Supplementary Note 1: BAC selection strategy and assembly of the rhesus MSY sequence

The CHORI-250 BAC library had been fingerprinted and assembled into fingerprint contigs at the Michael Smith Genome Sciences Centre (http://www.bcgsc.ca/downloads/rhesusmap.tar.gz). We identified Y chromosome fingerprint contigs by searching for contigs containing multiple BACs with end sequences that did not match the female whole-genome shotgun sequence. We then verified the male-specificity of these contigs using +/- PCR assays on male and female genomic DNA and selected tiling paths of clones for sequencing. We used high-density filter hybridization with pools of overgo probes to identify clones from the CHORI-250 and RMAEX libraries to fill gaps.

The assembled rhesus MSY sequence spans 11.0 megabases (Mb) in three contigs (Supplemenary File 1). The largest contig (7.9 Mb) is anchored by the pseudoautosomal region, which was confirmed to be located at the distal end of the chromosome by extended metaphase fluorescence in situ hybridization (EM-FISH) (Supplementary Fig. 2). The smallest contig (1.1 Mb) is bounded by the centromere. The rhesus Y is acrocentric, as determined by EM-FISH (Supplementary Fig. 3). Therefore, the third contig (1.9 Mb) is located between the 7.9 Mb and 1.1 Mb contigs. We determined the orientation of the 1.9 Mb contig by EM-FISH (Supplementary Fig. 4) and radiation hybrid mapping (Supplementary Fig. 5 and Supplementary File 2). We conclude that our assembled sequence is nearly complete based on the small size of the gaps between the three contigs as determined by interphase FISH (Supplementary Fig. 6).

Supplementary Note 2: Determination of X-Y ancestral gene content within strata 1-5.

For this analysis, we used the human X-linked protein-coding gene set from Ensembl. We ordered the genes according to their position on the human X chromosome, which is an approximation of the gene order on the ancestral chromosome, and divided the genes into strata according to boundaries given in Lahn and Page[1] and Ross, *et al.*[2]. We then determined whether each gene was present in the X-Y common ancestor or added to the X chromosome after X-Y differentiation according to the following criteria (see Supplementary Table 4):

1. From Bellott *et al.*[3]: Autosomal progenitors of the human X and Y chromosomes are represented in chicken chromosomes 1 and 4, so a gene with a homolog in a syntenic location within one of these chicken chromosomes is ancestral. (Note: all strata 4 and 5 genes have homologs on chicken 4).

2. From Bellott *et al.*[3] If a gene is absent from chicken 1 or 4, it may have been added to the X or lost in chicken. If such a gene is present in a syntenic location in at least one outgroup species (*Xenopos tropicalis* or *Anolis carolinensis*), it is ancestral.

3. The remaining genes were presumably added to the X sometime during mammalian evolution. We determined the approximate timing of introduction to the X by searching for syntenic loci in various other mammalian X chromosome sequences: rhesus, marmoset, mouse, rat, dog, bovine and opossum. Specifically, we used TBLASTX to search for hits for each gene in question as well as its two neighboring genes. (Note: All other X chromosome assemblies, with the exception of mouse, are draft sequences, so some genes will be absent because of missing data.) For strata 1 and 2, which formed prior to the eutherian (placental mammals) – metatherian (marsupial mammals) split, a gene was considered ancestral only if it was found in opossum. For stratum 3, which was formed prior to the eutherian radiation, a gene was considered ancestral if it was found in at least two non-primate species.

Each MSY gene and pseudogene was in turn considered ancestral if its X-linked counterpart was classified as ancestral. All but two MSY genes, *TSPY* and *AMELY*, have X-linked counterparts with homologs on chicken 1 and 4. A homolog of *TSPYL2*, which is the X counterpart of *TSPY*, is present in *Xenopus tropicalis* in the same

sequence contig as homologs of two syntenic genes: *KDM5C* and *GPR173*. A homolog of *AMELX* is present in *Xenopus tropicalis* in the same sequence contig as homologs of two syntenic genes: *ARHGAP6* and *CXorf22*.

Supplementary Note 3: Human/chimpanzee/rhesus ancestral MSY gene content.

We deduce that within strata 1-4, the MSY of the human/chimpanzee/rhesus common ancestor had the same set of ancestral genes as that of the present-day rhesus and human MSY's and, therefore, there has been no gene loss in either species within these strata. While it is formally possible that the human/chimpanzee/rhesus common ancestor had additional ancestral genes and the same ancestral genes were lost independently in each lineage, we believe that this is highly unlikely for the following reason: The X-degenerate regions of the three species are composed of the same segments of X-homologous sequence (Supplementary Figs. 8 and 9), implying that the deletion events that removed all but the remaining few X-homologous genes occurred in the common ancestor of these species. The likelihood of the alternative scenario – that the same exact series of deletion events occurred independently in each lineage – is extremely low.

Supplementary Note 4: Discussion of dN/dS values for ancestral genes.

The dN/dS ratios for the ancestral genes were calculated from alignments of rhesus and human homologs and are given in Supplementary Table 3. These values range widely, from 0.09 for *NLGN4Y* to 1.33 for *AMELY*. Of the 19 ancestral genes, 12 display evidence for purifying selection, with dN/dS ratios of 0.33 or less and statistically significant deviation from neutrality by Fisher's exact test. Of the remaining 7 genes, only *AMELY* has a dN/dS ratio greater than 0.7. This gene is one of the smallest ancestral genes so the confidence intervals for this calculation are large (0.61-2.98). Therefore, we would not conclude that this gene has experienced the effects of positive selection or relaxed constraint. We also calculated dN/dS for *MXRA5Y*, which is intact in rhesus but a pseudogene in human. As expected, the dN/dS ratio at this locus is high (0.99) reflecting its neutral evolution in the human lineage. We performed sliding window dN/dS analyses to determine if selection was operating differently at distinct locations with each gene, and the results are shown in Supplementary Figure 16. In

most cases, the dN/dS value varies along the length of the gene. However, most genes also have an extended region that displays a very low dN/dS value, confirming the operation of purifying selection on these genes.

Supplementary Note 5: Determining time of pseudogene inactivation.

Within the older strata, five of six pseudogenes display at least one inactivating mutation shared between rhesus, human and chimpanzee, so their inactivation predates the OWM-ape split (see Supplementary Fig. 15). We cannot date with certainty the inactivation of the four pseudogenes within stratum 5 or the sixth pseudogene in the older strata, however, because these loci have been deleted outright in rhesus. All five of these pseudogenes are located within a single 280-kb region in human, so they were likely removed by one large deletion in the rhesus lineage.

LITERATURE CITED

[1]   Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964-967 (1999).

[2]   Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325-337 (2005).

[3]   Bellott, D. W. *et al.* Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* **466**, 612-616 (2010).