Published in final edited form as:

Mamm Genome. 2012 August ; 23(7-8): 454–466. doi:10.1007/s00335-012-9403-5.

A pronounced evolutionary shift of the pseudoautosomal region boundary in house mice

Michael A. White^{1,2}, Akihiro Ikeda¹, and Bret A. Payseur¹

¹Laboratory of Genetics, 425-G Henry Mall, University of Wisconsin, Madison, WI 53706

Abstract

The pseudoautosomal region (PAR) is essential for the accurate pairing and segregation of the X and Y chromosomes during meiosis. Despite its functional significance, the PAR shows substantial evolutionary divergence in structure and sequence between mammalian species. An instructive example of PAR evolution is the house mouse *Mus musculus domesticus* (represented by the C57BL/6J strain), which has the smallest PAR among those that have been mapped. In C57BL/6J, the PAR boundary is located just ~700 kb from the distal end of the X chromosome, whereas the boundary is found at a more proximal position in *Mus spretus*, a species that diverged from house mice 2–4 million years ago. Here, we use a combination of genetic and physical mapping to document a pronounced shift in the PAR boundary in a second house mouse subspecies, *Mus musculus castaneus* (represented by the CAST/EiJ strain), ~430 kb proximal of the *M. m. domesticus* boundary. We demonstrate molecular evolutionary consequences of this shift, including a marked lineage-specific increase in sequence divergence within *Mid1*, a gene that resides entirely within the *M. m. castaneus* PAR but straddles the boundary in other subspecies. Our results extend observations of structural divergence in the PAR to closely related subspecies, pointing to major evolutionary changes in this functionally important genomic region over a short time period.

Keywords

pseudoautosomal region; house mouse; *Mid1*

Introduction

Proper pairing and segregation of homologous chromosomes during meiosis is essential for the production of viable gametes. Crossing over physically links homologous chromosomes, helping to ensure the fidelity of this process. Although autosomes may recombine throughout their lengths, the X and Y chromosomes are mostly distinct, lacking the sequence similarity required for exchanging material. For the X and Y chromosomes, recombination is restricted to the pseudoautosomal region (PAR), a short stretch of sequence homology (Burgoyne 1982; Cooke et al. 1985; Simmler et al. 1985) where an obligate crossover occurs during meiosis (Keitges et al. 1985; Rouyer et al. 1986; Soriano et al. 1987). Inhibition of recombination in the PAR can cause sex chromosome aneuploidy (Shi et al. 2001) and meiotic arrest (Gabriel-Robez et al. 1990; Burgoyne et al. 1992; Mohandas et al. 1992).

Corresponding Author: Michael A. White, mawhite@fhcrc.org.

²Present Address: Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98102

Despite its central role in pairing and segregation of the sex chromosomes, the PAR evolves rapidly. Compared to the rest of the genome, PAR sequence divergence is highly elevated between closely related species (Salido et al. 1996; Perry and Ashworth 1999; Schiebel et al. 2000; Filatov and Gerrard 2003; Montoya-Burgos et al. 2003; Huang et al. 2005; Bussell et al. 2006; Kasahara et al. 2010) and even among individuals of the same species (Schiebel et al. 2000). The PAR also varies in size in a manner consistent with progressive degradation over evolutionary time (Graves 1995; Graves et al. 1998; Lahn and Page 1999; Skaletsky et al. 2003; Bergero and Charlesworth 2009). This pattern is most clearly documented in eutherian mammals, where considerable variation in PAR size has been found (Perry et al. 2001; Ross et al. 2005; Raudsepp and Chowdhary 2008; Young et al. 2008; Das et al. 2009). In all cases, the PAR boundaries have migrated from an ancestral location at the *amelogenin* loci (Iwase et al. 2003; Iwase et al. 2007) to various distal locations on the X chromosome.

Comparisons among taxa with known PAR boundaries have revealed little variation between closely related species. The PAR boundaries of orangutans, gorillas, chimpanzees, and humans are identical (Ellis et al. 1990), despite being separated by 4.1 million years of evolution (Hobolth et al. 2007). PAR boundaries are also identical across ruminants, spanning ~20.7 million years of evolution (Van Laere et al. 2008). The identification of recently diverged species or subspecies that differ in PAR boundary position would provide a much-needed framework for understanding the evolutionary processes responsible for variation in this important genomic region.

House mouse subspecies (*Mus musculus*) provide a unique system for studying variation in the PAR and examining its functional consequences. Across species with mapped boundaries, the smallest PAR is found in house mice, where exchange between the X and Y chromosomes is restricted to a ~700 kb region on the distal end of the X chromosome (Perry et al. 2001). This large boundary shift is thought to have occurred during the last 2 – 4 million years (Perry and Ashworth 1999) since the common ancestor of house mice split from *Mus spretus* (She et al. 1990; Suzuki et al. 2004). The current house mouse PAR boundary maps to intron three of the *Midline 1* gene (*Mid1*; synonyms: *Fxy*, *Trim18*) (Palmer et al. 1997), whereas in *M. spretus*, rat, and humans, *Mid1* is entirely X-linked (Dal Zotto et al. 1998; Perry and Ashworth 1999; Perry et al. 2001). This substantial change may be the result of a translocation from the ancestral *Mid1* position to the current location spanning the PAR boundary (Galtier 2004). Some intrasubspecific variation in PAR boundary position has been found among classical inbred mouse strains (primarily *M. m. domesticus* in origin) and has been associated with large repetitive DNA elements (Kipling et al. 1996a; Kipling et al. 1996b). Nevertheless, it remains unclear whether the boundary varies among subspecies.

We use a combination of genetic and physical mapping to document a dramatic shift of close to 430 kb in the PAR boundary of *M. m. castaneus* relative to other subspecies of house mice, despite the short divergence time separating these lineages. We describe molecular evolutionary consequences of this boundary shift, including a localized increase in sequence divergence between *M. m. castaneus* and other mice at a gene near the *M. m. domesticus* PAR boundary. We propose that the *M. m. castaneus* PAR boundary is evolutionarily derived, and discuss the implications of a highly variable PAR for male fertility in house mice.

Materials & Methods

Genetic Mapping of the PAR Boundary

Three wild-derived inbred strains purchased from The Jackson Laboratory (www.jax.org), *Mus musculus castaneus* (CAST/EiJ), *M. m. domesticus* (WSB/EiJ), and *M. m. musculus*

(PWD/PhJ) were used to map the PAR boundary. Parents were crossed in reciprocal directions to generate F1 hybrids (*M. m. musculus*^{PWD} × *M. m. castaneus*^{CAST}; *M. m. castaneus*^{CAST} × *M. m. musculus*^{PWD}; *M. m. domesticus*^{WSB} × *M. m. castaneus*^{CAST}; and *M. m. castaneus*^{CAST} × *M. m. domesticus*^{WSB}). Mice were housed in the University of Wisconsin School of Medicine and Public Health mouse facility according to animal care protocols approved by the University of Wisconsin Animal Care and Use Committee.

The *M. m. castaneus*^{CAST} boundary was localized by amplifying ~1 kb fragments spaced regularly throughout the distal end of the X chromosome in *M. m. domesticus*^{WSB} males, *M. m. castaneus*^{CAST} males, (*M. m. castaneus*^{CAST} × *M. m. domesticus*^{WSB})F1 males, and (*M. m. domesticus*^{WSB} × *M. m. castaneus*^{CAST})F1 males using polymerase chain reaction (PCR) amplification (primers in Table S1). Genomic DNA was extracted from liver with the Wizard Genomic DNA Purification Kit (Promega) following protocols recommended by the manufacturer. Each fragment was Sanger sequenced (using the PCR primers as sequencing primers) to identify all polymorphisms in the interval. Heterozygous polymorphisms in (*M. m. domesticus*^{WSB} × *M. m. castaneus*^{CAST})F1 males denoted an interval within the pseudoautosomal region (PAR), where homologous sequences from the *M. m. domesticus*^{WSB} X chromosome and the *M. m. castaneus*^{CAST} Y chromosome were amplified. Intervals outside of the PAR were hemizygous, only amplifying from the *M. m. domesticus*^{WSB} X chromosome.

The PAR boundaries of *M. m. musculus*^{PWD} and *M. m. domesticus*^{WSB} were mapped to 10 kb intervals by comparing against large deletions present in introns of *Mid1* on the distal end of chromosome X in *M. m. castaneus*^{CAST}. The deletions were inferred from failed PCR amplification in *M. m. castaneus*^{CAST} despite successful amplification in *M. m. domesticus*^{WSB} and *M. m. musculus*^{PWD} using identical primers. The deletions were also visible in the recently completed *M. m. castaneus*^{CAST} genome sequence (Keane et al. 2011) (<http://www.sanger.ac.uk/resources/mouse/genomes/>) as a complete absence of reads. Fragments within the deleted regions were PCR amplified in *M. m. domesticus*^{WSB}, *M. m. musculus*^{PWD}, and the F1 hybrids (primers in Table S1); accurate amplification was verified by Sanger sequencing. The PAR was identified when there was positive amplification in F1 hybrids with *M. m. castaneus*^{CAST} as the female parent. Amplification in these cases was from the PAR of the *M. m. domesticus*^{WSB} or *M. m. musculus*^{PWD} Y chromosome.

Physical Mapping of the *M. m. castaneus* PAR Boundary

A bacterial artificial chromosome (BAC) library of *M. m. castaneus*^{CAST} (CHORI-26; Children's Hospital Oakland Research Institute) was used to identify the sequence position of the PAR boundary. The BAC library was screened with a 451 bp DNA probe that localized to the approximate PAR boundary identified from the F1 hybrids (166,054,512 bp - 166,054,962 bp NCBI build 37.2 of the mouse genome; forward primer: 5' TGCCCATTTGTGTCCACTGATGCC 3', reverse primer: 5' CCCACAATTGCTTTTCACACACAAC 3'). Approximately 25 ng of the probe was labeled with [P32] dCTP (Perkin Elmer) using the Rediprime II Random Prime Labeling System (Amersham, GE Healthcare). The membranes were prehybridized for 30 minutes in hybridization solution (4X SSC, 1% milk powder, 1% SDS, 10X Denhardt's Reagent, 37 µg/ml salmon sperm DNA; Sigma) at 65°C, followed by hybridization with the probe overnight at 65°C. After hybridization, the membranes were washed twice in wash buffer I (2X SSC, 0.1% SDS) at 65°C for 30 minutes, once in wash buffer II (0.5X SSC, 0.1% SDS) at 65°C for 15 minutes, and visualized with Hyblot CL film (Denville Scientific). Five clones were identified (CH26-362J9, CH26-287P17, CH26-273M10, CH26-467K7, and CH26-462C12) and were end-sequenced with T7 and SP6 primers. The end sequences were mapped to the C57BL/6J genome using the UCSC BLAST-like alignment tool (BLAT; Kent 2002). The end sequences were deposited in GenBank under accession numbers JM426711-JM426720.

One clone from the X chromosome (CH26-467K7) and one clone from the Y chromosome (CH26-462C12) were shotgun sequenced using the TOPO Shotgun Subcloning Kit (Invitrogen). BAC colonies were grown overnight in 100 ml of LB medium and DNA was isolated using the Qiagen Plasmid Midi Kit with the standard protocol except the volumes of the lysis reagents were doubled to increase the efficiency of isolating low-copy plasmids. Approximately 10 µg of BAC DNA was sheared in a nebulizer (Invitrogen) with compressed nitrogen (10 psi, 60 seconds). About 2 µg of sheared DNA was used in subsequent reactions. Positive colonies were verified for inserts by colony PCR, amplifying with M13 forward and reverse primers. Inserts from 342 colonies from the Y chromosome clone and 66 colonies from the X chromosome clone were Sanger sequenced directly from the colony PCR amplification by diluting the reaction 1:20, using 1 µl as template, and sequencing with M13 forward or reverse primers. The shotgun sequences were deposited in GenBank under accession numbers JN650207-JN650208.

BAC shotgun sequence reads were mapped to the mouse genome reference sequence (NCBI build 37.2) using BLAT with default parameters. All reads were trimmed from the 3' end for a maximum sequence length of 1 kb. The highest scored BLAT alignment was chosen for each BAC read. Y chromosome BAC reads ended at a more distal location on the mouse reference X chromosome than the X chromosome BAC reads, localizing the PAR boundary to a narrow region. From this location, the X and Y chromosome BACs were sequenced at regular intervals, walking proximally until sequence identity between the two chromosomes was lost. The PAR boundary was verified with multiple primers, which sequenced in both directions across the boundary (primers in Table S1). The mouse reference genome was searched for homologous sequence with 500 bp from either side of the PAR boundary using BLAT. These sequences were also searched for interspersed repeats and low-complexity DNA sequences using RepeatMasker (A.F.A Smit, R. Hubley, P. Green, <http://repeatmasker.org>).

Molecular Evolution of the PAR

Alignments between C57BL/6J and rat from X chromosome positions 165,980,014 – 166,428,730 were extracted from whole-genome alignments between the two species (Keane et al. 2011). Consensus sequences of *M. m. castaneus*^{CAST}, *M. m. domesticus*^{WSB}, *M. m. spretus* (SPRET/EiJ), and a second strain of *M. m. musculus* (PWK/PhJ) (the *M. m. musculus* strain used to map the PAR boundary did not have genome sequence available) were added to the alignment (Keane et al. 2011). Missing sequences in the house mouse subspecies were filled with N's. Some exonic sequences were missing from *M. m. castaneus* and *M. m. spretus* in the next-generation sequencing (Keane et al. 2011). Using Sanger sequencing, these regions were found to be false negatives and were filled in for *M. m. castaneus*. Missing sequence in *M. m. spretus* was filled using a previously sequenced coding region (GenBank accession number AF186460). The first phylogenetic analysis of *Mid1* contained all exonic and intronic sequence from the start of exon 2 (the translation start site) through the end of exon 10. To investigate whether molecular evolutionary effects of PAR movement were visible in other strains of *M. m. castaneus*, we sequenced exons 2, 5, 6, and 7 from *Mid1* in the *M. m. castaneus* (CIM) strain established from mice in India (provided by Francois Bonhomme, University of Montpellier). The *M. m. castaneus*^{CIM} exon sequences were deposited in GenBank under accession number JN651093. A second phylogenetic analysis involving *M. m. castaneus*^{CIM} was conducted. This analysis was restricted to the four sequenced exons of *Mid1* (only exons 2, 5, 6, and 7 were sequenced from CIM; primers in Table S1). The best fitting model of molecular evolution was selected using MrModelTest (Nylander 2004; <http://www.abc.se/~nylander/mrmodeltest2/mrmodeltest2.html>). The highest scoring model was selected based upon Akaike's information criterion (Posada and Buckley 2004) (GTR+Gamma for the phylogenetic

analysis without CIM and GTR for the phylogenetic analysis with CIM). The phylogenetic tree was estimated using Mr.Bayes (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) with four Markov chains running for 2,000,000 generations (two simultaneous runs). The first 25% of trees were discarded as burn-in. Topology and branch length priors were left at default settings.

Pairwise sequence divergence between each house mouse subspecies and *M. spretus* was examined across the region in 10 kb windows sliding at 1 kb increments. Distance was estimated in each window using PAUP* version 4.0b10 (Swofford 2002) with an HKY85 model (Hasegawa et al. 1985), which accounts for variable rates of transitions and transversions as well as unequal base frequencies. Distances followed the same patterns across the pseudoautosomal region when alternative models of molecular evolution were used. G/C nucleotide content was quantified in 50 kb windows from the *M. m. castaneus*^{CAST} PAR boundary to the transcription start site of *Mid1*.

The molecular evolution of *Mid1* was measured in both introns and exons. Intron lengths were inferred for each strain by calculating the total length of consensus sequences from the next-generation sequencing that mapped to the mouse reference genome (Wellcome Trust Sanger Institute; <http://www.sanger.ac.uk/resources/mouse/genomes/>). Missing sequencing reads were assumed to be deletions relative to the other subspecies. To verify this assumption, PCR fragments were amplified across the two introns with largest deletions (*M. m. castaneus*: introns 2 and 3) and compared to the inferred intron length (primers in Table S1). Amplification of the correct interval was confirmed by Sanger sequencing the PCR product (using the PCR amplification primers as sequencing primers). High levels of sequence divergence present in the next-generation sequencing of *M. m. castaneus*^{CAST} were verified by Sanger sequencing in a subset of the ten exons (exons 2, 5, 6, 7; primers in Table S1). Rates of nonsynonymous and synonymous divergence in exons of *Mid1* were calculated using MEGA version 5.0 (Tamura et al. 2007) under a Pamilo-Bianchi-Li model (Pamilo and Bianchi 1993) that allows for different substitution rates for transitions and transversions. GC-biased molecular evolution in the exons was quantified by calculating the proportion of substitutions from *M. spretus* that resulted in a G or C nucleotide (using *M. spretus* as a proxy for the ancestral state). The total proportion of G or C nucleotides was also quantified within regions of *Mid1* introns that were alignable between all taxa (GC_i) as well as the total proportion of GC nucleotides in the third codon positions throughout *Mid1* exons (GC3). Degenerate nucleotides were ignored in all GC analyses.

Results

Mapping the PAR Boundary

In crosses between *M. m. domesticus*^{WSB} and *M. m. castaneus*^{CAST}, we identified genotypes suggestive of pseudoautosomal inheritance at SNPs that are entirely X-linked in C57BL/6J (Perry et al. 2001). Markers located outside the pseudoautosomal region (PAR) at 166.32 Mb and 166.36 Mb (all positions based upon NCBI build 37.2 of the mouse reference genome) were hemizygous in (*M. m. castaneus*^{CAST} × *M. m. domesticus*^{WSB})F1 males but were heterozygous in (*M. m. domesticus*^{WSB} × *M. m. castaneus*^{CAST})F1 males, indicating a proximal shift in the region of X homology found on the *M. m. castaneus*^{CAST} Y chromosome. To determine the extent of the shifted PAR boundary in *M. m. castaneus*^{CAST}, we sequenced ~1 kb fragments moving proximal along the X chromosome (Figure 1). Every interval was heterozygous in (*M. m. domesticus*^{WSB} × *M. m. castaneus*^{CAST})F1 males until approximately 166.00 Mb, with the exception of one small region at 166.05 Mb. After 166.00 Mb, all intervals were hemizygous for the *M. m. domesticus*^{WSB} X chromosome, indicating non-PAR sequence. In the reciprocal F1s that lacked an extended PAR, all regions were hemizygous, matching the X chromosome

sequence of *M. m. castaneus*^{CAST}. Additionally, polymorphisms between the X and Y chromosomes increased in frequency towards 166.00 Mb in the *M. m. castaneus*^{CAST} inbred male parent, a pattern observed at PAR boundaries of other species (Van Laere et al. 2008). These results localized the PAR of *M. m. castaneus*^{CAST} to approximately 166.00 Mb.

To fine-map the PAR boundary, we isolated clones from a *M. m. castaneus*^{CAST} bacterial artificial chromosome library (BAC) that screened positive for a DNA probe within the PAR, near the boundary (~166.06 Mb). We end-sequenced five positive clones and mapped these to the mouse reference genome (C57BL/6J; NCBI build 37.2) using the UCSC BLAST-like alignment tool (BLAT) (Kent 2002). Four clones originated from the X chromosome, with end-sequences that shared greater than 95.8% sequence identity with the X chromosome. A fifth clone spanned the PAR boundary on the Y chromosome. One end of this clone mapped to 166.07 Mb of the mouse reference genome (98.2% sequence identity) whereas the other end had 100% sequence identity with unmapped Y chromosome sequence.

We shotgun sequenced the Y chromosome clone and generated 305.37 kb of sequence (342 reads) that was mostly homologous to the X and Y chromosomes of the reference mouse genome. All contigs that showed sequence identity with the X chromosome (140.31 kb of sequence, 152 reads) mapped between 165.98 and 166.07 Mb of the reference mouse genome (Figure 1). The other contigs had the highest sequence similarity to unmapped segments of the Y chromosome (157.52 kb of sequence, 179 reads) or to a small proportion of the autosomes (7.54 kb of sequence, 11 reads). In comparison, shotgun sequencing reads from one of the X chromosome clones spanned both sides of the boundary (65.70 kb of sequence, 66 reads; Figure 1). The end of shotgun sequencing reads from the Y chromosome clone at 165.98 Mb, in conjunction with the switch from heterozygous to hemizygous genotypes in F1 hybrids (at 166.00 Mb), localized the *M. m. castaneus*^{CAST} PAR boundary to 165.98 Mb, 430 kb proximal of the C57BL/6J boundary (Figure 2) (Perry et al. 2001).

To precisely position the PAR boundary, we used chromosome walking in the Y chromosome clone and in one X chromosome clone from the BAC screening. We sequenced from the most proximal X chromosome shotgun sequencing read and found an abrupt transition from nearly complete X and Y chromosome sequence homology to a complete absence of homology at position 165,980,014 of the reference mouse genome (Figure 3). We used BLAT to identify regions of the mouse reference genome sequence that matched the 250 bp of non-homology between the X and Y chromosomes. As expected, the X chromosome sequence mapped to a position beginning at 165,980,013 in the reference mouse genome (99.6% homology). The Y chromosome sequence shared 100% sequence identity with unmapped Y chromosome contigs and some regions across the autosomes. We used RepeatMasker to identify any repetitive elements within the 250 bp that would result in alignment to multiple regions of the genome. The Y sequence was identical to a long interspersed repetitive element (LINE1; 99.6% sequence identity) that is present throughout the mouse genome.

M. m. castaneus^{CAST} has several large deletions within the introns of *Mid1*. We used these deletions to map approximate PAR boundaries in *M. m. domesticus*^{WSB} and *M. m. musculus*^{PWD}. In F1 hybrids where *M. m. castaneus*^{CAST} is the maternal parent, amplification can only occur from the Y chromosome PARs in *M. m. domesticus*^{WSB} or *M. m. musculus*^{PWD} (amplification cannot occur from the deleted portion of the *M. m. castaneus*^{CAST} X chromosome). Using this logic, we assayed for amplification at regular intervals along the deletions. In both subspecies, amplification began at locations that were in close proximity of the C57BL/6J boundary (*M. m. domesticus*^{WSB}: 166.40 – 166.41 Mb; *M. m. musculus*^{PWD}: 166.39 – 166.40 Mb; Figure 4). The *M. m. musculus*^{PWD} boundary

falls within intron 2 of *Mid1*, whereas the *M. m. domesticus*^{WSB} boundary falls within intron 2, exon 3, or intron 3.

Sequence Divergence within a Larger PAR

Elevated rates of sequence evolution have been documented within the PAR and have been attributed to high rates of gene conversion (Perry and Ashworth 1999; Montoya-Burgos et al. 2003; Galtier 2004; Yi et al. 2004; Huang et al. 2005). In C57BL/6J, *Midline 1* (*Mid1*) spans the PAR boundary (Palmer et al. 1997) and exhibits substantially higher divergence in the PAR side of the gene than in the exons that reside outside of the PAR (Perry and Ashworth 1999; Montoya-Burgos et al. 2003; Huang et al. 2005). In *M. m. castaneus*^{CAST}, *Mid1* is completely encompassed by the PAR, suggesting that the entire gene may show an elevated rate of evolution. We first evaluated the phylogenetic relationship of the *M. m. castaneus*^{CAST} *Mid1* sequence to that in other house mice, *M. spretus*, and rat (all exonic and intronic sequence from exon 2 – exon 10). We recovered strong support for a phylogeny placing *M. spretus* and rat as outgroups to house mice. Within house mice, *M. m. castaneus*^{CAST} was sister to *M. m. musculus*^{PWK}, C57BL/6J, and *M. m. domesticus*^{WSB}. Divergence between *M. m. castaneus*^{CAST} and *M. spretus* was much higher than C57BL/6J-*M. spretus* divergence, *M. m. domesticus*^{WSB}-*M. spretus* divergence or *M. m. musculus*^{PWK}-*M. spretus* divergence (Figure 5), indicating an increase in evolutionary rate in the *M. m. castaneus*^{CAST} lineage.

To determine whether evolutionary rates were elevated differentially throughout the gene, we partitioned the *Mid1* coding sequence into groups of exons and calculated synonymous and non-synonymous divergence. Compared to other house mice, *M. m. castaneus*^{CAST} had higher synonymous divergence and slightly higher non-synonymous divergence from *M. spretus* across all groups of exons (Table 1). *Mid1* still showed constrained coding evolution in *M. m. castaneus*^{CAST}, with Ka/Ks ratios less than 1, suggesting that the gene remains functional in this subspecies. Consistent with previous studies, elevated divergence in C57BL/6J was restricted to exons within the PAR (exons 4–10). This restricted pattern was also evident in *M. m. domesticus*^{WSB} and *M. m. musculus*^{PWK}, matching the approximate PAR boundaries we mapped in these subspecies.

We searched for substitution biases toward G and C as well as reductions in intron size, two additional patterns of molecular evolution that were previously associated with the PAR (Duret et al. 1995; Montoya-Burgos et al. 2003). To measure intron length in *Mid1*, we quantified the total length of sequencing reads that were mapped to the C57BL/6J reference sequence in each subspecies. The total read lengths of each intron for *M. m. domesticus*^{WSB} and *M. m. musculus*^{PWK} closely matched those of C57BL/6J (Table 2), with the smallest introns in the PAR (introns 4–10). In *M. m. castaneus*^{CAST}, all introns showed some reduction in size; the most striking reductions occurred in introns two, three, and six where greater than 80% of the introns were missing. The total length of mapped reads in introns may be biased if high levels of divergence prevent alignment to the reference genome. We verified the accuracy of the next-generation sequencing by PCR amplifying sequences across introns two and three of *M. m. castaneus*^{CAST}. Both PCR-amplified introns closely matched the estimates from next-generation sequencing (intron two: ~4.5 kb PCR, 6.7 kb next-generation; intron three: ~3.2 kb PCR, 1.9 kb next-generation). We quantified a bias toward G and C substitutions in the coding region of *Mid1*. In all pairwise comparisons with *M. spretus*, nearly every substitution was a base change to G or C (*M. m. castaneus*^{CAST}/*M. spretus*: 240/241; *M. m. musculus*^{PWK}/*M. spretus*: 131/132; *M. m. domesticus*^{WSB}/*M. spretus*: 128/129). Furthermore, in *M. m. castaneus* the proportion of G and C nucleotides was elevated throughout *Mid1* in third codon positions (GC3) and within introns (GCi) downstream of the transcription start site (Table S2; Table S3), where the gene is fully encompassed by the PAR. These patterns were restricted to the 3' end of the gene in the

other subspecies, identical to the patterns observed in C57BL/6J (Perry and Ashworth 1999; Montoya-Burgos et al. 2003; Galtier 2004; Huang et al. 2005).

To determine whether elevated sequence divergence persists through the entire *M. m. castaneus*^{CAST} PAR, we conducted a sliding window analysis of pairwise divergence from the PAR boundary (position 165,890,014) through the end of *Mid1* (position 166,428,730). Divergence was indistinguishable in the three subspecies comparisons with *M. spretus* until the start of the *Mid1* coding sequence (Figure 6). Additionally, we examined GC-biased substitution in the larger *M. m. castaneus*^{CAST} PAR in 50 kb increments across the region. There were no clear differences in GC substitution bias across the intervals, with proportions ranging from 0.49–0.54. There were low levels of missing sequencing reads, opposite of what was observed in the *Mid1* introns, in the region proximal to the *Mid1* transcription start site (143 kb of total sequence; *M. m. domesticus*^{WSB}: 481 bp missing; *M. m. musculus*^{PWK}: 28 bp missing; *M. m. castaneus*^{CAST}: 1,191 bp missing; *M. spretus*: 6,116 bp missing). The larger PAR in *M. m. castaneus*^{CAST} seems to have influenced the molecular evolution of the *Mid1* gene, but not the PAR sequence proximal to it.

To evaluate whether the high sequence divergence we observed in *M. m. castaneus*^{CAST} was representative of the *M. m. castaneus* subspecies, we sequenced four exons in a second strain of *M. m. castaneus* (CIM) from India. We found similarly high sequence divergence in the coding region of *Mid1* (Figure 7). Divergence was again elevated at both the 3' ends and 5' ends of the gene. Additionally, we found a comparable reduction in the length of introns (intron one: ~3.5 kb, intron two: ~2.0 kb). Although we did not directly estimate the position of the PAR boundary in *M. m. castaneus*^{CIM}, the molecular evolution of *Mid1* suggests a larger PAR may also be present in this strain. Application of a similar mapping strategy to hybrids between *M. m. castaneus*^{CIM} and *M. m. domesticus*^{WSB} would clarify the boundary position.

Discussion

Sequence Divergence within the PAR

Several distinct patterns of molecular evolution have been observed within PARs, including high rates of nucleotide divergence, GC bias in nucleotide substitutions, and reductions in intron size (Salido et al. 1996; Perry and Ashworth 1999; Schiebel et al. 2000; Filatov and Gerrard 2003; Montoya-Burgos et al. 2003; Huang et al. 2005; Bussell et al. 2006; Kasahara et al. 2010). These patterns have been connected to high rates of recombination within the PAR (stemming from the obligate crossover that occurs in the region during male meiosis) and the biased fixation of GC alleles from recombination-associated gene conversion (Strathern et al. 1995; Rattray et al. 2001; Marais 2003; Duret and Galtier 2009). All three of the genes known to be located in the PAR in house mice show substantial divergence from rat and humans, biases toward G and C substitutions, and reductions in intron size (*Sts* (Salido et al. 1996; Filatov and Gerrard 2003; Kasahara et al. 2010), *Mid1* (Perry and Ashworth 1999; Montoya-Burgos et al. 2003; Galtier 2004; Huang et al. 2005), and *Asmt* (Kasahara et al. 2010)). Importantly, *Mid1* only exhibits high divergence at the 3' end of the gene that resides within the PAR in C57BL/6J. In contrast, we demonstrated high divergence throughout the entire coding sequence of the gene in *M. m. castaneus*^{CAST}, similar to the level of divergence separating rat and *Mus*. In addition, pairwise divergence between *M. m. castaneus*^{CAST} and *M. spretus* is higher than all other subspecies comparisons with *M. spretus*. Combined, these molecular signatures suggest the entire *Mid1* gene of *M. m. castaneus*^{CAST} has resided within the PAR for a longer period of time, has been subject to higher rates of evolution, or both.

The high sequence divergence within *Mid1* may instead be attributed to duplications of *Mid1* exons within *M. m. castaneus*^{CAST}. Each duplicated exon would accumulate independent substitutions, increasing the apparent level of divergence if sequenced together. Data from C57BL/6J suggest that *Mid1* is subject to frequent spontaneous duplications caused by a high rate of unequal crossovers (Dal Zotto et al. 1998). However, if duplications were driving sequence divergence, we would expect to see high heterozygosity in the Sanger sequences. Instead, most substitutions were clearly homozygous in both strains of *M. m. castaneus* (CAST and CIM; data not shown), suggesting that the elevated divergence was not caused by comparing paralogous genes.

Despite elevated divergence in *Mid1*, the intergenic sequence proximal to the gene resembled that of non-PAR divergence from other house mice. In other species, GC substitution bias decreases towards the PAR boundary (Huang et al. 2005; Bussell et al. 2006; Chen et al. 2006; Raudsepp and Chowdhary 2008). In *M. m. castaneus*^{CAST}, we observed an abrupt decline in GC substitution bias and nucleotide divergence immediately proximal to the coding sequence of *Mid1*, with levels remaining low throughout the 385 kb to the PAR boundary. Furthermore, although it is difficult to reliably measure indels using next-generation sequencing assemblies, there was little missing sequence proximal to the *Mid1* transcription start site. The dichotomy in sequence divergence between the proximal and distal segments of the *M. m. castaneus*^{CAST} PAR suggests the proximal segment is a recent translocation of X chromosome sequence to the Y chromosome. The second PAR in humans was formed by recombination between LINE elements (Kvaløy et al. 1994) that led to translocations between non-homologous regions of the X and Y (Freije et al. 1992; Charchar et al. 2003). An alternative possibility is that the proximal PAR in *M. m. castaneus*^{CAST} is a region of ancestral homology between the X and Y chromosome and sequence divergence has been suppressed. However, this scenario seems unlikely based upon the abrupt shift in sequence divergence between the proximal and distal halves of the *M. m. castaneus*^{CAST} PAR. GC content would likely remain high in the proximal half of the PAR after the boundary shift in *M. m. musculus* and *M. m. domesticus* if the region is ancestral. In humans, GC content remains high throughout the entire *XG* gene, even though the proximal portion is no longer within the PAR (Galtier 2004).

Evolution of the House Mouse PAR Boundary

The PAR boundary of house mice is thought to have originated from a single rearrangement of *Mid1* from an ancestral, X-linked location to the current position spanning the PAR boundary of C57BL/6J (Perry and Ashworth 1999; Perry et al. 2001; Galtier 2004). We describe similar PAR boundaries within *Mid1* in another representative of *M. m. domesticus* (WSB/EiJ), and in a different subspecies of house mouse, *M. m. musculus* (PWD/EiJ). Assuming a subspecies tree that places *M. m. musculus* and *M. m. castaneus* as sister taxa (White et al. 2009; Keane et al. 2011), the most parsimonious evolutionary scenario is that the ancestral PAR boundary was located within *Mid1*. Following this rearrangement, there was an independent shift of the PAR boundary to a more proximal location in the *M. m. castaneus* lineage (Figure 8). Under this scenario, the *Mid1* gene would be positioned further from the PAR boundary in *M. m. castaneus*^{CAST} where recombination rates would be higher (Burgoyne 1982; Rouyer et al. 1986). The accompanying higher rate of gene conversion would result in higher sequence divergence within *Mid1*. Consistent with this model, we observed a dramatic lineage-specific increase in the evolution of *Mid1* in two independent strains of *M. m. castaneus* from different geographic regions (Thailand and India). Evidence for this model could be evaluated further by mapping the boundary position in a more closely related ancestor than *M. spretus*, such as *Mus spicilegus*.

Alternatively, variation in PAR boundaries among strains may reflect ancestral polymorphism rather than subspecies differences. House mouse subspecies exhibit high

levels of incomplete lineage sorting across the genome (Geraldes et al. 2008; White et al. 2009; Keane et al. 2011), consistent with a recent divergence time of ~500,000 generations (She et al. 1990; Boursot et al. 1996; Suzuki et al. 2004; Salcedo et al. 2007; Geraldes et al. 2008). Polymorphism surveys of more individuals from each subspecies are needed to determine the nature of the PAR boundary differences documented here.

Implications of PAR Variation in House Mice

Variation in PAR structure has been linked to meiotic failures and reproductive isolation in hybrids between *M. spretus* and *M. m. domesticus* (C57BL/6J). Heterozygosity in the PAR causes premature dissociation of the sex chromosomes and sterility in F1 males (Guénet et al. 1990; Matsuda et al. 1991; Matsuda et al. 1992; Hale et al. 1993). Although the boundary structure has diverged substantially in the two species (Palmer et al. 1995; Rugarli et al. 1995; Dal Zotto et al. 1998; Perry and Ashworth 1999), the precise boundary and overall size of the PAR has not been identified in *M. spretus*. Furthermore, the exact structural variant responsible for meiotic arrest has not been determined. The threshold of divergence needed to reduce the efficacy of XY chromosome pairing and to trigger meiotic arrest remains unknown.

The *M. m. castaneus*^{CAST} PAR exhibits multiple types of structural variation. In addition to an alternative PAR boundary, we detected over 50 kb of sequence that was deleted from the introns of *Mid1* relative to *M. m. domesticus*. Together, there has been divergence in the location and amount of homology from *M. m. domesticus* available for XY pairing. Despite high divergence, F1 sterility has not been observed in multiple crosses with this strain. *M. m. castaneus*^{CAST} has frequently been used for genetic mapping in crosses with classical inbred strains of mice (mostly of *M. m. domesticus* origin; e.g. Janaswami et al. 1997; Anunciado et al. 2000; Ishikawa et al. 2000; Lyons et al. 2003; Lyons et al. 2004; Yi et al. 2006). Backcross and intercross populations have been regularly established, indicating the F1 hybrids between these two subspecies are effectively fertile. In most crosses, reproductive parameters have not been examined, making it difficult to determine whether subfertility exists. In the few instances where meiotic phenotypes were quantified, high levels of premature XY dissociation were noted in hybrids with *M. m. castaneus* (Matsuda et al. 1982), including the strain used in this study (CAST/EiJ; White et al. 2012). These crosses also do not address whether sterility occurs in generations beyond the F1. In F2 hybrids between *M. m. castaneus*^{CAST} and *M. m. domesticus* (WSB/EiJ), hybrid male sterility was strongly associated with the PAR (White et al. 2012). Sterility in these cases could result from unequal crossing over between structurally different PARs (Harbers et al. 1986; Kipling et al. 1996a; Kipling et al. 1996b; Dal Zotto et al. 1998), resulting in duplications or deletions in the region.

Variation in the PAR boundary may also have consequences for the expression of a key developmental patterning gene, *Mid1*. *Mid1* is broadly expressed throughout mouse development (Quaderi et al. 1997; Dal Zotto et al. 1998) and has conserved roles in midline patterning in humans (Gaudenz et al. 1998) and mice (Lancioni et al. 2010). In C57BL/6J, the first three 5' exons are missing on the Y chromosome and appear to be transcribed exclusively from the X chromosome (Palmer et al. 1997). In *M. m. castaneus*^{CAST}, all exons are present on the X and Y chromosomes. This raises the possibility that *Mid1* can be transcribed from both chromosomes, potentially doubling gene dosage in males. Furthermore, recombination in F1 hybrids between *M. m. castaneus*^{CAST} and other subspecies would create *Mid1* transcripts that contain a subset of highly diverged *M. m. castaneus*^{CAST} exons. *Mid1* is known to have a large number of tissue-specific splice variants (Winter et al. 2004). Combining exons from different subspecies may result in transcripts with altered tissue expression. *M. m. castaneus*^{CAST} parents and hybrids provide

useful resources to explore how *Mid1* regulation is affected by the evolution of PAR boundary position.

Acknowledgments

We thank Beth Dumont for useful discussions on meiosis and recombination. We thank Francois Bonhomme and Annie Orth for providing the CIM strain. This research was funded by NSF Grant DEB 0918000. M.A.W. was supported by an NLM training grant in Computation and Informatics in Biology and Medicine to the University of Wisconsin (NLM 2T15LM007359).

References

- Anunciado RV, Imamura T, Ohno T, Horio F, Namikawa T. Developing a new model for non-insulin dependent diabetes mellitus (NIDDM) by using the Philippine wild mouse, *Mus musculus castaneus*. *Exp Anim*. 2000; 49(1):1–8. [PubMed: 10803355]
- Bergero R, Charlesworth D. The evolution of restricted recombination in sex chromosomes. *Trends in ecology & evolution*. 2009; 24(2):94–102. [PubMed: 19100654]
- Boursot P, Din W, Anand R, Darviche D, Dod B, VonDeimling F, Talwar G, Bonhomme F. Origin and radiation of the house mouse: Mitochondrial DNA phylogeny. *J Evolution Biol*. 1996; 9(4):391–415.
- Burgoyne PS. Genetic homology and crossing over in the X and Y chromosomes of Mammals. *Hum Genet*. 1982; 61(2):85–90. [PubMed: 7129448]
- Burgoyne PS, Mahadevaiah SK, Sutcliffe MJ, Palmer SJ. Fertility in mice requires X-Y pairing and a Y-chromosomal "spermiogenesis" gene mapping to the long arm. *Cell*. 1992; 71(3):391–398. [PubMed: 1423603]
- Bussell JJ, Pearson NM, Kanda R, Filatov DA, Lahn BT. Human polymorphism and human-chimpanzee divergence in pseudoautosomal region correlate with local recombination rate. *Gene*. 2006; 368:94–100. [PubMed: 16356662]
- Charchar FJ, Svartman M, El-Mogharbel N, Ventura M, Kirby P, Matarazzo MR, Ciccodicola A, Rocchi M, Esposito M, Graves JAM. Complex events in the evolution of the human pseudoautosomal region 2 (PAR2). *Genome Research*. 2003; 13(2):281–286. [PubMed: 12566406]
- Chen J-F, Lu F, Chen S-S, Tao S-H. Significant positive correlation between the recombination rate and GC content in the human pseudoautosomal region. *Genome*. 2006; 49(5):413–419. [PubMed: 16767166]
- Cooke HJ, Brown WR, Rappold GA. Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature*. 1985; 317(6039):687–692. [PubMed: 2997619]
- Dal Zotto L, Quaderi NA, Elliott R, Lingerfelter PA, Carrel L, Valsecchi V, Montini E, Yen CH, Chapman V, Kalcheva I, et al. The mouse *Mid1* gene: implications for the pathogenesis of Opitz syndrome and the evolution of the mammalian pseudoautosomal region. *Hum Mol Genet*. 1998; 7(3):489–499. [PubMed: 9467009]
- Das PJ, Chowdhary BP, Raudsepp T. Characterization of the bovine pseudoautosomal region and comparison with sheep, goat, and other mammalian pseudoautosomal regions. *Cytogenetic and Genome Research*. 2009; 126(1–2):139–147. [PubMed: 20016163]
- Duret L, Galtier N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annual review of genomics and human genetics*. 2009; 10:285–311.
- Duret L, Mouchiroud D, Gautier C. Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. *Journal Of Molecular Evolution*. 1995; 40(3):308–317. [PubMed: 7723057]
- Ellis N, Goodfellow PN. The mammalian pseudoautosomal region. *Trends Genet*. 1989; 5(12):406–410. [PubMed: 2696184]
- Ellis N, Yen P, Neiswanger K, Shapiro LJ, Goodfellow PN. Evolution of the pseudoautosomal boundary in Old World monkeys and great apes. *Cell*. 1990; 63(5):977–986. [PubMed: 2124175]
- Filatov DA, Gerrard DT. High mutation rates in human and ape pseudoautosomal genes. *Gene*. 2003; 317(1–2):67–77. [PubMed: 14604793]

- Freije D, Helms C, Watson MS, Donis-Keller H. Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science*. 1992; 258(5089):1784–1787. [PubMed: 1465614]
- Gabriel-Robez O, Rumpler Y, Ratomponirina C, Petit C, Levilliers J, Croquette MF, Couturier J. Deletion of the pseudoautosomal region and lack of sex-chromosome pairing at pachytene in two infertile men carrying an X;Y translocation. *Cytogenetics and cell genetics*. 1990; 54(1–2):38–42. [PubMed: 2249473]
- Galtier N. Recombination, GC-content and the human pseudoautosomal boundary paradox. *Trends Genet*. 2004; 20(8):347–349. [PubMed: 15262406]
- Gaudenz K, Roessler E, Quaderi N, Franco B, Feldman G, Gasser DL, Wittwer B, Horst J, Montini E, Opitz JM, et al. Opitz G/BBB syndrome in Xp22: mutations in the MID1 gene cluster in the carboxy-terminal domain. *American Journal Of Human Genetics*. 1998; 63(3):703–710. [PubMed: 9718340]
- Geraldes A, Basset P, Gibson B, Smith KL, Harr B, Yu H-T, Bulatova N, Ziv Y, Nachman MW. Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Molecular Ecology*. 2008; 17(24):5349–5363. [PubMed: 19121002]
- Graves JA. The evolution of mammalian sex chromosomes and the origin of sex determining genes. *Philos Trans R Soc Lond, B, Biol Sci*. 1995; 350(1333):305–311. discussion 311–302. [PubMed: 8570696]
- Graves JA, Wakefield MJ, Toder R. The origin and evolution of the pseudoautosomal regions of human sex chromosomes. *Human molecular genetics*. 1998; 7(13):1991–1996. [PubMed: 9817914]
- Guénet JL, Nagamine C, Simon-Chazottes D, Montagutelli X, Bonhomme F. Hst-3: an X-linked hybrid sterility gene. *Genet Res*. 1990; 56(2–3):163–165. [PubMed: 2272506]
- Hale DW, Washburn LL, Eicher E. Meiotic abnormalities in hybrid mice of the C57BL/6J×Mus spretus cross suggest a cytogenetic basis for Haldane's rule of hybrid sterility. *Cytogenet Cell Genet*. 1993; 63(4):221–234. [PubMed: 8500353]
- Harbers K, Soriano P, Müller U, Jaenisch R. High frequency of unequal recombination in pseudoautosomal region shown by proviral insertion in transgenic mouse. *Nature*. 1986; 324(6098):682–685. [PubMed: 3025741]
- Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal Of Molecular Evolution*. 1985; 22(2):160–174. [PubMed: 3934395]
- Hobolth A, Christensen OF, Mailund T, Schierup MH. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet*. 2007; 3(2):e7. [PubMed: 17319744]
- Huang S-W, Friedman R, Yu N, Yu A, Li W-H. How strong is the mutagenicity of recombination in mammals? *Molecular Biology and Evolution*. 2005; 22(3):426–431. [PubMed: 15496551]
- Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 2001; 17(8):754–755. [PubMed: 11524383]
- Ishikawa A, Matsuda Y, Namikawa T. Detection of quantitative trait loci for body weight at 10 weeks from Philippine wild mice. *Mamm Genome*. 2000; 11(10):824–830. [PubMed: 11003694]
- Iwase M, Kaneko S, Kim H, Satta Y, Takahata N. Evolutionary history of sex-linked mammalian amelogenin genes. *Cells, tissues, organs*. 2007; 186(1):49–59. [PubMed: 17627118]
- Iwase M, Satta Y, Hirai Y, Hirai H, Imai H, Takahata N. The amelogenin loci span an ancient pseudoautosomal boundary in diverse mammalian species. *Proc Natl Acad Sci USA*. 2003; 100(9):5258–5263. [PubMed: 12672962]
- Janaswami PM, Birkenmeier EH, Cook SA, Rowe LB, Bronson RT, Davisson MT. Identification and genetic mapping of a new polycystic kidney disease on mouse chromosome 8. *Genomics*. 1997; 40(1):101–107. [PubMed: 9070925]
- Kasahara T, Abe K, Mekada K, Yoshiki A, Kato T. Genetic variation of melatonin productivity in laboratory mice under domestication. *Proc Natl Acad Sci USA*. 2010; 107(14):6412–6417. [PubMed: 20308563]
- Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*. 2011; 477(7364):289–294. [PubMed: 21921910]

- Keitges E, Rivest M, Siniscalco M, Gartler SM. X-linkage of steroid sulphatase in the mouse is evidence for a functional Y-linked allele. *Nature*. 1985; 315(6016):226–227. [PubMed: 3858674]
- Kent WJ. BLAT--the BLAST-like alignment tool. *Genome Research*. 2002; 12(4):656–664. [PubMed: 11932250]
- Kipling D, Salido EC, Shapiro LJ, Cooke HJ. High frequency de novo alterations in the long-range genomic structure of the mouse pseudoautosomal region. *Nat Genet*. 1996a; 13(1):78–80. [PubMed: 8673108]
- Kipling D, Wilson HE, Thomson EJ, Lee M, Perry J, Palmer S, Ashworth A, Cooke HJ. Structural variation of the pseudoautosomal region between and within inbred mouse strains. *Proc Natl Acad Sci USA*. 1996b; 93(1):171–175. [PubMed: 8552598]
- Kvaløy K, Galvagni F, Brown WR. The sequence organization of the long arm pseudoautosomal region of the human sex chromosomes. *Human molecular genetics*. 1994; 3(5):771–778. [PubMed: 8081364]
- Lahn BT, Page DC. Four evolutionary strata on the human X chromosome. *Science*. 1999; 286(5441):964–967. [PubMed: 10542153]
- Lancioni A, Pizzo M, Fontanella B, Ferrentino R, Napolitano LMR, De Leonibus E, Meroni G. Lack of *Mid1*, the mouse ortholog of the *Opitz syndrome* gene, causes abnormal development of the anterior cerebellar vermis. *The Journal of Neuroscience*. 2010; 30(8):2880–2887. [PubMed: 20181585]
- Lyons MA, Wittenburg H, Li R, Walsh KA, Korstanje R, Churchill GA, Carey MC, Paigen B. Quantitative trait loci that determine lipoprotein cholesterol levels in an intercross of 129S1/SvImJ and CAST/Ei inbred mice. *Physiological genomics*. 2004; 17(1):60–68. [PubMed: 14701919]
- Lyons MA, Wittenburg H, Li R, Walsh KA, Leonard MR, Korstanje R, Churchill GA, Carey MC, Paigen B. *Lith6*: a new QTL for cholesterol gallstones from an intercross of CAST/Ei and DBA/2J inbred mouse strains. *The Journal of Lipid Research*. 2003; 44(9):1763–1771.
- Maraes G. Biased gene conversion: implications for genome and sex evolution. *Trends in genetics*. 2003; 19(6):330–338. [PubMed: 12801726]
- Matsuda Y, Hirobe T, Chapman VM. Genetic basis of X-Y chromosome dissociation and male sterility in interspecific hybrids. *Proc Natl Acad Sci USA*. 1991; 88(11):4850–4854. [PubMed: 2052565]
- Matsuda Y, Imai HT, Moriwaki K, Kondo K, Bonhomme F. X-Y chromosome dissociation in wild derived *Mus musculus* subspecies, laboratory mice, and their F1 hybrids. *Cytogenet Cell Genet*. 1982; 34(3):241–252. [PubMed: 7140375]
- Matsuda Y, Moens PB, Chapman VM. Deficiency of X and Y chromosomal pairing at meiotic prophase in spermatocytes of sterile interspecific hybrids between laboratory mice (*Mus domesticus*) and *Mus spretus*. *Chromosoma*. 1992; 101(8):483–492. [PubMed: 1424992]
- Mohandas TK, Speed RM, Passage MB, Yen PH, Chandley AC, Shapiro LJ. Role of the pseudoautosomal region in sex-chromosome pairing during male meiosis: meiotic studies in a man with a deletion of distal Xp. *Am J Hum Genet*. 1992; 51(3):526–533. [PubMed: 1496984]
- Montoya-Burgos JI, Boursot P, Galtier N. Recombination explains isochores in mammalian genomes. *Trends Genet*. 2003; 19(3):128–130. [PubMed: 12615004]
- Palmer S, Perry J, Ashworth A. A contravention of Ohno's law in mice. *Nat Genet*. 1995; 10(4):472–476. [PubMed: 7670497]
- Palmer S, Perry J, Kipling D, Ashworth A. A gene spans the pseudoautosomal boundary in mice. *Proc Natl Acad Sci USA*. 1997; 94(22):12030–12035. [PubMed: 9342357]
- Pamilo P, Bianchi NO. Evolution of the *Zfx* and *Zfy* genes: rates and interdependence between the genes. *Molecular Biology and Evolution*. 1993; 10(2):271–281. [PubMed: 8487630]
- Perry J, Ashworth A. Evolutionary rate of a gene affected by chromosomal position. *Curr Biol*. 1999; 9(17):987–989. [PubMed: 10508587]
- Perry J, Palmer S, Gabriel A, Ashworth A. A short pseudoautosomal region in laboratory mice. *Genome Research*. 2001; 11(11):1826–1832. [PubMed: 11691846]
- Posada D, Buckley T. Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*. 2004; 53(5):793–808. [PubMed: 15545256]

- Quaderi NA, Schweiger S, Gaudenz K, Franco B, Rugarli EI, Berger W, Feldman GJ, Volta M, Andolfi G, Gilgenkrantz S, et al. Opitz G/BBB syndrome, a defect of midline development, is due to mutations in a new RING finger gene on Xp22. *Nature Genetics*. 1997; 17(3):285–291. [PubMed: 9354791]
- Rattray AJ, McGill CB, Shafer BK, Strathern JN. Fidelity of mitotic double-strand-break repair in *Saccharomyces cerevisiae*: a role for SAE2/COM1. *Genetics*. 2001; 158(1):109–122. [PubMed: 11333222]
- Raudsepp T, Chowdhary BP. The horse pseudoautosomal region (PAR): characterization and comparison with the human, chimp and mouse PARs. *Cytogenetic and Genome Research*. 2008; 121(2):102–109. [PubMed: 18544933]
- Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003; 19(12):1572–1574. [PubMed: 12912839]
- Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP, et al. The DNA sequence of the human X chromosome. *Nature*. 2005; 434(7031):325–337. [PubMed: 15772651]
- Rouyer F, Simmler MC, Johnsson C, Vergnaud G, Cooke HJ, Weissenbach J. A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. *Nature*. 1986; 319(6051):291–295. [PubMed: 3941746]
- Rugarli EI, Adler DA, Borsani G, Tsuchiya K, Franco B, Hauge X, Distech C, Chapman V, Ballabio A. Different chromosomal localization of the *Clcn4* gene in *Mus spretus* and C57BL/6J mice. *Nat Genet*. 1995; 10(4):466–471. [PubMed: 7670496]
- Salcedo T, Gerald A, Nachman MW. Nucleotide variation in wild and inbred mice. *Genetics*. 2007; 177(4):2277–2291. [PubMed: 18073432]
- Salido EC, Li XM, Yen PH, Martin N, Mohandas TK, Shapiro LJ. Cloning and expression of the mouse pseudoautosomal steroid sulphatase gene (*Sts*). *Nat Genet*. 1996; 13(1):83–86. [PubMed: 8673109]
- Schiebel K, Meder J, Rump A, Rosenthal A, Winkelmann M, Fischer C, Bonk T, Humeny A, Rappold G. Elevated DNA sequence diversity in the genomic region of the phosphatase *PPP2R3L* gene in the human pseudoautosomal region. *Cytogenetics and cell genetics*. 2000; 91(1–4):224–230. [PubMed: 11173861]
- She JX, Bonhomme F, Boursot P, Thaler L, Catzeflis F. Molecular phylogenies in the genus *Mus*: Comparative analysis of electrophoretic, *scnDNA* hybridization, and mtDNA RFLP data. *Biol J Linn Soc*. 1990; 41(1–3):83–103.
- Shi Q, Spriggs E, Field LL, Ko E, Barclay L, Martin RH. Single sperm typing demonstrates that reduced recombination is associated with the production of aneuploid 24,XY human sperm. *American journal of medical genetics*. 2001; 99(1):34–38. [PubMed: 11170091]
- Simmler MC, Rouyer F, Vergnaud G, Nyström-Lahti M, Ngo KY, de la Chapelle A, Weissenbach J. Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature*. 1985; 317(6039):692–697. [PubMed: 2997620]
- Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*. 2003; 423(6942):825–837. [PubMed: 12815422]
- Soriano P, Keitges EA, Schorderet DF, Harbers K, Gartler SM, Jaenisch R. High rate of recombination and double crossovers in the mouse pseudoautosomal region during male meiosis. *Proc Natl Acad Sci USA*. 1987; 84(20):7218–7220. [PubMed: 3478692]
- Strathern JN, Shafer BK, McGill CB. DNA synthesis errors associated with double-strand-break repair. *Genetics*. 1995; 140(3):965–972. [PubMed: 7672595]
- Suzuki H, Shimada T, Terashima M, Tsuchiya K, Aplin K. Temporal, spatial, and ecological modes of evolution of Eurasian *Mus* based on mitochondrial and nuclear gene sequences. *Molecular Phylogenetics and Evolution*. 2004; 33(3):626–646. [PubMed: 15522792]
- Swofford, DL. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Sunderland, MA: Sinauer Associates; 2002.

- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*. 2007; 24(8):1596–1599. [PubMed: 17488738]
- Van Laere A-S, Coppieters W, Georges M. Characterization of the bovine pseudoautosomal boundary: Documenting the evolutionary history of mammalian sex chromosomes. *Genome Research*. 2008; 18(12):1884–1895. [PubMed: 18981267]
- White MA, Ané C, Dewey CN, Larget BR, Payseur BA. Fine-scale phylogenetic discordance across the house mouse genome. *PLoS Genet*. 2009; 5(11):e1000729. [PubMed: 19936022]
- White MA, Stubbings M, Dumont BL, Payseur BA. Genetics and evolution of hybrid male sterility in house mice. *Genetics*. 2012 In press.
- Winter J, Lehmann T, Krauss S, Trockenbacher A, Kijas Z, Foerster J, Suckow V, Yaspo M-L, Kulozik A, Kalscheuer V, et al. Regulation of the MID1 protein function is fine-tuned by a complex pattern of alternative splicing. *Hum Genet*. 2004; 114(6):541–552. [PubMed: 15057556]
- Yi N, Zinniel DK, Kim K, Eisen EJ, Bartolucci A, Allison DB, Pomp D. Bayesian analyses of multiple epistatic QTL models for body weight and body composition in mice. *Genetical research*. 2006; 87(1):45–60. [PubMed: 16545150]
- Yi S, Summers TJ, Pearson NM, Li W-H. Recombination has little effect on the rate of sequence divergence in pseudoautosomal boundary 1 among humans and great apes. *Genome Research*. 2004; 14(1):37–43. [PubMed: 14672979]
- Young AC, Kirkness EF, Breen M. Tackling the characterization of canine chromosomal breakpoints with an integrated in-situ/in-silico approach: the canine PAR and PAB. *Chromosome research*. 2008; 16(8):1193–1202. [PubMed: 19005636]

Figure 1.

Localization of the PAR boundary in *M. m. castaneus*^{CAST}. The PAR boundary was coarsely mapped by genotyping (*M. m. domesticus*^{WSB} × *M. m. castaneus*^{CAST}) F1 males across the distal end of the X chromosome (A). The PAR was identified by the presence of heterozygous genotypes (red). Hemizygous genotypes (blue) were detected outside of the region at ~166.00 Mb. The boundary region was fine-mapped by shotgun sequencing an X chromosome and Y chromosome bacterial artificial chromosome (BAC) from a *M. m. castaneus*^{CAST} male library that spanned the boundary region (B). Shotgun sequencing fragments from the X chromosome BAC spanned the boundary (black dots; 66 total reads). Shotgun sequencing reads from the Y chromosome were split. 152 reads mapped to the distal side of the boundary, whereas the remainder of the reads had the highest sequence similarity to unmapped regions of the mouse reference Y chromosome (179 reads) or a small proportion of the autosomes (11 reads). Centromere (cen.) and telomere (tel.) ends of the chromosome are indicated.

C57BL/6J

CAST/EiJ

Figure 2.

A shifted PAR boundary in *M. m. castaneus*^{CAST}. The PAR boundary of *M. m. castaneus*^{CAST} is located approximately 430 kb upstream of the boundary in the reference mouse genome (C57BL/6J). The red bar denotes regions of X and Y chromosome homology. Black vertical bars denote the 10 exons of *Mid1*, which span the boundary in C57BL/6J (intron sizes are not drawn to scale between the two subspecies). The PAR boundary in *M. m. castaneus* is proximal to the *Mid1* transcription start site. The centromere (cen.) and telomere (tel.) ends of the X chromosome are indicated.

Figure 3.

Sequence alignment spanning the PAR boundary of *M. m. castaneus*^{CAST}. Sequence homology is depicted by black boxes. Sequence homology between the X and Y chromosomes is lost across the border of the PAR, where bases highlighted in black represent random sequence identity. Position 1 corresponds to 165,980,262 and position 500 corresponds to 165,979,764 of the X chromosome of the mouse reference genome.

Figure 4. Approximate PAR boundaries in *M. m. musculus*^{PWD} and *M. m. domesticus*^{WSB}. Distal regions of the X chromosome were amplified in parents and F1 hybrids every 10 kb using PCR. *M. m. castaneus*^{CAST} sequences could not be amplified from these regions due to large deletions in this strain. PAR sequence was identified when F1 hybrids with *M. m. musculus*^{PWD} or *M. m. domesticus*^{WSB} Y chromosomes showed positive amplification. In these hybrids, amplification can only occur from the PAR of the Y chromosome. X chromosome position is shown in Mb.

Figure 5.

Bayesian phylogenetic analysis of the exon and intron sequence of *Midline1* (*Mid1*; exon 2 through the end of exon 10). *M. m. castaneus*^{CAST} shows substantial divergence from the remainder of house mice in this region. Branch lengths and the scale bar indicate nucleotide substitutions per site. Posterior probabilities supporting each clade are shown at the nodes.

Figure 6.

Pairwise sequence divergence between each of the three house mouse subspecies and *M. spretus*. A sliding window analysis was conducted along the sequence from the *M. m. castaneus*^{CAST} PAR boundary to the end of the coding region of *Midline1* (*Mid1*) (exon 10). Genetic distance was calculated using an HKY85 molecular model of evolution. The beginning of the coding region and the transcription start site of *Mid1* are indicated by grey dashed lines. The PAR boundaries of *M. m. castaneus*^{CAST} and C57BL/6J are indicated by red dashed lines. *M. m. castaneus*^{CAST} divergence increases dramatically at the beginning of the *Mid1* coding region.

Figure 7.

Bayesian phylogenetic analysis of exons 2, 5, 6, and 7 of *Midline1* (*Mid1*). Two strains of *M. m. castaneus* (CAST and CIM) exhibit substantial divergence from the remainder of house mice in this region. Branch lengths and the scale bar indicate nucleotide substitutions per site. Posterior probabilities supporting each clade are shown at the nodes.

Figure 8.

Two alternative evolutionary scenarios leading to a shifted PAR boundary in *M. m. castaneus*^{CAST}. Assuming a subspecies tree placing *M. m. musculus* and *M. m. castaneus* as sister subspecies (White et al. 2009; Keane et al. 2011), the PAR boundary could have shifted independently into *Midline1* (*Mid1*) in the two lineages leading to *M. m. musculus* and *M. m. domesticus* (A). A more parsimonious scenario is that the house mouse ancestral boundary resided within *Mid1* and moved proximally only in the *M. m. castaneus* lineage (B). The red and blue bars denote regions of X and Y chromosome homology of the distal and proximal PAR boundaries, respectively. Arrowheads denote changes in PAR boundary position. Branches are not drawn to scale. The centromere (cen.) and telomere (tel.) ends of the X chromosome are indicated.

\$watermark-text

\$watermark-text

\$watermark-text

Synonymous and non-synonymous divergence within the *Midline1 (Midl)* coding region.

TABLE 1

Exons	N ^a	Comparison	Ka	SE	Ks	SE	Ka/Ks
2-3	245	Cast./Spret.	0.032	±0.009	0.125	±0.027	0.256
		Musc./Spret.	0.002	±0.002	0.019	±0.014	0.105
		Dom./Spret.	0.002	±0.002	0.019	±0.014	0.105
		BL6/Spret.	0.002	±0.002	0.019	±0.014	0.105
4-6	170	Cast./Spret.	0.025	±0.008	0.416	±0.079	0.060
		Musc./Spret.	0.025	±0.003	0.132	±0.033	0.189
		Dom./Spret.	0.025	±0.003	0.132	±0.033	0.189
		BL6/Spret.	0.025	±0.003	0.132	±0.033	0.189
7-10	235	Cast./Spret.	0.079	±0.017	0.857	±0.213	0.092
		Musc./Spret.	0.049	±0.009	0.597	±0.095	0.082
		Dom./Spret.	0.049	±0.009	0.597	±0.095	0.082
		BL6/Spret.	0.049	±0.009	0.597	±0.095	0.082

^aTotal number of codons.

\$watermark-text

\$watermark-text

\$watermark-text

TABLE 2

Midline1 (Mid1) intron length (bp) in *Mus musculus musculus* (Muse.), *Mus musculus domesticus* (Dom.), *Mus musculus castaneus* (Cast.), and C57BL/6J.

Intron	Musc.	Dom.	Cast.	C57BL/6J	% Reduction ^a
1	47,031	47,034	38,010	47,044	19.2
2	38,393	38,377	6,664	38,393	82.6
3	10,763	10,763	1,929	10,763	82.1
4	4,195	4,386	1,769	4,855	63.6
5	1,129	1,922	775	1,922	59.7
6	915	915	70	915	92.3
7	754	754	583	754	22.7
8	1,461	1,461	1,371	1,461	6.2
9	1,096	1,096	301	1,096	72.5
Total	105,737	106,708	51,472	107,203	52.0

^aPercent reduction of *M. m. castaneus* CAST introns from C57BL/6J.