suggests that high levels of Y-Y gene conversion observed in humans may have been selected to oppose the Y degeneration. We also studied with a similar approach the evolution of ampliconic regions on the Y chromosomes and found that the fixation of many copies at once is unlikely, which suggest these regions probably evolved gradually unless selection for increased dosage favored large-scale duplication events. Exploring the parameter space showed that Y–Y gene conversion may be beneficial in most mammalian species, which is consistent with recent data in chimpanzees and mice.

Key words: ampliconic genes, gene conversion, sex chromosomes, Muller's ratchet, recombination modifier.

Introduction

The human sex chromosomes originated ~150 millions years ago from a pair of regular autosomes after the emergence of the male determinant gene *Sry* (Lahn and Page 1999; Veyrunes et al. 2008). The Y chromosome became recombinationally inert in several steps—possibly involving inversions on the Y—and has currently only two small regions called pseudoautosomal that concentrate all the meiotic crossing-over events in males (Lahn and Page 1999). The suppression of crossing-over on the Y probably evolved with accumulation of sex-antagonistic (good for male and bad for female) genes on that chromosome (Charlesworth et al. 2005). The absence of crossing-over has been widely studied theoretically and it leads to inefficient selection and reduced polymorphism, the so-called Hill–Robertson effects

(for a review see Charlesworth B and Charlesworth D 2000 and Gordo and Charlesworth 2001). The sequencing and annotation of the human X and Y chromosomes, respectively, in 2005 and 2003 revealed the extent of the effects of the absence of crossing-over on the Y. The comparison with the X showed that the male-specific region of the Y (MSY) lost about 97% of the genes initially present. It accumulated large amounts of repeats (~80% of its current DNA content), which mostly turn into heterochromatin (~60% of the MSY), and was widely rearranged. The pseudoautosomal regions (PAR), however, have perfectly normal characteristics compared with the rest of the genome (Skaletsky et al. 2003; Ross et al. 2005).

A big surprise came out of the sequencing of the human Y chromosome. In the MSY region, most of the genes

belong to nine gene families. These genes—called ampliconic genes—undergo gene conversion within each gene family, which was unexpected because MSY was believed to be recombinationally inert and gene conversion is a form of recombination (Rozen et al. 2003). Moreover, based on the comparison of interparalogues divergence and human–chimp divergence, they could estimate the level of gene conversion and it was found to be extraordinarily high, about 1,000-fold the genome average (Rozen et al. 2003; Bosch et al. 2004). These nine gene families show structures that seem to promote gene conversion: many of them are included in large palindromes and others in inverted repeats and tandem arrays, which could explain the very high rate observed (Rozen et al. 2003). Although some cases of translocation from autosomes to the Y have been described (Bhowmick et al. 2007; Yu et al. 2008), most of the ampliconic genes have originated from one parental gene that was already present on the autosomal pair that gave rise to the human XY (Bhowmick et al. 2007). They thus amplified specifically on the Y and gave sometimes very large families (one of them has 35 copies). Interestingly, all the ampliconic genes are expressed in testis and many of them are involved in spermatogenesis (Skaletsky et al. 2003; Rozen et al. 2003). All this suggests that duplication and gene conversion may have protected these genes (important for male functions) from degeneration (Charlesworth 2003; Hawley 2003; Rozen et al. 2003). A recent study suggests that this beneficial effect may have been strong because gene conversion is associated with some cost caused by unequal crossing over resulting in gene loss (Lange et al. 2009).

The amplification seems to have occurred independently in several lineages because the genes that were amplified are not the same in humans, chimpanzees, and mice, which is consistent with the degeneration of the Y being an ongoing process that started 150 Ma and continued independently in the different mammalian lineages (Bhowmick et al. 2007; Alföldi 2008; Hughes et al. 2010). The chimpanzee ampliconic regions are larger than the human ones and it has been suggested that this could be because selection (through sperm competition) is stronger in chimps and this may have caused stronger Hill–Robertson effects, and hence the need for a more efficient protection (Hughes et al. 2010). Several cases of gene conversion between X and Y copies of some sex-linked genes have also been reported in the literature (Pecon Slattery et al. 2000; Iwase et al. 2003; Marais and Galtier 2003; Rosser et al. 2009). There is thus a growing belief that gene conversion could slow down degeneration of Y-linked genes. However, nonrecombining regions of the genome are expected to accumulate duplicate genes because of inefficient selection to eliminate quickly duplicates due to small population size (Lynch and Conery 2003; Lynch 2007), and the idea that gene conversion can oppose Y degeneration has been criticized (Graves 2004). It

is therefore unclear whether the evolution of ampliconic regions has been driven by positive selection or drift. Earlier works on the effect of gene conversion on the molecular evolution of gene families have been done (Nagylaki and Petes 1982; Nagylaki 1984; Walsh 1985; Nagylaki and Barton 1986; Hurst and Smith 1998), but these have not included Hill–Robertson effects and are not directly relevant for the evolution of the Y chromosome. We therefore lack theoretical work on whether gene conversion can oppose Y degeneration at all.

Here, we used a theoretical framework for studying the evolution of recombination modifiers (Keightley and Otto 2006; Gordo and Campos 2008) and adapted it for gene conversion modifiers in order to understand under which conditions a gene conversion modifier could be advantageous. We focused on Y–Y gene conversion because current data suggest that Y–Y gene conversion is much more frequent than X–Y gene conversion and has potentially a stronger impact on Y chromosome evolution. We considered a haploid population of Y chromosomes evolving under the Wright–Fisher model, and introduced a Y variant to study its fate through Monte Carlo simulations with varying population genetics parameters (e.g., population size, deleterious mutation rate, and gene conversion rate). Mutations effects on fitness are drawn from a gamma distribution, as suggested from recent data (Keightley and Eyre-Walker 2007). Importantly, we gathered from the literature realistic values for these parameters in humans to clearly test whether Y variants could get fixed with higher probability than that expected from a random neutral process in humans. We ran three distinct sets of simulations. In a first step, we introduced a variant with gene conversion in a population of Y with 1/3 of duplicate genes (as in the human Y), and we found that gene conversion is clearly advantageous when considering biologically reasonable parameters for humans. In the second set of simulations, we introduced a single variant with duplicates in a population of Y chromosomes carrying only single-copy genes and found that large duplicates are not easily fixed in a population under parameters reasonable for humans. We therefore suggest that copies are more likely to get fixed one after another to ultimately reach a large copy number because this scenario involves the fixation of nearly neutral events and is more likely than a one-step amplification. In the last set of simulations, we introduced a single variant with both duplication and gene conversion in a population of Y with single-copy genes and found that selection against duplication was fairly strong and overcame the selective advantage of gene conversion for large-scale duplication events, which again supported the idea that the ampliconic regions observed in humans must have evolved step by step. Taking this and variations around these simulations, we discuss how ampliconic regions could have evolved in humans and other mammals.

Material and Methods

Model for the Gene Conversion Modifier

In our forward Monte Carlo computer simulations, the population included N haploid individuals (Y chromosomes), and evolved under the standard Wright–Fisher model, of constant population size and no overlapping generations. Each individual was represented by a Y chromosome with L genes. We assumed that a given fraction of genes, f_{GC}, had a certain number of duplicates, C. Thus, the entire Y chromosome included $G = L + f_{GC} \cdot L (C - 1)$ loci. Gene conversion occurred only between duplicates of a given gene. Mutations occurred during reproduction and were assumed to follow a Poisson distribution, as commonly done in population genetics. Mutations appeared at a constant rate per Y chromosome per generation, U (see below). All mutations were considered deleterious and the effect on fitness of a mutation depended on the position where it occurred: if it occurred on a single-copy gene then its effect was $-s_d$; if it took place in a multicopy gene it had a smaller effect $-s_d/C$, where C is the number of copies of that gene. In any case, s_d was randomly generated from a gamma distribution:

$$g(x; \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha - 1} e^{-\beta x} \quad \text{for } x > 0$$

of fixed mean value α/β, where α is the shape parameter and β is the scale parameter. After mutation, selection occurred and those most adapted individuals had a greater chance of producing more offspring at the next generation. The fitness of an individual with a Y chromosome carrying k deleterious mutations on single-copy genes and m deleterious mutations in duplicated genes is given by:

$$w_{km} = \prod_{i=1}^{k} (1 - s_{d_i}) \prod_{j=1}^{m} \left(1 - \frac{1}{C} s_{d_j}\right)$$

where s_{d_i} and s_{d_j} are taken from a gamma distribution as indicated above.

Initially, the population evolved in the absence of gene conversion, that is, only under mutation, selection, and genetic drift, for a period of T_{eq} generations, where an equilibrium regime was reached (T_{eq} was set to N, which is sufficient to reach an equilibrium, data not shown). Then, an individual was chosen at random to carry the modifier allele, which allowed it to perform gene conversion. Most of the time, this individual went extinct stochastically, but occasionally it spread through the population to fixation, generating a population that had acquired the process of gene conversion. During the gene conversion phase, each individual carrying the modifier allele of gene conversion was affected by gene conversion at rate r_{GC}. During a gene conversion event, a duplicated gene was randomly selected and the mutations of a given locus, among the C available copies of that gene, were copied to one of the $C - 1$ remain-

ing ones. For each set of parameter values, we ran 20-times N simulations, where the fate of the modifier allele was followed. We then counted the number of simulations where gene conversion modifier was able to invade and fix and compared that number with the one expected under neutrality, where a random mutation has a likelihood of $1/N$ to become fixed. This way, we studied the conditions under which gene conversion is selected for.

Model for the Variant with Duplication

In the aforementioned model, we assumed that in the initial population of Y chromosomes a given fraction of the genes were duplicated. This allowed us to study the advantage of the gene conversion mechanism by itself, which required the existence of duplicated genes. In another set of simulations, we analyzed the sole effect of duplication. We asked under which conditions a duplication event was deleterious, neutral or even advantageous when it occurred in a population of Y chromosomes that was experiencing the evolutionary forces of mutation, selection, and drift. In this model, the initial population of Y chromosomes included genomes with single-copy genes only, experiencing, as before, mutation and selection against deleterious mutations but no gene conversion. The population consisted initially of haploid individuals with genome size L that evolved up to reaching an equilibrium between mutation, selection, and drift. We then introduced a variant with duplicated genes and a genome size of $G = L + f_{GC} \cdot L \cdot (C - 1)$. The mutation effects on the duplicated genes were reduced by a factor $1/C$, as previously. In order to keep the same value for the rate of mutation per gene, the genomic mutation rate of the variant was increased by a factor G/L, in comparison with genomes with single-copy genes only. We again ran 20-times N simulations and estimated the probability of fixation of the variant carrying a duplication as before.

Model for Modifier with Both Duplication and Gene Conversion

The simulation started with every individual consisting of genomes of same size L, and evolved under the Wright–Fisher model up to reaching a steady state regime. Then, a variant with duplication (genome size $G = L + f_{GC} \cdot L \cdot (C - 1)$) and gene conversion (rate r_{GC}) was introduced. This variant evolved up to either its loss or fixation in the population, and its probability of successfully invading was estimated as before.

Parameter Estimates for Humans

We collected from the literature values for the different parameters of the model (table 1). The effective population size is about 10,000 in humans (Eyre-Walker et al. 2002; Yu et al. 2002, 2003; Keightley et al. 2005), which gives 2,500 for the Y chromosome given that Y chromosomes

Table 1
Human Values for the Relevant Parameters of the Model

Y Population Genetics Parameters	Humans	References
N = population size[a]	2,500	Eyre-Walker et al. (2002), Yu et al. (2002), Yu et al. (2003), Keightley et al. (2005)
L = gene number	27	Skaletsky et al. (2003)
f_{GC} = fraction of duplicate genes	1/3	Skaletsky et al. (2003)
C = copy number	7	Skaletsky et al. (2003)
r_{GC} = gene conversion rate (per gene)[b]	0.4	Rozen et al. (2003)
U = deleterious mutation rate[c]	0.01	Nachman and Crowell (2000), Makova and Li (2002), Skaletsky et al. (2003)
s_d = selection coefficient[d]	α = 0.2, β = 2.5	Keightley and Eyre-Walker (2007)

[a] We took the estimated species population size (Eyre-Walker et al. 2002; Yu et al. 2002, 2003; Keightley et al. 2005) and divided it by 4 to get the Y population size.

[b] We took the gene conversion rate per nucleotide per generation (Rozen et al. 2003) and multiplied it by the average Y gene size (~2,000 bp) from ENSEMBL.

[c] We took the genome-wide deleterious mutation rate U_{tot} = 3 (Nachman and Crowell 2000) and divided it by the total number of genes in the human genome (~20,000 from ENSEMBL) to get u_{tot} the deleterious mutation rate per gene, then we got the mutation rate per gene in the male germline (where the Y stays generation after generation) u_{male}. We corrected u_{tot} by the male-mutation bias (b = 5; Makova and Li 2002) so that u_{male} = $2 \cdot b \cdot u_{tot}/(1 + b)$ and then multiplied this by the number of Y genes (78; Skaletsky et al. 2003) to get $U = u_{male} \times$ Ygene number, the deleterious mutation rate for the Y chromosome.

[d] We took the estimates for a gamma distribution that models the distribution of the fitness effects of mutations in humans (Keightley and Eyre-Walker 2007).

are as many as 1/4 of a typical autosome. Gene number (L), fraction of duplicate genes (f_{GC}) and copy number (C) were found to be, respectively, 27, 1/3, and 7 in the human Y (Skaletsky et al. 2003). The rate of Y–Y gene conversion has been estimated to be 2×10^{-4} conversions per duplicated nucleotide per generation for multigene families lying within palindromes (Rozen et al. 2003), and we multiplied this number by the mean human Y gene size (~2,000 bp, ENSEMBL GRcH37 for human data) to get a rate of gene conversion per gene (r_{GC}) of 0.4. We have estimated the deleterious mutation rate for the Y chromosome as follows. We took the genome-wide deleterious mutation rate U_{tot} = 3 (Nachman and Crowell 2000) and divided it by the total number of genes in the human genome (~20,000, ENSEMBL GRCh37 for human data) to get u_{tot} the deleterious mutation rate per gene (u_{tot} = U_{tot}/gene number). Assuming a sex ratio of 1, we had u_{male}=2 $b \cdot u_{tot}/(1 + b)$ with u_{male} being the deleterious mutation rate per gene in the male germline, b being the male-mutation bias and u_{male} = $b \cdot u_{female} \cdot b$ is 5 in humans (Makova and Li 2002) and multiplying u_{male} by 78, the total number of Y genes (Skaletsky et al. 2003), we obtained U = 0.01, the deleterious mutation rate for the Y chromosome. Selection coefficients followed a gamma distribution of parameters α = 0.2 and β = 2.5, which were estimated using human polymorphism data (Keightley and Eyre-Walker 2007).

Single Nucleotide Polymorphism Data Analysis
Raw single nucleotide polymorphisms (SNPs) data were collected from Hapmap version 2 (including Mormons (CEU), yorubas (YRI), hans (CHB) and japonese (JAP) population, see http://ftp.hapmap.org/frequencies/latest_phasell_ncbi_b36/fwd_strand/non-redundant/) and cross-linked with the NCBI database (NCBI b36.3 March 2009) to assign SNPs to Y genes. Information on gene position on the Y chromosome was also collected. The level of polymorphism

was computed as the number of SNPs/gene length. Gene length included coding and intron regions, which mean both neutral and selected polymorphism was analyzed here.

Results
Starting with a population of degenerating Y chromosomes with duplicate genes, we investigated whether a Y variant capable of gene conversion would outcompete the original Y more often than expected by chance. This allows us to estimate the advantage of gene conversion in a degenerating Y chromosome with duplicates. We then studied how the duplications could have evolved in the first place by introducing in a population of degenerating Y chromosomes with single-copy genes only, a variant with duplicate genes. This told us whether fixation of duplicate genes was neutral or counterselected and therefore allows us to quantify how easily they can evolve on a nonrecombining population of Y chromosomes. We also studied the case where the Y variant has duplicates undergoing gene conversion. This told us what were the chances for the duplication and gene conversion to be selected simultaneously. In all simulations, we counted the number of fixations of the variant and divided by the number of runs and compared this with the random expectation (1/N) to get the ratio of observed versus expected probability of fixation (Pfix$_{o/e}$). In all simulations, the default parameters are the values that we currently have for humans (Table 1), unless explicitly mentioned otherwise. In all the figures, each data point has been obtained with 20N independent simulation runs.

Evolution of Gene Conversion on the Y chromosome
Here we start with N Y chromosomes resembling the human Y (same gene number, % duplicates, and copy number) and we introduced a Y variant that has gene conversion within the copies of a duplicate gene. Using the human parameter
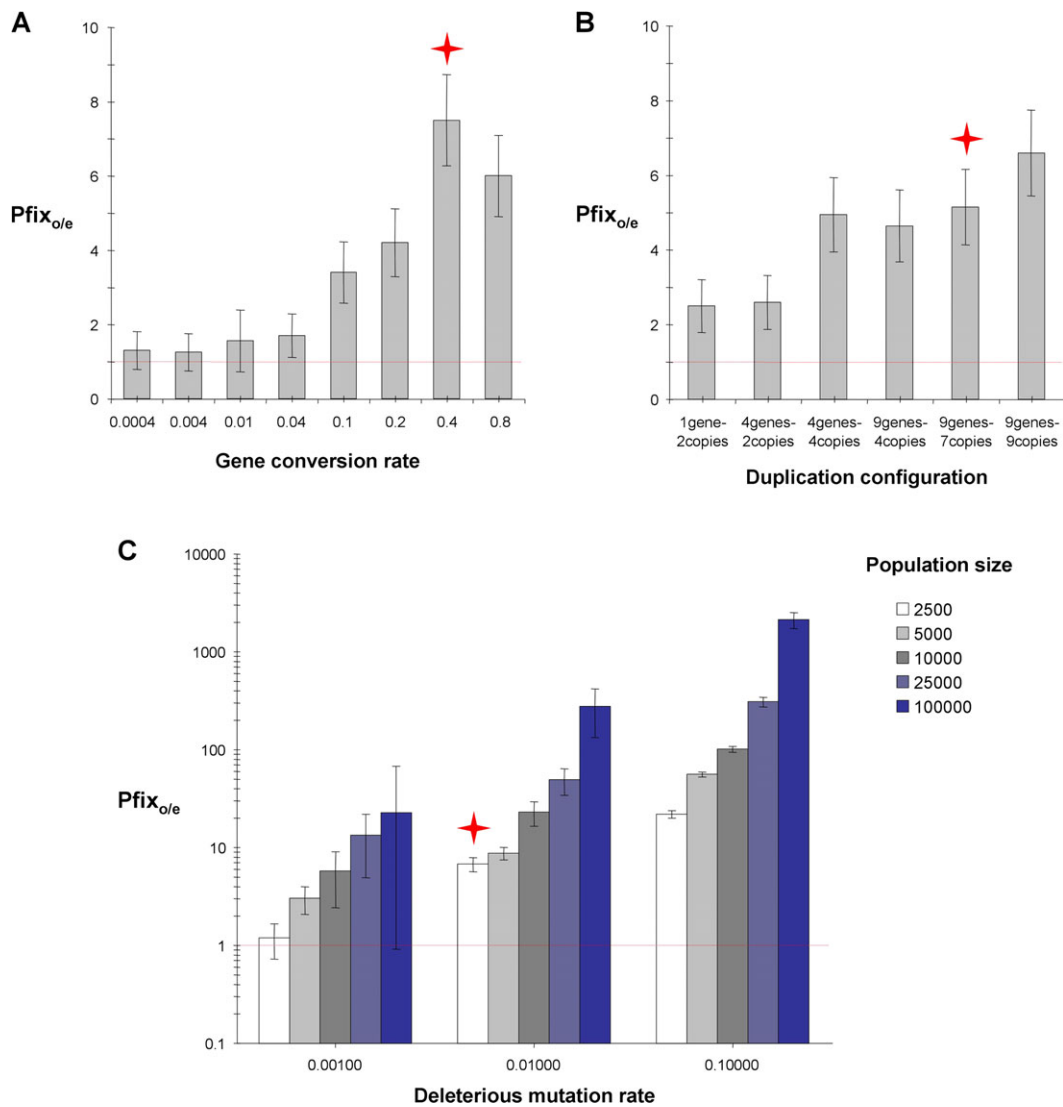
Fig. 1.—Evolution of Y variants for gene conversion. This figure shows results for simulations with introduction of variants with gene conversion in a population of Y chromosomes without gene conversion. $Pfix_{o/e}$ is the number of observed fixation divided by the number of replicates divided by $1/N$, the probability of fixation for a neutral variant. The red line shows the neutral expectation ($Pfix_{o/e} = 1$). The red star shows the simulations with human values for all parameters. Error bars are twice the standard error from the $20N$ replicates. When not specified, the parameters have the values shown in table 1 (from human data). The fixation of the Y variant was studied in different conditions. (*A*) Effect of the gene conversion rates. Here, different rates of gene conversion for the Y variants were tested. (*B*) Effect of the fraction of multicopy genes and the mean copy number (what we called duplication configuration). For instance, nine genes with seven copies each means that there are nine multicopy genes (with seven copies) and a total of $9 \times 7 + (27 - 9) = 81$ genes. Here, simulated Y chromosomes (both original population and variants) have different duplication configurations. (*C*) Effect of the population size and deleterious mutation rate. The *y* axis is in log scale. See text for more details.

values (Table 1), the simulations clearly show that the fixation of the gene conversion variant is much more likely than that of a neutral one ($Pfix_{o/e} = 7.5$, see fig. 1*A*), which suggests that gene conversion is advantageous for the human Y. Changing the rate of gene conversion shows that a high rate of gene conversion is needed for the variant to have an advantage. Indeed, modifiers with low rates of gene conversion do not show any significant advantage over the neutral expectation. Interestingly, the rate of gene conversion estimated for the human Y (0.4) is higher than the threshold for observing such an advantage, which suggests this high value may have been selected for (fig. 1*A*). When we analyze the dynamics of mean fitness of the Y chromosomal population as a strong gene conversion modifier increases in frequency we observe fitness does increase over time, which support the idea that gene conversion oppose Y degeneration (see fig. 2*B*). The analysis of the same dynamics for a weak gene conversion modifier shows that fitness does
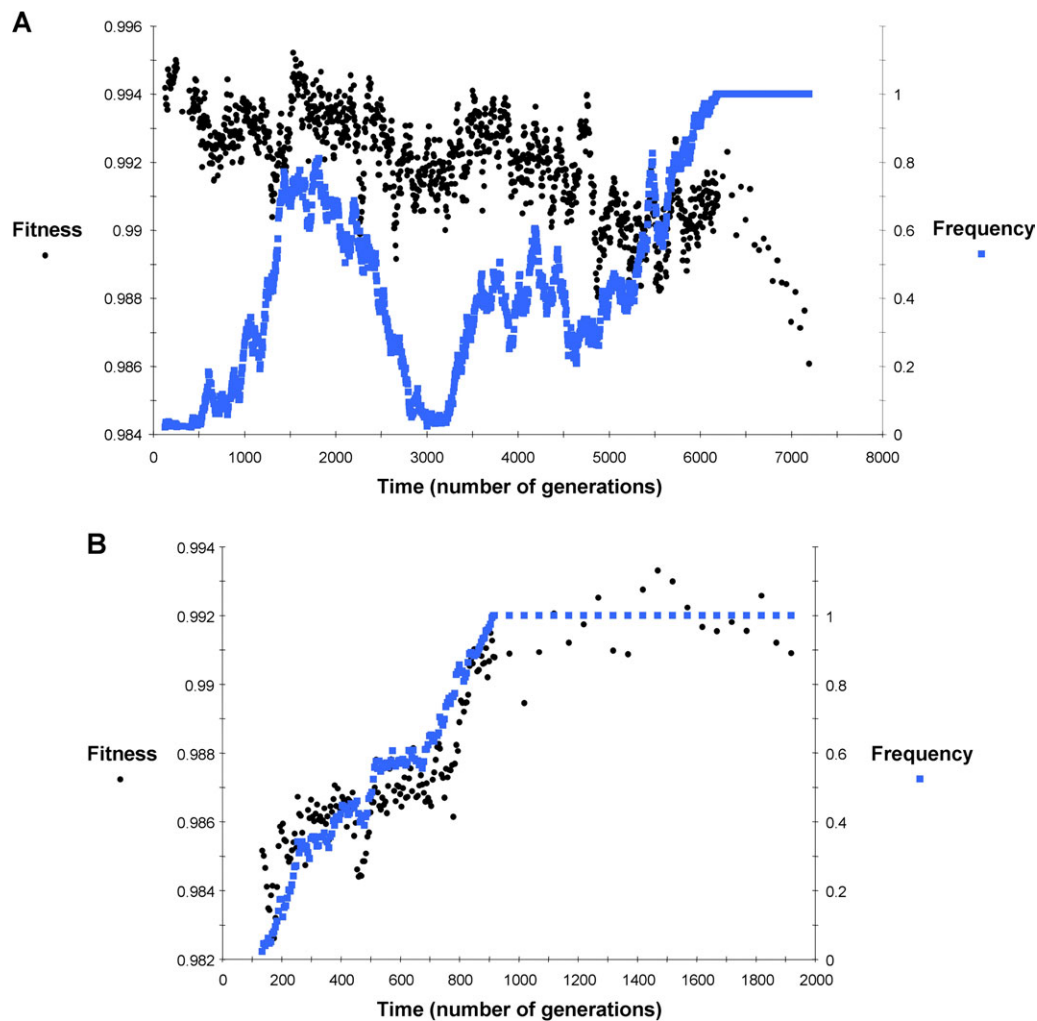
FIG. 2.—Dynamics of fixation of weak and strong gene conversion modifiers. Mean fitness of the population (black dots) and frequency of the Y variant in the population (blue squares) are shown over time (number of generations) for (A) the fixation of a Y variant with low rate of gene conversion (0.0004, weak modifier), and (B) the fixation of a Y variant with high rate of gene conversion (0.4, strong modifier). The x axis does not have the same scale in (A) and (B). Fixation time is much longer for the weak modifier compared with the strong one.

not increase over time and fixation takes much more time, which suggests it is mainly due to drift (see fig. 2A). This confirms the rate of gene conversion has to be high for its counter effect on Y degeneration to be seen. Next, we asked what would be the effect of the number of copies on the advantage of a gene conversion modifier because in the Y there are different numbers of copies in different genes. Figure 1B shows that the advantage increases as the number of gene duplicates increases in the Y, which may explain why the gene families in humans can be so large (mean copy number of 7). Changing the population size (N) and the deleterious mutation rates (U) show that the advantage of the variant increases with N and U (fig. 1C). This result is similar to that observed for the advantage of modifiers of the recombination rate when invading asexual populations (Gordo and Campos 2008). We thus expect that in

other mammals with large N and/or large U the ampliconic region be larger than in humans.

Evolution of Duplication on the Y Chromosome
Here, we start with N Y chromosomes resembling the human Y except that they only have single-copy genes and we introduced a Y variant that has duplicate genes. For instance, in the case of one duplicate with two copies, the variant has one of the Y gene with two copies. Using different duplication configurations (from one duplicate gene with two copies to nine duplicate genes with nine copies), we show that the fixation probability of the variant with duplications is close to that of a neutral variant for Y chromosomes with little duplication and it drops quickly as the duplication involves more genes and more copies and reaches 0 for humans values (see fig. 3A). In our simulations,
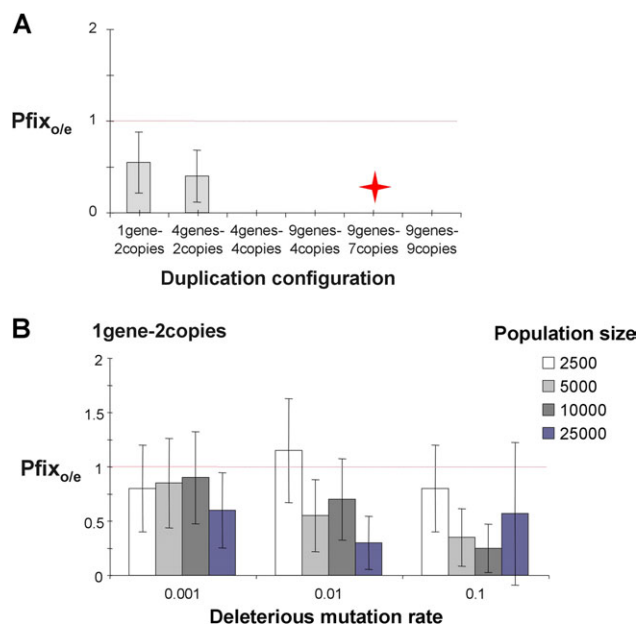
FIG. 3.—Evolution of Y variants for gene duplication. This figure shows results for simulations with introduction of variants for gene duplication in a population of Y chromosomes with single-copy genes only. See fig. 1 for definition of $Pfix_{o/e}$ and other details. (A) Fixation of Y variants with different fractions of multicopy genes and the mean copy numbers (duplication configurations). (B) The situation for the variant with just one duplicated gene (one gene, two copies) is studied in more details. Effects of population size and the rate of deleterious mutation are tested. See text for more details.
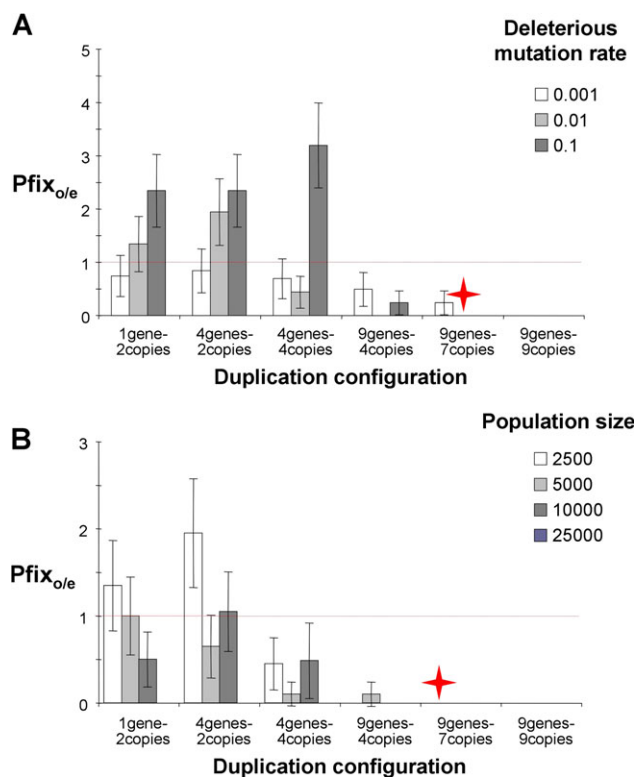

FIG. 4.—Evolution of Y variants for gene duplication and gene conversion. This figure shows results for simulations with introduction of variants having both gene duplication and gene conversion in a population of Y chromosomes with single-copy genes only (and no gene conversion). See fig. 1 for definition of $Pfix_{o/e}$ and other details. The fixation of Y variants with different fractions of multicopy genes and the mean copy numbers (duplication configurations) is studied. (A) Effect of deleterious mutation rate. (B) Effect of population size. See text for more details.

the fixation of duplicate seems unlikely unless it involves small-scale duplication events. We investigated the case with the highest probability of fixation (one duplicate gene with two copies) in more detail in fig. 3B. We got very similar results for a wide range of population sizes and deleterious mutation rates.

Joint Evolution of Duplication and Gene Conversion on the Y

We finally analyzed modifiers with two traits simultaneously, a duplication and the ability to do gene conversion. We start again with N Y chromosomes that have single-copy genes as in the previous section but this time we introduced a Y variant that has both duplicate genes and gene conversion between the copies. Figure 4 shows that the probability of fixation of the variant can exceed that of a neutral variant under some conditions but as soon as the duplication events involve more genes and more copies, the probability of fixation decreases whatever the deleterious mutation rate (fig. 4A) or population size (fig. 4B), mirroring the results obtained for duplication alone (see previous section). For low deleterious mutation rate (0.001), the probability tends to stay closer to neutrality all along the duplication configurations but also drops for large-scale duplication events (nine genes four cop-

ies and larger; fig. 4A). For higher mutation rates (0.01–0.1), the probability of fixation occasionally exceeds 1 for duplication events of intermediate size (fig. 4A). For small population size and small-scale duplication events, the probability of fixation also gets higher than 1 or close to 1 (fig. 4B). The fixation of the variant with both duplication and gene conversion is therefore likely for wider range of parameters compared with that of the variant with duplication only (see previous section). High deleterious mutation, small N and small- to intermediate-size duplication events are favorable conditions for the variant to spread. With parameter values reasonable for humans, the probability of fixation the whole system (ampliconic region and gene conversion) at once is null, which again favor the scenario of a gradual evolution of the structure observed in the human Y today.

Discussion

Our simulations with the gene conversion variant clearly show that in humans intra-Y gene conversion is

advantageous. However, our simulations with the duplication variant show fixation of many duplications at a time is not likely, which raises the question of how the duplications got fixed on the human Y in the first place. Interestingly, these simulations also show that small-scale duplication event (one gene gets two copies) can be fixed by chance and we speculate that the ampliconic regions in humans may have evolved by small steps. When duplication and gene conversion are considered simultaneously as in our simulations with the variant with both duplication and gene conversion, we find that small-scale to intermediate-scale duplication events can even have a fixation probability slightly higher than the neutral expectation. Our results suggest that the Y has accumulated duplicates gradually by drift or because duplication is in itself beneficial (e.g., there is selection on dosage) and the beneficial effect of gene conversion increased progressively as the ampliconic regions got bigger. Specific structures favoring gene conversion (e.g., palindromes) may have also been retained by selection in the process.

Effect of Gene Conversion on Y Degeneration

We do observe an advantage of gene conversion but how does that work? In our simulations, the Y chromosome was degenerating because of deleterious mutations reducing the effective population size of the Y. In this situation, the Y degenerates because it recurrently looses the class of least-loaded Y chromosomes (Muller's ratchet) or because the classes of Y chromosomes with deleterious mutations are bound to disappear and thus reduce the effective population size (background selection). With gene conversion, there is the possibility to restore the least-loaded haplotype for a gene family. If only a small fraction of the Y genes are affected by gene conversion, the advantage is small as suggested by our simulations with gene conversion variant on Y chromosome with different % of duplicate genes (see fig. 1B), but increases as the fraction of duplicate genes increases. In humans, this fraction is fairly high (1/3 of the genes) and most of the Y genes belong to gene families (>75%). A substantial part (the duplicate genes) of the human Y can therefore restore deleterious mutations-free haplotypes and the chromosome as a whole also tends to get deleterious mutations free. This way, gene conversion helps the Y chromosome as a whole to escape Muller's ratchet (and also Background selection) although only the genes undergoing gene conversion tend to avoid fixing deleterious mutations.

Evolution of High Y–Y Gene Conversion Rates

Our results show that Y–Y gene conversion opposes Y degeneration only when very frequent. Figure 1A suggests a threshold of 0.04 below which no effect is detectable. The estimates for Y–Y gene conversion in humans are close to 0.4 (Rozen et al. 2003), which is way above the threshold.

This suggests the observed high Y–Y gene conversion rate has been selected but another explanation is possible. Gene conversion could be intrinsically high on the Y chromosome for a mechanistic reason. A very interesting study rules out this explanation though: looking at some noncoding DNA repeats on the human Y, Bosch et al. (2004) could show that the rate of gene conversion is similar to that of the rest of the human genome and is 1,000-fold lower than the rate of gene conversion in the ampliconic regions. Palindromes, inverted repeats, and tandem arrays found in the ampliconic regions could explain why the rate of gene conversion is so high, and this in turn would explain why they have been retained.

Evolution of Y Ampliconic Regions

In our model, gene duplicates tend to buffer deleterious mutations (the coefficient selection of a mutation is divided by the copy number). Simulations show very clearly that gene duplication is at best neutral and often deleterious, which suggests that the fixation of a duplication event is not likely but can occur occasionally. This is in agreement with previous work showing that the duplicate genes accumulate in species with small effective population size and in regions of the genome where recombination is reduced, such as the Y chromosome (Lynch 2007). However, gene duplication can be advantageous. If a gene is under selection for increasing dosage (i.e., increasing expression level), then gene duplication will be favored by selection if the duplicates are functional and do increase expression level (see Conant and Wolfe 2008 and Innan and Kondrashov 2010 for review). This is a likely situation for some of the ampliconic Y genes. Many of these genes are expressed in testis and are involved in spermatogenesis (Skaletsky et al. 2003). In chimpanzee, the selection on producing sperm is very strong due to sperm competition in this species. Interestingly, the size of ampliconic regions is larger in chimps than in humans and this may be because of selection on dosage on the spermatogenesis genes (Hughes et al. 2010). If duplication is beneficial, fixation of duplicate will be easy, gene conversion will be strongly favored, and we expect a system of gene duplication and conversion to evolve more easily.

How to Survive on the Y Chromosome

If gene conversion is advantageous and allows the Y genes to escape from degeneration, why do not all the Y genes have duplicates? There are three possible explanations for this. As we have already mentioned, the evolution of ampliconic regions may be a gradual process and some genes may have not duplicated but may do so in the future. Evidence for this idea of stochastic recruitment of genes to contribute to ampliconic regions comes from the observation that the genes in the ampliconic regions are not the same in humans, chimps, and rodents (Bhowmick et al. 2007; Alföldi 2008; Hughes et al. 2010). A second possibility that we have just discussed

in the previous section is that some genes could be under selection on dosage and not others. Those that are under such selection will accumulate duplicate far more easily over time and under this condition, gene conversion will be strongly favored. A third possibility is that there may be different ways of surviving on the Y chromosome. The Y single-copy genes could be essential genes robust to complete inactivation because selection is so strong on these genes that the reduction of N due to absence of crossing-over is not enough to make selection completely inefficient, and they can survive without gene conversion. Interestingly, many of the Y single-copy genes are housekeeping genes and potentially under strong selection (Skaletsky et al. 2003; Wilson and Makova 2009). Some of these genes may have subfunctionalized X and Y copies that are both essential, and others may have neofunctionalized Y copies (with the X copy keeping the ancestral function) essential for male function as suggested by the comparison of expression patterns for X and Y single-copy genes (Wilson and Makova 2009).

Biased or Unbiased Y–Y Gene Conversion

An important point of our results is that we did not need to include biased gene conversion in the model. We see an advantage when gene conversion is unbiased. This can be surprising at first because several theoretical studies on the evolution of multigene families have shown that gene conversion only improves natural selection when gene conversion is biased and it disfavors the most common mutations, which are likely to be deleterious (Nagylaki and Petes 1982; Nagylaki 1984; Walsh 1985; Nagylaki and Barton 1986). However, this work did not model Muller's ratchet and only considered very low rates of gene conversion. In our simulations (fig. 1A), we show that the rate of gene conversion is a crucial parameter. For low rates, we get no advantage. We only start seeing a significant advantage for rather high rates (>0.01). The discrepancy may just be because these early works did not investigate high rates of gene conversion. This observation is important because we do not know whether gene conversion affecting the human Y chromosome is biased or not. We know that allelic recombination during meiosis is associated with biased gene conversion toward GC bases in mammalian genomes (Marais 2003; Duret and Galtier 2009). We also have evidence that ectopic recombination in autosomes could be also biased (Galtier 2003). We do not know however for the particular gene conversion going on the Y chromosome. At any rate, if Y–Y gene conversion is shown to be actually biased, this would make the advantage for gene conversion even stronger.

Possible Costs of Y–Y Gene Conversion

No costs of gene conversion have been included in our model. A recent paper has shown that ectopic crossovers do occur within the ampliconic regions in humans causing chromosomal rearrangements (e.g., large deletions) and in-

fertility (Lange et al. 2009). This suggests that the same mechanism can generate gene conversion and crossover (as in classical meiotic allelic recombination between homologues). The ectopic crossovers of course would counteract the beneficial effect of gene conversion and the net result would probably depend on how frequent are ectopic crossovers, which we currently do not know. When we have an estimate for this ratio (crossover/gene conversion), it will be interesting to include it in the model. Another issue is whether the same cost holds true for the other classes of ampliconic regions (tandem arrays and inverted repeats). Indeed, Lange et al. (2009) only investigated palindromic regions. An important point related to costs is that it may imply a balance between the advantage to increase the size of ampliconic regions because gene conversion opposes degeneration better and to reduce the size of such regions because large ampliconic regions increase the opportunity for ectopic crossovers and chromosomal rearrangements.

Y–Y Gene Conversion and Y Polymorphism

A Y chromosome that is degenerating through the accumulation of deleterious mutations is expected to have low levels of neutral polymorphism (see Charlesworth B and Charlesworth D 2000 and Gordo and Charlesworth 2001 for review). Reduced levels of neutral polymorphism have been repeatedly observed in Y or W chromosomes in many species (Yi and Charlesworth 2000; Filatov et al. 2001; Berlin and Ellegren 2004 and see the chapter on sex chromosomes in Lynch 2007 for review). In the human genome, the Y chromosome is the one with the lowest level of nucleotide diversity, and this has been taken as evidence for degeneration of the human Y (The International SNP Map Working Group 2001). However, previous work on a few markers did not report significantly reduced genetic diversity on the human Y (Nachman 1998). We collected all the SNP data available for the Y chromosome in Hapmap and computed the level of nucleotide polymorphism for the normally recombining PAR genes and nonrecombining MSY genes (subdivided in ampliconic genes and single-copy genes). We found that genetic diversity is significantly reduced at the ampliconic genes (no SNPs in the ampliconic regions) compared with the single-copy MSY genes (significant Mann–Whitney test, $P < 0.0001$; Table 2). It is not clear why this should be because we expect similar levels of genetic diversity for single-copy genes and multicopy genes under strong gene conversion (Innan 2009, and not shown simulations that we have performed). Interestingly, the levels of diversity were found to be very similar in PAR genes and single-copy MSY genes (no significant Mann–Whitney test, $P = 0.589$; Table 2). This means that the global reduction in genetic diversity observed of the human Y chromosome is mainly due to very low diversity in the ampliconic genes and not to degeneration. Indeed, the single-copy MSY genes have the same amount of genetic diversity compared

Table 2

Nucleotide Polymorphism Levels for the Human Y Genes

Y Genes Class	PAR Genes	MSY Genes[a]	
		Single-copy	Ampliconic
Gene number	19	19	27
Levels of polymorphism[b]	9.7×10^{-5}	8.4×10^{-5}	0
Mann–Whitney test (with single-copy genes)	Not significant	—	$P < 0.0001$

[a] PAR = pseudoautosomal region (recombining), MSY = male-specific region (non-recombining), MSY genes are further divided in single-copy and ampliconic genes (undergoing gene conversion).

[b] Hapmap data, see Material and Methods section.

with the PAR genes. This suggests that there is no or very little ongoing degeneration on the human Y chromosome. This could be because the human Y is already degenerate and gene density is too low for Hill–Robertson effects to be strong (Bachtrog 2008). Y–Y gene conversion in most of the MSY genes could also attenuate Hill–Robertson effects as discussed in the first section of the discussion and make degeneration undetectable.

The Advantage of Y–Y Gene Conversion in Other Species

Our results also have implications for other species than humans. In particular, fig. 1C shows that the advantage of doing gene conversion should be stronger in species with large N (but <100,000). The ampliconic regions are twice as larger in chimps than in humans and the intrapalindrome divergence is lower in chimps than in humans, which suggests a higher rate of gene conversion in chimps than in humans (Hughes et al. 2010). This fits with our predictions since the effective population size is known to be larger in chimps than in humans (Keightley et al. 2005). Preliminary data show that ampliconic regions have expanded to an incredible degree in mice with 95% of the mice Y being ampliconic (Alföldi 2008). The repeats in the mice Y show a very high sequence conservation, which suggests a very high rate of gene conversion. Mice effective population size is probably 10 times larger than that of humans (Keightley et al. 2005), and this fits again with our predictions. Interestingly, humans and chimps also differ in their mating system and this has been suggested to have an effect on the size of ampliconic regions (Hughes et al. 2010). Sperm competition is very strong in chimps and certainly stronger than in humans. An individual has a clear advantage in producing more sperm than its competitors in chimps. This may have been possible by increase of the dosage of sperm-related Y genes through increase of copy number. However, another effect could contribute to the pattern observed. Degeneration seems more pronounced in the chimp Y compared with the human Y (Hughes et al. 2010), which could be due to recurrent episodes of positive selection (and the concomitant selective sweeps) on the spermatogenesis genes located on the chimp ampliconic regions. This is consistent with our finding that protection against degeneration increases with gene copy number (see fig. 1B) and could contribute to the larger ampliconic regions in chimps than in humans. However, here fixation of beneficial mutations would be involved (not deleterious ones as in our model) and it would be interesting to investigate theoretically how gene conversion can protect Y genes against selective sweeps caused by fixation of beneficial mutations. Another evidence supporting the idea that evolution of ampliconic regions counteracts degeneration comes from birds. In chicken, HINTW is a multicopy gene on the W chromosome, which shows evidence for W–W gene conversion. This gene seems to share a very similar evolutionary dynamics compared with the mammalian Y ampliconic genes. However, this gene is not involved in spermatogenesis (the W is carried by females) and selection on dosage is not clear, which leaves the idea of W–W gene conversion opposing W degeneration the only explanation here (Backström et al. 2005). Further theoretical and empirical work is needed for studying gene conversion on nonrecombining sex chromosomes but our results suggest that preventing gene degeneration by gene conversion could be fairly common in organisms with sex chromosomes and with small to intermediate population sizes such as mammals, birds, and possibly other vertebrates.

Literature Cited

Alföldi JE. 2008. Sequence of the mouse Y chromosome. [thesis]. Department of Biology, [Cambridge (MA)]: Massachusetts Institute of Technology.

Backström N, Ceplitis H, Berlin S, Ellegren H. 2005. Gene conversion drives the evolution of HINTW, an ampliconic gene on the female-specific avian W chromosome. Mol Biol Evol. 22(10):1992–1999.

Bachtrog D. 2008. The temporal dynamics of processes underlying Y chromosome degeneration. Genetics. 179(3):1513–1525.

Berlin S, Ellegren H. 2004. Chicken W: a genetically uniform chromosome in a highly variable genome. Proc Natl Acad Sci U S A. 101(45):15967–15969.

Bhowmick BK, Satta Y, Takahata N. 2007. The origin and evolution of human ampliconic gene families and ampliconic structure. Genome Res. 17(4):441–450. Epub 21 December 2006.

Bosch E, Hurles ME, Navarro A, Jobling MA. 2004. Dynamics of a human interparalog gene conversion hotspot. Genome Res. 14(5):835–844.

Charlesworth B. 2003. The organization and evolution of the human Y chromosome. Genome Biol. 4(9):226.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci. 355(1403):1563–1572.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. Heredity. 95(2):118–128.

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet. 9(12):938–950.

Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genomics Hum Genet. 10:285–311.

Eyre-Walker A, Keightley PD, Smith NG, Gaffney D. 2002. Quantifying the slightly deleterious mutation model of molecular evolution. Mol Biol Evol. 19(12):2142–2149.

Filatov DA, Laporte V, Vitte C, Charlesworth D. 2001. DNA diversity in sex-linked and autosomal genes of the plant species Silene latifolia and Silene dioica. Mol Biol Evol. 18(8):1442–1454.

Gailtier N. 2003. Gene conversion drives GC content evolution in mammalian histones. Trends Genet. 19(2):65–68.

Gordo I, Campos PR. 2008. Sex and deleterious mutations. Genetics. 179(1):621–626.

Gordo I, Charlesworth B. 2001. Genetic linkage and molecular evolution. Curr Biol. 11(17):R684–R686.

Graves JA. 2004. The degenerate Y chromosome–can conversion save it? Reprod Fertil Dev. 16(5):527–534.

Hawley RS. 2003. The human Y chromosome: rumors of its death have been greatly exaggerated. Cell. 113(7):825–828.

Hughes JF, et al. 2010. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. Nature. 463(7280):536–539.

Hurst LD, Smith NGC. 1998. The evolution of concerted evolution. Proc R Soc Lond B. 265:121–127.

Innan H. 2009. Population genetic models of duplicated genes. Genetica. 137(1):19–37.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet. 11(2):97–108.

The International SNP Map Working Group. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 409(6822):928–933.

Iwase M, et al. 2003. The amelogenin loci span an ancient pseudoautosomal boundary in diverse mammalian species. Proc Natl Acad Sci U S A. 100(9):5258–5263.

Keightley PD, Eyre-Walker A. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. Genetics. 177(4):2251–2261.

Keightley PD, Lercher MJ, Eyre-Walker A. 2005. Evidence for widespread degradation of gene control regions in hominid genomes. PLoS Biol. 3(2):e42.

Keightley PD, Otto SP. 2006. Interference among deleterious mutations favours sex and recombination in finite populations. Nature. 443(7107):89–92.

Lahn BT, Page DC. 1999. Four evolutionary strata on the human X chromosome. Science. 286:964–967.

Lange J, et al. 2009. Isodicentric Y chromosomes and sex disorders as byproducts of homologous recombination that maintains palindromes. Cell. 138(5):855–869.

Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates Inc.

Lynch M, Conery JS. 2003. The origins of genome complexity. Science. 302(5649):1401–1404.

Makova KD, Li WH. 2002. Strong male-driven evolution of DNA sequences in humans and apes. Nature. 416(6881):624–626.

Marais G. 2003. Biased gene conversion: implications for genome and sex evolution. Trends Genet. 19(6):330–338.

Marais G, Galtier N. 2003. Sex chromosomes: how X-Y recombination stops. Curr Biol. 13(16):R641–R643.

Nachman MW. 1998. Y chromosome variation of mice and men. Mol Biol Evol. 15(12):1744–1750.

Nachman MW, Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans. Genetics. 156(1):297–304.

Nagylaki T. 1984. The evolution of multigene families under intrachromosomal gene conversion. Genetics. 106(3):529–548.

Nagylaki T, Barton N. 1986. Intrachromosomal gene conversion, linkage, and the evolution of multigene families. Theor Popul Biol. 29(3):407–437.

Nagylaki T, Petes TD. 1982. Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. Genetics. 100(2):315–337.

Pecon Slattery J, Sanner-Wachter L, O'Brien SJ. 2000. Novel gene conversion between X-Y homologues located in the nonrecombining region of the Y chromosome in Felidae (Mammalia). Proc Natl Acad Sci U S A. 97(10):5307–5312.

Ross MT, et al. 2005. The DNA sequence of the human X chromosome. Nature. 434:325–337.

Rosser ZH, Balaresque P, Jobling MA. 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(1):130–134.

Rozen S, et al. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature. 423(6942):873–876.

Skaletsky H, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature. 423:825–837.

Veyrunes F, et al. 2008. Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. Genome Res. 18(6):965–973.

Walsh JB. 1985. Interaction of selection and biased gene conversion in a multigene family. Proc Natl Acad Sci U S A. 82(1):153–157.

Wilson MA, Makova KD. 2009. Evolution and survival on eutherian sex chromosomes. PLoS Genet. 5(7):e1000568.

Yi S, Charlesworth B. 2000. Contrasting patterns of molecular evolution of the genes on the new and old sex chromosomes of Drosophila miranda. Mol Biol Evol. 17(5):703–717.

Yu N, Fu YX, Li WH. 2002. DNA polymorphism in a worldwide sample of human X chromosomes. Mol Biol Evol. 19(12):2131–2141.

Yu N, et al. 2003. Low nucleotide diversity in chimpanzees and bonobos. Genetics. 164(4):1511–1518.

Yu YH, Lin YW, Yu JF, Schempp W, Yen PH. 2008. Evolution of the DAZ gene and the AZFc region on primate Y chromosomes. BMC Evol Biol. 8:96.

Associate editor: Laurence Hurst