

Acquisition of ampliconic sequences marks a selfish mouse *t*-haplotype

Received: 25 February 2025

Accepted: 10 November 2025

Published online: 01 December 2025

 Check for updates

Callie M. Swanepoel¹, Gaojianyong Wang², Lucy Zhang¹, Björn Brändl³, Hermann Bauer⁴, Pavel Tsaytler⁴, Franz-Josef Müller^{2,3}, Bernhard G. Herrmann⁴ & Jacob L. Mueller¹ ✉

Mendelian genetics posits equal transmission of alleles, but selfish alleles can bias the transmission of large genomic regions or entire chromosomes. One long-standing question is how transmission bias evolves to encompass large genomic regions. *Mus musculus* (house mouse) *t*-haplotypes exhibit up to 99% transmission bias from heterozygous males and harbor selfish alleles genetically linked to large inversions spanning the proximal half of chromosome 17. Here, by generating a high-quality, single-haplotype assembly of a *t*-haplotype, we reveal the evolution of eight large amplicons with known and candidate selfish alleles as a distinct genetic feature. Three amplicons are conserved in closely related *Mus* species, and two have known selfish alleles in the oldest inversion, implicating amplicons and an inversion drove the origins of a selfish chromosome 17 ~3MYA. The remaining *t*-haplotype amplicons harbor gene families expressed predominantly in haploid spermatids, newly acquired retrogenes, and the most differentially expressed genes in wild-type/*t*-haplotype spermatids. Targeted deletion of a ~1.8 Mb amplicon with candidate selfish alleles on the *t*-haplotype reduces selfish transmission in heterozygous males by ~3%. Notably, the evolution of selfish allele-containing amplicons and inversions on the *t*-haplotype parallels mammalian sex chromosome evolution as signatures of selfish transmission. We propose amplicon acquisition and large inversions initiate evolutionary arms races between selfish haplotypes and serve as genome-wide signatures of selfish transmission.

Mendelian genetics is a game of fair play, where parental alleles have an equal chance of being transmitted to offspring. Selfish alleles do not play fair; they bias their transmission to offspring at >50% frequencies. A single allele¹ or multiple genetically linked alleles in a haplotype can cooperate to be selfish^{2–5}. However, the origin and accumulation of cooperative selfish alleles into a selfish haplotype remains poorly understood.

Mus musculus chromosome 17 *t*-haplotypes are ideal for understanding how multiple alleles, spanning the proximal half of a

chromosome, evolved to cooperate in selfish transmission^{6–8}. Over approximately 3 million years (MY), large inversions suppressed meiotic crossing over across a ~40 Mb region of *Mus* chromosome 17, termed *t*-complex, and gave rise to the genetically linked *t*-haplotype^{9–14}. Selfish *t*-haplotype alleles cooperate in haploid spermatids of heterozygous (*wt/t*) male mice to bias up to ~99% *t*-haplotype transmission to offspring^{8,15,16}. The extreme transmission bias of the *t*-haplotype is countered by homozygous (*t/t*) mice exhibiting embryonic lethality¹⁷ or male sterility¹⁸ that prevent fixation of the

¹Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI, USA. ²Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Berlin, Germany. ³Department of Psychiatry and Psychotherapy, Christian-Albrecht University of Kiel, Kiel, Germany. ⁴Department of Developmental Genetics, Max Planck Institute for Molecular Genetics, Berlin, Germany. ✉ e-mail: jacobmu@umich.edu

t-haplotype in *M. musculus*. Understanding how and when the *t*-haplotype sequences evolved selfish transmission is a nearly century-old question hindered by the lack of a high-quality *t*-haplotype sequence assembly. Previous sequencing of *t*-haplotypes^{19,20} could not generate accurate assemblies due to the high sequence identity with the *wt* chromosome, repetitive sequences, and embryonic lethality of *t/t* mice.

To understand how the *t*-haplotype evolved selfish transmission, we generated a high-quality, single-haplotype assembly of a *t*-haplotype using homozygous (*t/t*) mouse embryonic stem cells (mESCs). Our sequencing of the *t*-haplotype revealed three key findings. First, the *t*-haplotype acquired eight amplicons (arrays of palindromic and tandem >8 kb segmental duplications with >99% identity) harboring known selfish alleles and candidate selfish alleles based on haploid spermatid expression. Second, we reconstructed the origins of the *t*-complex and how it evolved through a series of inversions and amplicon acquisition. Third, we generated the first targeted mutation on a *t*-haplotype—a 1.8 Mb deletion of a *t*-haplotype amplicon with a candidate selfish gene family—and found reduced *t*-haplotype selfish transmission. These three findings unveil amplicons and inversions as the most prominent features of selfish haplotype evolution on *Mus* chromosome 17, strikingly similar to features of selfish alleles on mammalian sex chromosomes. Via the lens of the *t*-haplotype, we posit that amplicons and large inversions are common genomic signatures of selfish haplotypes.

Results

Five inversions suppress *t*-haplotype crossing over

To define the precise genomic architecture of the *t*-haplotype, we generated a haplotype-resolved assembly of the *t^{us}*-haplotype. *t^{us}* is ideal for delineating the evolution of the *t*-haplotype because it is a full-length haplotype (i.e., ~40 Mb) and exhibits all phenotypes associated with the *t*-haplotype^{21–23}. To generate a de novo, high-quality assembly of the *t^{us}*, we used homozygous (*t^{us}/t^{us}*) genomic DNA from mESCs and combined PacBio HiFi, Oxford Nanopore (ONT), and BioNano long-molecule sequencing and mapping technologies. Our *t^{us}* assembly comprises 99.76 Mb across three contiguous gapless scaffolds, with scaffold order and orientation validated by Hi-C contact mapping (Fig. 1, Supplementary Fig. 1A, B).

Our final *t^{us}* assembly defines the precise inversion landscape harboring *t*-haplotype selfish alleles. Previous recombination mapping of rare crossovers between the *t*-haplotype and *wt* chromosome 17 predicted four large, non-overlapping inversions within the *t*-complex^{9–13}, but the precise boundaries of each inversion were not known. By comparing our *t^{us}* assembly to *M. musculus* chromosome 17 (*wt*), we confirmed four predicted inversions (1,3,4,5)^{9–13} and discovered one novel, overlapping inversion (2) (Fig. 1, Supplementary Fig. 1A, Supplementary Table 1). Each inversion was confirmed by PCR and Sanger sequencing of the junctions (Supplementary Fig. 1C, Supplementary Table 2). We find two segments of non-inverted sequence, between inversion 3 and 4 and inversion 4 and 5, encompassing ~2 Mb of possible substrate for rare *t*-haplotype cross-over events with the *wt*

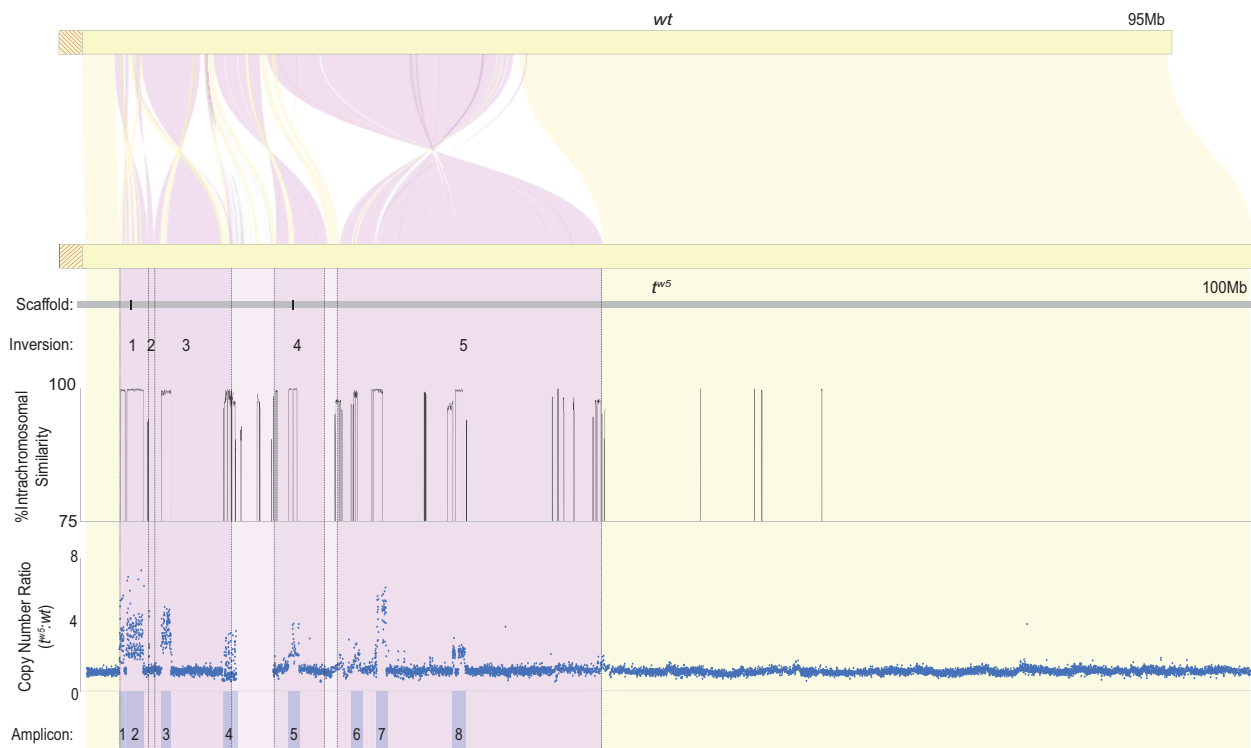


Fig. 1 | Five inversions of the mouse chromosome 17 *t^{us}* haplotype. Syntenic plot comparison of the mm10 reference chromosome 17 (*wt*; top) and *t^{us}* sequence assembly (bottom) with syntenic blocks in yellow and inverted regions in purple. Red diagonal lines represent the centromere. Vertical dotted lines represent inversion junctions. Three non-overlapping hybrid scaffolds (4 Mb, 14 Mb, and 81 Mb) are represented by a gray line with black bars denoting two physical gaps between scaffolds. Five inversions are numbered between the *t^{us}* and *wt* chromosomes—a plot of intrachromosomal sequence similarity to identify the *t^{us}* ampliconic sequence. The *t^{us}* sequence was divided into 50-kb sliding windows

with a 1 kb step size and compared to the *t^{us}* assembly. The highest non-self-sequence similarity was plotted for each window, and all values >75% are shown. *t^{us}* copy number estimates relative to *wt* (Y-axis) are based upon read-depth analysis of whole-genome short-read sequencing data from *t^{us}/t^{us}* mESCs and C57BL/6J adult mice mapped to our *t^{us}* assembly (X-axis). Each blue dot represents copy number estimates in a 3 kb window of unique sequence. Copy number ratios >1 represent a sequence with increased copy numbers on the *t^{us}*. We identified eight sequence blocks of increased copy number on the *t^{us}* (amplicons shaded blue; Amp1–8) relative to *wt*. Source data are provided as a Source data file.

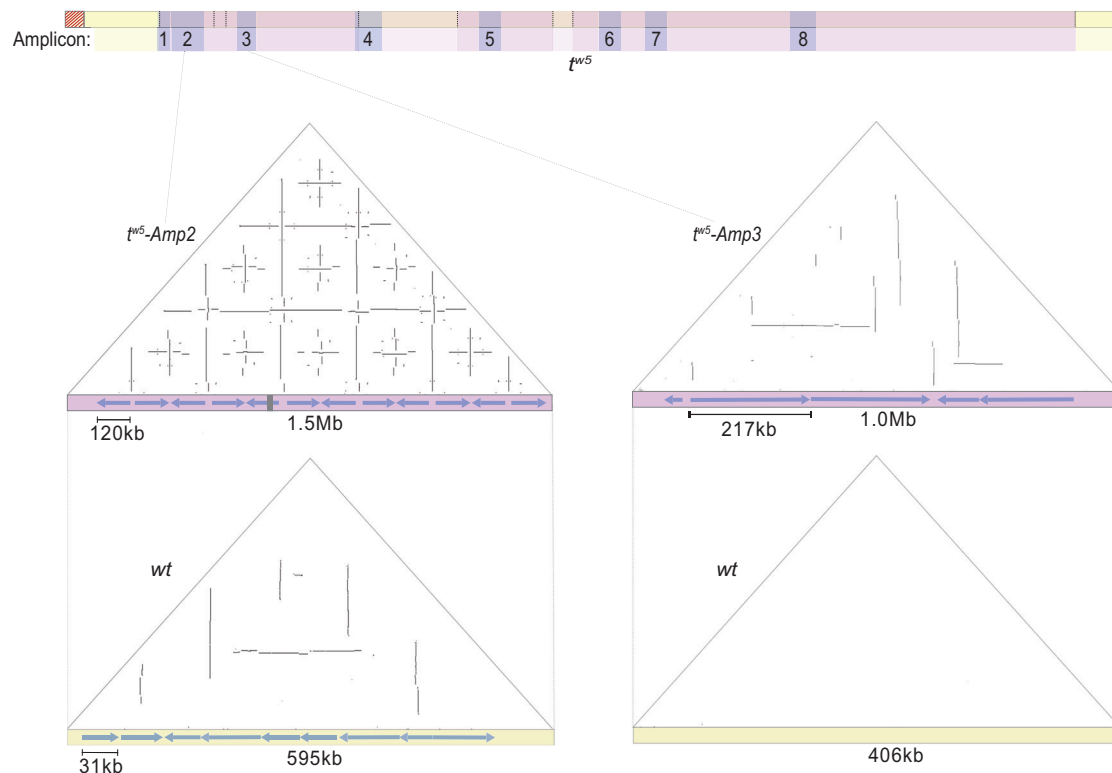


Fig. 2 | t^{ws} sequence is amplified relative to the wt . A schematic of the t^{ws} sequence with amplicons shaded blue and dotted lines representing inversion junctions. Below are self-symmetry dot-plots highlighting palindromic (vertical lines) and tandem (horizontal lines) segmental duplications within t^{ws} -Amp2 (top left) and the orthologous wt region (bottom left) that is amplified to a lower copy

number. Each dot represents a 100 bp window of 100% nucleotide identity. The X-axis is the sequence, 5' to 3', and blue arrows represent the positions and orientations of each >8 kb, >99% identical amplicon. The gray bar denotes a physical gap in the assembly. Dot-plots on the right compare t^{ws} -Amp3 (top), uniquely amplified on the t^{ws} , to the orthologous wt region (bottom).

allele (estimated one cross-over per 500–1000 offspring²⁴) to produce partial t -haplotypes^{11,13,25–28}. Altogether, the t^{ws} spans ~41 Mb of meiotic cross-over suppressed sequence, which is ~6.8 Mb (~20%) longer than the homologous wt region of the C57BL/6J chromosome 17 (Supplementary Table 3).

Eight distinct amplicons on the t -haplotype

Regions of the genome where meiotic crossovers are suppressed accumulate repetitive sequences and undergo single-copy sequence degeneration^{29–31}. We asked whether these repetitive sequence types account for the ~6.8 Mb difference in t^{ws} total sequence relative to wt . We find the t^{ws} has ~11.7 Mb more repetitive sequence as compared to wt , including 6.5 Mb of amplicons across eight regions (Amp1–8) with >99% intrachromosomal sequence identity (Fig. 1, Supplementary Fig. 2A, B, Supplementary Fig. 3, Supplementary Table 4), 2.8 Mb of satellite sequence (Supplementary Fig. 2C and Supplementary Table 3), and 2.1 Mb of transposable elements (Supplementary Table 3). Since the t^{ws} has 11.7 Mb more repetitive sequence than wt , this suggests non-repetitive sequences degenerated on the t^{ws} to explain the overall ~6.8 Mb difference in t^{ws} total sequence length. Indeed, the t^{ws} has 3.8 Mb of single-copy sequence with low alignment coverage relative to wt (Supplementary Fig. 1D and Supplementary Table 3), indicating loss or highly diverged sequence. Thus, the t^{ws} sequence appears to have accumulated repetitive sequences and degenerated single-copy sequence, consistent with known patterns of chromosomal regions with suppressed meiotic crossing over^{30,31}.

Since amplicons are the most distinct sequence on the t -haplotype relative to wt , we further characterized their structure and copy number. To quantify t^{ws} sequence amplification, we determined each amplicon's relative copy number and structural arrangement on the t^{ws} versus wt chromosomes (Fig. 1 and Supplementary Fig. 2B). We found

that all eight t -haplotype amplicons have a higher copy number relative to wt . Since all eight t -haplotype amplicons have a higher copy number in t^{ws} relative to the wt homologous sequence, we asked if the amplicons are t^{ws} -specific or due to the increased copy number of shared amplicons. Five of the eight amplicons (Amp3, Amp5, Amp6, Amp7, and Amp8) are uniquely amplified on t^{ws} and single copy on wt (Figs. 1 and 2 and Supplementary Fig. 3). Three of the eight amplicons (Amp1, Amp2, and Amp4) are also amplified on wt , but with a lower copy number. The t^{ws} -specific amplicons could have been gained on the t -haplotype or lost from the wt chromosome, while the amplicons shared between t^{ws} and wt may represent more ancestral amplicons.

Conserved amplicons in the oldest inversion point to the origin of the t -haplotype

To reconstruct the evolutionary history of t -haplotype amplicons, we asked whether amplicons evolved on chromosome 17 in closely related *Mus* species. We estimated chromosome 17 copy number in *M. caroli*, *M. spretus*, and *M. spicilegus* (Fig. 3A and Supplementary Fig. 4A). We found five t -haplotype amplicons are acquired recently only on the t -haplotype, as they are not amplified in any of the four *Mus* species examined and are most closely related to *M. musculus* (Supplementary Fig. 5). Relative to *M. caroli*, we find the three amplicons shared between *M. musculus* t -haplotype and wt are conserved in *M. spretus* and *M. spicilegus* (Fig. 3A, B and Supplementary Fig. 4A, C). Two of the three conserved amplicons, Amp2 and Amp4, follow an alternative branching pattern reflecting a more complex evolutionary history (Supplementary Fig. 5). Comparison of repeat unit sequences within a haplotype (i.e., t^{ws} versus t^{ws}) than between haplotypes (i.e., t^{ws} versus *M. musculus*), supports ongoing intrachromosomal gene conversion for Amp1 and Amp2 (Supplementary Fig. 5)^{32,33}, but not Amp4, perhaps because it overlaps within a non-inverted region permitting inter-

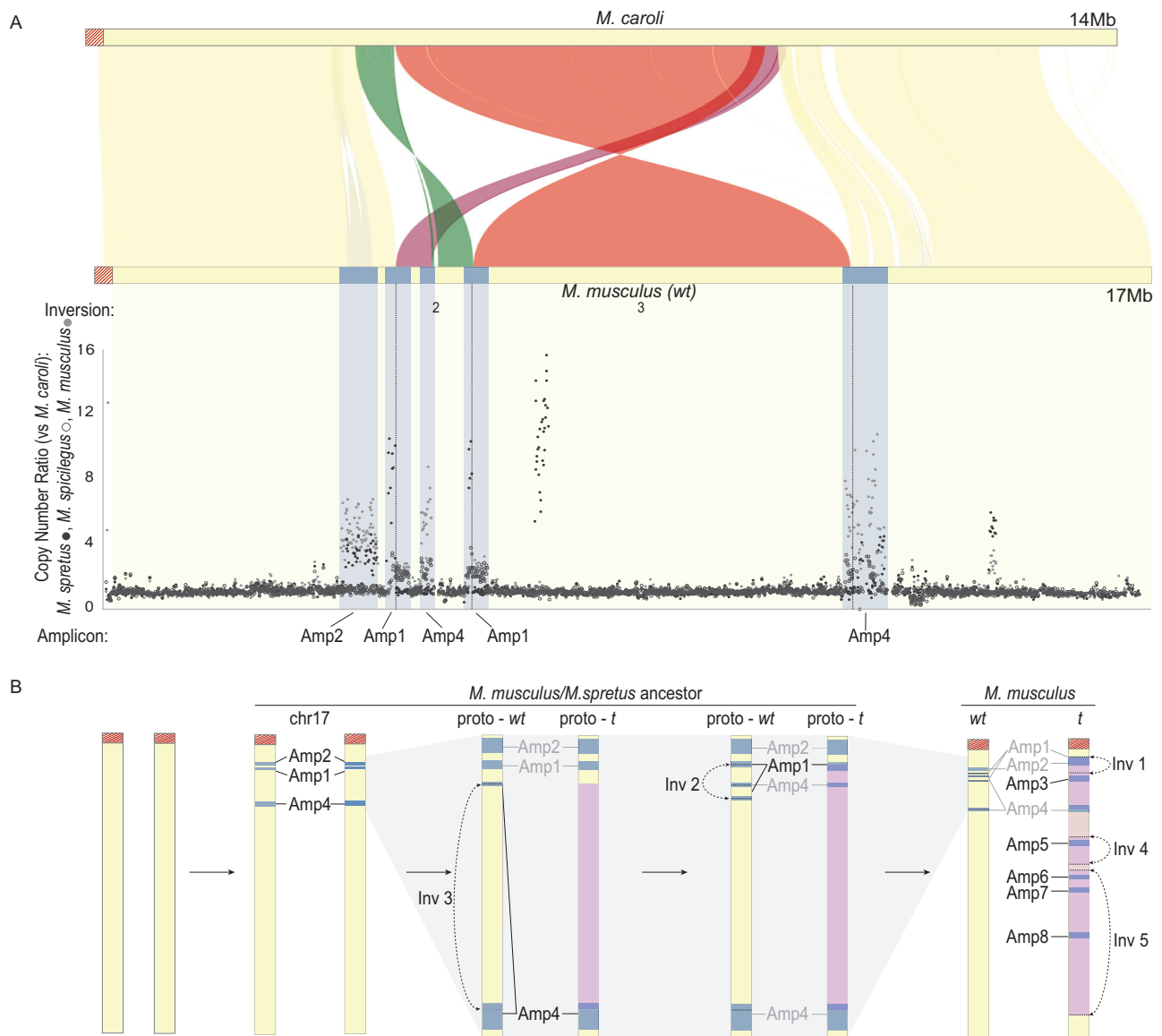


Fig. 3 | Conserved amplification genetically linked to the *t*-complex oldest inversion. **A** (Top) Syntenic plot comparison of the first 14 Mb of *M. caroli* chr17 and the first 17 Mb of *M. musculus* chr17. Syntenic blocks are shaded in yellow, and inverted regions are shaded in green, orange, or maroon. Gray shading indicates amplified sequence (Amp2) on *M. musculus* chr17. Blue shading along the *M. musculus* chromosome represents regions of conserved amplification. Dotted lines represent inversion junctions. (Bottom) Copy number estimates across the first 17 Mb of *M. spretus*, *M. musculus*, and *M. spicilegus* chr17 relative to *M. caroli* (Y-axis). Each dot represents a copy number estimate across a 3 kb window of

repeatmasked sequence. Conserved amplified regions are labeled below the X-axis. **B** A schematic representation of step-wise inversions (Inv) and amplicon acquisition on *wt* and *t*-haplotypes. Gray background shading represents a zoomed-in view of the region encompassing Inv3 and Inv2. Black dotted arrows denote inversions, and blue-shaded regions denote amplicons. Horizontal dotted lines represent inversion junctions. Along the *t* chromosome, dark purple shading represents inverted regions, and light purple shading represents non-inverted regions between *t* and *wt*. Red diagonal lines represent the centromere. Source data are provided as a Source data file.

chromosomal recombination (Fig. 1). Two of the three conserved amplicons (Amp1 and Amp4) harbor *Tagap* and *Smok2* gene families, which have been previously implicated to have selfish *t*-haplotype alleles^{34,35} (Fig. 4A, Supplementary Fig. 4A, C and Supplementary Table 4). The third conserved amplicon (Amp2) harbors *Dynlt1* and *Tmem181a* gene families that are unexplored for their contribution to selfish transmission (Fig. 4A, Supplementary Fig. 4A, C and Supplementary Table 4). While the selfish *t*-haplotype is only observed in *M. musculus*, the conserved amplification of known gene families with selfish alleles suggests selfish transmission initiated in a common ancestor of *M. spretus* and *M. musculus*.

If conserved amplicons initiated selfish transmission, we predict they would be associated with the most ancestral *t*-complex

inversions. To identify the oldest inversion, we compared the order and orientation of chromosome 17 assemblies for *M. spretus*, *M. spicilegus*, *M. musculus*, and the *t^{us}* relative to *M. caroli* (Fig. 3 and Supplementary Fig. 4B). We find inversions 1, 4, and 5 are unique to *t^{us}*, supporting their recent origin on the *t*-haplotype (Fig. 3B and Supplementary Fig. 4B–D). Inversions 2 and 3 are derived (inverted) in *M. spicilegus* and *M. musculus* and ancestral (non-inverted) on the *t^{us}* relative to *M. caroli* (Fig. 3 and Supplementary Fig. 4B–D). Thus, inversions 2 and 3 evolved on an independent haplotype from inversions 1, 4, and 5 (Fig. 3B and Supplementary Fig. 4D). This finding is consistent with previously observed crossing over in the region spanning inversions 2 and 3 between a *t*-haplotype and *M. spretus* chromosome 17, but not between *M. musculus* (*wt*) and *M. spretus*

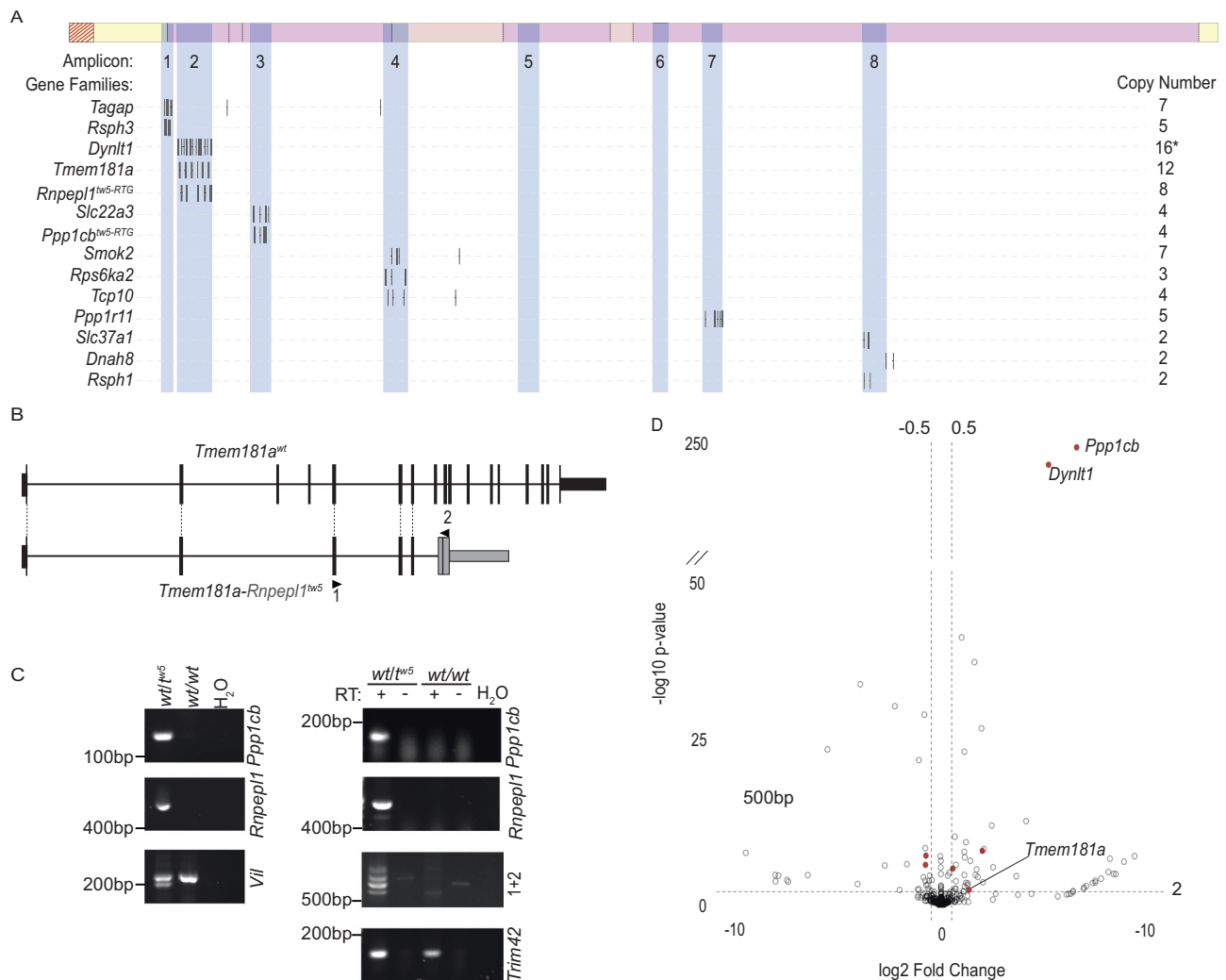


Fig. 4 | *t^{us}* amplicons with known and candidate gene families involved in selfish transmission. **A** The *t^{us}* with ampliconic regions shaded in blue and numbered below. Below are the positions of ampliconic gene families with haploid spermatid expression (black ticks) and their copy numbers. **B** Our gene annotation schematic representation of *Tmem181a-Rnpepl1^{tw5}* fusion gene structure as compared to *Tmem181a^{wt}*. Black vertical bars represent exons, and horizontal lines represent introns. Dotted lines connect orthologous exons between genes. The regions of *Rnpepl1* that retrotransposed to the *t^{us}* include part of exon 10, exon 11, and the 3' UTR and are shaded in gray. Arrows with numbers denote positions of RT-PCR primers used in (C). **C** PCR (left) validation of *Ppp1cb* and *Rnpepl1* retrotransposition from *t^{us}* but not *wt* genomic DNA. *Vil* is an internal positive control to genotype *t^{us}* (two-bands) versus *wt* (single-band). RT-PCR (right) validation of

retrotransposed *Ppp1cb*, *Rnpepl1*, and *Tmem181a-Rnpepl1* testis expression. Four different bands amplified across the *Tmem181a-Rnpepl1* fusion junction were cloned and sequenced for validation. *Trim42* is a positive control for spermatid-expressed genes. + and – denote RT positive and no RT controls. PCR and RT-PCR experiments were repeated independently at least twice with similar results. **D** Volcano plot of differentially expressed genes ($-\log_{10} p$ value > 0.01; \log_2 fold change > 0.5) in *t^{us}/wt* versus *wt/wt* haploid spermatids. Filled red dots represent *t^{us}* ampliconic gene families with differential expression. Diagonal lines represent a break in the y-axis. *p* values were calculated using a two-sided Wald test and adjusted for multiple comparisons using the Benjamini–Hochberg method to control the false discovery rate. Source data are provided as a Source data file.

chromosome 17¹⁰. To achieve the order and orientation of inversions 2 and 3, inversion 3 must precede inversion 2, which we confirmed via PCR of the derived (*M. spicilegus* and *M. musculus*) and ancestral (*M. spretus* and *t^{us}*) inversion junctions (Supplementary Fig. 4C). Based on these findings, we conclude that inversion 3 is the most ancestral inversion, followed by inversion 2. Notably, *Smok2* and *Tagap*, gene families with known selfish *t*-haplotype alleles, map to the breakpoints of inversion 3 and inversion 2. Our finding of conserved amplification of *Smok2* and *Tagap* gene families being linked to the oldest inversions (2 and 3) suggests they initiated the evolution of selfish transmission.

t-haplotype amplicons harbor known and new candidate selfish alleles

Smok2 and *Tagap* (formerly named *Tagap1*)³⁶ are among five genes previously implicated to have selfish alleles (*Fgd2*, *Nme3*, *Tiam2*,

Tagap, and *Smok^{Tcr}*)^{34–37} based on sufficiency studies with partial *t*-haplotypes. Of these five genes, *Smok^{Tcr}* and *Tagap* are the only genes within *t^{us}* amplicon gene families (i.e., *Smok2* and *Tagap* gene families; Fig. 4A and Supplementary Table 4). Nonetheless, the discovery of known selfish alleles within *t^{us}* amplicons suggests other amplicon gene families have alleles contributing to *t*-haplotype selfish transmission.

We asked whether the remaining *t^{us}* amplicons are potential regions for selfish allele evolution based on the presence of genes with haploid spermatid expression, the cell type where known selfish alleles act molecularly³⁸. To identify new candidate *t*-haplotype amplicon gene families contributing to selfish transmission, we gene-annotated each amplicon and examined haploid spermatid gene expression. We annotated the amplicons for gene orthologs to the *wt* chromosome 17 and newly acquired genes within the *t^{us}* amplicons. We identified 12 *t^{us}*

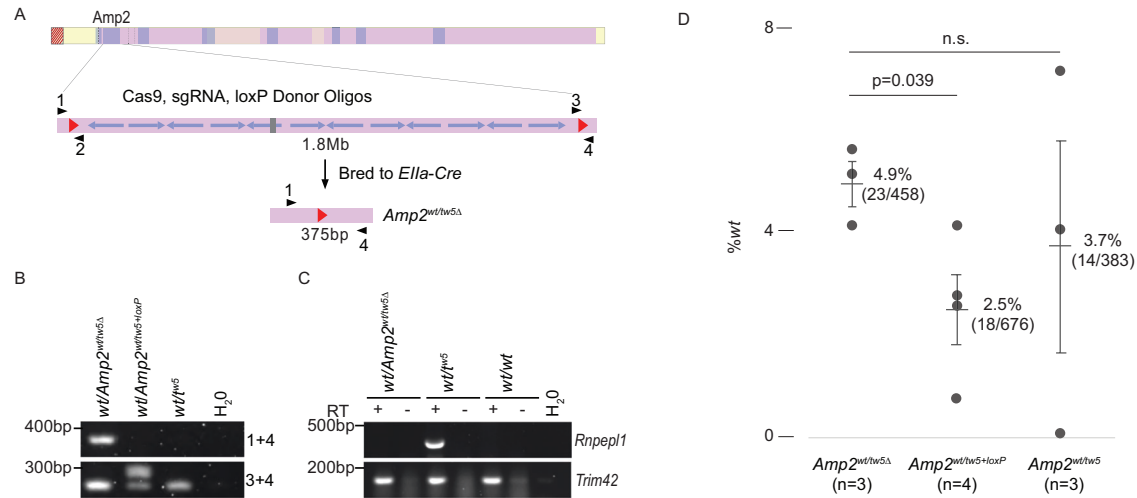


Fig. 5 | Deletion of t^{us} -Amp2 reduces selfish transmission. **A** A schematic representation of the t^{us} with ampliconic regions shaded in blue. Below is a schematic of the t^{us} -Amp2 region targeted for deletion via sequential integration of loxP sites (red arrows) using sgRNA-guided Cas9 and germline deleted by crossing to mice carrying an *Ella-Cre* transgene. Black arrows with numbers denote primer positions for genotyping. **B** PCR validation of the *Amp2^{tusΔ}* from genomic DNA. **C** RT-PCR validation of the *Amp2^{tusΔ}* using primers specific to the *Rnpepl1^{tus-RTG}* within t^{us} -Amp2. PCR and RT-PCR experiments were repeated independently at

least twice with similar results. **D** *Amp2^{wt/tusΔ}* ($n = 3$), *Amp2^{wt/tusΔ+loxP}* ($n = 4$) and *Amp2^{wt/tusΔ}* ($n = 3$) males were mated against WT females, and progeny were genotyped for *Amp2^{tusΔ}* and t^{us} versus *wt* alleles. The average percent *wt* transmission (center bar), standard error, and total number of *wt* alleles/total progeny genotyped (in parentheses) are shown. *p* values were calculated using a two-sided likelihood ratio test comparing nested binomial generalized linear mixed models with a binomial family. Each dot represents the percent *wt* alleles transmitted from a single stud male. Source data are provided as a Source data file.

amplicon gene families with *wt* orthologs and *wt/t^{us}* haploid spermatid expression (Fig. 4A, D and Supplementary Table 4). To identify newly acquired t^{us} genes, we generated a de novo t^{us} gene annotation and revealed four t^{us} -specific retrotransposed genes (*tus*-RTG) (Supplementary Fig. 5), two of which were previously proposed *t*-haplotype retrogenes, but their copy number and gene structure were not characterized³⁹. Two t^{us} retrogenes, *Rnpepl1^{tus-RTG}* and *Ppp1cb^{tus-RTG}*, are within amplicons and have *wt/t^{us}* haploid spermatid expression (Fig. 4A, C and Supplementary Fig. 6). Two *Rnpepl1^{tus-RTG}* copies have predicted open reading frames (ORFs) containing the last two exons and 3'UTR of the *Rnpepl1* progenitor gene, and two copies are predicted to form fusion transcripts with *Tmem181a* (Fig. 5B, C). However, PCR cloning and Sanger sequencing of *Tmem181a-Rnpepl1^{tus-RTG}* fusion transcripts reveal alternate splicing within *Rnpepl1^{tus-RTG}* and no apparent open reading frame, suggesting *Rnpepl1^{tus-RTG}* leads to loss of some *Tmem181a* functional copies. In contrast to the gene fusion associated with *Smok2³⁵*, the *Tmem181a-Rnpepl1^{tus-RTG}* gene fusion appears to be disruptive. All 14 spermatid-expressed amplicon gene families have gene copies with intact ORFs, and 13/14 are transcribed from the t^{us} allele (Supplementary Table 5). At least five spermatid-expressed gene families have predicted functions in sperm motility^{35,40-43} (Supplementary Table 5), which is known to be impaired in t^{us}/wt males⁴⁴⁻⁴⁶. Overall, the spermatid expression and predicted function of *t*-haplotype amplicon gene families make them exciting candidates for selfish alleles.

Differential expression of *t*-haplotype amplicon gene families in haploid spermatids

Previously implicated *t*-haplotype selfish alleles are differentially expressed in *wt/t^{us}* versus *wt/wt* testes^{34,36,37,47}. Thus, if t^{us} amplicon gene families have a role in selfish transmission, they may be differentially expressed in *wt/t^{us}* versus *wt/wt* haploid spermatids. We find *Ppp1cb^{tus-RTG}* and *Dynlt1*, both within t^{us} amplicons, are the top two most over-expressed genes relative to *wt* (Fig. 4D and Supplementary Fig. 7A, B). *Ppp1cb^{tus-RTG}* arose from a chromosome 5 progenitor copy of *Ppp1cb*, and acquired spermatid expression since *Ppp1cb* is only expressed in spermatogonia and spermatocytes⁴⁸ (Fig. 4D and

Supplementary Fig. 7A, B). Whether the *Ppp1cb^{tus-RTG}* acquired novel spermatid function remains to be tested. *Dynlt1* is the most amplified gene family on t^{us} and is within conserved Amp2 (Figs. 2, 3 and 4A and Supplementary Table 4). Because there is an assembly gap in the t^{us} -Amp2 region where the *Dynlt1* gene family resides, we used digital droplet PCR (ddPCR) to determine that *Dynlt1^{tus}* has 16 copies and *Dynlt1^{wt}* has four copies (Supplementary Fig. 8). The ~5-fold increase in *Dynlt1* RNA expression in *wt/t^{us}* spermatids is concordant with a 4X increase in *Dynlt1^{tus}* DNA copy number. Similar to previously identified *t*-haplotype selfish alleles^{34,37}, differential expression of *Ppp1cb* and *Dynlt1^{tus}* amplicon gene families indicates their potential contribution to selfish transmission.

A *t*-haplotype amplicon encoding differentially expressed genes contributes to selfish transmission

To address the functional contribution of *t*-haplotype amplicons to selfish transmission, we tested the necessity of t^{us} -Amp2, harboring candidate selfish gene families: *Dynlt1*, *Rnpepl1^{tus-RTG}*, *Tmem181a*, and *Tmem181a-Rnpepl1^{tus-RTG}* (Fig. 5). We selected t^{us} -Amp2 because it is one of the three conserved amplicons in *M. spretus* (Fig. 3 and Supplementary Fig. 4), is the most amplified region on the t^{us} (~16 copies) (Figs. 2 and 4A and Supplementary Fig. 8), harbors the second most over-expressed gene family (*Dynlt1*) in haploid spermatids from *wt/t^{us}* versus *wt/wt* males (Fig. 4D and Supplementary Fig. 7B), and *Dynlt1* encodes dynein protein, a protein family known to be involved in sperm motility and sperm axoneme formation^{3,40,41}.

We generated a targeted ~1.8 Mb deletion of t^{us} -Amp2 and assessed if selfish transmission of t^{us} from *Amp2^{wt/tusΔ}* males is reduced. We observe significantly decreased t^{us} transmission from 97.5% in *Amp2^{wt/tusΔ}* to 95.1% upon Amp2 deletion in *Amp2^{wt/tusΔ}* males ($p = 0.039$), which share the same pedigree and thus have similar genetic backgrounds (Fig. 5 and Supplementary Fig. 9). We also observe reduced, but statistically not significant transmission of t^{us} between *Amp2^{wt/tusΔ}* and *Amp2^{wt/tusΔ}* ($p = 0.495$) males. However, *Amp2^{wt/tusΔ}* has a different genetic background (Fig. 5 and Supplementary Fig. 9). Genetic background is known to contribute substantially to variation in t^{us} transmission⁴⁹. Therefore, to assay the effect of genetic manipulation

on t^{w5} transmission, it is essential to test males with a similar genetic background, ideally littermates. We also observe reduced, but statistically not significant, transmission of a wt -*Amp2* deletion (*Amp2^{wtΔ}*) (Supplementary Fig. 10, $p = 0.664$). We could not test transmission from *Amp2^{wtΔ/tw5Δ}* or *Amp2^{wtΔ/tw5Δ}* males, because they are inviable. Together, our data support the contribution of *Amp2* to selfish transmission, but we cannot exclude the possibility of genetic background effects on transmission from *Amp2^{wt/tw5Δ}*, *Amp2^{wt/tw5-loxp}*, and *Amp2^{wt/tw5}* males. We consider *Dynl1* the most likely gene family contributing to t^{w5} -*Amp2*-mediated selfish transmission based on its evolution, expression, and predicted protein function. We cannot exclude the possibility of *Tmem181a* contributing to selfish transmission since it is upregulated in haploid spermatids from wt/t^{w5} versus wt/wt (Fig. 4D). Since *Rnpepl1^{tw5-RTG}* and the *Tmem181a-Rnpepl1^{tw5-RTG}* fusion transcripts lack open reading frames, they are less likely contributors to selfish transmission. Regardless of which gene family contributes to selfish transmission, the deletion of t^{w5} -*Amp2* and wt -*Amp2* supports amplicons as potential candidates for selfish transmission.

Discussion

Our study of a t -haplotype revealed amplicons with known and candidate selfish t -haplotype alleles, suggesting amplicon acquisition is a key genomic signature of selfish haplotype evolution. Most t -haplotype amplicons harbor spermatid-expressed genes, thus potentially being selfish. Indeed, spermatid-expressed ampliconic genes arose with the initiation and evolution of the t -complex, and deletion of t^{w5} -*Amp2* reduces selfish transmission. The ~3% contribution of t^{w5} -*Amp2* can significantly impact t -haplotype transmission over evolutionary time. The t^{w5} -*Amp2* region likely cooperates with other selfish t -haplotype alleles, culminating in extreme selfish t -haplotype transmission. This cooperation is supported by t -haplotype alleles in inversion 1 (where t^{w5} -*Amp2* is located) and inversion 5 additively influencing selfish transmission⁵⁰, and partial t -haplotypes without inversion 1 reducing selfish transmission^{8,15}. Our studies lay the foundation for future functional studies to address the cooperation of known and candidate selfish t -haplotype alleles.

Based on the evolutionary history of the t -complex, we propose inversions, selfish alleles, and amplification initially arose on the ancestral wt haplotype and were countered by subsequent inversions and amplicon acquisition on the t -haplotype (Fig. 3 and Supplementary Fig. 4). We speculate that an evolutionary arms race for selfish transmission driven by inversions and amplicon acquisition continually shapes *Mus* chromosome 17 evolution, leaving the intriguing question of whether independent inversions or amplicons evolved in other *M. musculus* t -haplotypes or *Mus* species. Our high-quality t -haplotype sequence and cross-species comparisons provide initial insight into this question (e.g., Supplementary Fig. 4 showing an independently acquired amplicon in *M. spretus* and additional inversion in *M. spicilegus*) and are a necessary starting point for future comprehensive sequence comparisons.

Evolution of the t -haplotype shares striking similarities with mammalian sex chromosome evolution. The mammalian X and Y chromosomes diverge by suppression of crossing over (likely via inversions)³⁰, independently acquire multiple amplicons^{51–54}, and exhibit selfish transmission due to an evolutionary arms race between spermatid-expressed ampliconic gene families^{55,56}. These combined features are considered unique to mammalian sex chromosomes; however, our study finds the t -haplotype shares these genomic features despite >150 MY difference in age. Unlike the mammalian X and Y chromosomes, recent t -haplotype evolution and sequence similarity to wt allow a more precise reconstruction of its origin. Evolutionary parallels between mammalian sex chromosomes and the t -haplotype suggest large inversions and amplicon acquisition may be genome-wide features of past or ongoing evolutionary arms races for selfish transmission.

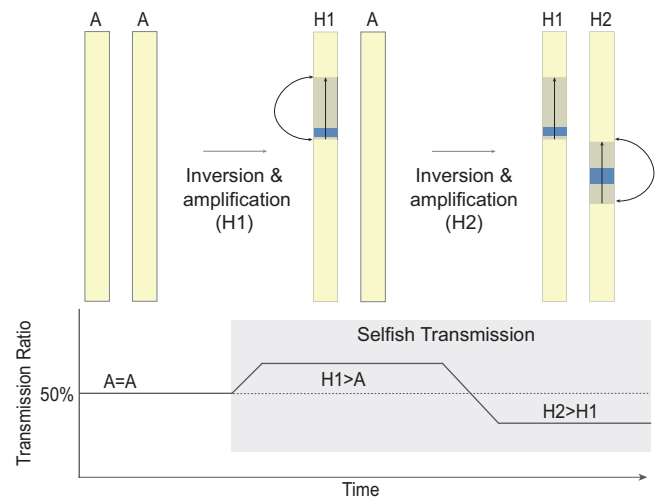


Fig. 6 | A model of a selfish transmission arms race. A model of homologous chromosomes that undergo inversions (shaded gray with black arrows) and amplicon acquisition (blue) to gain selfish transmission. “A” represents the ancestral chromosome, and “H1” and “H2” represent independent haplotypes competing for selfish transmission. The graph below models the change in transmission ratio with the acquisition of inversions and amplicons on each haplotype over evolutionary time.

Beyond the t -complex and mammalian sex chromosomes, do other large inversions harbor selfish amplicons? Selfish amplicons within large inversions are historically difficult to detect due to challenges in (1) resolving large inversions⁵⁷, (2) identifying amplicons^{58–60}, and (3) tracking transmission biases not associated with overt phenotypes (e.g., sex for X and Y chromosomes, and tailless T/t mice). High-quality human and mouse genome assemblies reveal 49 and 16 amplicons with spermatid-expressed gene families, respectively, representing potential regions of ongoing or past selfish transmission (Supplementary Fig. 11). At least one human autosomal amplicon with spermatid-expressed genes is within a known polymorphic inversion under selection, like the t -haplotype⁶¹ (Supplementary Fig. 11). Amplicon-containing inverted haplotypes with selfish transmission may thus be more common than previously thought. Amplicons can also harbor non-germline expressed genes (e.g., MHC locus⁶² and olfactory receptors on the t -haplotype) (Supplementary Fig. 11), indicating amplicons may evolve in response to degeneration of an inverted chromosomal region, such as to facilitate recombination to repair deleterious alleles or increase gene expression by having multiple copies. Our high-quality assembly and analyses of a mouse t -haplotype are a road map for studying the roles of amplicons in additional haplotypes with selfish transmission across species.

We propose a model of how selfish amplicons and large inversions shape chromosome evolution (Fig. 6). Selfish amplicons within an inverted haplotype (H1) can become fixed in the population or engage in an evolutionary arms race for selfish transmission with the ancestral haplotype (A) or new haplotypes (H2), as with the t -haplotype or sex chromosomes. Whether the outcome is fixation or an ongoing arms race, amplicons and large inversions reshape chromosome evolution. Over evolutionary time, the fair play of many Mendelian genes may be impacted by episodes of selfish haplotype foul play.

Methods

Animal use

All animal experiments were conducted in compliance with The University of Michigan’s IACUC board (PRO00011089).

t^{w5} sequencing

The mouse t^{w5} haplotype was sequenced using a combination of PacBio HiFi, Oxford Nanopore, and BioNano optical mapping. All sequencing data were generated using high molecular weight (HMW) DNA isolated from passage 19 or lower of male t^{w5}/t^{w5} mESCs on a C57BL/6J genetic background. t^{w5} inversion junction breakpoints were validated by PCR from DNA of t^{w5}/wt mice as well as from t^{w5}/t^{w5} mESCs (Supplementary Fig. 1C), indicating the five inversions are detectable in vivo in mice that have been propagated in the lab for many generations. Our data are consistent with previous reports of at least four t -haplotype inversions relative to wt , thus confirming the genetic integrity of the t^{w5} haplotype structure in our ES cell culture.

PacBio HiFi. PacBio HiFi sequencing was conducted with HMW DNA isolated from 2 to 4×10^6 t^{w5}/t^{w5} mESCs per replicate using Analytik Jena's smart-m-Prep kit. Two to five micrograms of HMW DNA was sheared to 20–30 kb fragments using the MegaRuptor (Speed 31/32, $T = 120$ min) and subjected to 1x AmpurePB cleanup. PacBio circular consensus/HiFi-libraries were prepared from 4 μ g of fragmented DNA following PacBio's SMRTbell Express Template Preparation protocol. Size selection was performed using a BluePippin with an HSLF chip, and DNA between 10–50 kb was eluted and isolated using AmpurePB beads. The resulting libraries range between 20 and 28 kb fragments (Fragment Analyzer) and were loaded on four PacBio SequelII SMRT cells using adaptive loading. After 30 h of data acquisition between 39 and 142 Gb (38.89, 141.72, 104.41, and 133.3 Gb) of unique Molecular Data were yielded per sample (260.91, 541.38, 539.61, 581.01 Gb Total Bases).

Oxford nanopore. Nanopore ultra-long DNA sequencing was performed using UHMW DNA isolated from the Nanobind CBB Big DNA kit (Circulomics, cat. # NB-900-001-01). Briefly, three aliquots of 6×10^6 t^{w5}/t^{w5} mESCs were subjected to DNA extraction following the manufacturer's protocol. The eluates were then quantified using the Qubit dsDNA BR kit, and 40 μ g of UHMW DNA was subjected to the Nanobind UL Library Prep kit (Circulomics, cat. # NB-900-601-01) in conjunction with ONT's Ultra-Long DNA Sequencing kit (Oxford Nanopore, cat. # SQK-ULK001). Tagmentation was performed using 6 μ l of FRA enzyme and subsequent ligation of the RAP-F sequencing adapter for 1 h at room temperature. The final sequencing libraries were loaded onto R9.4.1 MinION flow cells (Oxford Nanopore, FLO-MINI06) for N50 fragment length estimation and then processed on two R9.4.1 PromethION flow cells (Oxford Nanopore, FLO-PRO002) for 72 h. A nuclease flush (Oxford Nanopore, EXP-WSH003) was performed every 16 h according to the manufacturer's protocol, and a fresh sequencing library was loaded. A total of 121 Gb (~48x coverage) basecalled bases with N50 read length of ~90 kb were obtained and further subjected to genome assembly.

BioNano optical mapping. Ultra-HMW DNA was extracted from frozen cells using the "Bionano Prep SP Frozen Cell Pellet DNA Isolation Protocol v2". The long DNA molecules were labeled with the DLE-1 enzyme according to the "Bionano Prep Direct Label and Stain (DLS) Protocol." The labeled molecules were loaded into a single flowcell of a Saphyr Chip G1.2. The chip was run on the Bionano Saphyr Optical Mapping Instrument twice, collecting the images of 955.75 Gbp molecules greater than 150 kb with a minimum of 9 labels. These data were assembled using Bionano Solve v.3.6.1 software and optArguments_nonhaplotype_DLE1_saphyr.xml parameters into 739 genome maps with an N50 of 99.13 Mb and a total length of 2725.521 Mb.

t^{w5} chromosome-level assembly

A chromosome-level assembly was generated from PacBio HiFi, ONT, and Bionano long molecule technologies. Briefly, Guppy v6.3.7 (Oxford Nanopore Technologies, UK) was used for basecalling ONT

raw reads, utilizing the configuration file dna_r9.4.1_450bps_hac.cfg. The generated sequences are named ONT data. To obtain PacBio HiFi consensus reads, we followed the recommended workflow of DeepConsensus⁶³, where the subreads of each run were partitioned into 500 bp segments. We obtained the draft consensus sequences using ccs v6.3.0 (PacificBiosciences, USA) and aligned the subreads to the draft consensus sequences using actc v0.3.1 (PacificBiosciences, USA) for each segment of partitioned data. We then applied DeepConsensus⁶³ v1.0.0 to the draft consensus sequences and aligned the subreads of each 500 bp segment to generate more accurate consensus sequences. The generated accurate consensus sequences from DeepConsensus⁶³ are named HiFi data. We used Verkko⁶⁴ v1.3, using both HiFi and ONT data with default parameters to generate a de novo assembly.

Assembled sequence contigs were placed onto BioNano optical map scaffolds using Bionano Solve v.3.6.1 software and default hybridScaffold_DLE1_config parameters, resolving conflicts in both the optical map and sequence assemblies to generate large contigs. Three non-overlapping contigs comprising the t^{w5} chromosome were identified using BLAST⁶⁵ to compare all contigs to the proximal 40 Mb of mm10 chr17 that was repeat masked using default parameters of RepeatMasker⁶⁶. The order and orientation of each contig was determined based upon shared repeat content at the 3' and 5' ends of each scaffold and independently validated using Hi-C scaffolding (Supplementary Fig. 1A). While there are no sequence gaps within scaffolds, there are two remaining physical gaps between scaffolds represented by 25 kb of N's.

Validation of t^{w5} assembly

To independently validate the t^{w5} assembly, we use PacBio CLR to generate a de novo assembly and Hi-C to confirm the order and orientation of the contigs (Supplementary Fig. 1A).

PacBio CLR. PacBio CLR sequencing was performed with 15 μ g of HMW DNA from t^{w5}/t^{w5} mESCs. For library preparation, HMW DNA was sheared to a target size of >30 kb using the MegaRuptor (Speed 3, $t = 159$ m49 s) and cleaned up using AmpurePB beads. PacBio CLR libraries were prepared from 4 μ g of fragmented DNA following PacBio's SMRTbell Express Template Preparation protocol. Size selection was performed using a BluePippin with an HSLF chip, and 30 kb DNA fragments were eluted and cleaned up using AmpurePB beads. The final library fragment size was 43 kb, according to Fragment Analyzer. Two flow cells were run with 15 h acquisition time on the PacBio SequelII, yielding 78 and 142 Gb unique data (98 and 165 Gb total data).

Hi-C. Hi-C was performed using 2×10^6 t^{w5}/t^{w5} mESCs. The cells were collected by trypsinization, washed with PBS, counted, and snap-frozen using liquid nitrogen. Crosslinking was performed using 2% formaldehyde according to the crosslinking protocol for cryopreserved cells provided by the Arima-HiC Kit user guide (A160134_v01). Following the manufacturer's protocol, proximally ligated DNA was generated using the Arima-HiC Kit (A510008). Hi-C library preparation was conducted using the KAPA HyperPrep Kit (Roche, 07962312001) and KAPA single-indexed adapter (Roche, 08005702001; index 4: TGACCA) according to Arima's Library preparation user guide. The Covaris S220 sonicator was used for DNA shearing. The library was quantified using the Qubit DNA HS assay (Life Technologies), and the library size was validated using DNA HS Bioanalyzer chips (Agilent). Paired-end sequencing was performed on the NovaSeq 6000 with 2x 100 bp read length. The ligation sites of HiC were obtained using pre-set parameters in the Arima pipeline.

Dot-plots. Two-dimensional (2D) dot-plots comparing chromosome assemblies from *M. musculus* (GRCm38-mm10), *M. spicilegus* (MUSP714_v2, GCA_03336285.2), *M. spretus*⁶⁷ (SPRET_Eij_v3,

GCA_921997135.2), and *M. caroli*⁶⁸ (CAROLI_EIJ_v1.1, GCA_900094665.2), and self-symmetry dot-plots showing internal duplications were generated from a custom Perl script⁶⁹.

Mapping t-complex inversion boundaries

To precisely map inversion boundaries, chromosomal regions spanning t^{us} inversion breakpoint junctions were estimated based on 2D dot-plots, repeat masked using default parameters on RepeatMasker⁶⁶, and compared to the mm10 chr17 sequence using BLAT⁷⁰. Inversion boundaries were experimentally validated via PCR using primers flanking predicted t^{us} inversion boundaries compatible with both t^{us} and mm10 chr17 sequences, but only amplified from t^{us} genomic DNA (gDNA) (Supplementary Fig. 1C and Source data File).

Intrachromosomal identity, TE, and GC content

Sliding windows analyses were conducted to estimate percent intrachromosomal identity (50 kb window, 1 kb step), TE (200 kb window, 10 kb step), and GC content (100 kb window, 10 kb step). To calculate intrachromosomal identity, BLAST⁷⁰ was used to compare all windows to the t^{us} assembly, and the top filtered results were >10 kb, not non-self-alignments. To calculate TE content, the t^{us} assembly was repeat masked using the default parameters of RepeatMasker⁶⁶, and percent SINE, LINE, and ERV elements were calculated for each window using a custom Perl script. Similarly, a custom Perl script calculated the percent GC content for each window. Source data are provided in the Source data file.

Sequence degeneration estimates

Single-copy sequence degeneration on the t^{us} was estimated via alignment to the first 50 Mb of mm10 chr17 (excluding the centromere) using Minimap2 v.2.14⁷¹ and identifying regions with low t^{us} alignment coverage. Coverage was determined using bedtools⁷², and the sum of non-overlapping, low-coverage base pairs was calculated using a custom Python script. Low t^{us} coverage regions were plotted relative to *wt* using a sliding window (100 kb window, 10 kb step). Source data are provided in the Source data file.

Copy number estimates

Copy number estimates were estimated using Illumina short-read sequencing data generated from homozygous t^{us}/t^{us} mESCs and previously published C57BL/6j⁷³ (SRR7511358), *M. spretus*⁶⁷ (ERR9880927), *M. caroli*⁶⁸ (ERR133992), and *M. spicilegus*⁷⁴ (SRR6356165) genomic DNA. We analyzed the short-read sequencing data via fastCN⁷⁵, a pipeline that estimates copy number based on read depth within 3 kb, repeat-masked windows along a reference genome. Thus, fastCN copy number estimation is independent of duplication size or structural variation. Briefly, reads were mapped to the t^{us} or mm10 chr17 assemblies in which RepeatMasker⁶⁶, Tandem Repeat Finder⁷⁶, and 50-mers with an occurrence greater than 50 were masked. Read depth was normalized to account for GC content, averaged in windows containing 3 kb of unmasked positions, and converted to copy number estimates based on a defined single-copy sequence in the mm10 reference assembly. Amplicon gene copy numbers were estimated using the average fastCN copy number across all 3 kb windows spanning the mm10 gene coordinates. fastCN gene copy number estimates are supported by ddPCR copy number estimates (Supplementary Fig. 8) and the mm10 reference genome. Source data are provided in the Source data file.

Illumina short-read sequencing

According to the manufacturer's protocol, mate-pair sequencing libraries were prepared from t^{us}/t^{us} mESC gDNA using the Illumina Nextera - MatePair kit and sequenced on an Illumina HiSeq2500.

Amplicon phylogenetic analysis

To conduct a sequence-based phylogenetic analysis of t^{us} amplicons, we first used minimap2 (v.2.24) to perform pair-wise alignments

between t^{us} amplicon regions and each *Mus* species' chromosome 17 assembly. From each pair-wise alignment, we identified an orthologous sequence that has been duplicated on the t^{us} , defined as a repeat unit. We used Mauve⁷⁷ to conduct multiple sequence alignments (MSAs) of t^{us} repeat units across *Mus* species and the t^{us} . From each MSA, we extracted and concatenated locally collinear blocks greater than 10 kb that include sequence from all *Mus* species and the t^{us} (except the MSA for Amp1, which excludes *M. spicilegus*). PhyML (v. 3.3.20180621)⁷⁸ was used to generate maximum-likelihood Neighbor-Joining phylogenetic trees with 100 bootstraps under the Jukes-Cantor model of nucleotide substitution. Phylogenetic trees were visualized and transformed to cladograms using FigTree (v1.4.4). It is important to note that while our copy number analysis supports conserved amplification of Amp1, Amp2, and Amp4 in *M. spretus* and *M. spicilegus*, their assemblies are single-copy and are thus likely misassembled. These misassemblies limit the interpretation of our phylogenetic analysis.

Pachytene spermatocyte and round spermatid FACS and RNA-seq library preparation

A previously published protocol was followed to isolate pachytene spermatocytes (4N) and round spermatids (1N) from adult testes⁷⁹. We dissociated cells from a pair of testes by enzymatic treatment with Collagenase type IA (Worthington Biochemical), DNase I (Invitrogen), and Trypsin (Gibco). Cell suspensions were passed through 70 μ m and 40 μ m cell strainers and incubated with Hoechst 33342 (Thermo Fisher) for DNA content and propidium iodide for cell viability. Cell sorting was performed on the MoFlo Astrios Cell Sorter. The purity of sorted cells was determined via fluorescence microscopy, visual inspection of 100 cells' morphology, and γ H2AX (Abcam, ab17722) immunofluorescence for pachytene spermatocytes or nuclear staining with DAPI for round spermatids. The purity of sorted cells used for subsequent RNA-seq was >85%.

Cells sorted for RNA sequencing were sorted directly into TRIzol LS (Thermo Fisher Scientific) for total RNA extraction. RNA was treated with DNase using the RNeasy Mini Kit and the RNase Free DNase Set (QIAGEN) to remove any contaminating genomic DNA. RNA sequencing libraries were generated using samples with >7 RIN scores and >5 ng total RNA. mRNA-seq libraries were generated using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, E7490) and NEBNext UltraExpress RNA Library Prep Kit (NEB, E3330). Pachytene spermatocyte mRNA-seq libraries were generated from three *wt/wt* and two age-matched *wt/t^{us}* biological replicates. Round spermatid mRNA-seq libraries were generated from three biological replicates of *wt/wt* and *wt/t^{us}* males. All libraries were sequenced on a single Illumina NovaSeq Flowcell (~25 million reads per sample on average). Alignments were performed with Kallisto⁸⁰ using the RefSeq Index, and differential expression was estimated using DESeq2⁸¹ and plotted using ggplot2⁸².

To detect allele-specific transcription, RNA-seq reads from *wt/t^{us}* and *wt/wt* round spermatids were trimmed using Cutadapt v2.3⁸³. FastQC v0.11.8 was used to ensure the quality of data⁸⁴. FastQ Screen was used to screen for contaminating sequence reads⁸⁵. Reads were mapped to the reference genome GRCm38, using STAR v2.7.8a⁸⁶. Aligned reads were visualized using the Integrative Genomics Viewer (IGV; v2.19.4)⁸⁷ and mm10 NCBI RefSeq gene annotation. t^{us} -specific variants were identified in aligned reads from *wt/t^{us}* spermatids and are not detectable in *wt/wt* spermatids (Supplementary Fig. 7E).

t^{us} gene annotation

To generate gene annotations of the t^{us} sequence, we combined predicted gene annotations based on our sorted pachytene spermatocyte and round spermatid RNA-seq data with the mm10 RefSeq gene set. The RNA-seq reads were first normalized in Trinity⁸⁸ and then aligned to make second-strand cDNA predictions using t^{us} , the genome assembly as a guide, and an intron max length of 10 kb. We aligned the

Trinity-based and mm10 RefSeq-based cDNAs to the t^{us} assembly via BLAT⁶⁵ using a mindentity=92. We retained only the best hits with a minimum of 80% coverage using the Augustus⁵⁹ perl script filterPSL.pl, and generated gff hints files for Augustus with the blat2hints.pl script. The Trinity-based RNA-seq and mm10 RefSeq gff hint files were combined and used in Augustus to annotate softmasked t^{us} sequence using Augustus with parameters `-species=humans, softmasking=1, --UTR=on` with configuration file `extrinsic.M.RM.E.W.P.cfg`. The Augustus gff outputs were converted to fasta files with `gffread`⁹⁰. To identify t^{us} amplified gene families with *wt* orthologs, we compared the CDS of all annotated mm10 chr17 protein-coding genes within syntenic regions of estimated t^{us} amplicons (based on fastCN and self-symmetry triangular dot-plots). Genes with ≥ 2 t^{us} orthologs were considered ampliconic.

ORF analysis

To define ORFs of spermatid-expressed t^{us} amplicon gene families, `blat` was used to compare the coding sequence of mm10 annotated genes to the t^{us} assembly. t^{us} genes with the same ORF as mm10 annotated genes were confirmed using `Expasy`⁹¹. For t^{us} amplicon gene families that did not share the same ORF as mm10 annotated genes, we identified their predicted ORFs from our de novo gene annotation.

Generation of an $Amp2^{t^{us}}$ and $Amp2^{wt\Delta}$ mouse lines

To generate mice with a -1.8 Mb deletion of t^{us} , `Amp2`, `loxP` sites were sequentially integrated on the 5' and 3' flanking regions of `Amp2` via CRISPR. sgRNAs were designed to target unique sequence flanking `Amp2` on the t^{us} , selected based on the UCSC Genome Browser CRISPR Targets track scores, and synthesized by Synthego (Supplementary Table 7). sgRNA, Cas9 mRNA (IDT Alt R S.p. Cas9 Nuclease V3), and a single-stranded oligo donor carrying the `loxP` sequence (IDT; Supplementary Table 7) were mixed for microinjection at a concentration of 30, 50, and 10 ng/ μ l, respectively. CRISPR sgRNAs, Cas9, and `loxP` donor oligos were pronuclear injected into zygotes from 8–12-week-old C57BL/6J (JAX stock #000664) X 129S4/SvJae $Amp2^{wt/t^{us}}$ males (JAX stock #009104 and Max Plank Institute for Molecular Genetics) crossed to 20-week-old F1 females from a C57BL/6J X SJL cross (JAX stock #100012) to generate $Amp2^{wt/t^{us-LoxP}}$ mice. Since 5' and 3' `loxP` sites were integrated sequentially, $Amp2^{wt/t^{us-LoxP}}$ mice were crossed to 20-week-old C57BL/6J X SJL F1 females (JAX stock #100012) for one more time. $Amp2^{wt/t^{us-LoxP}}$ (floxed) mice were then crossed to `Ella-Cre` (JAX stock #003724 backcrossed onto C57BL/6J) to generate $Amp2^{wt/t^{us\Delta}}$ mice (Fig. 5). The transmission distortion assay used $Amp2^{wt/t^{us-LoxP}}$ and $Amp2^{wt/t^{us\Delta}}$ males sired from the same grandparents. $Amp2^{wt/t^{us}}$ mice were alternately crossed onto C57BL/6J (JAX stock #000664) and 129S4/SvJae (JAX stock #009104) genetic backgrounds each generation to improve fertility for line maintenance. Thus, $Amp2^{wt/t^{us-LoxP}}$ and $Amp2^{wt/t^{us\Delta}}$ males have a genetic background that differs substantially from $Amp2^{wt/t^{us}}$ males by having been crossed to SJL and `Ella-Cre` carrying mice (Supplementary Fig. 9). Mice were genotyped by extracting DNA from tail or ear biopsy using NaOH extraction⁹² and primers flanking the 5' and 3' `loxP` sites (Supplementary Table 7).

To generate mice with a -500 kb deletion of *wt-Amp2*, CRISPR and sgRNA targeting the 5' and 3' flanking regions of *wt-Amp2*, and a homologous donor oligo were used. sgRNAs were designed to target unique sequence flanking `Amp2` on the *wt* using the CRISPR/Cas9 10K Tracks in the UCSC genome browser (Supplementary Table 7). sgRNA, Cas9 mRNA (IDT Alt R S.p. Cas9 Nuclease V3), and a single-stranded oligo donor carrying homology arms to the cut sites (IDT; Supplementary Table 7) were mixed for microinjection at a concentration of 30, 50, and 10 ng/ μ l, respectively. Pronuclear microinjection was performed on zygotes derived from 20-week-old C57B/6J/SJL F1 males and females (JAX stock #100012). Founder mice carrying the deletion were genotyped by extracting DNA from tail or ear biopsy using NaOH extraction⁹² and primers flanking the 5' and 3' sgRNA cut sites

(Supplementary Fig. 10A, Supplementary Table 7). Source data are provided in the Source data file.

Transmission distortion assay

To assess the transmission frequency of t^{us} or *wt* alleles, we genotyped offspring resulting from $Amp2^{wt/t^{us\Delta}}$, $Amp2^{wt/t^{us-LoxP}}$, and $Amp2^{wt/t^{us\Delta}}$ males bred with 8–27 weeks old CD1 females (Charles River stock #022) for $Amp2^{t^{us\Delta}}$, *wt* versus t^{us} , and $Amp2^{wt\Delta}$ alleles (Supplementary Table 7). At least three male mice between 8 and 27 weeks old were bred for each genotype. At least 100 offspring were sired from $Amp2^{wt/t^{us\Delta}}$, $Amp2^{wt/t^{us-LoxP}}$, and $Amp2^{wt/t^{us}}$ males (except for one male that produced 98 offspring), and at least 80 offspring were sired from $Amp2^{wt\Delta}$ males (Supplementary Table 6).

RT-PCR

Total RNA was extracted using `Trizol` (Life Technologies) according to the manufacturer's protocol. Ten μ g of total RNA was DNase-treated using `Turbo DNase` (Life Technologies) and reverse transcribed using `Superscript II` (Invitrogen) using oligo (dT) primers following the manufacturer's protocol. Intron-spanning primers were used to perform RT-PCR on adult testis cDNA for *Ppp1cb*, *Rnpepl1*, *Tmem181a*, *Rnpepl1*, and *Trim42*, a round spermatid-specifically expressed gene and internal positive control. Source data are provided in the Source data file.

ddPCR

Droplet digital PCR (ddPCR) was conducted to estimate *Dynl1* copy on the t^{us} and *wt* haplotypes experimentally. According to the manufacturer's protocol, HMW DNA was extracted from snap-frozen spleen tissue using the `MagAttract HMW DNA Kit` (QIAGEN). Primers and TaqMan custom probes unique to *Dynl1* and *Sox9*, a single copy internal control, were designed to amplify -150 bp products (Integrated DNA Technologies; Supplementary Table 7). Products were amplified from two technical replicates for each DNA sample according to the Droplet Digital PCR Applications Guide (Bio-Rad) on the QX200 Droplet Generator and QX200 Droplet Reader and quantified using `QuantaSoft` Software.

Defining human and mouse genome-wide amplicons

Genome-wide segmental duplications were collected from the UCSC Genome Table Browser for human (GRCh38/hg38) and mouse (GRcm38/mm10) genomes. Segmental duplications were filtered by size (≥ 10 kb), identity ($\geq 99\%$), and distance between duplications (< 1.5 Mb). We define this filtered set of segmental duplications as genome-wide amplicons. All UCSC RefSeq-annotated genes within amplicons were collected from the UCSC Genome Table Browser. To determine whether ampliconic genes are spermatid-expressed, we aligned RNA-seq data from sorted human (SRR1977587, SRR1977610, SRR1977618) and mouse (SRR1977639) spermatids⁹³ to their respective RefSeq indices using `Kallisto`⁸⁰, and defined expression as genes with ≥ 10 tpm.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The primary read files for the t^{us} sequence assembly, mate-pair sequencing, and RNA sequencing, and t^{us} assembly have been deposited in the National Center for Biotechnology Information (NCBI) database under [BioProject ID PRJNA1142398](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1142398). Publicly available genome assemblies used in this study can be found with the following accession codes: *M. spicilegus* (MUSP714_v2, [GCA_03336285.2](https://www.ncbi.nlm.nih.gov/assembly/GCA_03336285.2)), *M. spretus* (SPRET_Eij_v3, [GCA_921997135.2](https://www.ncbi.nlm.nih.gov/assembly/GCA_921997135.2)), and *M. caroli* (CAR-OLI_Eij_v1.1, [GCA_900094665.2](https://www.ncbi.nlm.nih.gov/assembly/GCA_900094665.2)). Publicly available Illumina WGS data

used in this study can be found with the following accession codes: C57BL/6j (SRR7511358), *M. spretus* (ERR9880927), *M. caroli* (ERR133992), and *M. spicilegus* (SRR6356165). Publicly available RNA sequencing data used in the study can be found with the following accession codes: human round spermatid (SRR1977587, SRR1977610, SRR1977618), mouse round spermatid (SRR1977639). Source data are provided as a Source data file. Source data are provided with this paper.

Code availability

Custom codes developed for data analysis and visualization are publicly available on the Zenodo repository (<https://doi.org/10.5281/zenodo.17296126>).

References

- Nuckolls, N. L. et al. wtf genes are prolific dual poison-antidote meiotic drivers. *eLife* **6**, <https://doi.org/10.7554/eLife.26033> (2017).
- Burt, A. & Trivers, R. *Genes in Conflict: The Biology of Selfish Genetic Elements* (Belknap Press of Harvard University Press, 2006).
- Herrmann, B. G. & Bauer, H. The mouse t-haplotype: a selfish chromosome—genetics, molecular mechanism and evolution. in *Evolution of the House Mouse* (Baird, S. J. E., Macholán, M., Munclinger, P. & Piálek, J.) 297–314 (Cambridge University Press, 2012).
- Larracuente, A. M. & Presgraves, D. C. The selfish segregation distorter gene complex of *Drosophila melanogaster*. *Genetics* **192**, 33–53 (2012).
- Lindholm, A. K. et al. The Ecology and evolutionary dynamics of meiotic drive. *Trends Ecol. Evol.* **31**, 315–326 (2016).
- Chesley, P. & Dunn, L. C. The inheritance of taillessness (anury) in the house mouse. *Genetics* **21**, 525–536 (1936).
- Dunn, L. C. & Bennett, D. A new case of transmission ratio distortion in the house mouse. *Proc. Natl. Acad. Sci. USA* **61**, 570–573 (1968).
- Lyon, M. F. Transmission ratio distortion in mouse t-haplotypes is due to multiple distorter genes acting on a responder locus. *Cell* **37**, 621–628 (1984).
- Artzt, K., Shin, H. S. & Bennett, D. Gene mapping within the T/t complex of the mouse. II. Anomalous position of the H-2 complex in t haplotypes. *Cell* **28**, 471–476 (1982).
- Hammer, M. F., Schimenti, J. & Silver, L. M. Evolution of mouse chromosome 17 and the origin of inversions associated with t haplotypes. *Proc. Natl. Acad. Sci. USA* **86**, 3261–3265 (1989).
- Herrmann, B. et al. Genetic analysis of the proximal portion of the mouse t complex: evidence for a second inversion within t haplotypes. *Cell* **44**, 469–476 (1986).
- Shin, H. S., Flaherty, L., Artzt, K., Bennett, D. & Ravetch, J. Inversion in the H-2 complex of t-haplotypes in mice. *Nature* **306**, 380–383 (1983).
- Silver, L. M. & Artzt, K. Recombination suppression of mouse t-haplotypes due to chromatin mismatching. *Nature* **290**, 68–70 (1981).
- Lyon, M. F., Evans, E. P., Jarvis, S. E. & Sayers, I. t-Haplotypes of the mouse may involve a change in intercalary DNA. *Nature* **279**, 38–42 (1979).
- Lyon, M. F. Transmission ratio distortion in mice. *Annu. Rev. Genet.* **37**, 393–408 (2003).
- Seitz, A. W. & Bennett, D. Transmission distortion of t-haplotypes is due to interactions between meiotic partners. *Nature* **313**, 143–144 (1985).
- Chesley, P. Development of the short-tailed mutant in the house mouse. *J. Exp. Zool.* **70**, 429–459 (1935).
- Lyon, M. F. Male sterility of the mouse t-complex is due to homozygosity of the distorter genes. *Cell* **44**, 357–363 (1986).
- Runge, J.-N., Ullrich, K. & Lindholm, A. K. De novo assembly of the selfish t supergene reveals a deleterious evolutionary trajectory. *bioRxiv* <https://doi.org/10.1101/2024.09.15.613113> (2024).
- Harr, B. et al. Genomic resources for wild populations of the house mouse, *Mus musculus* and its close relative *Mus spretus*. *Sci. Data* **3**, 160075 (2016).
- Dunn, L. C. & Bennett, D. A comparison of the effects, in compounds, of seven genetically similar lethal T alleles from populations of wild house mice. *Genetics* **45**, 1531–1538 (1960).
- Kuretake, S., Maleszewski, M., Tokumasu, A., Fujimoto, H. & Yanagimachi, R. Inadequate function of sterile tw5/tw32 spermatozoa overcome by intracytoplasmic sperm injection. *Mol. Reprod. Dev.* **44**, 230–233 (1996).
- Garside, W. & Hillman, N. The in vivo and in vitro transmission frequencies of the tw5-haplotype in mice. *Genet. Res.* **53**, 21–24 (1989).
- Silver, L. M. Mouse t haplotypes. *Annu. Rev. Genet.* **19**, 179–208 (1985).
- Howard, C. A., Gummere, G. R., Lyon, M. F., Bennett, D. & Artzt, K. Genetic and molecular analysis of the proximal region of the mouse t-complex using new molecular probes and partial t-haplotypes. *Genetics* **126**, 1103–1114 (1990).
- Lyon, M. F. & Meredith, R. Investigations of the nature of T-alleles in the mouse. III. Short tests of some further mutant alleles. *Heredity* **19**, 327–330 (1964).
- Herrmann, B. G., Barlow, D. P. & Lehrach, H. A large inverted duplication allows homologous recombination between chromosomes heterozygous for the proximal t complex inversion. *Cell* **48**, 813–825 (1987).
- Vojtiskova, M., Viklicky, V., Voracova, B., Lewis, S. E. & Gluecksohn-Waelsch, S. The effects of a t-allele (tAE5) in the mouse on the lymphoid system and reproduction. *J. Embryol. Exp. Morphol.* **36**, 443–451 (1976).
- Charlesworth, B., Sniegowski, P. & Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**, 215–220 (1994).
- Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).
- Charlesworth, B. & Charlesworth, D. The degeneration of Y chromosomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**, 1563–1572 (2000).
- Rozen, S. et al. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**, 873–876 (2003).
- Swanepoel, C. M., Gerlinger, E. R. & Mueller, J. L. Large X-linked palindromes undergo arm-to-arm gene conversion across mus lineages. *Mol. Biol. Evol.* **37**, 1979–1985 (2020).
- Bauer, H., Willert, J., Koschorz, B. & Herrmann, B. G. The t complex-encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. *Nat. Genet.* **37**, 969–973 (2005).
- Herrmann, B. G., Koschorz, B., Wertz, K., McLaughlin, K. J. & Kispert, A. A protein kinase encoded by the t complex responder gene causes non-Mendelian inheritance. *Nature* **402**, 141–146 (1999).
- Bauer, H., Veron, N., Willert, J. & Herrmann, B. G. The t-complex-encoded guanine nucleotide exchange factor Fgd2 reveals that two opposing signaling pathways promote transmission ratio distortion in the mouse. *Genes Dev.* **21**, 143–147 (2007).
- Bauer, H. et al. The nucleoside diphosphate kinase gene Nme3 acts as quantitative trait locus promoting non-Mendelian inheritance. *PLoS Genet.* **8**, e1002567 (2012).
- Veron, N. et al. Retention of gene products in syncytial spermatids promotes non-Mendelian inheritance as revealed by the t complex responder. *Genes Dev.* **23**, 2705–2710 (2009).
- Kelemen, R. K., Elkrewi, M., Lindholm, A. K. & Vicoso, B. Novel patterns of expression and recruitment of new genes on the t-haplotype, a mouse selfish chromosome. *Proc. Biol. Sci.* **289**, 20211985 (2022).

40. Indu, S. et al. Aberrant Expression of Dynein light chain 1 (DYNLT1) is Associated with Human Male Factor Infertility. *Mol. Cell. Proteom.* **14**, 3185–3195 (2015).
41. Kagami, O. et al. A dynein light chain of sea urchin sperm flagella is a homolog of mouse Tctex 1, which is encoded by a gene of the t complex sterility locus. *Gene* **211**, 383–386 (1998).
42. Lindemann, C. B. Structural-functional relationships of the dynein, spokes, and central-pair projections predicted from an analysis of the forces acting within a flagellum. *Biophys. J.* **84**, 4115–4126 (2003).
43. Sironen, A., Shoemark, A., Patel, M., Loebinger, M. R. & Mitchison, H. M. Sperm defects in primary ciliary dyskinesia and related causes of male infertility. *Cell Mol. Life Sci.* **77**, 2029–2048 (2020).
44. Winkler, L. & Lindholm, A. K. A meiotic driver alters sperm form and function in house mice: a possible example of spite. *Chromosome Res.* **30**, 151–164 (2022).
45. Pilder, S. H. et al. The molecular basis of “curlicue”: a sperm motility abnormality linked to the sterility of t haplotype homozygous male mice. *Soc. Reprod. Fertil. Suppl.* **63**, 123–133 (2007).
46. Olds-Clarke, P. & Johnson, L. R. t haplotypes in the mouse compromise sperm flagellar function. *Dev. Biol.* **155**, 14–25 (1993).
47. Charron, Y. et al. Two isoforms of the RAC-specific guanine nucleotide exchange factor TIAM2 act oppositely on transmission ratio distortion by the mouse t-haplotype. *PLoS Genet.* **15**, e1007964 (2019).
48. Ernst, C., Eling, N., Martinez-Jimenez, C. P., Marioni, J. C. & Odom, D. T. Staged developmental mapping and X chromosome transcriptional dynamics during mouse spermatogenesis. *Nat. Commun.* **10**, 1251 (2019).
49. Gummere, G. R., McCormick, P. J. & Bennett, D. The influence of genetic background and the homologous chromosome 17 on t-haplotype transmission ratio distortion in mice. *Genetics* **114**, 235–245 (1986).
50. Redkar, A. A., Si, Y., Twine, S. N., Pilder, S. H. & Olds-Clarke, P. Genes in the first and fourth inversions of the mouse t complex synergistically mediate sperm capacitation and interactions with the oocyte. *Dev. Biol.* **226**, 267–280 (2000).
51. Hughes, J. F. et al. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* **483**, 82–86 (2012).
52. Hughes, J. F. et al. Sequence analysis in *Bos taurus* reveals pervasiveness of X-Y arms races in mammalian lineages. *Genome Res.* **30**, 1716–1726 (2020).
53. Mueller, J. L. et al. Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat. Genet.* **45**, 1083–1087 (2013).
54. Skaletsky, H. et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
55. Kruger, A. N. et al. A neofunctionalized X-linked ampliconic gene family is essential for male fertility and equal sex ratio in mice. *Curr. Biol.* **29**, 3699–3706.e5 (2019).
56. Cocquet, J. et al. A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse. *PLoS Genet.* **8**, e1002900 (2012).
57. Porubsky, D. et al. Recurrent inversion polymorphisms in humans associate with genetic instability and genomic disorders. *Cell* **185**, 1986–2005 e1926 (2022).
58. Bailey, J. A., Church, D. M., Ventura, M., Rocchi, M. & Eichler, E. E. Analysis of segmental duplications and genome assembly in the mouse. *Genome Res.* **14**, 789–801 (2004).
59. Kelley, D. R. & Salzberg, S. L. Detection and correction of false segmental duplications caused by genome mis-assembly. *Genome Biol.* **11**, R28 (2010).
60. Vollger, M. R. et al. Long-read sequence and assembly of segmental duplications. *Nat. Methods* **16**, 88–94 (2019).
61. Stefansson, H. et al. A common inversion under selection in Europeans. *Nat. Genet.* **37**, 129–137 (2005).
62. Rogers, J. H. & Willison, K. R. A major rearrangement in the H-2 complex of mouse t haplotypes. *Nature* **304**, 549–552 (1983).
63. Baid, G. et al. DeepConsensus improves the accuracy of sequences with a gap-aware sequence transformer. *Nat. Biotechnol.* **41**, 232–238 (2023).
64. Rautiainen, M. et al. Telomere-to-telomere assembly of diploid chromosomes with Verkko. *Nat. Biotechnol.* **41**, 1474–1482 (2023).
65. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
66. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0 (2013–2015).
67. Lilue, J. et al. Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci. *Nat. Genet.* **50**, 1574–1583 (2018).
68. Thybert, D. et al. Repeat associated mechanisms of genome evolution and function revealed by the *Mus caroli* and *Mus pahari* genomes. *Genome Res.* **28**, 448–459 (2018).
69. Kuroda-Kawaguchi, T. et al. The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat. Genet.* **29**, 279–286 (2001).
70. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
71. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
72. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
73. Sarsani, V. K. et al. The genome of C57BL/6J “Eve”, the mother of the laboratory mouse genome reference strain. *G3* **9**, 1795–1805 (2019).
74. Couger, M. B., Arevalo, L. & Campbell, P. A high quality genome for *mus spicilegus*, a close relative of house mice with unique social and ecological adaptations. *G3* **8**, 2145–2152 (2018).
75. Pendleton, A. L. et al. Comparison of village dog and wolf genomes highlights the role of the neural crest in dog domestication. *BMC Biol.* **16**, 64 (2018).
76. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
77. Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403 (2004).
78. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
79. Gaysinskaya, V., Soh, I. Y., van der Heijden, G. W. & Bortvin, A. Optimized flow cytometry isolation of murine spermatocytes. *Cytom. A* **85**, 556–565 (2014).
80. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
81. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
82. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* 2nd edn. *Use R!* (Springer International Publishing: Imprint: Springer, 2016).
83. Martin, C. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, <https://doi.org/10.14806/ej.17.1.200> (2011).
84. Andrews, S., Gilley, J. & Coleman, M. P. Difference Tracker: ImageJ plugins for fully automated analysis of multiple axonal transport parameters. *J. Neurosci. Methods* **193**, 281–287 (2010).
85. Wingett, S. W. & Andrews, S. FastQ Screen: a tool for multi-genome mapping and quality control. *F1000Res* **7**, 1338 (2018).

86. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
87. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
88. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
89. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
90. Perteza, G. & Perteza, M. GFF utilities: GffRead and GffCompare. *F1000Research* **9**, <https://doi.org/10.12688/f1000research.23297.2> (2020).
91. Gasteiger, E. et al. ExpASY: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788 (2003).
92. Truett, G. E. et al. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques* **29**, 52–54 (2000).
93. Lesch, B. J., Silber, S. J., McCarrey, J. R. & Page, D. C. Parallel evolution of male germline epigenetic poisoning and somatic development in animals. *Nat. Genet.* **48**, 888–894 (2016).

Acknowledgements

The authors want to acknowledge the Genome Technology Access Center at the McDonnell Genome Institute at Washington University School of Medicine for assistance with genomic analysis, the University of Michigan Flow Cytometry Shared Resource Laboratory for FACS, the Advanced Genomics Core for RNA Sequencing, and the Transgenic Animal Model Core for generating mouse deletion lines. The authors are grateful to Christian Rohrandt for his invaluable input on nanopore data analysis and preprocessing. The authors thank the MPIMG IT and Seq-Core facility for Mate-Pair and PacBio library preparation and sequencing, particularly for the support from Thomas Kreitler. The authors thank M. Meisler for sharing *Ella-Cre* mice, T. Keane for sharing a draft chr17 *M. spretus* assembly, D. Tautz, L. Odenthal-Hesse, and K. Ulrich for sharing *M. spicilegus* and *M. spretus* tissue and a *M. spicilegus* chr17 assembly, and D.W. Bellott for technical advice. The authors thank M. Arlt, D.W. Bellott, E. Clowney, J. Kidd, A. Lawson, J. Moran, T. Wilson, and Y. Yamashita for their comments. This work was supported by the National Institutes of Health grants HD094736 (J.L.M.), HD104339 (C.M.S.), HD079342 (C.M.S.) and GM007544 (C.M.S.), National Science Foundation grant 1941796 (J.L.M.), the Max Planck Society (G.W., H.B., P.T., and B.G.H.), BMBF (IntraEpiGliom, FKZ 13GW0347C; G.W. and F.J.M.), BMBF (P4D, FKZ 01EK2204C; F.J.M.), Deutsche Forschungsgemeinschaft (German Research Foundation) EXC 22167-390884018 and DFG CRC-1665 –515637292 (B.B. and F.J.M.).

Author contributions

C.M.S. performed overall project coordination, experimental and computational analyses, and figure generation. G.W. performed long-read

(Oxford Nanopore and PacBio HiFi) data processing, integrated long-read and Hi-C genome assemblies, and manually curated the t^{w5} region. L.Z. performed ddPCR and genotyping for transmission distortion assays. B.B. prepared DNA for ultra-long Nanopore and PacBio HiFi sequencing. H.B. generated the t^{w5}/t^{w5} mouse embryonic stem cell (mESC) line. P.T. conducted Hi-C analysis from t^{w5}/t^{w5} mESCs. F.J.M., B.G.H., and J.L.M. provided project leadership and coordination. C.M.S. and J.L.M. wrote the manuscript with contributions from other authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-66616-9>.

Correspondence and requests for materials should be addressed to Jacob L. Mueller.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025