RESEARCH ARTICLE

SEX DETERMINATION

Sex chromosome transformation and the origin of a male-specific X chromosome in the creeping vole

Matthew B. Couger[1]†, Scott W. Roy[2,3]†, Noelle Anderson[3], Landen Gozashti[4], Stacy Pirro[5], Lindsay S. Millward[6], Michelle Kim[7], Duncan Kilburn[7], Kelvin J. Liu[7], Todd M. Wilson[8], Clinton W. Epps[6], Laurie Dizney[9], Luis A. Ruedas[10], Polly Campbell[11]*

The mammalian sex chromosome system (XX female/XY male) is ancient and highly conserved. The sex chromosome karyotype of the creeping vole (*Microtus oregoni*) represents a long-standing anomaly, with an X chromosome that is unpaired in females (X0) and exclusively maternally transmitted. We produced a highly contiguous male genome assembly, together with short-read genomes and transcriptomes for both sexes. We show that *M. oregoni* has lost an independently segregating Y chromosome and that the male-specific sex chromosome is a second X chromosome that is largely homologous to the maternally transmitted X. Both maternally inherited and male-specific sex chromosomes carry fragments of the ancestral Y chromosome. Consequences of this recently transformed sex chromosome system include Y-like degeneration and gene amplification on the male-specific X, expression of ancestral Y-linked genes in females, and X inactivation of the male-specific chromosome in male somatic cells. The genome of *M. oregoni* elucidates the processes that shape the gene content and dosage of mammalian sex chromosomes and exemplifies a rare case of plasticity in an ancient sex chromosome system.

Sex chromosomes have arisen many times in evolution. A new pair of sex chromosomes generally originates when an autosomal locus acquires a sex-determining function (*1*). What happens next differs substantially across taxa. At one extreme, turnover of sex chromosomes can be rapid and even polymorphic within species [e.g., frogs (*2*, *3*) and fishes (*4*, *5*)]; at the other, the sex chromosome systems of taxa such as mammals, birds, and *Drosophila* show marked evolutionary stability (*6*).

More than 150 million years of evolution have shaped the identity and dosage of genes on mammalian sex chromosomes since they evolved from an autosomal pair in the ancestor of eutherian (placental) and metatherian (marsupial) mammals (*7*–*9*). Yet the defining features of the X and Y chromosomes arose very early in their evolution and are highly conserved across extant species. On the eutherian Y chromosome, fewer than 5% of genes survived the massive degeneration that followed the progressive suppression of recombination with its once homologous partner, the X chromosome (*10*, *11*). Between 9 and 16 of these ancient genes persist in any single lineage. Of the 10 such genes present in rodents, 5 (*Sry*, *Ddx3y*, *Usp9y*, *Uty*, and *Zfy*) are common to all eutherian species with sequenced Y chromosomes (*8*, *11*, *12*). Surviving Y-linked genes are not specifically expressed in the testes, nor are their functions exclusive to spermatogenesis. Instead, their expression profile is broad, and functions include fundamental cellular processes such as transcription, translation, and chromatin modification (*8*, *11*).

In contrast to the gene-depleted Y chromosome, more than 90% of ancestral genes survive on mammalian X chromosomes (*13*, *14*). Most of these genes are subject to dosage compensation by female X inactivation, the transcriptional silencing of one X chromosome in female somatic cells that balances the dosage of most X-linked genes between XY males and XX females. Although the initial driver of the evolution of X inactivation remains controversial (*15*), it is clear that silencing of X-linked genes evolved in parallel with early gene loss on the Y (*16*). Because a critical minority of X-linked genes escape silencing, two X chromosomes are essential for female-typical development, whereas more than one X chromosome leads to anomalous development in males (*17*, *18*). In this study, we show that these ancient and highly conserved properties of the X and Y chromosomes can be lost, swapped between chromosomes, and even reversed between the sexes, all on a short evolutionary time scale.

The creeping vole (*Microtus oregoni*) is one of very few mammals with an atypical sex chromosome system (*19*, *20*). Fifty years after Susumo Ohno described the cytological details of this system (*21*), it remains an unsolved puzzle in sex chromosome evolution. Figure 1 depicts the notable features of the system, highlighting differences between standard XY systems (Fig. 1A) and the *M. oregoni* system as defined by Ohno (Fig. 1B). Like other therian mammals, *M. oregoni* has two sex chromosomes, termed X and Y by Ohno (though shown here to have more complex chimeric origins). However, the number of sex chromosomes is different between diploid germline and somatic cells (*21*). Even more unusual, the distribution of sex chromosomes across cell types is reversed between females and males. Adopting Ohno's designations, females have paired sex chromosomes in the germ line ($2n$ = 18, XX) but not in the soma ($2n$ = 17, X0), whereas males have paired sex chromosomes in somatic cells ($2n$ = 18, XY) but a single unpaired chromosome in the germ line ($2n$ = 17, Y0), the latter of which leads to production of gametes bearing either a Y chromosome or no sex chromosome (*21*, *22*). Ohno demonstrated that the two X chromosomes in female meiosis are one and the same, the products of mitotic nondisjunction after DNA replication in the germ line (*22*, *23*). Because the X chromosome was never observed in the male germ line, Ohno inferred that a similar process of X nondisjunction produced XXY and Y0 germ cells, with the former rapidly eliminated by apoptosis (*21*, *22*). The reason the latter (which would be expected to lack essential X-linked genes) would preferentially survive remained unknown.

Ohno's observations imply that the *M. oregoni* X chromosome is effectively nonrecombining, is hemizygous in the somatic cells of both sexes, and can respond only to selection on allelic transmission in females. Moreover, these pronounced changes in X chromosome biology are evolutionarily recent: Whereas the diversification of *Microtus* [~70 species in 2 million years (*24*)] is characterized by dynamic karyotype evolution, all other species are XX/XY (*25*). The value of *M. oregoni* as a rare natural experiment in mammalian sex chromosome evolution has received surprisingly little attention (*26*), and the genetics of this system are completely unstudied. Here we present a chromosome-scale assembly of the *M. oregoni* genome. Analysis of this genome shows a number of unusual features (schematically summarized in Fig. 1C), demonstrating that the distribution of sex chromosomes and sex-linked genes between males and females is even more surprising than Ohno supposed.

Genome reorganization and Y-derived genes in females

We first used read cloud genome assembly (*27*) with Hi-C chromosome scaffolding (*28*, *29*) to sequence and assemble a male *M. oregoni*

[1]Department of Thoracic Surgery, Brigham and Women's Hospital, Boston MA, 02115, USA. [2]Department of Biology, San Francisco State University, San Francisco, CA 94117, USA. [3]Department of Molecular and Cell Biology, University of California, Merced, Merced, CA 95343, USA. [4]Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA. [5]Iridian Genomes, Inc., Bethesda, MD 20817, USA. [6]Department of Fisheries and Wildlife, Oregon State University, Corvallis, OR 97330, USA. [7]Circulomics Inc., Baltimore, MD 21202, USA. [8]US Forest Service, PNW Research Station, Corvallis, OR 97331, USA. [9]Department of Biology, University of Portland, Portland, OR 97203, USA. [10]Department of Biology and Museum of Natural History, Portland State University, Portland, OR 97207, USA. [11]Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, Riverside, CA 92521, USA.
†These authors contributed equally to this work.
*Corresponding author. Email: polly.campbell@ucr.edu

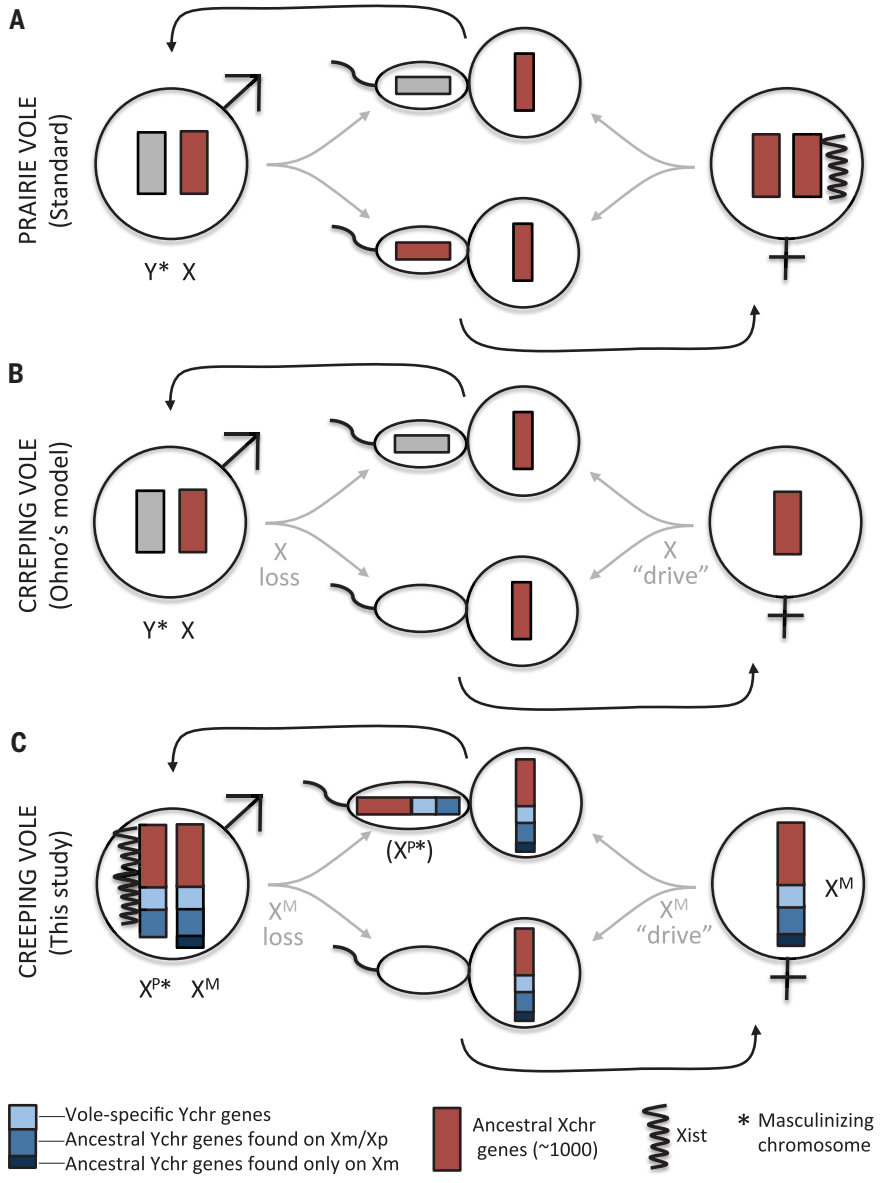


Fig. 1. Atypical chromosome structures and cycles of the creeping vole (*Microtus oregoni*). (**A**) Standard XY chromosome cycle as seen in the prairie vole (*M. ochrogaster*). (**B**) Atypical chromosome cycles of *M. oregoni*, showing chromosomal loss in spermatogenesis and chromosomal "drive" (transmission to 100% of oocytes). Ohno referred to male-specific and sex-shared sex chromosomes as Y and X, respectively (*21*). (**C**) Summary of the present results, showing X-Y hybrid structures of the two sex chromosomes (which we call X^P and X^M), X^M-specific ancestral Y chromosomal genes, and silencing of the X^P by *Xist* in males. Ychr, Y chromosome; Xchr, X chromosome.

somatic genome. This assembly comprised 10 scaffolds, a number consistent with the diploid karyotype as defined by Ohno (*21*) ($2n$ = 18, XY), suggesting that most scaffolds represented complete chromosomes (tables S1 and S2). We produced high-quality gene models by using de novo assembled and genome-aligned RNA sequencing (RNA-seq) reads for multitissue male and female *M. oregoni* transcriptomes, identified genes with one-to-one homologs in the prairie vole (*Microtus ochrogaster*) female genome assembly (Mi_och 1.0), and compared their distribution across chromosomes. Nine of the scaffolds contained one-to-one gene homologs originating from multiple *M. ochrogaster* autosomes. Homology searches against the *M. ochrogaster*, mouse, and rat genomes unequivocally identified the remaining scaffold as homologous to the X chromosome.

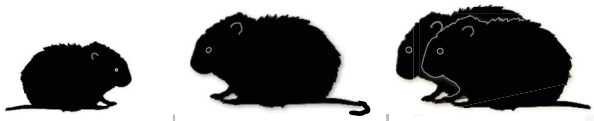
Notably, the assembly did not contain a scaffold homologous to other mammalian Y chromosomes. Instead, eight ancestrally Y-linked genes—*Ddx3y*, *Eif2s3y*, *Kdm5d*, *Sry*, *Ube1y*, *Usp9y*, *Uty*, and *Zfy*—were placed on the X chromosome scaffold. Blast searches for additional Y-linked genes recovered *Rbmy*. Thus, of the 10 rodent ancestral Y-linked genes, only *Tspy1* was missing. *Tspy1* could not be recovered from related species [*Microtus agrestis* (field vole) and *Myodes glareolus* (bank vole)], suggesting ancestral loss. We used polymerase chain reaction (PCR) to confirm the presence of these Y-derived genes in additional male *M. oregoni* samples, including females as a control. Unexpectedly, all nine genes amplified in both sexes. By contrast, amplification was limited to males in the sister species *Microtus longicaudus* (long-tailed vole) (*30*) and in *M. ochrogaster* (Fig. 2). Thus, recent genome reorganization has exposed female *M. oregoni* to Y-derived genes that have been evolving independently in males for up to 150 million years.

Two X chromosomes in males

To better understand the relationship between sex and sex chromosome identity, we produced short-read assemblies (table S3) for somatic tissues from additional *M. oregoni* males and females (n = 2 animals per sex). On the basis of karyotypic studies (*21*, *22*), somatic cells for both males (XY) and females (X0) have a single X chromosome and should therefore both have an X-linked read depth that is half that of the autosomes. As expected, females exhibited an ~50% reduction in X-linked relative to autosome-linked read depth, comparable to that in the XY male *M. agrestis* genome (Fig. 3A). By contrast, X-linked read depth in *M. oregoni* males was indistinguishable from autosomal read depth and was similar to that found in the XX female *M. ochrogaster* genome (Fig. 3A). The presence of two X chromosome haplotypes in males was further supported by the observation of many heterozygous X-linked single-nucleotide polymorphisms (SNPs) in males but not in females (Table 1).

In most mammals, X chromosome expression is balanced between the sexes by the silencing of one X in female somatic cells. Initiation of this process relies on *Xist* noncoding RNA, which is highly expressed from the inactive X chromosome in XX females (*31*). Notably, we found a sex-reversed pattern in *M. oregoni*, with *Xist* detected in male but not female transcriptomes. We confirmed this result with reverse transcription (RT)–PCR in different individuals sampled from multiple populations (n = 10 per sex) (Fig. 3B). Inclusion of *M. longicaudus* confirmed that this reversal of sex-specific expression arose in the *M. oregoni* lineage (Fig. 3B).

The doubling of X chromosome haplotypes and the high frequency of X-linked heterozygous sites shared across males can be reconciled with the *M. oregoni* karyotype if the chromosome called the Y by Ohno is largely derived from a complete copy of the ancestral X chromosome. Male-limited *Xist* expression provides strong support for this inference. Hereafter, we refer to these X-derived chromosomes according



Ancestral Y-linked genes	*M. oregoni*		*M. longicaudus*		*M. ochrogaster*	
	♂	♀	♂	♀	♂	♀
Eif2s3y	✓	✓	✓	✗	✓	✗
Zfy	✓	✓	✓	✗	✓	✗
Usp9y	✓	✓	✓	✗	✓	✗
Ddx3y	✓	✓	✓	✗	✓	✗
Uty	✓	✓	✓	✗	✓	✗
Ube1y	✓	✓	✓	✗	✓	✗
Kdm5d	✓	✓	✓	✗	✓	✗
Rbmy	✓	✓	✓	✗	✓	✗
Sry	✓	✓	✓	✗	✓	✗

Fig. 2. Distribution of ancestral Y-linked genes between males and females in *M. oregoni* and closely related species. The nine genes were recovered from both male (n = 3) and female (n = 2) genomes in *M. oregoni*. PCR results showed that Y genes amplified in all male and female *M. oregoni* but only in male *M. longicaudus* (n = 4 per sex) and *M. ochrogaster* (n = 3 per sex).

to their parental origin: X^M is present in the somatic cells of both sexes but is maternally transmitted; X^P is male specific and is paternally transmitted. We designate the latter as a second X rather than a Y because it is recognized as such by the cellular machinery of male *M. oregoni*.

A haplotype-aware long-read assembly of the *M. oregoni* genome

We next sought to understand the longer-range structure of the X^P and X^M chromosomes. Previous karyotypic analysis has shown that X^M contains two large blocks of constitutive heterochromatin that are not apparent on X^P (*25*, *32*), resulting in a size difference between the two sex chromosomes that explains Ohno's designation of X^P as the Y chromosome (*21*). Because our original male genome was based on short reads, we could not accurately resolve this repetitive content. With the goal of resolving repetitive structures and accurately discriminating X^M and X^P alleles, we produced an additional male somatic genome by using high-coverage Hifi PacBio reads, ultralong Nanopore reads, and Bionano optical mapping. We produced PacBio circular consensus sequencing (CCS) libraries and assembled the genome by using a haplotype-aware assembler (*33*). To better resolve ancestral X- and Y-derived genes on X^M and X^P, we used a similar approach (CCS plus Bionano mapping) to produce a male *M. ochrogaster* genome.

Our *M. oregoni* long-read assembly was a marked improvement on our short-read–based assembly (tables S4 to S7). The autosomal primary contigs had an N50 statistic value of 42 Mb, whereas the sex chromosomal primary contigs had an N50 of 7.74 Mb. The more fragmented nature of X^M and X^P contigs was likely due to the large amount of repetitive content on both chromosomes (see below).

Separate assembly and comparison of candidate X^P and X^M haplotypes

To distinguish X^M and X^P contigs within the X-derived parts of the chromosomes, we turned to the SNPs that we had identified from the two male and two female short-read genomes. Sites where X^P and X^M differ are expected to lead to cases in which females show the same single SNP (X^M), whereas males are heterozygous for the same two variants (X^P and X^M). Indeed, 91.6% (2648 of 2892) of observed X-linked SNPs showed exactly this pattern (Table 1). For each contig, we identified sites containing either putative X^P or X^M variants. Different sites on the same contig showed clear correspondence, with 91.3% of the contigs (63 of 69) harboring multiple sites showing either all X^P or all X^M variants. This pattern supports the inference that these SNPs reflect allelic differences between X^P and X^M and allowed us to parse individual contigs between the two chromosomes. Additional candidate X^M and X^P contigs were then identified by searching for pairs of regions with multigene homology to the *M. ochrogaster* X chromosome. This yielded partial candidate assemblies of the X-derived portions of X^M and X^P, spanning a total of 127.5 and 82.5 Mb, respectively. Blast searches of the X^P and X^M haplotypes confirmed that most genes within this region were homologous to X-linked genes on the *M. ochrogaster* and mouse X chromosomes, indicating a dearth of gene movement to the X-derived parts of these chromosomes (Table 2). Figure 3C shows the aligned candidate X^M (yellow) and X^P (green) haplotypes. The colinearity between the two haplotypes is clearly visible from the many parallel arcs connecting homologous X^M and X^P sequences (dark lines).

Transposable element proliferation on X^P and X^M

Scrutiny of X chromosomal haplotypes revealed a significantly elevated density of transposable elements. We focused on the five most commonly found transposable elements, as judged by number of Blast hits. These included multiple L1_Mur2 family long interspersed nuclear elements as well as MYSERV6 elements. We identified putative transposase genes and assessed the fraction of each contig constituted by these genes. Relative to autosomes, putative transposase genes made up a larger percentage of X chromosomal sequences (4.1 and 5.0% of X^M and X^P sequences, respectively, compared with 1.3 to 2.0% for the autosomes; $P < 0.001$ by bootstrapping; Fig. 3D). Notably, elevated transposable element content was broadly distributed across X^P and X^M (outer histogram in Fig. 3C).

Complex divergent structures of the Y-derived regions on X^M and X^P

The presence of Y-derived genes in both sexes and the replacement of the ancestral Y chromosome by a second X-like chromosome in males could be explained by Y-to-X gene movement followed by Y chromosome loss. At least one other *Microtus* species carries pseudogenized copies of *Sry* on the X chromosome (*34*), and Y-to-X translocations preceded loss of the Y chromosome in two more distantly related vole species (*35*). In this case, we would expect to find Y-derived genes integrated into an otherwise X-derived chromosome. Alternatively, chromosomal breakage could have resulted in X-Y fusion, in which instance we would expect to detect a distinct region of Y-derived gene content. Given the high rate of karyotype evolution in *Microtus* and the substantially reduced diploid number in *M. oregoni* compared with close relatives (*25*), this scenario is also plausible.

We found some support for an X-Y fusion in both X^P and X^M sequences, albeit with distinct structural patterns and differences in gene order, identity, and copy number between the two chromosomes (Fig. 4). In particular, a probable break point was captured on a

6.5-Mb X^P scaffold, in which a repetitive array of nine ancestral X-linked genes was followed by an interval containing ancestral Y genes as well as genes found within an interval on the *M. ochrogaster* Y (Fig. 4A). The scaffold ended with >1.5 Mb of ampliconic content comprising 143 copies of an autosomal retrogene, *Rpl21*, with 3 copies of *Sry* embedded in the second half of the array. All *Rpl21* copies shared the same 8–base pair deletion in coding sequence and subsequent premature stop, indicating pseudogenization before amplification. X^M contigs with putative homology to this scaffold were highly fragmented and did not contain the ancestral Y genes or the *Rpl21* array (fig.

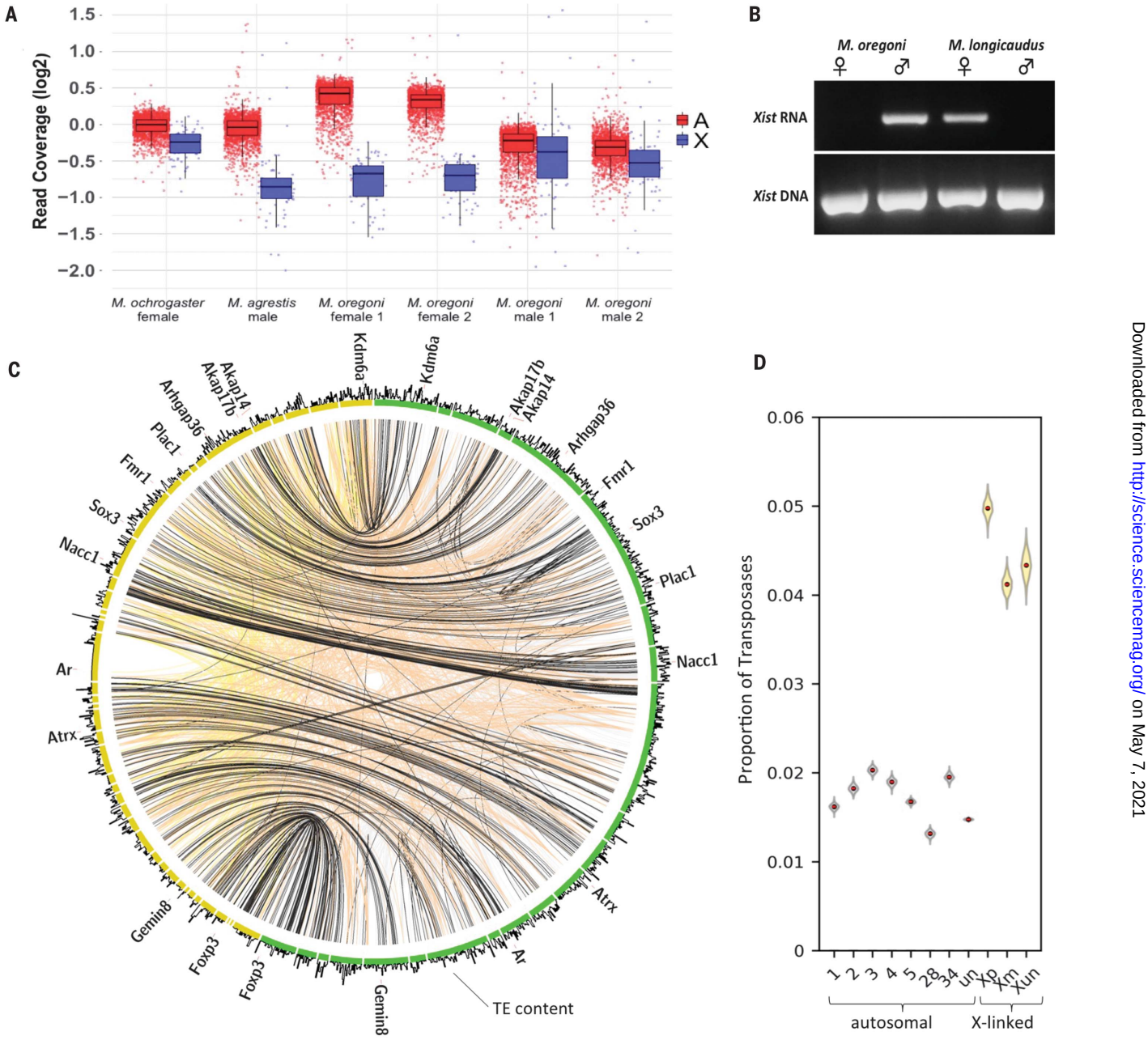


Fig. 3. Separate X^P and X^M haplotypes in male *M. oregoni*. (**A**) Doubled X chromosomal DNA sequencing read coverage in *M. oregoni* males suggests two X chromosomes. Read coverage comparisons are shown for autosomal (red) and X-linked (blue) genes for males and females of three species of *Microtus*. Related species show the expected reduction of X-linked coverage for males (*M. agrestis*) but not females (*M. ochrogaster*), whereas *M. oregoni* shows the reversed pattern. (**B**) *Xist* expression is sex-reversed in *M. oregoni*. *Xist* is expressed in male but not female *M. oregoni* (RT-PCR, replicated in n = 10 per sex); female-limited expression in *M. longicaudus* is typical of other eutherian mammals (RT-PCR, n = 1 per sex). *Xist* DNA amplifies in both sexes in both species (PCR, n = 1 per sex per species). (**C**) Circos plot of assembled X^P (green) and X^M (yellow) chromosomes. Black linkages denote homologies between low–copy number genes, showing clear conservation of synteny; colored linkages represent homologies for genes with higher copy numbers, showing more complex patterns of gene amplification (yellow and green show within-chromosome homologies; orange shows between-chromosome homologies). The histogram plot around the outside of the circle indicates transposase density. (**D**) Higher transposable element density on X chromosomes. The fraction of contig sequence covered by transposase genes is shown. Numbers denote the seven autosomal scaffolds that are nearly the length of chromosomes. un, unscaffolded autosomal sequence; Xun, unassigned X-derived sequence. Violin plots show variation over bootstrap replicates.

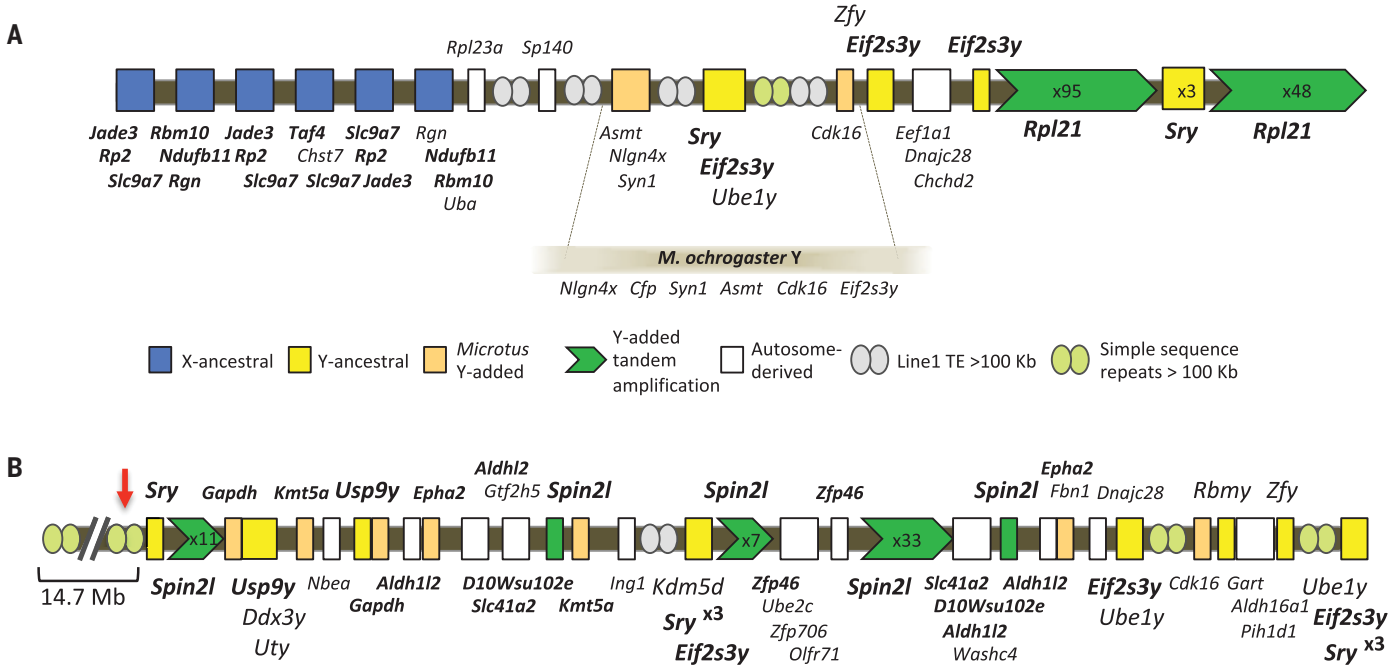


Fig. 4. Complex structure of Y-derived intervals on X^P and X^M. Cartoons of Y-derived scaffolds showing gene order and distributions of transposable elements and simple sequence repeats on (**A**) X^P and (**B**) X^M. Multicopy genes and genes amplified in tandem arrays are in bold; ancestral Y genes are in a larger font. *Microtus* Y-added genes are found on the *M. ochrogaster* Y chromosome but are not reported from other sequenced Y chromosomes. These include *Spin2l*, a single-copy autosomal gene with two translocated copies on the *M. ochrogaster* Y. The X^P scaffold is 6.5 Mb; the dashed lines mark an interval with putative homology to an interval on the *M. ochrogaster* Y chromosome. The X^M scaffold is 20.6 Mb; the red arrow indicates the start of a 14.7-Mb simple sequence repeat expansion. See also fig. S1. TE, transposable element.

S1A). Instead, all nine ancestral Y genes were concentrated on a large X^M scaffold that included 53 copies of a member of the spindlin gene family homologous to *Spin2l* in *M. ochrogaster*, as well as a 14.7-Mb span of simple sequence repeats that likely represents one of the heterochromatic regions that are specific to X^M (Fig. 4B). We found more than 100 Mb of smaller, currently unscaffolded contigs composed exclusively of this simple sequence repeat. We speculate that this large tract of heterochromatin forms a boundary between Y- and X-derived genes on X^M. We identified an X^P contig with putative homology to the genic interval shown in Fig. 4B that contained a subset of the same genes but was not colinear (fig. S1B). Notably, three ancestral Y genes (*Kdm5d*, *Rbmy*, and *Uty*) placed on this X^M scaffold were missing from all X^P contigs, suggesting absence from the X^P.

Taken together, our reconstructions of Y-derived intervals on X^M and X^P suggest an initial breakage-and-fusion event during male meiosis. This initial event was likely followed by a period during which both ancestral and newly formed sex chromosomes segregated together through the germ line. Additional breakage-and-fusion events and nonhomologous recombination before exclusion of X^M from male meiosis are non–mutually exclusive explanations for the partially shared gene content and structural differences between X^M and X^P. The excess of transposable elements in these Y-derived intervals (Fig. 4), and throughout X^M and X^P (Fig. 3, C and D), suggests a candidate mechanism for inferred breakage events.

The fate of Y-derived genes

We next studied the ancestral Y-linked genes. Given that these genes have a long history of being male specific, and that the X^M is transmitted only through females, we expected that inactivating mutations in the ancestral Y genes would be enriched on X^M relative to X^P. Using subassemblies of contigs from male and female short-read genomes, together with the male long-read genome and sex-specific transcriptomes, we reconstructed complete exon-intron structures for *Eif2s3y*, *Ddx3y*, *Kdm5d*, and *Uty* (Table 3). *Ddx3y* and *Kdm5d* were expressed in both sexes; *Uty*, *Ube1y*, and *Sry* were specific to male transcriptomes (Table 3). Notably, evidence for pseudogenization was not chromosome specific. For example, a nonsense mutation in *Ube1y* exon 15 was present in all reads within both male and female genome samples, consistent with presence on both X^M and X^P. Similarly, we found three *Ddx3y* transcripts (two female, one male) with different retained introns, all of which produced frameshifts and premature stop codons. Thus, we found no evidence for differential survival of ancestral Y genes in male versus female *M. oregoni*.

Both gene model–based annotation of Y-derived genes on X^M and X^P (Fig. 4) and read depth–based comparisons between female and male genomes (see below) provided evidence for copy number differences between the sex chromosomes. Given the primary function of *Sry* in initiating the male developmental pathway, the presence of multiple copies of *Sry* on a maternally transmitted X chromosome is particularly surprising. To validate and extend this result, we assayed *Sry* copy number in *M. oregoni* sampled from two populations and in *M. ochrogaster*, a species with multiple *Sry* copies on the Y (*36*) but none on the X. Results confirmed multiple copies of *Sry* in *M. oregoni* females, and the relative copy number for both populations was comparable to that of male *M. ochrogaster* (Fig. 5A). Notably, *Sry* copy number was highly elevated in male *M. oregoni*, with between-population differences suggestive of ongoing amplification of this gene on X^P.

To better understand the distribution of functional *Sry* copies between males and females, we retrieved and aligned all *Sry* sequences from contigs robustly assigned to X^M or X^P in our long-read male *M. oregoni* genome. We recovered 23 copies of *Sry* from this genome, including six pseudogenes, all assigned to X^P. The remainder had intact coding regions with seven copies assigned to X^M and 10 to X^P. All but two of the intact copies differed by at least one nonsynonymous substitution, with a minimum identity of 81.6% at the protein level. Phylogenetic analysis of

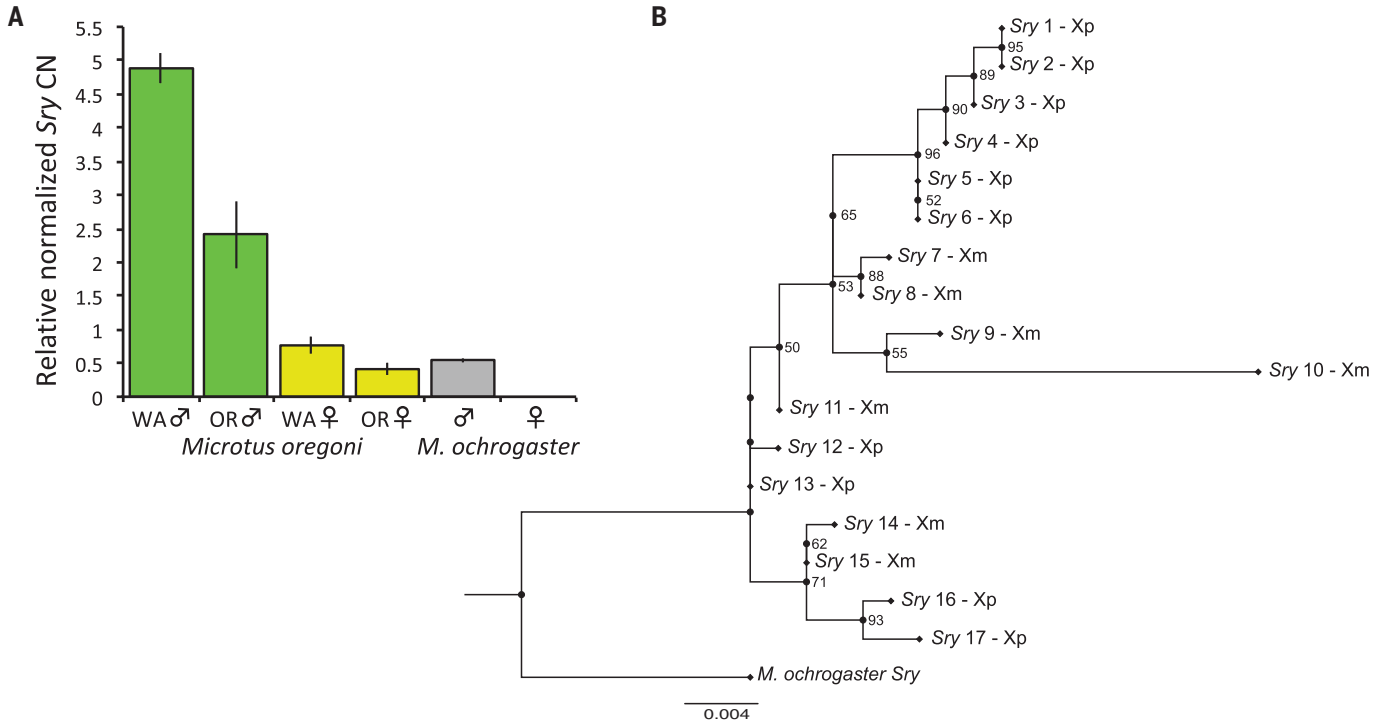


Fig. 5. Amplification and diversification of *Sry* in *M. oregoni*. (**A**) Relative copy number (CN) assay for *Sry*, showing differences between *M. oregoni* males from northern Washington (WA, *n* = 7) and Oregon (OR, *n* = 4), equivalent CN in female *M. oregoni* (WA, *n* = 12; OR, *n* = 3) and male *M. ochrogaster* (*n* = 3), and no signal in female *M. ochrogaster* (*n* = 3). Error bars indicate SEM. (**B**) Phylogenetic tree of 17 functional *Sry* copies on X^M and X^P in a single male *M. oregoni*, with *Sry* from *M. ochrogaster* as the outgroup. Bootstrap values were generated from 10,000 pseudoreplicates; values under 50 are not shown.

intact *Sry* copies suggests rapid diversification through amplification in the *M. oregoni* lineage, with some evidence for divergence between sequences on the two chromosomes (Fig. 5B).

Gene amplification across the X^P and X^M chromosomes

Copy number amplifications are a common feature of nonrecombining chromosomes, and amplification is thought to promote gene longevity on mammalian Y chromosomes (*11*). We therefore asked whether the copy number amplifications observed in the Y-derived genomic regions represented a more general feature of the X^P and X^M chromosomes. Given the likely incompleteness of the X^P and X^M haplotypes, we first evaluated gene amplification at a chromosome scale by mapping the short-read data from two female and two male *M. oregoni* to the (largely unamplified) *M. ochrogaster* short-read–based assembly (Mi_och 1.0) and assessing read coverage. Read coverage in *M. oregoni* females, and a single *M. agrestis* male, closely tracked that of the female *M. ochrogaster* genome (Fig. 6A). By contrast, *M. oregoni* males showed extended regions of amplification ranging from ~100 kb to ~1.5 Mb, with an estimated two- to eightfold elevation in read coverage. No similar amplifications were found on other chromosomes in *M. oregoni*. Comparisons across all four *M. oregoni* short-read genomes confirmed the lack of amplification in both females and showed largely indistinguishable patterns of amplification in the two males. A single observed region of amplification relative to *M. ochrogaster* (Fig. 6A) was shared between males and females, suggesting a shared X^P and X^M amplification.

We also compared read coverage for individual genes (Fig. 6B and table S8). These results showed the expected pattern for *M. oregoni*, with most ancestral X-linked genes (blue dots) showing male/female coverage ratios of ~2 for most X-derived genes (blue line), ~1 for putatively X^M-specific genes (*Kdm5d*, *Rbmy*, and *Uty*), and >2 for *Sry*. This analysis also suggested that X^P copy number is roughly equal to that on X^M for two additional ancestral Y genes (*Zfy* and *Eif2s3y*) and is higher for three genes (*Ube1y*, *Usp9y*, and *Ddx3y*). Consistent with the results in Fig. 4A, we also found evidence for elevated X^P copy number for many X-linked genes (dots above the blue line).

Increased protein sequence evolution on X^P and X^M chromosomes

Another common feature of nonrecombining chromosomes is accelerated rates of nonsynonymous change, generally attributed to reduced selective efficiency. We studied the candidate X^M-X^P differences (Table 1), which constitute most observed SNPs. We directionalized those changes as either X^M or X^P mutational changes. Consistent with inefficient selection due to lack of recombination, putative X^P changes had a highly elevated ratio of nonsynonymous to synonymous changes (dN/dS = 0.63 compared with 0.21 for fixed interspecific difference with *M. ochrogaster*; $P < 10^{-5}$). Moreover, the normalized incidence of SNPs leading to premature stop codons was 40 times that for species differences (0.46 versus 0.012). Rates of nonsynonymous changes and premature stop codons were not significantly different for various classes of genes that might be expected to experience differential selection on male- versus female-transmitted chromosomes (e.g., genes involved in gametogenesis and sex-biased genes; table S9). However, the pattern of putative fixed changes on X^M was very different from that on X^P: dN/dS was only moderately elevated relative to interspecific differences (dN/dS = 0.37), and no premature stop codons were observed.

Somatic inactivation of the X^P chromosome

Finally, given our finding of male-limited *Xist* expression in *M. oregoni*, we were interested in understanding X chromosome expression in XX males. In eutherian mammals, female somatic cells randomly silence one X chromosome in early development, a choice that is preserved in daughter cells. Using the 2648 putative X^P/X^M SNP differences, we quantified

allelic representation in RNA-seq datasets from somatic tissues (seminal vesicles, heart, brain, and liver) collected from two male *M. oregoni*. Notably, across seven of eight datasets, 99.4% of informative reads supported exclusive expression of X^M alleles, including ≥80% reads for each of the 134 genes individually (data S1). Thus, X chromosome inactivation occurs but is nonrandom in male *M. oregoni*. The four informative SNPs within the Y-derived chromosomal region also showed X^M-specific expression. This suggests that transcriptional silencing extends across much or all of the physical X^P chromosome and provides additional justification for referring to it as an X rather than a Y.

Notably, the seminal vesicle transcriptome was the only dataset with a few exceptions to X^P silencing. The expression of four genes (*Mbtps2*, *Prkx*, *LOC101999900*, and *Spin4*) was strongly X^P biased, potentially indicating local reactivation of the X^P in association with male reproduction. Consistent with X inactivation mediated by the expression of *Xist* from the otherwise silenced X, only X^P *Xist* variants were found in the transcriptomes (100% of reads supporting the X^P variant for all 66 SNP sites).

Nonrandom X chromosome inactivation in a eutherian mammal is surprising but could be explained in *M. oregoni* by *Xist* loss of function on X^M. We therefore evaluated the genomic *Xist* alleles on X^P and X^M but found no obvious functional sequence differences between the haplotypes in transcribed regions. However, comparing the *Xist* minimal promoter sequence with those of the mouse and *Microtus levis* (southern vole), we identified two X^M-specific point mutations in conserved *Xist* promoter elements (*37*, *38*) with likely functional effects. One mutation is located in conserved element VI, which contains a binding motif for a CTF/NF1 transcription factor (Fig. 7). Mutations in CTF/NF1 binding sites have been shown to decrease transcriptional activity in other promoters (*39*, *40*). Most notably, we found an A>G mutation at the transcription initiation site in conserved element I, only present in the X^M *Xist* allele (Fig. 7). In an in vitro assay using synthetic promoters with directed mutations in this transcriptional initiator element, introduction of a construct containing the same A>G mutation reduced transcriptional activity by 97% and reduced initiation-related transcription factor binding (*41*). Thus, the few nucleotide differences in the X^M *Xist* promoter may provide a functional explanation for nonrandom X chromosome inactivation in male *M. oregoni*, likely mediated at the level of *Xist* transcription initiation on X^M.

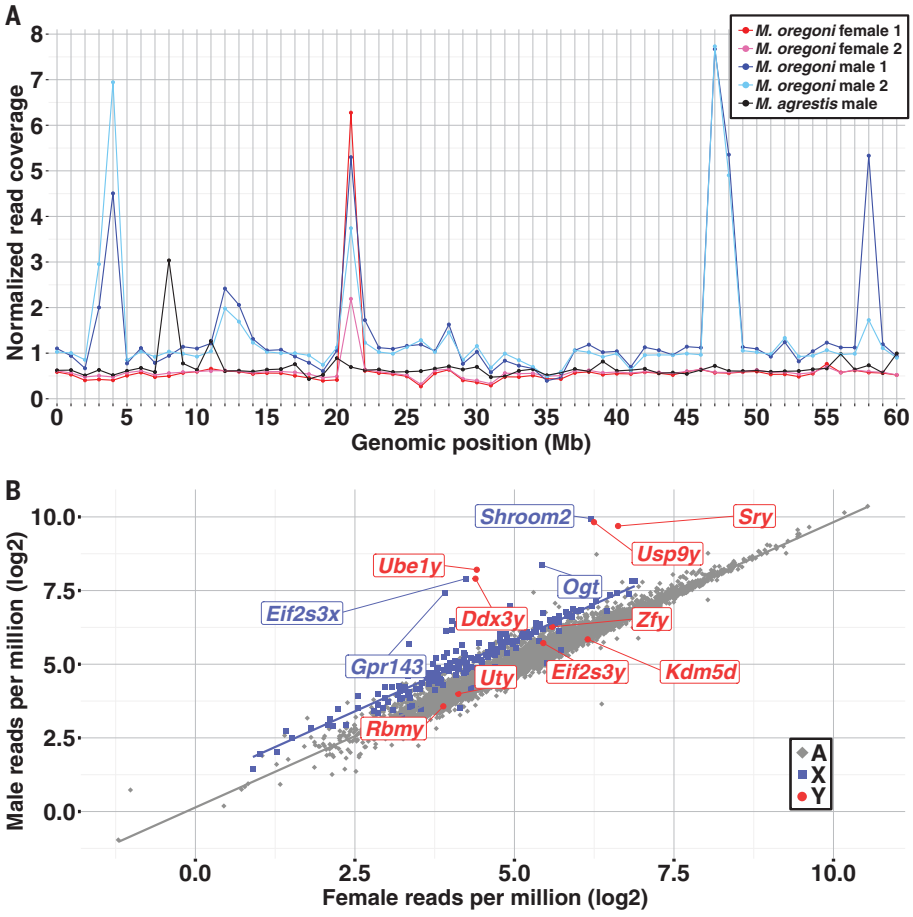


Fig. 6. Amplification of X^P- and X^M-linked genes. (**A**) Amplified regions on the male-specific *M. oregoni* X chromosome. DNA read density is shown in 1-Mb windows across the X chromosome of *M. ochrogaster*. Two *M. oregoni* males are shown in shades of blue, two females in shades of red, and male *M. agrestis* in black. Read density is normalized to the genome-wide average density for each species and to the read density of female *M. ochrogaster* for each window. (**B**) Read density for autosomal (A, gray), X-linked (X, blue), and ancestral Y chromosome (Y, red) genes in *M. oregoni*. Ancestral Y-linked genes and X-linked genes with elevated copy numbers in males are labeled, and regression lines for autosomal and X-linked genes (excluding outliers) are shown.

Discussion

We report substantial sex chromosome turnover in the creeping vole, including (i) presence of two X-like chromosomes in males; (ii) *Xist*-mediated silencing of the male-specific X chromosome in male somatic tissues; and (iii) a complete suite of ancestral Y-derived genes on the maternally transmitted X chromosome (summarized in Fig. 1C). Sex chromosomes in this species have undergone radical restructuring that includes homogenization of gene content between males and females, scrambling of gene order, and divergent molecular evolution. Both the maternally transmitted (X^M) and the male-limited (X^P) chromosomes carry full sets of X chromosome genes joined to core ancestral Y-linked genes. The presence of multiple functional copies of *Sry* on both sex chromosomes raises the question of why X^M0 individuals do not develop as males. It will be interesting to learn how masculinization is avoided. Even more surprising, at least two ancestral Y genes are expressed in females, three may be exclusively transmitted through females, and none show evidence of X^M-specific degeneration. Given that X chromosome genes with surviving Y homologs are overrepresented among the X inactivation escape genes common to mice and humans (*42*, *43*) and are disproportionately likely to be haploinsufficient in humans (*44*), it is possible that the functional Y gene copies retained on X^M compensate for reduced dosage of their X-linked homologs in X0 females. On the other hand, independent loss of most Y-linked genes in two rodent species in which both sexes are X0 [*Ellobius lutescens* (*35*) and *Tokudaia osimensis* (*45*)] suggests that such dosage constraints are surmountable. Alternatively, ancestral Y-linked gene copies on the X^M could be essential for somatic functions in males, particularly given X^P silencing.

In species with ancient, heteromorphic sex chromosomes (e.g., *Drosophila* and mammals),

the rare events that give rise to a new sex chromosome often involve chromosomal fusion (*46*, *47*). Indeed, the male-specific sex chromosome in *M. oregoni* was originally interpreted as an X-Y fusion (*48*, *49*), a hypothesis dismissed by Ohno (*21*). Features of the genome both support and complicate this hypothesis. Concentration of Y-derived genes within specific chromosomal regions suggests the possibility of an X-Y fusion, and we find a point of contact between X- and Y-derived genes on X^P. However, other features, including structural differences between Y-related regions on the X^M and X^P chromosomes and the apparent absence of some Y-derived genes from the X^P, point to a more complex history.

Perhaps the most likely scenario involves initial creation of the proto-X^P through an X-Y fusion during male meiosis. The Y-derived sequence could then have been transferred from the proto-X^P chromosome to a "standard" X chromosome through nonhomologous recombination during spermatogenesis. However, we cannot exclude the possibility that the X^P and X^M independently acquired sets of Y-linked genes directly from a more typical Y chromosome.

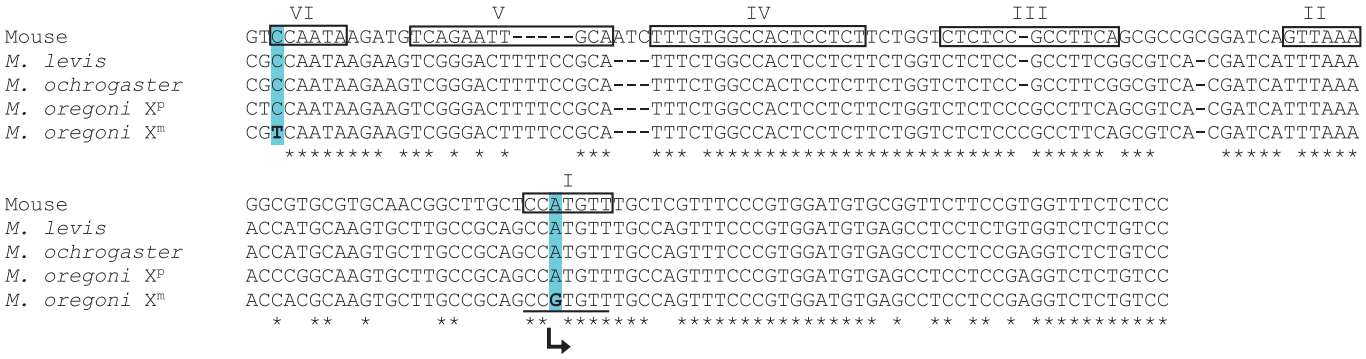


Fig. 7. Multiple mutations in core promoter elements of *Xist* on X^M. The sequence alignment of the *Xist* minimal promoter shows disruptions in conserved promoter elements I to VI (boxes). The underline and arrow represent the transcription start sequence and site, respectively, in element I. Blue shading indicates conserved sites at which X^M allele mutations (bold) potentially affect transcription of *Xist* in *M. oregoni* males. Asterisks indicate fully conserved bases.

Table 1. X-linked polymorphism types and rates, and measures of evolutionary rate for *M. oregoni* males and females and between *M. oregoni* and *M. ochrogaster*. The second column indicates whether *M. ochrogaster* exhibits the sex-specific or shared sequence. Synon., synonymous; Nonsynon., nonsynonymous; CI, confidence interval; –, not applicable.

M. oregoni sequences	*M. ochrogaster* sequence	Candidate explanation	Sites	% of Intraspecific SNPs	Synon.	Nonsynon.	Stop	dS (95% CI)	dN/dS (95% CI)	dSTOP/dS (95% CI)
Monomorphic	Alternative	Species difference	9145	–	5385	3748	12	0.024 (0.024–0.025)	0.21 (0.21–0.22)	0.01 (7×10^{-3}–0.02)
Male-specific (both males heterozygous)	Shared	X^P mutation	2030	70.2%	649	1324	57	2.9×10^{-3} (2.7×10^{-3}–3.2×10^{-3})	0.63 (0.57–0.69)	0.46 (0.35–0.61)
Male-specific (both males heterozygous)	Male-specific	X^M mutation	618	21.4%	280	338	0	1.3×10^{-3} (1.1×10^{-3}–1.4×10^{-3})	0.37 (0.32–0.44)	0.0 (0.0–0.15)
Singleton in female	Shared	X^M mutation	20	0.7%	11	9	0	5×10^{-5} (2×10^{-5}–8×10^{-5})	0.252 (0.11–0.6)	0.0 (0.0–3.85)
Singleton in male (heterozygous)	Shared	X^M or X^P mutation	124	4.3%	36	84	4	1.6×10^{-4} (1.1×10^{-4}–2.2×10^{-4})	0.718 (0.48–1.1)	0.580 (0.24–1.71)
Various	–	Various	100	3.5%	33	64	3	1.5×10^{-4} (1×10^{-4}–2×10^{-4})	0.597 (0.39–0.90)	0.474 (0.18–1.64)

Table 2. Conservation of X-derived gene content on the *M. oregoni* sex chromosomes, X^M and X^P. "Total genes" denotes the subset of X-derived protein coding genes that could be assigned with confidence to X^M and X^P.

	M. oregoni	*M. ochrogaster*		Mouse	
	Total genes	X	Autosome	X	Autosome
X^M	272	225	47	187	41
X^P	335	264	71	215	65

Table 3. Ancestral Y-linked gene reconstruction and expression in male and female *M. oregoni*. ?, not all exons found; RI, retained introns in all transcripts; stop, one or more nonsense mutations.

Gene	Intact in genomes		Detected in transcriptomes		Functional copy	
	♂	♀	♂	♀	♂	♀
Eif2s3y	✓	✓	X	X	✓	✓
Zfy	?	?	X	X	?	?
Usp9y	?	?	X	X	?	?
Ddx3y	✓	✓	✓	✓	RI	RI
Uty	✓	✓	✓	X	✓	✓
Ube1y	?	?	✓	X	RI, stop	stop
Kdm5d	✓	✓	✓	✓	✓	✓
Rbmy	✓	✓	X	X	✓	✓
Sry	✓	✓	✓	X	✓	✓

Notably, whereas X^P exhibits the hallmarks of degradation of nonrecombining chromosomes (*46*, *50*), X^M appears somewhat less transformed, with less-pronounced increase in dN/dS, lower transposable element density, and less genomic amplification. This pattern could be explained if recombination cessation of the X^P occurred first—for instance, during an intermediate stage in which the X^P segregated like a Y chromosome in a standard XY system. However, more effective purifying selection on the X^M could also be due to lack of expression of X^P-linked genes and/or a higher mutation rate on X^P due to male-driven mutation (*51*–*53*). Indeed, the existence of nonrandom X inactivation raises the question of why the X^P remains largely intact. Given that X-linked genes are highly expressed during the early stages of mammalian spermatogenesis (*7*, *54*), we speculate that large-scale deletions on X^P are opposed by selection on germline functions in male *M. oregoni*.

Collectively, these results provide insight into the essential and expendable properties of mammalian sex chromosomes, demonstrating that ancestrally male-specific genes can be accommodated in female genomes, while supporting the proposed dosage sensitivity of ancient X-Y gene pairs (*8*, *11*). The *M. oregoni* sex chromosome karyotype has intrigued evolutionary biologists for decades (*21*, *26*, *55*–*57*). Our resolution of this puzzle reveals a pattern of sex chromosome transformation that was previously unknown in mammals and lays the foundation for using *M. oregoni* as a model system for sex chromosome evolution and function.

REFERENCES AND NOTES

1. J. J. Bull, *Evolution of Sex Determining Mechanisms* (Benjamin/Cummings, 1983).
2. Á. S. Roco *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4752–E4761 (2015).
3. N. Rodrigues, Y. Vuille, A. Brelsford, J. Merilä, N. Perrin, *Heredity* **117**, 25–32 (2016).
4. J. E. Mank, J. C. Avise, *Sex Dev.* **3**, 60–67 (2009).
5. J. Kitano, C. L. Peichel, *Environ. Biol. Fishes* **94**, 549–558 (2012).
6. D. Bachtrog *et al.*, *PLOS Biol.* **12**, e1001899 (2014).
7. L. Potrzebowski *et al.*, *PLOS Biol.* **6**, e80 (2008).
8. D. Cortez *et al.*, *Nature* **508**, 488–493 (2014).
9. R. A. Close, M. Friedman, G. T. Lloyd, R. B. J. Benson, *Curr. Biol.* **25**, 2137–2142 (2015).
10. J. F. Hughes *et al.*, *Nature* **483**, 82–86 (2012).
11. D. W. Bellott *et al.*, *Nature* **508**, 494–499 (2014).
12. B. M. Skinner *et al.*, *Genome Res.* **26**, 130–139 (2016).
13. D. W. Bellott *et al.*, *Nature* **466**, 612–616 (2010).
14. J. L. Mueller *et al.*, *Nat. Genet.* **45**, 1083–1087 (2013).
15. F. Lin, K. Xing, J. Zhang, X. He, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11752–11757 (2012).
16. E. Pessia, T. Makino, M. Bailly-Bechet, A. McLysaght, G. A. B. Marais, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5346–5351 (2012).
17. X. Deng, J. B. Berletch, D. K. Nguyen, C. M. Disteche, *Nat. Rev. Genet.* **15**, 367–378 (2014).
18. F. Printzlau, J. Wolstencroft, D. H. Skuse, *J. Neurosci. Res.* **95**, 311–319 (2017).
19. R. Matthey, *Experientia* **12**, 337–338 (1956).
20. R. Jiménez, F. J. Barrionuevo, M. Burgos, *Sex Dev.* **7**, 147–162 (2013).
21. S. Ohno, J. Jainchill, C. Stenius, *Cytogenetics* **2**, 232–239 (1963).
22. S. Ohno, C. Stenius, L. Christian, in *Chromosomes Today*, C. D. Darlington, K. R. Lewis, Eds. (Oliver and Boyd, 1966), vol. 1, pp. 182–187.
23. S. Ohno, in *Congenital Malformations, Papers and Discussion Presented at the 2nd International Conference* (New York, 1964), pp. 40–42.
24. G. G. Musser, M. D. Carleton, in *Mammal Species of the World: A Taxonomic and Geographic Reference*, D. E. Wilson, D. M. Reeder, Eds. (Smithsonian Institution Press, 2005), pp. 955–1188.
25. W. S. Modi, *Syst. Biol.* **36**, 109–136 (1987).
26. B. Charlesworth, N. D. Dempsey, *Heredity* **86**, 387–394 (2001).
27. N. Spies *et al.*, *Nat. Methods* **14**, 915–920 (2017).
28. E. Lieberman-Aiden *et al.*, *Science* **326**, 289–293 (2009).
29. J. N. Burton *et al.*, *Nat. Biotechnol.* **31**, 1119–1125 (2013).
30. S. Fink, M. C. Fischer, L. Excoffier, G. Heckel, *Syst. Biol.* **59**, 548–572 (2010).
31. I. Okamoto, A. P. Otte, C. D. Allis, D. Reinberg, E. Heard, *Science* **303**, 644–649 (2004).
32. B. L. Libbus, L. A. Johnson, *Cytogenet. Cell Genet.* **47**, 181–184 (1988).
33. S. Nurk *et al.*, *Genome Res.* **30**, 1291–1305 (2020).
34. J. A. Marchal, M. J. Acosta, M. Bullejos, R. Díaz de la Guardia, A. Sánchez, *Genomics* **91**, 142–151 (2008).
35. E. Mulugeta *et al.*, *Genome Res.* **26**, 1202–1210 (2016).
36. M. J. Acosta *et al.*, *Sex Dev.* **4**, 336–347 (2010).
37. S. A. Sheardown, A. E. T. Newall, D. P. Norris, S. Rastan, N. Brockdorff, *Gene* **203**, 159–168 (1997).
38. T. B. Nesterova *et al.*, *Genome Res.* **11**, 833–849 (2001).
39. B. W. Konicek, X.-L. Xia, M. A. Harrington, *DNA Cell Biol.* **14**, 961–969 (1995).
40. J. M. Giger, P. W. Bodell, K. M. Baldwin, F. Haddad, *Exp. Physiol.* **94**, 1163–1173 (2009).
41. R. Javahery, A. Khachi, K. Lo, B. Zenzie-Gregory, S. T. Smale, *Mol. Cell. Biol.* **14**, 116–127 (1994).
42. F. Yang, T. Babak, J. Shendure, C. M. Disteche, *Genome Res.* **20**, 614–622 (2010).
43. B. P. Balaton, C. J. Brown, *Trends Genet.* **32**, 348–359 (2016).
44. J. F. Hughes, D. C. Page, *Annu. Rev. Genet.* **49**, 507–527 (2015).
45. A. Kuroiwa, Y. Ishiguchi, F. Yamada, A. Shintaro, Y. Matsuda, *Chromosoma* **119**, 519–526 (2010).
46. D. Bachtrog, *Nat. Rev. Genet.* **14**, 113–124 (2013).
47. C. Murata, Y. Kuroki, I. Imoto, A. Kuroiwa, *Chromosome Res.* **24**, 407–419 (2016).
48. M. J. D. White, *Proc. Zool. Calcutta, Mookerjee Memor. Vol.* 113–114 (1957).
49. R. Matthey, *Experientia* **14**, 240–241 (1958).
50. B. Charlesworth, D. Charlesworth, *Philos. Trans. R. Soc. London Ser. B* **355**, 1563–1572 (2000).
51. J. F. Crow, *Nat. Rev. Genet.* **1**, 40–47 (2000).
52. J. Shendure, J. M. Akey, *Science* **349**, 1478–1483 (2015).
53. M.-C. Grégoire *et al.*, *Cell. Mol. Life Sci.* **75**, 2859–2872 (2018).
54. S. H. Namekawa *et al.*, *Curr. Biol.* **16**, 660–667 (2006).
55. E. Lifschytz, D. L. Lindsley, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 182–186 (1972).
56. L. D. Hurst, A. Pomiankowski, *Genetics* **128**, 841–858 (1991).
57. C.-I. Wu, A. W. Davis, *Am. Nat.* **142**, 187–212 (1993).

ACKNOWLEDGMENTS

We are grateful to T. Manning for advice and assistance with field work and to W. Timp for use of the PromethION instrument. We thank P. Blood, R. Acosta, and other members of the Pittsburgh Supercomputing Center for assistance in implementing assembly programs on Bridges. We thank J. Schafer at the Oklahoma State University High Performance Computing Center for implementation of software and computational jobs on the cluster. P.C. thanks J. M. Campbell for help translating R. Matthey and dedicates this work to her memory. **Funding:** This work was funded by National Science Foundation IOS 1558109 and start-up funds from the University of California, Riverside (P.C.), National Science Foundation MCB 1616878 (S.W.R.), and National Institutes of Health R44GM134994 (K.J.L.). **Author contributions:** Conceptualization: P.C., M.B.C., and S.W.R.; Field collections: P.C., L.S.M., T.M.W., C.W.E., L.D., and L.A.R.; Assembly, annotation, and SNP calling: M.B.C.; High–molecular weight DNA extraction and Nanopore data production: M.K., D.K., and K.J.L.; Illumina sequencing: S.P.; Data analysis and validations: S.W.R., M.B.C., N.A., L.G., and P.C.; Writing: P.C., S.W.R., M.B.C., and N.A. **Competing interests:** K.J.L., M.K., and D.K. are employees of Circulomics Inc. K.J.L. owns stock in Circulomics. K.J.L. and D.K. have awarded and pending patents related to the methods used. **Data and materials availability:** Raw data are deposited in the NCBI Sequence Read Archive (Bioproject IDs PRJNA361280 and PRJNA505851). The genome assembly is available at NCBI (accession number JAGKIF000000000). Custom scripts are available at https://github.com/lgozasht/Microtus-sex-chromosome-evolution.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/372/6542/592/suppl/DC1
Materials and Methods
Fig. S1
Tables S1 to S13
References (*58*–*91*)
Data S1

View/request a protocol for this paper from *Bio-protocol*.

22 January 2021; accepted 7 April 2021
10.1126/science.abg7019

Sex chromosome transformation and the origin of a male-specific X chromosome in the creeping vole

Matthew B. Couger, Scott W. Roy, Noelle Anderson, Landen Gozashti, Stacy Pirro, Lindsay S. Millward, Michelle Kim, Duncan Kilburn, Kelvin J. Liu, Todd M. Wilson, Clinton W. Epps, Laurie Dizney, Luis A. Ruedas and Polly Campbell

Science **372** (6542), 592-600.
DOI: 10.1126/science.abg7019

Mystery solved?

Chromosomal sex determination arises when an autosomal locus acquires a sex-determining function. In some taxa, this process occurs often. The XY system in mammals, however, has been evolutionarily stable across a wide array of species. Fifty years ago, a variation on this norm was described in the creeping vole (*Microtus oregoni*), but the details have remained mostly unknown. Couger *et al.* sequenced the sex chromosomes in this species and found that the Y chromosome has been lost, the male-determining chromosome is a second X that is largely homologous to the female X, and both the maternally inherited and male-specific sex chromosomes carry vestiges of the ancestral Y.

Science, this issue p. 592