Expression evolution of ancestral XY gametologs across all major groups of placental mammals

Mónica Martínez-Pacheco¹, Mariela Tenorio¹, Laura Almonte¹, Vicente Fajardo¹, Alan Godínez¹, Diego Fernández¹, Paola Cornejo-Páramo¹, Karina Díaz-Barba¹, Jean Halbert², Angelica Liechti², Tamas Székely³, Araxi O. Urrutia³, Diego Cortez¹,⁵

1. Center for Genome Sciences, UNAM, CP62210 Cuernavaca, Mexico.
2. Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland
3. Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK.
4. Ecology Institute, UNAM, Mexico.
5. Corresponding author (dcortez@ccg.unam.mx).

Email addresses: monicamp@ccg.unam.mx, yoye.maye@hotmail.com, lauraalm@lcg.unam.mx, vfajardo@lcg.unam.mx, vladimiralan@comunidad.unam.mx, Dfernand@lcg.unam.mx, pcornejo@lcg.unam.mx, karinadb@lcg.unam.mx, Jean.Halbert@unil.ch, angeliechti@hotmail.com, T.Szekely@bath.ac.uk, A.Urrutia@bath.ac.uk, dcortez@ccg.unam.mx

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Abstract

Placental mammals present 180 million-year-old Y chromosomes that have retained a handful of dosage-sensitive genes. However, the expression evolution of Y-linked genes across placental groups has remained largely unexplored. Here, we expanded the number of Y gametolog sequences by analyzing ten additional species from previously unexplored groups. We detected seven remarkably conserved genes across 25 placental species with known Y repertoires. We then used RNA-seq data from 17 placental mammals to unveil the expression evolution of XY gametologs. We found that Y gametologs followed, on average, a 3-fold expression loss and that X gametologs also experienced some expression reduction, particularly in primates. Y gametologs gained testis-specificity through an accelerated expression decay in somatic tissues. Moreover, despite the substantial expression decay of Y genes, the combined expression of XY gametologs in males is higher than that of both X gametologs in females. Finally, our work describes several features of the Y chromosome in the last common mammalian ancestor.

Keywords

Placental mammals; Sex chromosomes; Y chromosome; Gene expression levels; Dosage compensation mechanisms.

Significance Statement

Previous studies unveiled many features of the evolution of X and Y chromosomes in placental mammals. Nevertheless, the general patterns of gene expression evolution of XY gametologs across placental species remain largely unexplored outside humans and rodents. To close this knowledge gap, in this study we explored the conservation, expression evolution, and dosage compensation mechanisms of ancestral XY gametologs across all major groups of placental mammals. We uncovered important evolutionary processes that have shaped the expression patterns of sex chromosomes in the last mammalian ancestor.
Introduction

Sex chromosomes in marsupial and placental mammals originated from a pair of autosomes (Bachtrog 2013) following the emergence of the sex-determining gene SRY (Sinclair, et al. 1990) on the proto-Y chromosome. The protein coded by the SRY gene was capable to directly regulate the expression of SOX9 (Sekido and Lovell-Badge 2008; Li, et al. 2014) and, therefore, initiate the signaling cascade to develop testis. The origin of the sex chromosomes in therians occurred approximately 180 million years ago (ma) (Cortez, et al. 2014) and, since then, recombination has been suppressed along the majority of the X and Y chromosomes (Lahn and Page 1999), giving rise to the male-specific region of the Y chromosome (MSY) (Skaletsky, et al. 2003). Consequently, lack of homologous recombination prompted the rapid loss of genetic content of the Y chromosome (Charlesworth and Charlesworth 2000; Hughes, et al. 2012) and the evolution of mechanisms that could restore expression balance between males (one active X chromosome) and females (two active X chromosomes). Dosage balance on the X chromosome was later achieved by the recruitment of a long non-coding RNA (Xist) capable of mediating the inactivation of one X chromosome in females (Plath, et al. 2002).

Lack of homologous recombination between Y and X chromosomes also allowed the proliferation of repeated sequences on the Y chromosome (Skaletsky, et al. 2003). Accumulation of a large number of duplicated elements (Bachtrog 2013) resulted in complex genomic sequences, which are difficult to analyze and assemble. Several Y chromosomes have been sequenced though: human (Skaletsky, et al. 2003), chimpanzee (Hughes, et al. 2005; Hughes, et al. 2010), gorilla (Tomaszkiewicz, et al. 2016), macaque (Hughes, et al. 2012), mouse (Soh, et al. 2014), pig (Skinner, et al. 2016), and horse (Janecka, et al. 2018). Specific efforts also recovered Y genes from cat and dog (Pearks Wilkerson, et al. 2008; Li, et al. 2013), bull and marmoset (Bellott, et al. 2014), and some Y scaffolds in the polar bear (Bidon, et al. 2015), the grey wolf (Smeds, et al. 2019) and the red fox (Rando, et al. 2019). Also, a subtraction approach developed to work with male and female transcriptome data allowed the reconstruction of Y-linked gene repertoires in ten species of placental, marsupial and monotreme mammals (Cortez, et al. 2014).

Thus far, analyses of the MSY in placental mammals have revealed the presence of large palindromic structures (Skaletsky, et al. 2003), Y-linked genes conserved across species (Hughes, et al. 2012; Cortez, et al. 2014; Janecka, et al. 2018), presence of ampliconic gene families with large number of
copies (Goto, et al. 2009; Soh, et al. 2014; Ghenu, et al. 2016; Brashear, et al. 2018; Vegesna, et al. 2019), gene functions related to spermatogenesis (Colaco and Modi 2018; Liu 2019), and retention of ancestral X-Y gene pairs (i.e. XY gametologs) that are under purifying selection (Wilson and Makova 2009), code for proteins with regulatory functions and are haploinsufficient (Bellott, et al. 2014; Cortez, et al. 2014).

Haploinsufficiency means that the expression level of one copy does not produce sufficient protein to realize a biological function. We expect, therefore, some functional redundancy between X and Y gametologs in males to overcome haploinsufficiency, whereas the expression of an active X gametolog is complemented (at least in humans and mouse) by the expression of an X gametolog that escapes X chromosome inactivation in females (Bellott, et al. 2014; Tukiainen, et al. 2017). Notably, expression levels of X-linked genes in humans, including X gametologs, are affected by the presence of multiple copies of X and Y chromosomes (Raznahan, et al. 2018), indicating dynamic co-expression networks between sex chromosomes that regulate gene expression.

Several studies have examined the evolutionary constraints and expression of X- and Y-linked genes in primates and mouse (Lingenfelter, et al. 2001; Xu and Disteche 2006; Isensee, et al. 2008; Xu, et al. 2008; Wilson Sayres and Makova 2013; Slavney, et al. 2016; Godfrey, et al. 2020). Nevertheless, the general patterns of gene expression evolution of XY gametologs across major placental groups have remained largely unexplored. In this study, we expanded the number of species with Y-linked genes by recovering Y gametologs from ten additional species of placental mammals. We used extensive RNA-seq data from 17 species from all major groups of placental mammals to unveil the expression evolution of ancestral XY gametologs, evaluate potential dosage compensation mechanisms and defined attributes already present on the Y chromosome of the mammalian ancestor.

Materials and Methods

Data collection
We collected published RNA-seq data for nine placental mammals: human, macaque, rabbit, mouse lemur, sheep, panda, rat, mouse, and marmoset, (Supplementary Table 1). Also, we generated strand-specific RNA-seq libraries (using the Illumina TruSeq Stranded mRNA Library protocol) for
eight placental mammals: tree shrew, hamster, guinea pig, cow, pig, hedgehog, armadillo, and tenrec; at least one female and one male brain/cerebellum, heart, liver, kidney, and gonads. Each library was sequenced on Illumina HiSeq 2500 platforms at the Lausanne Genomic Technologies Facility. Libraries were sequenced as 100-nt single-end with 34 million reads on average per sample and >93% of reads with Q > 35 of mean quality (Supplementary Table 1). We also generated extra RNA-seq data for sheep. Besides, we generated male DNA-seq libraries for seven placental species: hamster, guinea pig, rabbit, sheep, hedgehog, armadillo, and tenrec using the Illumina TruSeq DNA protocol for short insert size (target size 400-450 nt). We sequenced 100-nt paired-end DNA-seq libraries on an Illumina HiSeq 2500 sequencers (112 million reads on average). For mouse lemur, panda, and tree shrew we worked with the available genomic published data (Supplementary Table 1). We also collected published RNA-seq data for three outgroup species: chicken, platypus, and A. carolinensis (Supplementary Table 1).

**Reported Y-linked genes**

We collected the nucleotide sequences of ancestral Y-linked genes (or transcripts) from human, macaque, marmoset, mouse, rat, guinea pig, rabbit, cow, sheep, panda, pig, and hedgehog directly from scientific publications, the Ensembl database (release 96; www.ensembl.org) or the NCBI-Genbank (www.ncbi.nlm.nih.gov/genbank) database (Supplementary Table 2).

**Assembly of Y-linked transcripts**

To assemble Y-linked transcripts in placental mammals we used a subtraction approach based on male/female RNA-seq data. Subtraction approaches have been used to find sex-linked sequences in mammals/birds (Cortez, et al. 2014), the green anole A. carolinensis (Marin, et al. 2017), casque-headed lizards (Acosta, et al. 2019), water skinks (Cornejo-Paramo, et al. 2020), plants (Akagi, et al. 2014) and insects (Carvalho and Clark 2013). This method allows the recovery of nucleotide sequences and the identification of sex-linked genes that are necessary for evolutionary and functional analyses. We used the subtraction approach to recover Y-linked transcripts in mouse lemur, tree shrew, hamster, guinea pig, rabbit, sheep, panda, hedgehog, armadillo, and tenrec. Specifically, for each species, we first removed reads with ambiguous nucleotides (N). Next, we concatenated male RNA-seq reads from male tissues into one file and aligned these reads using Hisat2 (default options; v2.0.2)(Kim, et al. 2015) against both the female reference genome downloaded from the Ensembl database (release 96; www.ensembl.org) and the de novo reconstructed female transcriptome; reads not mapping were selected; female transcriptomes were
obtained using female tissues (brain/cerebellum, kidney, heart, liver, and ovary) with Trinity (v2.0.2, default options). We also used the female RNA-seq data to build an index of 35 base pairs (bp) k-mers for each species, following a previous procedure (Akagi, et al. 2014). We calculated the frequency of these 35bp k-mers and removed those showing frequencies below ten; we did not consider rare k-mers as part of the overall signature of the female transcriptome. We used Bowtie2 (2.1.0) (Langmead and Salzberg 2012) to align the more abundant 35bp k-mers to the male reads that did not align with the female transcriptomes (with no mismatches and no indels allowed); we selected those male reads with no successful alignments. With the few remaining reads, we assembled a male-biased transcriptome with Trinity (v2.0.2, default options) (Grabherr, et al. 2011). We obtained around 20,000 transcripts per species showing biased expression in male tissues. These transcripts with male-biased expression could belong to autosomes or the X chromosome, and may not necessarily represent Y-linked genes. To test for Y-linked sequences, we used blastN (Altschul, et al. 1990) to perform sequence searches in male/female genomic data. Female genomic reads were downloaded from NCBI (www.ncbi.nlm.nih.gov); raw reads from the female reference genome projects. Moreover, we produced male genomic data for specific species. Male transcripts were marked as Y-linked when showing 100% identity over 90% or more of their sequence length aligned against the male genomic data and no significant alignments against the female genomic data. Transcripts present in both the male transcriptome and the male genome were considered as Y-linked. These Y-linked sequences were further examined based on alignments with Muscle (Edgar 2004) of X-Y pairs in each species and Y-Y alignments across placental species. Recovering ancestral Y gametologs using subtraction-based approaches is easier because these genes display a larger number of nucleotide differences with X gametologs. Note that subtraction methods are more accurate when the Y genes show an overall nucleotide sequence similarity below 98% compared to the X gametolog nucleotide sequence (Cortez, et al. 2014). Consequently, recent Y-linked genes on the MSY may not be identified. Group- and species-specific Y transcripts were not included in this study. As an additional control, we ran the subtraction-based method with male/female RNA-seq data from human, marmoset, mouse, rat, pig, cow, and opossum, for which Y sequences have been obtained using targeted direct-sequencing methods. The results from the sequence comparison are detailed in Supplementary Table 3 and the sequence alignments are in Supplementary Material.

**Defining Y-linked gene absence**

In placental species with sequenced Y chromosomes, we defined the absence of a Y gene when it
was not annotated on the Y chromosome and BlastN/tBlastN searches using orthologous Y-linked genes from other species did not retrieve significant matches. For species without sequenced Y chromosomes, we searched male genomic reads with BlastN and tBlastN using orthologous Y-linked genes from other species as queries. When sequence searches retrieved matches for the X gametolog instead of the Y gametolog, we considered the Y gene to be absent. We did not assess species for which male genomic data was not available or was insufficient.

**Identification of Y and X gametologs identities**

To establish Y gene identity, we searched NCBI GenBank (http://www.ncbi.nlm.nih.gov/genbank) with blastN and blastX for the closest homologs in the Mammalian taxa and selected transcripts that coded for known Y or X-linked proteins. BlastX searches also allowed the identification of CDS regions. Other Y-linked sequences found, often representing pseudogenes, transposable elements, or potential long non-coding RNAs, were not considered in downstream analyses. Then, the best blastN match (usually around 92-95% identity over the entire sequence) onto the annotated X chromosome of the reference genomes was considered the X gametologs. Some reference genomes lacked annotations for X gametologs (RBMX in macaque, DDX3X in tree shrew, USP9X in the guinea pig, TMSB4X in sheep, RPS4X, RBMX, ZFX in hedgehog, RPS4X, RBMX, TSPLY2, TMSB4X, and ZFY in tenrec). In these specific cases, we included in the analyses the transcripts for the X gametolog from the complete transcriptome reconstruction using female tissues.

**Identification of retrogenes**

Retrogenes were identified following a previous procedure (Marques, et al. 2005). Briefly, we retrieved amino-acid sequences for X and Y gametologs from the 25 species listed in Figure 1. The protein sequences were then used as queries in searches against the 25 complete genomes (Ensembl database, release 96; www.ensembl.org) using tBLASTn; default options (Altschul, et al. 1997). Adjacent homology matches were merged, combining only all nearby matches, and we verified they lacked reported introns. We required that query and target sequences had >50% similarity on the amino acid level and over >80% of their length shared. For each species, we aligned retrogenes to the parental X and Y gametologs using PRANK (Loytynoja and Goldman 2005), codon-based option (-codon). Pair-wise $d_S$ values for each of the two following pairs, retrogene – parental X gametologs and retrogene – Y gametolog, were obtained using codeml, implemented in the PAML package; runmode = -2, seqtype = 1, CodonFreq = 2 (Yang 1997). X or Y gametologs showing the
lowest $d_s$ values were considered to be the parental gene for a given retrogene.

**Expression analyses**

We downloaded from the Ensembl (release 96; www.ensembl.org) database the complete transcriptome (cDNA files) for human, macaque, marmoset, tree shrew, mouse lemur, mouse, rat, hamster, guinea pig, rabbit, cow, sheep, pig, panda, hedgehog, armadillo, and tenrec. Ensembl transcriptomes for human, macaque, and pig already included Y-linked transcripts. For all other species, we added to the transcriptome files the Y-linked genes reported in scientific publications, the NCBI-Genbank database or obtained based on the subtraction approach. In specific cases, we added to the files a few X-linked transcripts missing (see above). Then, we aligned RNA-seq reads from brain/cerebellum, heart, kidney, liver, and gonads (testis and ovary) and estimated expression levels as transcripts-per-million using Kallisto (Bray, et al. 2016), specifying 100 bootstraps. All single-end libraries were produced at the Lausanne Genomic Technologies Facility (https://www.unil.ch/gtf/en/home.html) following the same protocol with an average fragment length of 300 ± 50 base pairs (information needed to run Kallisto; -l and -s parameters), which were verified using a Bioanalyzer. Given that Y-linked genes from many species were transcripts reconstructed from male RNA-seq data, we reasoned that expression analyses should be carried out at the level of transcripts for all species. XY gametologs show 92-95% identity over the entire sequence, so Kallisto is capable of assigning the correct reads to these genes. We tested for cross-gametolog mapping by in silico generating random RNA-seq reads of 100 nucleotides long based on the nucleotide sequences of X gametologs (100,000 reads) and the nucleotide sequences of Y gametologs (100,000 reads). We mapped the Y- and X-derived reads to the transcriptomes separately. In the hypothetical scenario where cross-gametolog mapping has an important effect, several reads would equally map to both gametologs and, therefore, we would see expression levels of X genes when mapping Y-derived reads, and expression of Y genes when mapping X-derived reads. We observed that cross-gametolog mapping is remarkably low in our set of ancestral XY gametologs (Supplementary Figure 1). When multiple individuals for males or females were available, we calculated median expression levels per tissue. We also combined median expression levels for the brain and cerebellum. Expression level normalization across samples was performed using a scaling procedure (Brawand, et al. 2011) that uses one-to-one orthologous genes expressed in all samples showing the lowest variance across samples (Supplementary Figures 2-3). RNA-seq data for panda was from particular tissues (tongue, pituitary, colon, etc.) and RNA-seq data for
mouse lemur was only available for males; this data was only included in the general Y expression profiles reported in Figure 2. Mapping RNA-seq reads to multicopy Y genes is a complex task, particularly, because many multicopy Y genes are 99-100% identical at the nucleotide level. In gene repertoires recovered using transcriptome data and subtraction approaches, we obtained one gene copy representing all copies of a multicopy Y gene. However, for species with fully sequenced Y chromosomes, multicopy Y genes are well annotated and may include between two and ten copies. So, for species with sequenced Y chromosomes, when multiple copies of a Y or X genes were present, we added the individual gene expression levels to obtain one value per gene. We performed the same strategy for retrogenes of a given X or Y parental gametolog when multiple retrogenes for a given parental gene were present in the same species. SRY was excluded from the analyses because this gene is primarily expressed during development. We performed a PCA analysis to determine whether samples grouped by tissue (no batch effects present) or by experiment (batch effects present). We used expression levels of 1-to-1 orthologs across species. We found that samples clustered by tissue, which is expected when batch effects are not present (Supplementary Figure 4). Comparisons of expression levels between current and ancestral states were carried out as previously described (Julien, et al. 2012; Cortez, et al. 2014). Specifically, to infer ancestral expression levels we exploited the fact that current sex chromosomes are derived from ancestral autosomes and, therefore, have autosomal counterparts in species with non-homologous sex chromosomes, which are informative concerning proto-sex chromosome expression patterns. We calculated ancestral sex chromosome expression levels as median expression levels of autosomal one-to-one orthologues of X genes in out-group species with different sex chromosomes systems: A. carolinensis, platypus and chicken. Ancestral inferred expression output values were calculated per one gene copy/allele, that is, the obtained values were divided by 2. We obtained 16 ancestral values, one for each XY gametolog pair. These ancestral values were then used against the expression levels of all current XY gametologs from the different placental species. Furthermore, we only analyzed the expression level of X gametologs in a given species when Y gametologs were present. The tissue specificity index (TSI) for a given gene was calculated as the expression level (TPM) in the tissue with the highest expression level divided by the sum of expression values in all tissues (Julien, et al. 2012). Expression values are listed in Supplementary Table 4a-d. Genes used in the different analyses are listed in Supplementary Table 5a-e. Furthermore, we verified the general trends of Y expression loss using two alternative metrics: FPKM values obtained from Hisat2 (default options; v2.0.2)-Cufflinks (default options; v2.2.1) (Trapnell, et al. 2013; Kim, et al. 2015) and
normalized using 1-to-1 orthologs and a scaling procedure (Brawand, et al. 2011), and raw count obtained from Kallisto (Bray, et al. 2016) and normalized using 1-to-1 orthologs and the TMM normalization in the EdgeR package (Robinson, et al. 2010) (Supplementary Figure 5-6).

**Statistical analyses**

All statistical analyses were performed using the R package, standard libraries. Data were plotted using the R package, “ggplot2" library (www.ggplot2.org). We calculated significant differences using the non-parametric Mann–Whitney U test, which was corrected for false discovery rate using the Benjamin–Hochberg correction at 0.05. We tested current/ancestral ratios and female/male ratios against theoretical distributions with either a fixed median of 0 (i.e., similar expression levels in current and ancestral states or between males and females) or -1 (i.e., lower expression levels in current versus ancestral states or males versus females). Ancestral state reconstruction was calculated in the R package using the libraries: phytools, ape, and maps. Life-history traits for placental mammals are reported in the Supplementary Table 6.

**Results**

**Conservation of ancestral Y gametologs across 25 species of placental mammals**

Previous studies reported Y-linked genes in 15 species of placental mammals (Figure 1) (Skaletsky, et al. 2003; Hughes, et al. 2005; Hughes, et al. 2012; Li, et al. 2013; Bellott, et al. 2014; Cortez, et al. 2014; Soh, et al. 2014; Bidon, et al. 2015; Skinner, et al. 2016; Tomaszkiewicz, et al. 2016; Janecka, et al. 2018). We recovered Y-linked genes from ten additional species of placental mammals (Figure 1, names marked with an asterisk; Supplementary Tables 1-3) based on a subtraction approach that uses male and female RNA-seq data (see Methods). We worked with species from lineages that were not previously explored, such as lemurs (primates), tree shrew (scandentia), rabbit (lagomorpha), hedgehog (insectivores), armadillo (xenarthra), and tenrec (afrotheria). Ultimately, we compiled a catalog of Y gametologs across 25 species covering the four major groups of placental mammals: euarchontoglires, laurasiatherian, xenarthra, and afrotheria (Figure 1).

We focused the analyses on the ancestral Y gametologs that originated ~180 (stratum 1) and ~116 (stratum 2) million years ago (ma) (Cortez, et al. 2014) before the radiation of the four major groups
of placental mammals, which took place 90-100 ma (estimates retrieved from the TimeTree database; www.timetree.org).

Earlier work (Lahn and Page 1999; Hughes, et al. 2012; Li, et al. 2013; Cortez, et al. 2014; Janecka, et al. 2018) identified 16 ancestral Y gametologs from strata 1-2 on the Y chromosomes of placental mammals (Figure 1). Our analyses in ten additional species, from five new orders, identified the same 16 ancestral gametologs, although we did not limit the analyses to these specific genes (see Methods). This result indicated that the ancestor of placental mammals already presented a reduced and very distinctive set of genes on its MSY.

We observed that Y gametologs have followed different evolutionary trajectories across the 25 species of placental mammals (Figure 1). Seven of them are present in all the analyzed species (SRY, RBMY, TSPY, DDX3Y, UTY, USP9Y, and ZFY), whereas others seem to have been lost one (EIF2S3Y), two (HSFY, UBA1Y, EIFA1Y, CULB4Y), three (TXLNGY), four (KDM5D) or five times (RPS4Y and TMSB4Y). From the distribution of Y gametologs across the species’ phylogeny, we could not distinguish a clear pattern of why some Y genes are more frequently lost than others. We, therefore, explored the possibility that expression level changes could play an important role in the loss of Y-linked genes.

Expression evolution of XY gametologs in 17 species of placental mammals
We collected or generated RNA-seq data for 17 species of placental mammals and estimated the expression levels for X and Y gametologs in males and females (Supplementary Table 1). We first analyzed whether XY gametologs, overall, experienced changes in expression levels following the recombination arrest between X and Y chromosomes. To do so, we compared for each pair of XY gameologs the current expression levels (median value across tissues) against ancestral expression levels (the expression on the proto-sex chromosomes). To obtain ancestral expression levels, we exploited the fact that current sex chromosomes derived from ancestral autosomes and have autosomal counterparts in species with non-homologous sex chromosomes (Julien, et al. 2012). Thus, we calculated ancestral expression levels based on the median value across the expression levels of 1-1 orthologs of XY gametologs in the platypus, chicken and Anolis carolinensis (Supplementary Table 4a-d).
We found that Y gametologs, by and large, showed an acute 3-fold decrease in expression compared to the ancestral expression levels, whereas X gametologs in males showed minor expression loss and, lastly, X gametologs in females were the only ones that maintained the ancestral expression levels (Figure 2a; Mann–Whitney U test against a distribution with fixed medians). This general pattern shows slight variations within the mammalian groups: Y gametologs show greater expression loss in rodents and afrotheria/xenarthra. Moreover, expression levels of X gametologs in both males and females are also depressed in primates and afrotheria/xenarthra (Figure 2b; Mann–Whitney U test against a distribution with fixed medians). Since gene expression levels across vertebrate species are more strongly correlated with tissue-specificity rather than with the species’ phylogeny (Brawand, et al. 2011; Merkin, et al. 2012; Cardoso-Moreira, et al. 2019; Naqvi, et al. 2019), we examined the evolution of XY gametologs in each tissue separately. We found that somatic tissues (brain, heart, kidney, and liver) follow the general pattern. That is, a 3-fold decrease in expression levels of Y gametologs, a minor expression loss of X gametologs in males and a stable, similar to the ancestral expression, for X gametologs in females (Figure 2c). In gonads, conversely, X gametologs also showed large expression reduction in males (Figure 2c), which is due to the transcriptional silencing of genes by Meiotic Sex Chromosome Inactivation (MSCI) mechanism (Turner 2007). We found that the general pattern of XY expression change is maintained, though we used maximum expression levels as metric instead of median values (Supplementary Figure 7).

Next, we analysed that expression levels of Y gametologs across individual species and observed an important variation, though in several instances the values fluctuate around a 3-fold decrease in expression levels (median values across tissues) compared to the ancestral expression levels (Figure 2d; Mann–Whitney U test between groups). Interestingly, species showing the greatest Y expression loss are not found scattered across the species’ tree but they cluster in two groups (Figure 2d); three rodent species from the same super-family (mouse, rat, and hamster; Muroidea) and the two afrotheria/xenarthra species (armadillo and tenrec). Moreover, the amount of expression loss of Y gametologs in placental species do not correlate with the species’ phylogeny (Phylogenetic Generalized Least Squares, PGLS, P > 0.05; Figure 2d), species’ life-history traits (PGLS P > 0.05; Supplementary Table 6), or the number of conserved ancestral Y gametologs (PGLS P > 0.05; Figure 1 and 2c).
Every species of placental mammals has been subjected to specific selection pressures during evolution, which, in turn, could have differentially affected the rate of expression decline of the Y chromosome. We reconstructed the hypothetical expression levels of Y gametologs in the ancestors of primates, rodents, laurasiatherios, and afrotherios/xenarthra. We found that the ancestor of primates and laurasiatherios showed ~3-fold expression loss, the ancestors of rodents presented ~4-fold expression loss, and the ancestor of afrotherios/xenarthra may have exhibited an acute ~8-fold expression loss. These results suggest that, 90-100 million years ago, the ancestor of placental mammals already featured a Y chromosome with less than half of its expression levels.

**Evolution of tissue-specificity**

Next, we examined the expression evolution of Y gametologs across placental species. Theory predicts that Y genes could likely evolve male-beneficial functions (Burgoyne 1987; Chandley and Cooke 1994; Skaletsky, et al. 2003) because they are, ultimately, male-specific genes. So, it was not surprising that Y gametologs in many species showed biased expression towards testis compared to both X gametologs and proto-sex chromosomes, which, generally, maintained ubiquitous expression across tissues (Figure 3a). We found that HSFY is the only Y gene with testis-specificity that retained the expression pattern of the proto-sex chromosomes. In all other instances, remarkably, Y gametologs gained testis-biased expression due to an accelerated expression loss in somatic tissues compared to the testis (Figure 3b). That is, Y gametologs that evolved testis-specific expression did not lose expression in this tissue but lost significant expression levels in somatic tissues (Figure 3b; Mann–Whitney U test against a distribution with a fixed median and between groups). Alternatively, Y gametologs that did not evolve testis-biased expression showed similar expression loss in both somatic tissues and testis (Figure 3b; Mann–Whitney U test against a distribution with a fixed median and between groups). Although some Y genes acquired testis-biased expression in specific species (i.e. CULB4Y in pig, armadillo, and tenrec, EIF1A1Y in marmoset, ZFY in human, mouse, rat, and hamster or USP9Y in mouse and rat; Figure 3a), our data suggests that at least RBMY, TSPY, and UBA1Y were already testis-specific in the ancestor of placental mammals.

At the gene level, we found that most Y gametologs have lost expression levels across tissues (Figure 4a; Mann–Whitney U test against a distribution with a fixed median), with TMSB4Y showing the greatest expression decline (Figure 4a–e). Remarkably, however, three Y gametologs (HSFY, EIF2S3Y,
and ZFY) have consistently maintained the ancestral expression levels across tissues, suggesting that selection has been particularly keen on maintaining the expression level of these three genes (Figure 4a-e, genes in blue; Mann–Whitney U test against a distribution with a fixed median). In agreement with the results shown in Figure 3, Y gametologs that have gained testis-specificity in placental species (i.e. TSPY, RBMY, UBA1Y, and CULB4Y) exhibited considerable expression decline in somatic tissues (Figure 4a-e, genes in pink) but only minor expression loss in testis. In general, X gametologs appear to have more conserved expression levels in males and females (Figure 4a), with relatively higher expression in females compared to males, independently on the tissue (Figure 4a-e).

**Dosage compensation mechanisms**

We recurrently observed significantly higher expression of X gametologs in females compared to the X gametologs in males across our analyses (Figures 2, 4 and 5a-d). This result expands on previous notions that in human and mouse a majority of X gametologs escape X chromosome inactivation in females (Bellott, et al. 2014; Tukiainen, et al. 2017). Furthermore, we estimated that overall in placental mammals the additional expression of X gametologs in females corresponded to a +20%/+30% increase, which is remarkably similar to the measured extra expression (+30%) of X-linked genes escaping X chromosome inactivation in humans (Tukiainen, et al. 2017). Interestingly, the combined expression of XY gametologs in males compared to the expression of X gametologs in females did not result in a balanced scenario where both sexes have similar expression levels (Figure 5a-d). Instead, we consistently observed greater expression levels of XY gametologs in males compared to females, regardless of the tissue that is analyzed (Figure 4b; Mann–Whitney U test against a distribution with a fixed median). In other words, despite the expression loss of Y genes, the sum of XY expression values is larger than the expression levels of the active X gametologs and the expression of the X gametologs escaping X chromosome inactivation. This observation strongly suggests that expression levels of X gametologs have also declined in both sexes. The pattern of male-biased expression levels for XY gametologs was also noted in humans across a variety of tissues (Tukiainen, et al. 2017).

Retrogenes allow genes to acquire new genomic locations, change tissue-specificity (Carelli, et al. 2016) or escape specific molecular processes such as X chromosome inactivation (Hurst, et al. 2015). In particular, it has been proposed that gene duplications, such as retrogenes, could compensate for the loss of Y genes (Hughes, et al. 2015) or compensate gene dosage between sexes (Andres, et
al. 2008) by balancing the expression decline of Y gametologs. Thus, we screened the genomes of placental mammals for retrogenes that derived from XY gametologs. We considered for further analyses only those retrogenes under purifying selection ($d_s < 1$) and showing active transcription (TPM > 1) (Carelli, et al. 2016). Four XY gametologs ($RPS4XY$, $TMSB4XY$, $EIFA1XY$, and $EIF2S3XY$) comprised ~95% of the total number of retrogenes (Figure 6a). We found that XY-derived retrogenes are lowly expressed (Supplementary Table 4a-d) and expression levels of parental XY gametologs remained similar between males and females regardless of whether the expression of retrogenes was taken into account or not considered at all (Figure 6b), suggesting that retrogenes are expressed at similar levels in both sexes and do not over-compensate Y expression specifically in males. We observed similar results when we limited the analyses to the expression levels of parental X gametologs and X-derived retrogenes (Figure 6c) or parental Y gametologs and Y-derived retrogenes (Figure 6d). These results indicate a minor role of retrogenes in dosage compensation.

**Discussion**

In this study, we explored and expanded the knowledge of the evolution of XY gametologs, particularly gene expression evolution, by analyzing species from the four major groups of placental mammals. We focused on ancestral XY gametologs that stopped recombination before the radiation of placental mammals; other recent studies investigated the variability of ampliconic families (Brashear, et al. 2018; Vegesna, et al. 2019). Our results suggested that the ancestor of placental mammals already presented a limited set of Y-linked genes. Some Y gametologs show remarkable conservation across placental species. These genes code for proteins with functions in transcription regulation or during spermatogenesis: $SRY$ is the master-sex regulator of testis development (Li, et al. 2014); $ZFY$ is necessary for spermatogenesis (Nakasuji, et al. 2017) and could be a key regulator of genome-wide dosage effects together with $ZFX$ (Raznahan, et al. 2018); $RBMY$, $TSPY$, $DDX3Y$, $USP9Y$, and $EIF2S3Y$ are essential for spermatogenesis (Colaco and Modi 2018); finally, $UTY$ is important in tumor suppression (Gozdecka, et al. 2018). Interestingly, in species such as the spiny rat *Tokudaia muenninki* (Murata, et al. 2016) and the mole vole *Ellobius* (Mulugeta, et al. 2016) that lost the placental Y chromosome, some of the eight conserved Y gametologs moved from the placental Y chromosome onto the new sex chromosomes before the placental Y chromosome was lost ($ZFY$, $EIF2S3Y$, $UTY$, $DDX3Y$, $USP9Y$ in *T. muenninki*, and $ZFY$, $EIF2S3Y$, and $USP9Y$ in *Ellobius* species). These results indicate that retention of these remarkably conserved Y-linked genes is more important than preserving entire sex chromosomes.
Previous work examined the expression patterns of X- and Y-linked genes in primates and rodents (Lingenfelter, et al. 2001; Xu and Distech 2006; Isensee, et al. 2008; Xu, et al. 2008; Wilson Sayres and Makova 2013; Cortez, et al. 2014; Slavney, et al. 2016; Vegesna, et al. 2019; Godfrey, et al. 2020). Some of these studies already reported that Y chromosomes in placental mammals had lost expression output. Nevertheless, our data allowed to examine the evolution of Y expression across placental groups and indicated that Y gametologs already showed significant expression decline in the ancestor of placental mammals and testis-specific genes. The reduction in expression levels could have originated through a gradual decaying process or by a swift expression decline followed by long-lasting expression stasis. We found that the expression of Y chromosomes is similar in most placental species, which would favor an evolutionary model where Y chromosomes followed swift expression decline and are currently under expression stasis. Rodents from the super-family Muroidea and the two species from afrotheria/xenarthra presented greater Y expression loss, which could result from accelerated group-specific Y decline. Moreover, we found that three Y gametologs have minor expression loss in placental species. It would appear that these Y genes have successfully purged deleterious mutations from their promoter regions and their functions and haploinsufficiency are probably crucial. One of these genes is ZFY, a potential genome-wide dosage effects regulator along with ZFX (Raznahan, et al. 2018). A second gene is EIF2S3Y, which is essential for spermatogenesis and embryogenesis (Mazeyrat, et al. 2001; Yamauchi, et al. 2014). The third gene is HSFY, which is also critical for spermatogenesis (Colaco and Modi 2018). Furthermore, we also uncovered that accelerated expression loss in somatic tissues compared to testis expression levels has been the main mechanism that allowed Y-linked genes to gain testis-specificity, as also reported in (Cortez, et al. 2014), and, likely, specialized in relevant functions during spermatogenesis (Colaco and Modi 2018; Liu 2019). Lastly, we found little evidence relating Y expression loss enabling Y gene loss and that retrogenes, in general, compensate for either expression or gene losses.

We considered that the haploinsufficient nature of XY gametologs (Bellott, et al. 2014; Cortez, et al. 2014) would be incompatible with the expression decline of Y gametologs because males, eventually, would lack sufficient proteins to maintain cellular homeostasis. One potential solution to this problem could have been to increase the expression levels of the X chromosome in males. However, X genes in placental mammals have reached maximum expression levels (Hurst, et al. 2015). The alternative was to coordinate the expression of Y-linked genes with that of X-linked genes. Recent
work supports this idea of active cross-regulation between X and Y chromosomes regarding gene expression levels (Raznahan, et al. 2018). This cross-regulation between sex chromosomes has the potential to explain that X gametologs lost expression output as a consequence of Y expression decay. Data retrieved from humans (Tukiainen, et al. 2017), and the results from this study, indicated that expression levels of XY gametologs in males are slightly higher than the combined expression levels of X gametologs in females. So, it would appear that XY gametologs are not perfectly balanced in placental mammals. This premise, however, assumes that XY gametologs show complete functional redundancy, which is likely not the case. Although X and Y gametologs may act on similar cellular processes or interact with similar protein partners, Y gametologs have likely evolved specialized functions and affinities, for example, in testis-related functions. The available experimental data supports this hypothesis given that in transgenic mice, only the overexpression of \textit{SOX3} and \textit{EIF2S3X} could functionally replace \textit{SRY} and \textit{EIF2S3Y}, respectively (Sutton, et al. 2011; Yamauchi, et al. 2016). Moreover, it has been shown that \textit{UTY} has lower demethylase activity compared to \textit{UTX}, owing to point substitutions affecting substrate binding (Walport, et al. 2014; Gozdecka, et al. 2018; Gazova, et al. 2019). It is likely, therefore, that other Y genes also exhibit diminished activities caused by the accumulation of point substitutions. So, an interesting hypothesis worth studying in future work is that the extra expression of Y gametologs compensates for the reduced Y activity/affinity. Further experiments will be required to test this hypothesis. Interestingly, in a recent study, it was found that Y gametologs can show increased expression levels in specific tissues in humans (Godfrey, et al. 2020). So, the mechanisms that maintain XY dosage between sexes and the general pattern of Y expression loss we observed across mammals can be reversed to meet yet unknown tissue-specific protein requirements.

\textbf{Declarations}

\textbf{Ethics approval and consent to participate}

ERC Ethics Screening panel (associated with ERC Consolidator Grant 615253) and ethic committee in Lausanne (authorization 504/12).

\textbf{Competing interests}
The authors declare that they have no competing interests

Funding
This research was supported by grants from the European Research Council (No. 615253) and Swiss National Science Foundation (No. 146474) to Professor Henrik Kaessmann. Frontiers in Science CONACYT grant (No. FC-2016 /1682) and a Royal Society Newton Advanced Fellowship (no. NA160564) to DC, TS, and AOU. NKFIH-2558-1/2015 and ÉLVONAL-KKP 126949 to TS; a NERC grant (NE/P004121/1), a Royal Society Dorothy Hodgkin Research Fellowship (DH071902) and a CONACyT-AMC-Royal Society Newton mobility grant to AOU; travel grants SEP-UNAM-FUNAM Programa de Capacitación en Métodos de Investigación/2016 to PCP and CRS and a special grant from the Undergraduate Programme on Genomic Sciences – UNAM to PCP.

Authors’ contributions
DC designed the study. JH. and AL. prepared the samples and produced the sequencing libraries. MM-P, MT, LA, VG, AG, DF, PC-P, KD, and DC performed the analyses. TS, AOU and DC supervised the study and wrote the paper. All authors read and approved the final manuscript.

Acknowledgements
We thank Professor Henrik Kaessmann for his constant support and positive input throughout the project. We also thank Prof. Kaessmann’s laboratory where RNA was extracted, libraries were prepared and many initial analyses and discussions were conducted. We also thank K. Harshman and the Lausanne Genomics Technology Facility for high-throughput sequencing support. We thank Dr. Athanasia Tzika, Dr. Michel C. Milinkovitch and Dr. Philipp Khaitovich for providing key samples.

Data access
This project has been deposited at NCBI-SRA database (www.ncbi.nlm.nih.gov/sra) under the accession BioProject PRJNA580502. Y-linked sequences found in this study are available in the Supplementary Material. Code and command line to apply the subtraction approach on male/female RNA-seq data are available in the Supplementary Material.

Supplementary Material

http://mc.manuscriptcentral.com/gbe
1) Supplementary Tables 1-6. Details about the samples used. XY genes used. Y genes obtained using the subtraction approach. Expression data used. Lists of genes used. Life-history traits used.

2) Supplementary Figures 1-7. Cross-gametolog mapping. RNA-seq data before and after normalization for 1-to-1 orthologs and for all genes. PCA analysis of expression data. Expression patterns using hisat2/cufflinks and read counts. Expression of XY gametologs using maximum expression across tissues.

3) Sequences of Y-linked transcripts found in this study.

4) Code and command line to apply the subtraction approach on male/female RNA-seq data.

5) Sequence alignments of Y sequences from the subtraction-based method against Y sequences from direct-sequencing methods for human, marmoset, mouse, rat, pig, cow, and opossum.

References
Acosta A, Suarez-Varon G, Rodriguez-Miranda LA, Lira-Noriega A, Aguilar-Gomez D, Gutierrez-Mariscal M, Hernandez-Gallegos O, Mendez-de-la-Cruz F, Cortez D. 2019. Corytophanids Replaced the Pleurodont XY System with a New Pair of XY Chromosomes. Genome Biol Evol 11:2666-2677.
Akagi T, Henry IM, Tao R, Comai L. 2014. Plant genetics. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. Science 346:646-650.
Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403-410.
Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.
Andres O, Kellermann T, Lopez-Giraldez F, Rozas J, Domingo-Roura X, Bosch M. 2008. RPS4Y gene family evolution in primates. BMC Evol Biol 8:142.
Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet 14:113-124.
Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho TJ, Koutseva N, Zaghlul S, Graves T, Rock S, et al. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature 508:494-499.
Bidon T, Schreck N, Hailer F, Nilsson MA, Janke A. 2015. Genome-Wide Search Identifies 1.9 Mb from the Polar Bear Y Chromosome for Evolutionary Analyses. Genome Biol Evol 7:2010-2022.
Brashear WA, Raudsepp T, Murphy WJ. 2018. Evolutionary conservation of Y Chromosome ampliconic gene families despite extensive structural variation. Genome Res 28:1841-1851.
Brawand D, Soumillon M, Necsulea A, Julien P, Csardi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, et al. 2011. The evolution of gene expression levels in mammalian organs. Nature 478:343-348.
Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525-527.
Burgoyne PS. 1987. The role of the mammalian Y chromosome in spermatogenesis. Development 101 Suppl:133-141.
Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, Liechti A, Ascencao K, Rummel C, Ovchinnikova S, et al. 2019. Gene expression across mammalian organ development. Nature 571:505-509.
Carelli FN, Hayakawa T, Go Y, Imai H, Warnefors M, Kaessmann H. 2016. The life history of retrocopies illuminates the evolution of new mammalian genes. Genome Res 26:301-314.
Carvalho AB, Clark AG. 2013. Efficient identification of Y chromosome sequences in the human and Drosophila genomes. Genome Res 23:1894-1907.
Chandley AC, Cooke HJ. 1994. Human male fertility--Y-linked genes and spermatogenesis. Hum Mol Genet 3 Spec No:1449-1452.
Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci 355:1563-1572.
Colaco S, Modi D. 2018. Genetics of the human Y chromosome and its association with male infertility. Reprod Biol Endocrinol 16:14.
Cornejo-Paramo P, Dissanyake DSB, Lira-Noriega A, Martinez-Pacheco ML, Acosta A, Ramirez-Suastegui C, Mendez-de-la-Cruz FR, Szekely T, Urrutia AO, Georges A, et al. 2020. Viviparous Reptile Regarded to Have Temperature-Dependent Sex Determination Has Old XY Chromosomes. Genome Biol Evol 12:924-930.
Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, Grutzner F, Kaessmann H. 2014. Origins and functional evolution of Y chromosomes across mammals. Nature 508:488-493.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797.
Gazova I, Lengeling A, Summers KM. 2019. Lysine demethylases KDM6A and UTY: The X and Y of histone demethylation. Mol Genet Metab 127:31-44.
Ghenu AH, Bolker BM, Melnick DJ, Evans BJ. 2016. Multicopy gene family evolution on primate Y chromosomes. BMC Genomics 17:157.
Godfrey AK, Naqvi S, Chmatal L, Chick JM, Mitchell RN, Gygi SP, Skaletsky H, Page DC. 2020. Quantitative analysis of Y-Chromosome gene expression across 36 human tissues. Genome Res. Goto H, Peng L, Makova KD. 2009. Evolution of X-degenerate Y chromosome genes in greater apes: conservation of gene content in human and gorilla, but not chimpanzee. J Mol Evol 68:134-144.
Gozdecka M, Meduri E, Mazan M, Tzelepis K, Dudek M, Knights AJ, Pardo M, Yu L, Choudhary JS, Metzakopan E, et al. 2018. UTX-mediated enhancer and chromatin remodeling suppresses myeloid leukemogenesis through noncatalytic inverse regulation of ETS and GATA programs. Nat Genet 50:883-894.
Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644-652.
Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Graves T, Fulton RS, Dugan S, Ding Y, Buhay CJ, Kremizki C, et al. 2012. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. Nature 483:82-86.
Hughes JF, Skaletsky H, Koutseva N, Pyntikova T, Page DC. 2015. Sex chromosome-to-autosome transposition events counter Y-chromosome gene loss in mammals. Genome Biol 16:104.
Hughes JF, Skaletsky H, Pyntikova T, Graves TA, van Daalen SK, Minx PJ, Fulton RS, McGrath SD, Locke DP, Friedman C, et al. 2010. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. Nature 463:536-539.
Hughes JF, Skaletsky H, Pyntikova T, Minx PJ, Graves T, Rozen S, Wilson RK, Page DC. 2005. Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. Nature 437:100-103.

Hurst LD, Ghanbarian AT, Forrest AR, consortium F, Huminiecki L. 2015. The Constrained Maximal Expression Level Owing to Haploidy Shapes Gene Content on the Mammalian X Chromosome. PLoS Biol 13:e1002315.

Isensee J, Witt H, Pregla R, Hetzer R, Regitz-Zagrosek V, Noppinger PR. 2008. Sexually dimorphic gene expression in the heart of men and women. J Mol Med (Berl) 86:61-74.

Janecka JE, Davis BW, Ghosh S, Paria N, Das PJ, Orlando L, Schubert M, Nielsen MK, Stout TAE, Brashier W, et al. 2018. Horse Y chromosome assembly displays unique evolutionary features and putative stallion fertility genes. Nat Commun 9:2945.

Julien P, Brawand D, Soumillon M, Necsulea A, Liechti A, Schutz F, Daish T, Grutzner F, Kaessmann H. 2012. Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. PLoS Biol 10:e1001328.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357-360.

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

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359.

Li G, Davis BW, Raudsepp T, Pearks Wilkerson AJ, Mason VC, Ferguson-Smith M, O'Brien PC, Waters PD, Murphy WJ. 2013. Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. Genome Res 23:1486-1495.

Liu WS. 2019. Mammalian Sex Chromosome Structure, Gene Content, and Function in Male Fertility. Annu Rev Anim Biosci 7:103-124.

Lingenfelter PA, Delbridge ML, Thomas S, Hoekstra HE, Mitchell MJ, Graves JA, Disteche CM. 2001. Expression and conservation of processed copies of the RBMX gene. Mamm Genome 12:538-545.

Loytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. Proc Natl Acad Sci U S A 102:10557-10562.

Marin R, Cortez D, Lamanna F, Pradeepa MM, Leushkin E, Julien P, Liechti A, Halbert J, Bruning T, Mossinger K, et al. 2017. Convergent origination of a Drosophila-like dosage compensation mechanism in a reptile lineage. Genome Res 27:1974-1987.

Marques AC, Dupanloup I, Vinckenbosch N, Reymond A, Kaessmann H. 2005. Emergence of young human genes after a burst of reposition in primates. PLoS Biol 3:e357.

Mazeyrat S, Saut N, Grigoriev V, Mahadevaiah SK, Ojarikre OA, Rattigan A, Bishop C, Eicher EM, Mitchell MJ, Burgoyne PS. 2001. A Y-encoded subunit of the translation initiation factor Eif2 is essential for mouse spermatogenesis. Nat Genet 29:49-53.

Merkin J, Russell C, Chen P, Burge CB. 2012. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science 338:1593-1599.

Mulgeta E, Wassenaar E, Sleddens-Linkels E, van IWF, Heard E, Grootejod JA, Just W, Gribnau J, Baarends WM. 2016. Genomes of Ellobius species provide insight into the evolutionary dynamics of mammalian sex chromosomes. Genome Res 26:1202-1210.

Murata C, Kuroki Y, Imoto I, Kuroiwa A. 2016. Ancestral Y-linked genes were maintained by translocation to the X and Y chromosomes fused to an autosomal pair in the Okinawa spiny rat Tokudaia muenninki. Chromosome Res 24:407-419.
Nakasuji T, Ogonuki N, Chiba T, Kato T, Shiozawa K, Yamatoya K, Tanaka H, Kondo T, Miyado K, Miyasaka N, et al. 2017. Complementary Critical Functions of Zfy1 and Zfy2 in Mouse Spermatogenesis and Reproduction. PLoS Genet 13:e1006578.

Naqvi S, Godfrey AK, Hughes JF, Goodheart ML, Mitchell RN, Page DC. 2019. Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. Science 365.

Pearks Wilkerson AJ, Raudsepp T, Graves T, Albracht D, Warren W, Chowdhary BP, Skow LC, Murphy WJ. 2008. Gene discovery and comparative analysis of X-degenerate genes from the domestic cat Y chromosome. Genomics 92:329-338.

Plath K, Mlynczak-Evans S, Nusinow DA, Panning B. 2002. Xist RNA and the mechanism of X chromosome inactivation. Nature Genet 36:233-278.

Rando HM, Walston WH, Johnson JL, Stutchman JT, Trut LN, Farre M, Kukekova AV. 2019. The Red Fox Y-Chromosome in Comparative Context. Genes (Basel) 10.

Raznahan A, Parikshak NN, Chandran V, Blumenthal JD, Clasen LS, Alexander-Bloch AF, Zinn AR, Wangsa D, Wise J, Murphy DGM, et al. 2018. Sex-chromosome dosage effects on gene expression in humans. Proc Natl Acad Sci U S A 115:7398-7403.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140.

Sekido R, Lovell-Badge R. 2008. Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. Nature 453:930-934.

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischau AM, Lovell-Badge R, Goodfellow PN. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346:240-244.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, 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.

Skinner BM, Sargent CA, Churcher C, Hunt T, Herrero J, Loveland JE, 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.

Slavney A, Arbiza L, Clark AG, Keinan A. 2016. Strong Constraint on Human Genes Escaping X-Inactivation Is Modulated by their Expression Level and Breadth in Both Sexes. Mol Biol Evol 33:384-393.

Smeds L, Kojola I, Ellegren H. 2019. The evolutionary history of grey wolf Y chromosomes. Mol Ecol 28:2173-2191.

Soh YQ, Alfoldi J, Pyntikova T, Brown LG, Graves T, Minx PJ, Fulton RS, Kremitzki C, Kouteva N, Mueller JL, et al. 2014. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. Cell 159:800-813.

Sutton E, Hughes J, White S, Sekido R, Tan J, Arboleda V, Rogers N, Knowler K, Rowley L, Eyre H, et al. 2011. Identification of SOX3 as an XX male sex reversal gene in mice and humans. J Clin Invest 121:328-341.

Tomaszkiewicz M, Rangavittal S, Cechova M, Campos Sanchez R, Fescemyer HW, Harris R, Ye D, O'Brien PC, Chikhi R, Ryder OA, 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.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 31:46-53.

Tukiainen T, Villani AC, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, Kirby A, et al. 2017. Landscape of X chromosome inactivation across human tissues. Nature 550:244-248.
Turner JM. 2007. Meiotic sex chromosome inactivation. Development 134:1823-1831.
Vegesna R, Tomaszkiewicz M, Medvedev P, Makova KD. 2019. Dosage regulation, and variation in gene expression and copy number of human Y chromosome ampliconic genes. PLoS Genet 15:e1008369.
Walport LJ, Hopkinson RJ, Vollmar M, Madden SK, Gileadi C, Oppermann U, Schofield CJ, Johansson C. 2014. Human UTY(KDM6C) is a male-specific N-methyl lysyl demethylase. J Biol Chem 289:18302-18313.
Wilson MA, Makova KD. 2009. Evolution and survival on eutherian sex chromosomes. PLoS Genet 5:e1000568.
Wilson Sayres MA, Makova KD. 2013. Gene survival and death on the human Y chromosome. Mol Biol Evol 30:781-787.
Xu J, Deng X, Disteche CM. 2008. Sex-specific expression of the X-linked histone demethylase gene Jarid1c in brain. PLoS One 3:e2553.
Xu J, Disteche CM. 2006. Sex differences in brain expression of X- and Y-linked genes. Brain Res 1126:50-55.
Yamauchi Y, Riel JM, Ruthig VA, Ortega EA, Mitchell MJ, Ward MA. 2016. Two genes substitute for the mouse Y chromosome for spermatogenesis and reproduction. Science 351:514-516.
Yamauchi Y, Riel JM, Stoytcheva Z, Ward MA. 2014. Two Y genes can replace the entire Y chromosome for assisted reproduction in the mouse. Science 343:69-72.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555-556.

Figure Legends

Figure 1. Distribution of ancestral Y genes in 25 placental mammals. Green squares represent the presence of a Y gene, whereas white squares represent the absence of a Y gene. Blue squares denote Y genes annotated as pseudogenes but with detectable expression levels. Yellow squares represent Y genes with unclear status due to lack of expression data for SRY (mostly expressed during development), lack of testis data and absence of testis-specific genes (RBMY, CUL4B, and TSPY) for mouse lemur, missing genes in published references (RBMY in cat and several Y genes in polar bear), unclear Y-specific transcripts (TMSB4Y in panda, TSPY in tenrec). The ten species of placental mammals for which Y-linked genes were obtained in this study have their names marked with an asterisk. The presence of XY-derived retrogenes is highlighted with an r. The phylogenetic tree highlights the major groups of placental mammals and their divergence time (in millions of years).

Figure 2. Expression levels of ancestral XY gametologs in 17 placental mammals. a) Boxplots representing the current/ancestral expression ratio of X and Y gametologs in males and females (n = 200 genes). b) Boxplots representing the current/ancestral expression ratio of X and Y gametologs in males and females according to the phylogenetic group: Pr data from primates including tree shrew (n = 60 genes); Ro data from rodents including rabbit (n = 52 genes); La data from laurasiatheria (n = 65 genes); At data from armadillo and tenrec (n = 23 genes). c) Boxplots representing the current/ancestral expression ratio of X and Y gametologs in males and females in brain, gonads, heart, kidney, and liver; data from panda was not included. a-c) Significant differences (Mann–Whitney U test): Benjamin–Hochberg-corrected P < 0.05 of ratios against a distribution with a fixed median of 0 (i.e., similar expression levels of current and ancestral states) or -1 (i.e., 2-fold
loss in expression levels between current and ancestral states). Gray filled squares denote nonsignificant differences, whereas yellow filled squares denote significant differences. Error bars, maximum and minimum values, excluding outliers. d) Dot plot representing median current/ancestral expression ratios of X and Y gametologs in males and females for each species (n = 200 genes). Species are ordered on the X-axis based on the Y gametologs values (decreasing order). Significant differences (Mann–Whitney U test) between groups. e) Dot plot representing the median current/ancestral expression ratio of Y gametologs according to the phylogeny of the species. Colored bars indicate the reconstructed ancestral expression levels for each group. Purple dots on the phylogenetic tree indicate species divergence times. a-b and e-d) Median expression levels across tissues (TPM) were used.

Figure 3. Patterns of tissue specificity of ancestral XY gametologs in 15 placental mammals. a) Tissue specificity index (TSI) for X and Y gametologs in males and females. Estimates of TSI in the proto-sex chromosomes are also shown. Genes with a TSI smaller than 0.6 are colored in grey, whereas genes with a TSI greater than 0.6 are colored based on the tissue showing the highest expression level. Data from panda and mouse lemur was not included. b) Boxplots representing the current/ancestral expression ratios of Y gametologs in testis and somatic tissues. Y gametologs were divided in two groups according to their testis-specificity (genes with a testis-specificity index above 0.6 and genes with a testis-specificity index below 0.6). HSFY was excluded from the analyses. Ancestral expression levels were calculated for testis and somatic tissues separately. Significant differences (Mann–Whitney U test): Benjamin–Hochberg-corrected P < 0.05 between groups and of ratios against a distribution with a fixed median of 0 (i.e., similar expression levels of current and ancestral states) or -1 (i.e., 2-fold loss in expression levels between current and ancestral states). Gray filled squares denote nonsignificant differences, whereas yellow filled squares denote significant differences. Error bars, maximum and minimum values, excluding outliers.

Figure 4. Expression levels of XY gametologs in different tissues. Dot plots representing median current/ancestral expression ratios of X and Y gametologs in males and females (n = 15-7 genes per gametolog; see Figure 1) in a) brain, b) heart, c) kidney, d) liver, and e) gonads (testis/ovary). a-e) Gametologs are ordered on the X-axis based on the Y gametologs values (decreasing order). Significant differences (Mann–Whitney U test): Benjamin–Hochberg-corrected P < 0.05 of Y gametologs ratios against a distribution with a fixed median of 0 (i.e., no difference in current/ancestral expression levels). Gray filled squares denote nonsignificant differences, whereas yellow filled squares denote significant differences. Given that different tissues were available, data from panda and mouse lemur was not included.

Figure 5. Female/Male expression ratios of XY gametologs in different tissues. Dot plots representing median female/male expression ratios of X gametologs (red dots), and X and Y gametologs (orange dots) in a) brain, b) heart, c) kidney, and d) liver; gonads were not analyzed because ovaries and testis represent different tissues. a-d) Gametologs are ordered on the X-axis based on the Xf ÷ (Xm + Y) values (decreasing order). Formulas indicate the operations performed with X and Y gametologs to calculate the ratios. Significant differences (Mann–Whitney U test): Benjamin–Hochberg-corrected P < 0.05 between groups (red vs. orange dots; excluding male-biased HSFXY) and of Xf ÷ (Xm + Y) ratios against a distribution with a fixed median of 0 (i.e., similar expression levels of current and ancestral states). Gray filled squares denote nonsignificant
differences, whereas yellow filled squares denote significant differences. Given that different tissues were available, data from panda and mouse lemur was not included.

**Figure 6. Expression levels of XY-derived retrogenes.**

* a) The number of XY-derived retrogenes. Brown bars show the total number of XY-derived retrogenes, orange bars show retrogenes with $d_S < 1$, and yellow bars show retrogenes with $d_S < 1$ and TPM > 1.

* b) Violin plots representing the expression ratios of X gametologs in females compared to the combined expression of X and Y gametologs in males. Purple violin plot incorporates the expression of XY-derived retrogenes (denoted by the letter $R$ in the formula).

* c) Same as in b) but for X gametologs and X-derived retrogenes. Gametologs were divided into two groups depending on whether $X_f/X_m$ ratios were above or below zero. Purple violin plot incorporates the expression of X-derived retrogenes (denoted by the letters $Rx$ in the formula).

* d) Same as in b) but for X gametologs in females, Y gametologs in males and Y-derived retrogenes. Gametologs were divided into two groups depending on whether $X_f/Y$ ratios were above or below zero. Purple violin plot incorporates the expression of Y-derived retrogenes (denoted by the letters $Ry$ in the formula).

* b-e) Significant differences (Mann–Whitney U test) between groups. Median expression levels across tissues (TPM) were used for XY gametologs; cumulative expression was considered for retrogenes. Data from panda, mouse lemur, and hedgehog was not included.
### FIGURES

**Figure 1.**

![Phylogenetic tree and table of dates](http://mc.manuscriptcentral.com/gbe). The phylogenetic tree and table of dates are used to illustrate the evolutionary relationships and timing of divergence events among various species. The tree is color-coded to represent different taxonomic groups, and the table provides specific dates for the events depicted in the tree.
Figure 3.

(a) [Genetic expression heatmap showing differentially expressed genes across various tissue types.]

(b) [Box plot illustrating log2-fold change in gene expression between testis and somatic tissues.]

**Y** genes that gained testis-specificity

**Y** genes that did not gain testis-specificity
Figure 4.
Figure 5.
| Stratum 1 | Stratum 2/3 |
|----------|-------------|
| (180mya) | (120mya)    |

|       | SRY | RPS4Y | HSFY | RBMY | CULB4Y | KDM5D | UBA1Y | TSPY | TXLNGY | DDX3Y | EIFA1Y | USP9Y | UTY | TMSB4Y | ZFY |
|-------|-----|-------|------|------|--------|-------|-------|------|--------|-------|--------|------|-----|--------|-----|
| human | r   |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| chimpanzee | r |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| gorilla | r   |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| orangutan | r |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| macaque | r   |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| marmoset | r   |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| mouse lemur* | r |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| tree shrew* | r |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| mouse | r   |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| rat | r   |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| hamster* | r |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| guinea pig* | r |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| rabbit* | r   |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| sheep* | r   |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| bull | r   |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| pig | r   |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| cat | r   |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| dog | r   |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| panda* | r |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| polar bear | r |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| horse | r   |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| hedgehog* | r |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| armadillo* | r |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| tenrec* | r   |       |      |      |        |       |       |      |        |       |        |      |     |        |     |
| elephant | r |       |      |      |        |       |       |      |        |       |        |      |     |        |     |

**Placental Radiation**:
- Primates
  - Scandentia
  - Laurasiatheria
  - Euarchontoglires
    - Rodents
      - Lagomorpha
    - Insectivores
      - Carnivores
    - Even-toed ungulates
      - Ruminants
      - Xenarthra
      - Afrotheria
females

log$_2$ (current / ancestral) expression

males

Y males

X males

X females

no loss

~3-fold loss

log$_2$ (current / ancestral) expression

no loss

~3-fold loss

log$_2$ (current / ancestral) expression

X females

X males

Y males

brain

gonads

heart

kidney

liver

log$_2$ (current / ancestral) expression

X females

X males

Y males

primates

rodents

laurasiatheria

afrotheria / xenarthra

primates

rodents

laurasia

therios

afrotheria / xenarthra

http://mc.manuscriptcentral.com/gbe
### a) Gene Gameologs

| Gene          | X Y Gametologs |
|---------------|----------------|
| COLBXY        |                |
| DDATXY        |                |
| EFTBXY        |                |
| HISFXY        |                |
| KDMDC         |                |
| KDM3FXY       |                |
| RBXY          |                |
| RPRXY         |                |
| TMSB4XY       |                |
| TSPY2/TSPY    |                |
| UBA1XY        |                |
| TXLNGXY       |                |
| TSPYL2/TSPY   |                |
| USP9XY        |                |
| ZFXY          |                |

**X Y gametologs**
- all tissues
- testis-specific
- brain-specific
- heart-specific
- liver-specific

**Proto-sex chromosomes**
- male
- female

### b) Expression Analysis

- **Log(current/ancestral) expression**
  - -1.5
  - 0
  - 5
  - -10

- **Testis-specificity**
  - **P = 0.031**
  - **P = 1.42e-03**
  - **P = 8.72e-14**

**Y genes that gained testis-specificity**
- testis
- somatic

**Y genes that did not gain testis-specificity**
- testis
- somatic

**No loss**
