Human Molecular Genetics, 2011, Vol. 20, No. 15 **3010–3021**
doi:10.1093/hmg/ddr204
Advance Access published on May 6, 2011

Association of *Sly* with sex-linked gene amplification during mouse evolution: a side effect of genomic conflict in spermatids?

Peter J.I. Ellis, Joanne Bacon and Nabeel A. Affara*

Mammalian Molecular Genetics Group, Department of Pathology, University of Cambridge, Tennis Court Rd, Cambridge CB2 1QP, UK

Received March 31, 2011; Revised and Accepted May 1, 2011

In common with other mammalian sex chromosomes, the mouse sex chromosomes are enriched for genes with male-specific function such as testis genes. However, in mouse there has been an unprecedented expansion of ampliconic sequence containing spermatid-expressed genes. We show via a phylogenetic analysis of gene amplification on the mouse sex chromosomes that multiple families of sex-linked spermatid-expressed genes are highly amplified in *Mus musculus* subspecies and in two further species from the Palaearctic clade of mouse species. Ampliconic X-linked genes expressed in other cell types showed a different evolutionary trajectory, without the distinctive simultaneous amplification seen in spermatid-expressed genes. The Palaearctic gene amplification occurred concurrently with the appearance of *Sly*, a Yq-linked regulator of post-meiotic sex chromatin (PMSC) which acts to repress sex chromosome transcription in spermatids. Despite the gene amplification, there was comparatively little effect on transcript abundance, suggesting that the genes in question became amplified in order to overcome *Sly*-mediated transcriptional repression and maintain steady expression levels in spermatids. Together with the known sex-ratio effects of Yq/*Sly* deficiency, our results suggest that *Sly* is involved in a genomic conflict with one or more X-linked sex-ratio distorter genes. The recent evolution of the novel PMSC regulator *Sly* in mouse lineages has significant implications for the use of mouse-model systems in investigating sex chromosome dynamics in spermatids.

INTRODUCTION

Mammalian sex chromosomes are subject to evolutionary pressures that substantially differ from those acting on the autosomes, affecting both their structure and function (1–4). Many of the unique structural features, such as reduced Y chromosome gene content and accumulation of repeat sequences on the Y, are explained by the addition–attrition hypothesis (5). In this model, successive rounds of gene acquisition onto the sex chromosomes are followed by degeneration of the Y copies due to the lack of recombination and consequent inefficient elimination of detrimental mutations. Functional specialization is driven by the unique selective dynamics of the X and Y: constant male selection leads to retention of spermatogenesis genes on the Y, while hemizygous exposure of recessive X alleles in males leads to male function accumulation on this chromosome also.

However, recent work on the human and mouse sex chromosomes, culminating in their full sequencing (6–8), has shown that large-scale inverted repeats and palindromes, initially thought to be a Y-specific adaptation allowing for intrachromosomal recombination and gene conversion (9) are also present on the X in both species (10). Since the X undergoes normal recombination during female meiosis, it seems unlikely that palindrome accumulation on the X relates to recombinational dynamics. Rather, since genes lying within these amplified regions are predominantly testis-specific and spermatid-expressed, we and others (10,11) have proposed that these structures have a role in transcriptional regulation. In particular, gene amplification

*To whom correspondence should be addressed. Tel: +44 1223333709; Fax: +44 1223333700; Email: na@mole.bio.cam.ac.uk

appears to promote reactivation of X and Y genes in post-meiotic spermatids (12), bypassing the generally repressive effect of post-meiotic sex chromatin (PMSC) (13). PMSC represents a partial persistence of meiotic sex chromosome inactivation (MSCI) (14) into post-meiotic germ cell stages, however, PMSC is 'leaky' in comparison to MSCI, with single-copy X-linked genes showing transcriptional reactivation in up to 18% of spermatids. Multiple-copy X-linked genes show a higher degree of reactivation, with RNA FISH signals detected in up to 97% of round spermatids (12). Copy number amplification on the sex chromosomes may thus be an adaptation to permit robust expression of sex-linked genes in spermatids.

In the mouse, gene amplification on the sex chromosomes has progressed to a far greater extent than in human, generating very large, highly complex amplicons with no direct structural parallel in the human sequence. Once again, the gene content of these amplicons is very significantly biased towards spermatid genes (10). We hypothesize that this mouse-specific gene amplification is due to *Sly,* a recently evolved gene on mouse Yq (15) which potentiates PMSC, i.e. *Sly* suppresses transcription from the sex chromosomes in spermatids (16). Under this model, the massive amplification of spermatid genes on mouse X and Y is at least partly an adaptation to counteract the repressive effects of *Sly* and maintain adequate expression levels of sex-linked spermatid genes. This model makes two key predictions. First, it predicts that copy number amplification of spermatid-expressed genes on the mouse X and Y will be associated with the presence of *Sly* on mouse Yq, while amplifications involving non-spermatid-expressed genes will be independent of *Sly*. Secondly, it predicts that gene amplifications occurring in response to *Sly* will not be associated with an equivalent increase in transcript levels. Those copy number amplifications occurring in the absence of *Sly* may or may not be associated with an increase in transcription, depending on the factors driving the amplification in question.

We therefore investigated the relative copy numbers of the murine sex-linked ampliconic genes in a series of different mouse species spanning the evolution of *Sly*. We address two classes of genes: mouse-specific genes, for which we sought to determine the point at which they arose during mouse evolution and their subsequent history of gene amplification; and genes that are amplified in mouse but low copy number in rat, for which we sought to determine the timing of significant amplification events. After identifying the major amplification events during mouse evolution, we then examined the expression levels of several of the genes concerned in the appropriate species.

Our study analysed eight different rodent species: the three commensal forms of *Mus musculus* (*M.m. musculus, M.m. domesticus* and *M.m. castaneus*), two further Palaearctic species (*Mus spretus* and *Mus spicilegus*), an Oriental species within the *Mus* subgenus (*Mus caroli*), a more distantly related species from the *Coelomys* subgenus (*Mus pahari*) and laboratory rat (*Rattus norvegicus*) as an outgroup. The phylogenetic relationships for all of these except *Mus spretus* are well established [see Fig. 1 and associated references (17–19)].

RESULTS

Gene panel selection and cross-species bio-informatic analysis

An initial list of ampliconic genes (defined for the purposes of this study as >10 copies with >90% mutual identity) on the laboratory mouse sex chromosomes was identified from published literature, in particular the extant work on Yq genes (15,20–22) and Mueller *et al.*'s (12) recent catalogue of X ampliconic genes. This resulted in a panel of 14 X-linked genes, 1 Yp-linked gene and 6 Yq-linked genes. In addition, from the *H2al1*-related family, we included the single-copy autosomal gene *H2al2* and the two-copy Yp-linked gene *H2al2y* (23). Single-copy X genes *HPRT* and *Eif2s3x* and the single-copy autosomal gene *Hk1* were included for use as reference controls.

The laboratory mouse nucleotide sequences for multiple members of each gene family (and the single copies of the controls) were then aligned to published rat sequence data in order to identify the rat homologue if present. The results of this comparison are summarized in Table 1. In all, 6 of 14 X-linked genes, one out of one Yp-linked gene and zero out of six Yq-linked genes had a rat homologue detectable by nucleotide sequence comparison. Following the database searches, primers were designed (see Supplementary Material, Table S1) for PCR and RT–PCR analysis (see Materials and Methods).

If no rat homologue could be identified at the nucleotide level, the known (or predicted) consensus protein sequence for each gene family was used to perform a PSI-BLAST against the nr protein database (24) in order to identify more substantially diverged relatives. This final step detected ampliconic copies of *Ssty* and *Sstx* on the rat X and Y chromosome which were not detected in the cross-species nucleotide comparison due to very substantial sequence divergence.

Figure 2 shows the resulting distance tree of *Sstx/y*-related protein sequences. The rat *Ssty* sequences cluster with laboratory mouse *Ssty1* and *Ssty2* sequences and the rat *Sstx* sequences with the laboratory mouse *Sstx* sequences. Significantly, however, the multiple rat *Ssty1* sequences were more closely related to each other than to any mouse sequence. Similarly, the multiple rat *Sstx* sequences were more closely related to each other than to mouse *Sstx*. In contrast, *Ssty* sequences from *M. spicilegus* fell within the laboratory mouse group and were clearly identifiable as *Ssty1* or *Ssty2* family members. The simplest explanation for this is that the common rodent ancestor had a single *Ssty* gene and a single *Sstx* gene, which were subsequently independently amplified in rat and mouse lineages. In the mouse lineage, *Ssty* amplification proceeded via an initial gene duplication and sequence divergence (giving *Ssty1* and *Ssty2* sub-families), prior to the expansion of both sub-families. The gene duplication creating *Ssty1* and *Ssty2* is deduced to pre-date the split between *M. spicilecus* and *M. musculus* lineages.

Appearance of new genes on the mouse sex chromosomes

For those genes not present in the rat genome, an initial screen was performed via endpoint (non-quantitative) PCR analysis, allowing us to identify the point at which particular genes

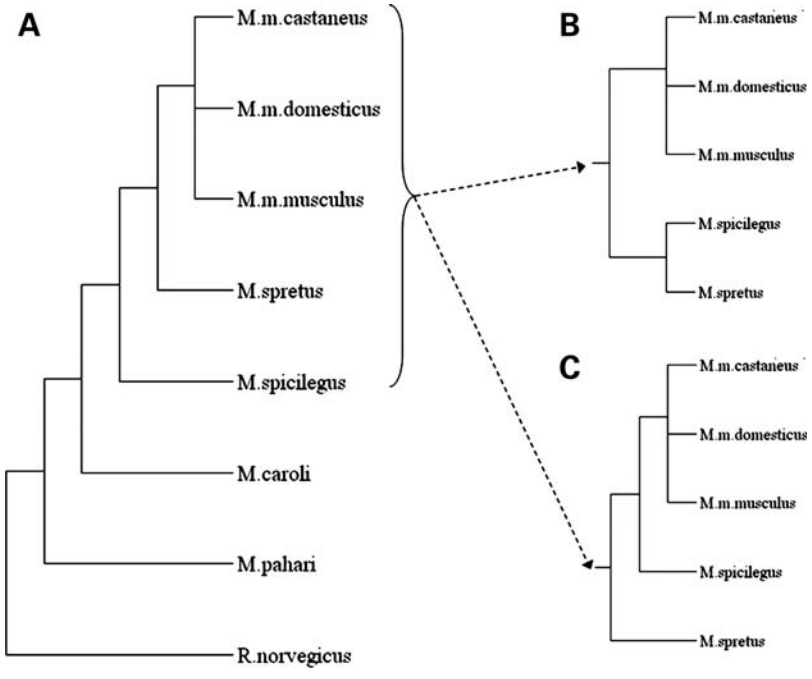

Figure 1. (**A**) Phylogeny of the species analysed in this study (branch lengths not to scale). (**B** and **C**) Alternative proposed phylogenies for the five species from the Palaearctic clade (comprising the *M. musculus* trio, *M. spretus* and *M. spicilegus*). The phylogenies differ only in the placement of *M. spretus*, which some studies place as a sister species to the *M. musculus* group, some as a sister group to *M. spicilegus* and some as an outgroup to *M. musculus* and *M. spicilegus* (17–19).

first arose in the mouse lineage (Fig. 3). These genes fall into three major groups: spindlin relatives (*Ssty1, Ssty2* and *Sstx*), relatives of the synaptonemal complex protein SCP3 (*Sly, Slx, Slxl1* and *Xlr*), histone 2A relatives (*H2al1*, *H2al2* and *H2al2y*) and a final set of genes whose coding potential is unclear or negligible (*Srsy*, *Asty* and *Orly*). This is a simple presence/absence call and does not give detailed information on copy number for the genes concerned. Testicular expression of genes arising during mouse evolution was confirmed via RT–PCR analysis of testis cDNAs from the species panel. Detailed primer design considerations are given in the Methods section.

Ssty1, Ssty2 and *Sstx* primers all gave no product from rat gDNA, indicating that our primers are specific for the mouse versions of these gene families. *Ssty1* and *Ssty2* were detected in gDNA and cDNA from the five Palaearctic species (*M. spicilegus, M. spretus* and the *M. musculus* group) only, while *Sstx* was detected in all mouse species including *M. pahari*. In *M. pahari*, *Ssty2* primers gave a weak band from gDNA which was not male-specific and presumably reflects some cross-reactivity with *Sstx* in this species. These data confirm the results of the PSI-BLAST search described earlier, and further suggest that the duplication giving rise to the *Ssty1* and *Ssty2* sub-families occurred subsequent to the split between *M. caroli* and Palaearctic species.

Slx and *Xlr* were detected in gDNA and cDNA from *M. caroli* onwards, whereas *Sly* and *Slxl1* arose later and are specific to the Palaearctic clade. This is consistent with our earlier study (22), which indicated that *Sly* arose as a fusion gene comprising the 5′ region of *Slx* and the 3′ region of *Xlr* and must therefore logically post-date both. Expression of *Sly* in Palaearctic species was confirmed by inter-exon RT–PCR using primers in exons 2 and 7. Further characterization of *Sly* transcript structure in each species is given in Supplementary Material, Figure S1.

The primer pair for the Yp-linked gene *H2al2y* amplified a male-specific product in gDNA from *M. spicilegus* and the three *M. musculus* species, however, expression was only detected in *M. musculus* species. X-linked (*H2al1*) members of this gene family are expressed in all species including rat (23). Primer pair *H2al1*(TR) is specific to the junction between amplicon units in the major X-linked tandem repeat tract of *H2al1* copies, and yielded a signal from *M. caroli* onwards. The autosomal copy *H2al2* arose via a mouse-specific X-to-autosome duplication/insertion (23), and although we were unable to design suitable primers for cross-species cDNA analysis, we were able to explicitly confirm the presence/absence of this gene by designing primers flanking the insertion site. Primer pair *H2al2*(insert), which detects the inserted gene copy, gives a product from *M. caroli* onwards. Conversely, primer pair *H2al2*(flank), which detects the uninserted autosomal sequence, yields a product in *M. pahari* and rat only.

Asty primers gave male-specific bands from *M. caroli* onwards, however, transcription was only reliably detected in *M. spretus* and *M. musculus* species. *Asty* primers also amplified a weak band in female *M.m. domesticus* DNA at higher cycle numbers (data not shown). Primer pair *Srsy*(1) detected a male-specific signal in the three *M. musculus* species only, while primer pair *Srsy*(2) gave a male-specific signal from these three species and *M. spicilegus*, but not *M. spretus*. Neither of the *Srsy* primer pairs detected any expression of this gene in testis cDNA from any species.

Table 1. Ampliconic genes present in 10 or more copies on laboratory mouse X or Y

	Chr.	Gene	Mouse copy number	Mouse testis expression	Rat homologue	Primers
Ampliconic genes (and relatives)	Yq	Srsy	High	None	—	Y[d,e]
		Asty	High	Premeiotic	—[b]	Y[d]
		Orly	High	Spermatid	—[b]	Y[f]
		Sly	High	Spermatid	—	Y
		Ssty1	High	Spermatid	prot	Y
		Ssty2	High	Spermatid		Y
	Yp	Rbmy	High	Premeiotic	Nuc	Y
		H2al2y	2	Spermatid	—	Y[d]
	X	Astx	Unknown	Unknown	—[b]	N[g]
		Xlr	12	None	—	Y
		Dmrtc1b	12	Premeiotic	Nuc	Y
		Sstx	∼10	Premeiotic and spermatid	prot	Y
		E330016L19Rik	∼21	Spermatid	—	N[h]
		Gmcl1l	∼21	Spermatid	Nuc	Y
		H2al1	∼15	Spermatid	nuc[c]	Y
		LOC665542	∼12	Spermatid	—	N[g]
		Slx	∼25	Spermatid	—	Y
		Slxl1	∼14	Spermatid	—	Y
		Ssxb	13	Spermatid	nuc	Y
		OTT	∼12	Multiple cell types including spermatid	nuc	Y
		Rhox4	28	Multiple cell types[a]	nuc	Y
		Srsx	∼14	Multiple cell types	—	N[g]
	Auto	H2al2	1	Spermatid	—	Y
Controls	X	Eif2s3x	1	Ubiquitous	nuc	Y
	X	HPRT	1	Ubiquitous	nuc	Y
	Auto	Hk1	1	Ubiquitous	nuc	Y

This table shows the chromosomal location, approximate copy number and testicular expression pattern in laboratory mouse, whether a rat homologue was identifiable via BLAST comparison of the nucleotide (nuc) or protein (prot) level, and whether primer pairs could be designed to selectively amplify the gene family. Copy numbers for X/Y genes are from Mueller *et al*. (12) and Ferguson *et al*. (23).

[a]*Rhox4* is used as a marker for the *Rhox* gene cluster, which has a diverse expression pattern including both spermatid-specific and non-spermatid-expressed members.

[b]No known open reading frame in mouse copy, protein BLAST not performed.

[c]Rat sequence too divergent to allow design of cross-reactive primers.

[d]Primers not suitable for qPCR.

[e]Two primer pairs designed detecting different subsets of Srsy sequences.

[f]Primers suitable for RT–PCR only.

[g]X/Y similarity too high to allow design of appropriate primers.

[h]Genome structure does not allow for design of appropriate primers.

For the chimeric locus *Orly*, it was not possible to design primer pairs which would amplify a specific product from genomic DNA, since this locus consists of rearranged partial copies of other Yq-linked genes *Ssty1*, *Asty* and *Sly* and has high sequence similarity to its parent genes (22). RT–PCR using primers specific for chimeric *Orly* transcripts showed transcription of multiple *Orly* splice isoforms in testis RNA from the three *M. musculus* species, but not in any other species tested. We cannot rule out the possibility that it is present but untranscribed in other species.

Copy number amplification of X- and Y-linked genes during mouse evolution

We carried out quantitative PCR for all genes for which suitable primers could be designed. Table 2 shows $\Delta\Delta C_t$ values for each gene in the panel relative to *M. musculus domesticus* and the corresponding estimates of absolute copy number. For each gene, we compared the estimated copy number in each species with the estimated copy number in rat, or to the earliest appearance in the phylogenetic record if the gene was not present in rat. A fold change of greater than 8 (i.e. a $\Delta\Delta C_t$ change ≥ 3) was taken as indicative of significant expansion (shown as bold in Table 2). The $\Delta\Delta C_t$ values for all three control genes were stable across all species ($\Delta\Delta C_t \leq 1$ relative to *M. musculus domesticus*) with the sole exception of the *Hk1* gene in *M. caroli,* which gave a somewhat weaker signal ($\Delta\Delta C_t = -2.2$). It may be that the sequence of the primer binding sites for this gene has diverged in this species (Fig. 4A). In all, qPCR data were obtained for 14 sex-linked ampliconic gene families: 11 of which are expressed in spermatids and three of which are not.

The three non-spermatid-expressed gene families (*Rbmy*, *Dmrtc1b* and *Xlr*) showed no consistent pattern of copy number changes during mouse evolution. *Rbmy* is amplified in all mouse species relative to rat, *Dmrtc1b* showed a stable copy number across all species including rat, while *Xlr* was absent in rat and *M. pahari* and showed stable copy number across the other six species in the study.

In contrast, of the 11 sex-linked, spermatid-expressed gene families examined, 10 show copy number expansion associated with the Palearctic clade of species. Four genes (*Ssty1, Ssty2, Sly* and *Slxl1*) are specific to this group, and absolute copy number estimates indicated that all four are present at

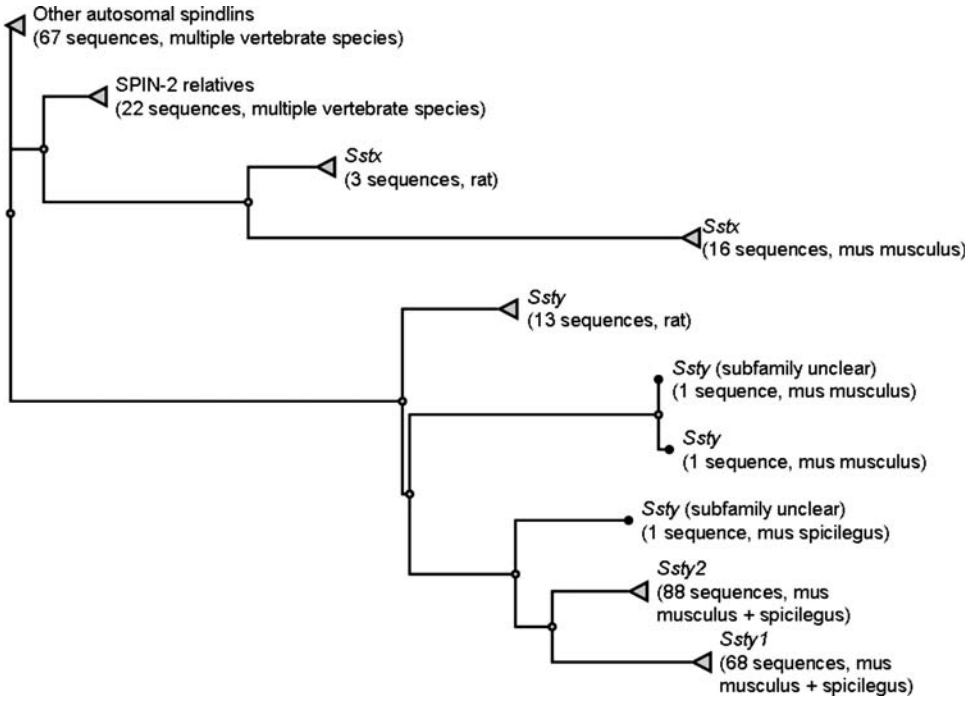

Figure 2. Distance tree of *Sstx/y* family members detected by PSI-PLAST. Autosomal spindlins are used to root the tree. The number of sequences in each branch of the tree reflects the number of sequences uploaded to public databases and should not be taken as indicative of actual copy number on the X and Y.

high copy number in all five Palaearctic species. Six further genes (*Gmcl1l, H2al1, Sstx, Ssxb, OTT* and *Rhox4*) are present in both Palaearctic and non-Palaearctic species, and become amplified specifically within the Palaearctic clade. *Slx* is the only exception to this rule, first appearing in *M. caroli* and showing limited expansion thereafter.

Figure 4 is a graphic representation of the copy number data for the six X-linked spermatid gene families showing copy number changes between Palaearctic and non-Palaearctic species, plus controls (Fig. 4A). Of these, three (*Ssxb, OTT* and *Rhox4*) became amplified at the root of the Palaearctic clade (Fig. 4B). *Ssxb* and *Rhox4*, in addition to the Palaearctic amplification, also showed high signal in *M. pahari*. Since the ancestral state in rat is low copy number and the copy number in *M. caroli* was also low for these two genes, this suggests that these two genes have also been independently amplified in *M. pahari.* An alternative possibility is that these two genes became amplified early on in mouse evolution and that copies were subsequently lost in *M. caroli*.

Sstx and *Gmcl1l* were both amplified in *M. spretus* and the three *M. musculus* species but not in *M. spicilegus* (Fig. 4C). Finally, *H2al1* appears to have undergone two rounds of gene amplification (Fig. 4D). The first of these occurred prior to the divergence of *M. caroli* and is likely to reflect the initial generation of the tandem repeat (as shown in the non-quantitative part of our study) and/or proliferation of interspersed copies on the X. The second expansion is much more recent, being specific to the three *M. musculus* species. This is likely to reflect expansion of the copy number within the tandem repeat. This is in agreement with our prior sequence-based analysis of this family (23), which showed an initial proliferation on the X concurrent with the divergence of *H2al1* and *H2al2,* followed by an evolutionarily recent expansion of the tandem repeat array (99–100% nucleotide identity between repeat units).

Expression analysis of amplified X-linked genes

We performed quantitative RT–PCR on testis RNA for the six spermatid-expressed genes showing copy number changes between Palaearctic and non-Palaearctic species, in order to determine whether the observed copy number increases led to increased transcript levels in testis. Expression levels were normalized using two autosomal single-copy spermatid-expressed genes, *Tnp1* and *Odf1*. Figure 5 shows the expression changes (using a $\Delta\Delta C_t$ scale) for these genes superimposed on the corresponding genomic DNA copy number data.

For the majority of gene amplifications occurring within the Palaearctic clade, copy number amplification was not accompanied by an equivalent increase in transcript levels. The only exception to this observation was *Ssxb*, where copy number and transcript level were both increased in Palaearctic species relative to *M. caroli*. The independent amplifications of *Ssxb* and *Rhox4* in *M. pahari* were both associated with marked changes in transcript level.

In particular, *Sstx* and *OTT* showed little change in expression across the species panel despite very significant changes in gene copy number. The amplification of *Gmcl1l* in *M. spretus* and *M.m. domesticus* and the amplification of *H2al1* in *M.m. domesticus* were associated with small increases in transcription, however, the transcriptional changes were considerably smaller than the changes in copy number. *Rhox4* showed no increase in transcript level

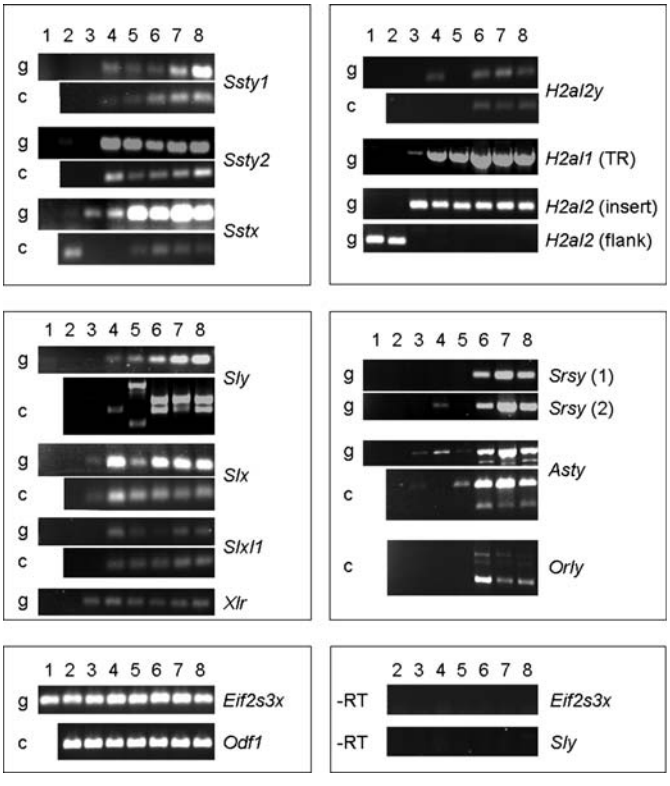

Figure 3. PCR screening of genomic DNA (g) and testis cDNA (c) for the presence of selected X- and Y-linked ampliconic genes in different mouse species. Upper left, spindlin relatives; upper right, histone 2A relatives; middle left, SCP3 relatives; middle right, non-coding genes; lower left, positive controls; lower right, negative controls. *Srsy*(1) and *Srsy*(2) indicate the results obtained with the two different primer pairs used. The multiple bands in the *Sly, Orly* and *Asty* cDNA assays represent splice variant isoforms. 1, *Rattus norvegicus*; 2, *M. pahari*; 3, *M. caroli*; 4, *M. spicilegus*; 5, *M. spretus*; 6, *M. musculus castaneus*; 7, *M. musculus domesticus*; 8, *M. musculus musculus*.

associated with the increased copy number in *M. spicilegus* and *M. spretus* relative to *M. caroli*. This gene did show an increase in transcription in *M.m. domesticus* relative to *M. spretus* and *M. spicilegus*, however, this did not correspond to a change in copy number.

DISCUSSION

This study presents quantitative PCR data on gene family expansion for 14 families of murine sex-linked ampliconic genes (Table 2) and qualitative data for a further three Y-linked families (*Asty*, *Orly*, *Srsy*) that were not amenable to quantitative analysis. Our findings are summarized in Figure 6. Quantitative transcriptional data were also obtained for selected genes showing significant changes in copy number across the panel of species.

We find that there was a significant burst of evolutionary innovation and gene expansion within the Palaearctic clade of mouse species. On the Y chromosome, this involved the differentiation of the pre-existing *Ssty* gene into *Ssty1* and *Ssty2* sub-families and the *de novo* acquisition of *Sly*. These genes are core constituents of the Yq-linked Huge Repeat Array (HRA) (8,22), and appear to be present at high copy number in all species within the Palaearctic clade. On the X chromosome, *Slxl1* arose *de novo* in this clade and amplified to high copy number, while *OTT, Rhox4* and *Ssxb* also became amplified. A second, more minor burst of gene acquisition /amplification is associated with the group of three *M. musculus* sub-species, involving the generation of the Yq-linked chimeric locus *Orly*, copy number expansion of the Yq-linked gene *Srsy*, expansion of the X-linked tandem repeat of *H2al1* copies and a further increase in the number of X-linked *Slx* copies. From the intensity of the non-quantitative *Asty* data (Fig. 3), it appears likely that this gene also is amplified in *M. musculus* sub-species. *Sstx* and *Gmcl1l* showed significant copy number expansion in *M. musculus* species and *M. spretus*.

In common with other studies (17–19), our results are ambiguous in regard to the phylogenetic placement of *M. spretus*. The results from the Y chromosome (*H2al2y*, *Srsy* primer pair 2 and sequencing of *Sly* cDNA) favour the phylogeny shown in Figure 1C, with *M. spicilegus* as a sister group to the *M. musculus* trio. In contrast, the results from the X chromosome (*Sstx* and *Gmcl1l*) favour the phylogeny shown in Figure 1A, with *M. spicilegus* basal to the *spretus/musculus* group. This could be explained by historical introgression of a *M. musculus*-related Y chromosome into the *M. spicilegus* population or an introgression of part or all of a *M. musculus*-related X chromosome into the *M. spretus* population.

Taken together, our data indicate a massive proliferation of ampliconic genes on both sex chromosomes, involving both the expansion of existing gene families and also the differentiation of new sub-families of ampliconic sequences. Importantly, this Palaearctic amplification appears to be limited to spermatid genes. The three ampliconic genes which are not expressed in spermatids (*Xlr*, *Rbmy* and *Dmrtc1b*) were not amplified in the Palaearctic clade relative to other species. In contrast, 10 out of 11 spermatid-expressed ampliconic genes show a Palaearctic amplification in copy number.

There are several possible scenarios to explain this coordinated amplification of genes across the sex chromosomes. First, for closely linked genes contained within the same amplicon, expansion of the ampliconic region will expand the copy number of all genes simultaneously. On the X chromosome, this applies to *Sstx* and *Gmcl1l*, both of which are located in amplicon Amp1 near the X centromere (12), and our results indicate that their copy numbers are indeed very closely correlated (Fig. 4C). On the Y chromosome, this would be expected to apply to the components of the HRA on Yq (8), *Ssty1*, *Ssty2*, *Sly*, *Asty*, *Srsy* and *Orly*. However, our results indicate that these five genes were not amplified strictly in lock-step with each other. In particular, *Orly* arose and became amplified substantially later than the other ampliconic genes on Yq. Moreover, variations in copy number of *Ssty1*, *Ssty2* and *Sly* between the three *M. musculus* sub-species are not identical, indicating that copy numbers of these genes have a degree of independence from each other via amplification/deletion of specific sub-regions of the HRA.

Co-amplification via linkage does not, however, apply to unlinked amplicons and hence cannot explain why X- and Y-linked genes show simultaneous amplification. Moreover,

Table 2. Mean $\Delta\Delta C_t$ values ± s.e.m relative to *M. musculus domesticus* for ampliconic X- and Y-linked genes

Category	Gene	$\Delta\Delta C_t$ Rat	Pahari	Caroli	Spicilegus	Spretus	Castaneus	Musculus	Domesticus
Y spermatid	*Ssty1*	—[a]	—[a]	—[a]	−1.95 ± 0.05 (30)	−1.78 ± 0.05 (30)	−1.43 ± 0.10 (40)	−0.48 ± 0.05 (80)	0.00 ± 0.07 (100)
	Ssty2	—[a]	—[a]	—[a]	−1.52 ± 0.05 (200)	−1.15 ± 0.05 (200)	−1.45 ± 0.05 (200)	−1.10 ± 0.20 (200)	0.00 ± 0.00 (500)
	Sly	—[a]	—[a]	—[a]	−2.04 ± 0.10 (10)	−0.16 ± 0.00 (40)	−0.11 ± 0.05 (40)	0.89 ± 0.05 (80)	0.00 ± 0.09 (50)
X spermatid	*Slx*	—[a]	—[a]	−0.75 ± 0.00 (30)	−2.55 ± 0.10 (9)	−3.13 ± 0.05 (6)	0.07 ± 0.10 (60)	0.97 ± 0.10 (100)	0.00 ± 0.00 (50)
	Slxl1	—[a]	—[a]	—[a]	−1.30 ± 0.10 (9)	−1.68 ± 0.05 (7)	−0.68 ± 0.10 (10)	0.17 ± 0.05 (30)	0.00 ± 0.05 (20)
	Gmcl1l	−6.75 ± 0.10 (1)	−5.62 ± 0.05 (2)	−5.97 ± 0.10 (2)	−5.42 ± 0.05 (3)	**−1.55 ± 0.05 (40)**	**−0.95 ± 0.05 (60)**	**−0.90 ± 0.10 (60)**	**0.00 ± 0.06 (100)**
	H2al1	−5.20 ± 0.77 (1)	−6.45 ± 0.05 (0.4)	−3.70 ± 0.15 (3)	−2.60 ± 0.05 (6)	−2.68 ± 0.00 (6)	**1.52 ± 0.56 (100)**	**−1.08 ± 0.05 (20)**	**0.00 ± 0.00 (40)**
	Sstx	—[a]	−7.15 ± 0.10 (0.6)	−4.80 ± 0.10 (3)	−5.35 ± 0.15 (2)	**−1.48 ± 0.05 (30)**	**−0.48 ± 0.00 (60)**	**−1.13 ± 0.15 (40)**	**0.00 ± 0.05 (90)**
	Ssxb	−5.23 ± 0.10 (1)	**3.40 ± 0.18[c] (400)**	−4.00 ± 0.12 (2)	**0.20 ± 0.15 (40)**	**−0.08 ± 0.00 (40)**	**0.12 ± 0.10 (40)**	**−0.13 ± 0.05 (30)**	**0.00 ± 0.15 (40)**
	OTT	−4.28 ± 0.15 (1)	−3.50 ± 0.00 (2)	−4.30 ± 0.25 (1)	**−0.50 ± 0.15 (10)**	−1.38 ± 0.15 (10)	**1.65 ± 0.17 (60)**	**1.62 ± 0.10 (60)**	**0.00 ± 0.10 (20)**
	Rhox4	−4.40 ± 0.18 (1)	**0.50 ± 0.00[c] (30)**	−3.45 ± 0.54 (2)	**−1.25 ± 0.05 (9)**	**−0.68 ± 0.05 (10)**	**0.42 ± 0.05 (30)**	**−0.73 ± 0.10 (10)**	**0.00 ± 0.10 (20)**
Y non-spermatid	*Rbmy*	−3.83 ± 0.10 (1)	**0.65 ± 0.05 (20)**	**0.75 ± 0.00 (20)**	**0.75 ± 0.05 (20)**	**0.02 ± 0.05 (10)**	**1.42 ± 0.05 (40)**	**2.02 ± 0.05 (60)**	**0.00 ± 0.00 (10)**
X non-spermatid	*Dmrtc1b*	−0.98 ± 0.51 (3)	—[b]	−0.95 ± 0.18 (3)	−0.80 ± 0.05 (3)	−0.73 ± 0.15 (4)	0.27 ± 0.05 (7)	—[b]	0.00 ± 0.05 (6)
	Xlr	—[a]	—[a]	−0.80 ± 0.00 (2)	−0.45 ± 0.05 (2)	−0.73 ± 0.00 (2)	−0.03 ± 0.05 (3)	0.17 ± 0.05 (3)	0.00 ± 0.05 (3)
Single copy controls	*Eif2s3x*	−0.78 ± 0.00 (1)	0.35 ± 0.00 (2)	0.55 ± 0.05 (2)	0.45 ± 0.05 (2)	0.47 ± 0.05 (2)	−0.13 ± 0.13 (2)	0.02 ± 0.05 (2)	0.00 ± 0.05 (2)
	Hprt	0.78 ± 0.04 (1)	−0.35 ± 0.03 (0.5)	−0.55 ± 0.03 (0.4)	−0.45 ± 0.05 (0.4)	−0.47 ± 0.05 (0.4)	0.13 ± 0.05 (0.7)	−0.02 ± 0.03 (0.6)	0.00 ± 0.03 (0.6)
	Hk1	−0.28 ± 0.00 (0.8)	−1.00 ± 0.10 (0.5)	−2.20 ± 0.10[d] (0.2)	−0.30 ± 0.10 (0.8)	0.10 ± 0.02 (1)	0.57 ± 0.00 (1)	0.22 ± 0.15 (1)	0.00 ± 0.12 (1)

Figures in parentheses indicate absolute copy number estimates to 1 significant figure. Bolded values indicate $\Delta\Delta C_t$ differences ≥3 relative to the earliest appearance in the record, i.e. an estimated 8-fold change in gene copy number.

[a]Gene not present in this form in this species.

[b]Data not obtained due to non-specific qPCR reaction in this species.

[c]Anomalously high signal possibly due to independent amplification of *Ssxb* and *Rhox4* in *M. pahari*.

[d]Anomalously low signal possibly due to sequence divergence of *Hk1* in *M. caroli*.

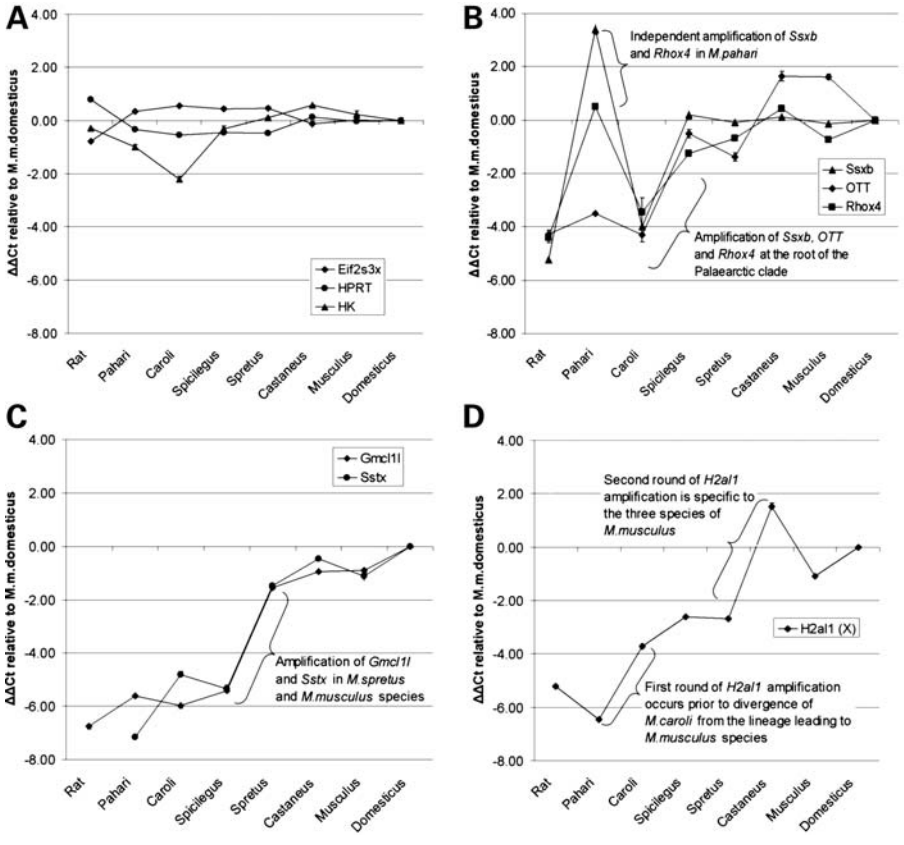

Figure 4. qPCR data for selected X- and Y-linked ampliconic genes in genomic DNA from the different mouse species surveyed. Numbers indicate ΔΔCt values relative to *M. musculus domesticus* for each gene. (**A**) *Eif2s3x*, *Hprt* and *Hk1* controls; (**B**) *Ssxb*, *OTT* and *Rhox4*; (**C**) *Gmcl1l* and *Sstx*; (**D**) *H2al1*.

it fails to explain why so many highly ampliconic structures are found on the X, some 22 in total (12). It is possible that all these amplifications are unrelated, however, this is implausible owing to the sex chromosome and cell-type specificity of the expansions. Given that the gene amplification selectively affects sex-linked genes expressed in spermatids, an alteration in sex chromosome dynamics appears the most likely explanation.

We have previously identified a genomic conflict (25) between the mouse X and Y chromosomes (11,26), whereby male mice with deletions on Yq show an offspring sex-ratio skew in favour of females and an upregulation of multiple X-linked genes in spermatids. Under the conflict model, the sex-ratio skew is mediated by 'selfish' X-linked spermatid-expressed genes favouring their own transmission, while one or more of the genes on Yq suppresses X transcription and restores a normal sex ratio. This 'arms race' between the chromosomes leads to expansion of copy number for both distorter and suppressor genes. Recent work has identified *Sly* as the key gene on Yq, since transgenic knockdown of *Sly* recapitulates all aspects of Yq deletion, including sperm head shape abnormalities, X chromosome depression in spermatids and sex-ratio skewing (16).

X/Y genomic conflict is also demonstrated by the chromosomal introgression dynamics in the hybrid zone between *M. musculus musculus* and *M. musculus domesticus*. Pure *musculus* and *domesticus* populations showed a census sex-ratio skew in favour of females, whereas a population bearing an introgressed *musculus* Y chromosome on a *domesticus* background showed a sex ratio near parity: i.e. a significant increase in the proportion of males (27). This is consistent with the data presented here (Table 2), showing that *Sly* has an increased copy number in *musculus* relative to *domesticus*. This is also a likely explanation for the asymmetric male sterility seen in crosses between *M.m. domesticus* (Lewes/EiJ) and *M.m. musculus* (PWK/PhJ) (28). In this cross, pairing of a *domesticus* Y chromosome (i.e. reduced *Sly* copy number) with a *musculus* X chromosome leads to widespread X overexpression in spermatids and sterility (29).

Among laboratory strains, the congenic HG.CAST-(D17Mit 196-D17Mit190) ($HQ17^{hg/hg}$) strain shows sex-ratio skewing in favour of females, however, in this case the causative locus is autosomal (30). In the case of an X/Y genomic conflict, once a sex-ratio distorter has arisen on a sex chromosome, autosomal genes will be selected to increase production of the rarer sex, hence the sex-ratio distortion in the $HQ17^{hg/hg}$ strain may relate to interspecific variation in such an autosomal modifier gene. More broadly, multiple X-linked and autosomal loci have been implicated in the male sterility observed in *M.m. musculus* x *M.m. domesticus* hybrids (31–33) and the hybrid breakdown seen in C57Bl/6J males consomic for the *M. molossinus* X chromosome (34). This heterogametic incompatibility [Haldane's rule (35)] is suggested to result in part from

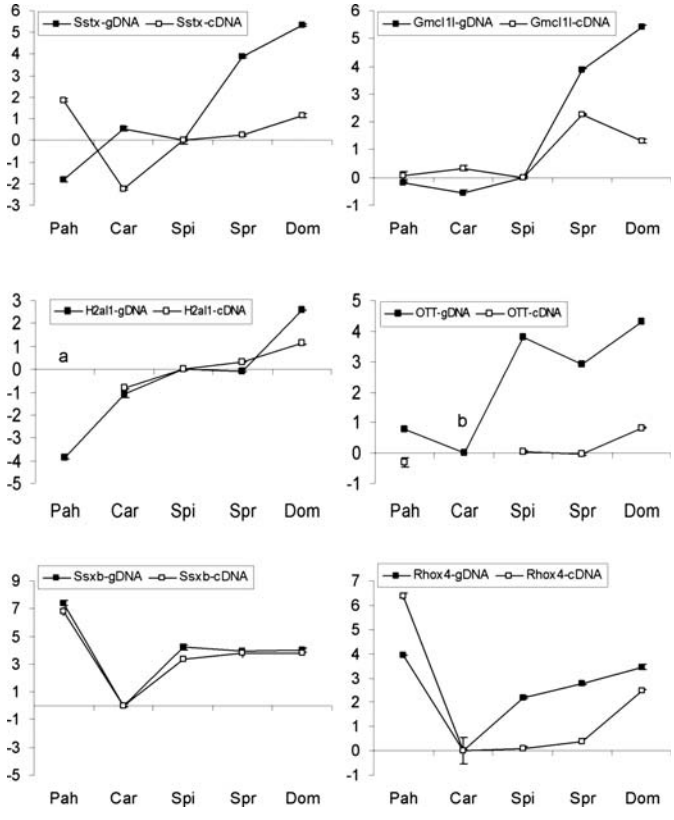

Figure 5. Open boxes: qPCR data from testis cDNA templates showing changes in testicular transcript levels for the six gene families showing Palaearctic amplification. Closed boxes: genomic DNA copy number levels for comparison. For genes amplified within the Palaearctic group (*Sstx, Gmcl1l* and *H2al1*), cDNA and gDNA values are shown relative to *M. spicilegus*. For genes amplified at the root of the Palaearctic clade, cDNA and gDNA values are shown relative to *M. caroli*, except in the case of *OTT*, where the cDNA signal was normalized to the median value across the data set due to signal drop-out in *M. caroli*. [a]*H2al1*was not detected in *M. pahari* testis cDNA. [b]*OTT* was not detected in *M. caroli* testis cDNA.

unbalancing of transmission ratio distortion genes and their suppressors (36). Intriguingly, one of the autosomal loci involved has been identified as *Prdm9*, a histone H3 trimethyltransferase which regulates MSCI (33).

While the genomic conflict hypothesis is attractive as an explanation for X/Y gene copy number expansion, a large question remains: why are so many disparate genes all amplified? The conflict itself should only lead to amplification of the distorter and repressor genes. Mueller *et al*. (12) identified 20 spermatid-specific multicopy gene families on the mouse X. Can these really all be involved in sex-ratio skewing?

We suggest rather that the breadth of gene amplification on the mouse X is due to *Sly*'s mechanism of action. Rather than being a targeted repressor of specific downstream genes, *Sly* acts via modulation of PMSC (16,37,38), globally repressing transcription of *all* sex chromosomal genes in spermatids. Many of these genes may thus be 'innocent bystanders' in regard to the conflict, while the identity of the 'true' target of *Sly* [i.e. the X-linked sex-ratio distorter(s) involved] remains to be established. Given this non-specific nature of *Sly*-mediated repression, copy number expansion of the 'bystander' X genes may be necessary simply in order to maintain adequate transcript levels in spermatids.

Our data show that the gene amplifications occurring in Palaearctic species [i.e. concurrently or subsequent to the appearance of *Sly*] are predominantly not associated with equivalent increases in transcript levels, i.e. there is a net decrease in the transcript level per genomic copy. Importantly, Mueller *et al.* (12) and our own work on the H2al family (23) show that multiple different copies of each ampliconic gene are expressed. This argues that the net decline in transcriptional activity per genomic copy reflects genuine transcriptional repression and not simply the accumulation of 'dead' non-transcribing pseudogene copies. These transcriptional data must be treated cautiously due to confounding factors such as changes in the cellular distribution of transcripts between species and changes in the proportion of spermatids in whole testis tissue. Nevertheless our results are in good agreement with the above hypothesis. In contrast, the independent amplifications of *Rhox4* and *Ssxb* in *M. pahari*, occurring in the absence of *Sly*, are associated with significant rises in transcript levels. These amplifications were presumably driven by other as yet undetermined factors.

In summary, there has been a profound modification of the sex chromosome transcriptome during mouse evolution, involving the proliferation of multiple ampliconic spermatid-specific genes. These ampliconic gene families account for ~18% of all protein-coding gene copies on the X chromosome (12) and almost 100% of the protein-coding gene copies on the Y chromosome—the handful of single-copy genes on Yp being vastly outnumbered by the ampliconic gene tracts on Yq. Surprisingly, however, this proliferation has had comparatively little effect on transcript levels for the genes assayed in this study. We propose that this relates to the emergence of the PMSC regulator *Sly* in the Palaearctic clade of mouse species, with amplified gene copy number being a means of evading *Sly*-mediated repression and maintaining transcription of sex-linked genes in spermatids. Investigation of the less highly ampliconic gene families on the mouse X (i.e those with 2–10 copies) will help test this proposal, however, this is not amenable to the cross-species PCR techniques used here and will require extensive sequencing work in multiple mouse species to determine the copy numbers of this class of genes.

MATERIALS AND METHODS

Sample collection

DNA was extracted using standard protocols from cryopreserved male/female liver samples from laboratory rat and from seven wild-derived mouse strains sourced from the Jackson Laboratory (jaxmice.jax.org): PWD/PhJ (*Mus musculus musculus*, strain number 004660), PERA/EiJ (*Mus musculus domesticus*, strain number 000930), CAST/EiJ (*Mus musculus castaneus*, strain number 000928), SPRET/EiJ (*Mus spretus*, strain number 001146), PANCEVO/EiJ (*Mus spicilegus*, strain number 001384), Mus pahari/EiJ (*Mus pahari*, strain number 002655) and CAROLI/EiJ *(Mus pahari*, strain number 000926). Briefly, the liver samples were homogenized in

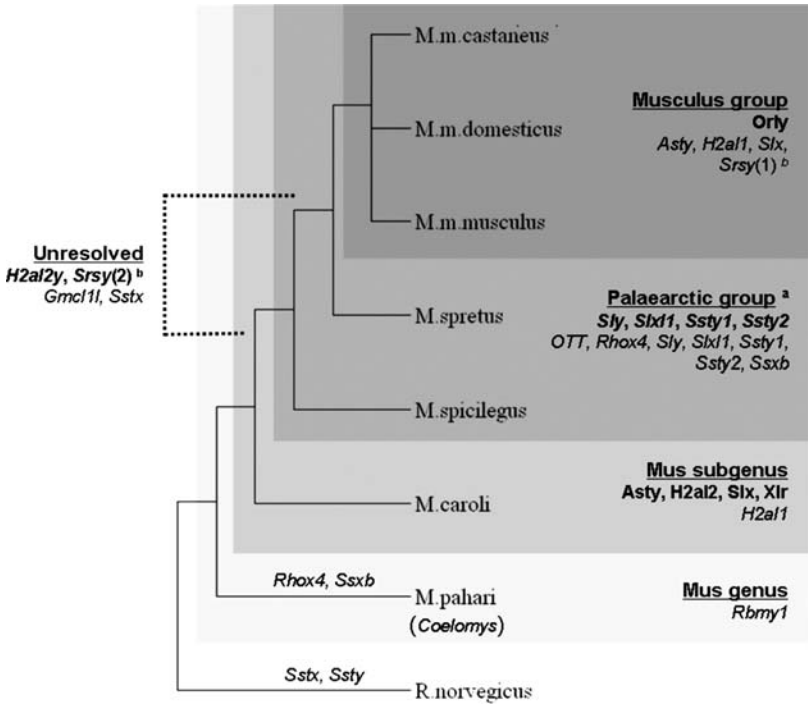

Figure 6. Timing of major gene acquisition/amplification events during mouse evolution in their phylogenetic context. Bolded text represents novel gene acquisition, italic text represents copy number amplification. The majority of events are consistent with the known phylogeny of these species and are shown to the right of the diagram. Four events remain unresolved due to the ambiguity over *M. spretus* placement (see Fig. 1 and Discussion) and are shown to the left of the diagram. Additionally, the species-specific independent amplifications in *M. pahari* (*Rhox4* and *Ssxb*) and rat (*Sstx* and *Ssty*) are indicated above the appropriate branches. [a]*Sly*, *Ssty1* and *Ssty2* are scored both as novel gene acquisitions in this clade and as gene amplifications, since absolute copy number estimates indicate high copy number in all Palaearctic species. [b]The two primer sets for *Srsy* gave differing results as discussed in the text.

DNA extraction buffer (50 mM Tris–HCl, pH 7.5, 1 mM EDTA, 5 mM $MgCl_2$, 50 mM NaCl, 5% glycerol, 0.1% Triton X-100, 0.1% beta-mercaptoethanol) and the cell nuclei pelleted by centrifugation. Nuclei were resuspended in homogenization buffer supplemented with EDTA (final concentration 10 mM) and sodium dodecyl sulphate (final concentration 1%) and digested with proteinase K overnight at 55°C. Isopropanol precipitation was used to concentrate the DNA, followed by resuspension in TE buffer (10 mM Tris pH 7.5, 1 mM EDTA).

RNA was extracted from cryopreserved testis tissue from the male individuals in each mouse line using TRI reagent (Sigma) according to the manufacturer's protocol.

Primer design

Primers were designed to detect each gene family based on the following criteria:

(1) If the gene family is present on both sex chromosomes, primer pairs should distinguish between the X and Y relatives (and/or autosomal copies).
(2) Primer pairs should detect all copies on the X (or Y) if possible. Where this is not possible, primers should detect at least 10 different copies.
(3) If a rat sequence is present with sufficient identity to the mouse sequence (~90% identity over ~100 bases), the primer pair should be able to amplify the rat sequence.
(4) Primer sequences should where possible be designed within a single exon, as exonic sequence is more likely to be conserved across species.
(5) The PCR product length should be between 75 and 150 base pairs, to allow use of quantitative PCR assays.

It was unfortunately not possible to satisfy all these criteria for each gene studied. For *Ssty1/2* and *Sstx*, sequence divergence precluded designing primers capable of cross-reacting with the mouse and rat genes. Therefore, in this case the primers were designed to amplify the laboratory mouse gene families and exclude the rat sequence. These primers should thus indicate the timing of amplification of *Ssty1/2/x* in the mouse lineage without the confounding factor of the independent gene amplification in the rat.

H2al genes form a complex family with a single copy on the rat X chromosome, 15 copies on the mouse X chromosome including an expanded tandem repeat of 10 copies, two copies on mouse Yp and a single copy on chromosome 2 derived from a mouse-specific duplication/insertion (23). We designed four sets of primers to determine the overall X copy number, the presence of Y copies, the presence of the tandem repeat amplicon and the presence of the autosomal copy.

For *Astx* and *Srsx*, it was not possible to design X-specific primers. For the Y counterparts *Asty* and *Srsy*, primers were designed to amplify a subset of Y-linked sequences, however, the resulting product size meant that these primer pairs were not suitable for quantitative use, and there may

also be other Y-linked sequences not detected by our primers. The predicted X-linked transcript *LOC665542* also proved to have highly similar sequence on the Y chromosome which precluded the design of specific primers. The related Y sequence is not annotated as containing any transcript, and there are no ESTs in dbEST matching *LOC665542*, suggesting that *LOC665542* is a mis-predicted transcript.

E330016L19Rik was not suitable for PCR primer design owing to its genomic structure, with very short exons. Since this transcript lies in the same X amplicon as *Slx* (Amp4) (12), its copy number is likely to be closely correlated with *Slx*. Finally, since *Orly* is a chimeric locus consisting of rearranged partial copies of other Yq-linked genes (22), it was not possible to design primers which would amplify a specific *Orly* product from genomic DNA. To test the presence/absence of *Orly*, we therefore carried out RT–PCR on testis RNA from the various mouse species using primers specific for the chimeric transcript.

PCR and RT–PCR

PCR reactions were carried out in a total volume of 10–20 μl using standard protocols. The polymerase used was HotStarTaq (QIAGEN), and 25 ng of genomic DNA was used as the template for each reaction. RT–PCR for *Orly* was carried out using the QIAGEN OneStep RT–PCR kit on 25 ng of total testis RNA. All other RT–PCRs used a two-step protocol as follows. First-strand cDNA was prepared from 1 μg of DNase I-treated testis total RNA from *M. pahari*, *M. caroli, M. spicilegus, M. spretus* and *M.m. domesticus*, in a reaction volume of 20 μl. This first strand cDNA was diluted to 200 μl with nuclease-free H_2O and 1 μl used as template for PCR reactions. Details of primer pairs, T_m and extension time are given in Supplementary Material, Table S1.

Quantitative PCR and RT–PCR

qPCR was performed on 25 ng of male genomic DNA for each gene/species combination using an iCycler (BioRad) and iQ SYBR Green Supermix (Bio-Rad). To enable use of consistent reference genes, an annealing temperature of 55°C was used for all qPCR reactions. The fluorescent signal threshold crossing point (C_t) for each gene was normalized to the average signal for the two reference single copy X genes, *Eif2s3x* and *Hprt*, giving a normalized signal of ΔC_t. To examine changes in gene copy number throughout mouse evolution, the ΔC_t values were normalized to the value for *Mus musculus domesticus*, thus giving the $\Delta\Delta C_t$ values shown in Table 2. Absolute copy number estimates were calculated as $2^{\Delta\Delta Ct}$ relative to the known copy number in rat where possible. For genes not present in rat or where the rat copy number is unknown, absolute copy number was estimated as $2^{\Delta Ct}$ relative to average signal for *Hprt1* and *Eif2s3x*. Quantitative transcriptional analysis followed the same protocol, using 1 μl of first-strand cDNA as the template for the qPCR reactions, and was normalized to the average signal of the control genes *Odf1* and *Tnp1*.

Data quality control

All qPCR reactions were performed in duplicate. Repeat data were obtained if any of the following conditions were met: (i) a C_t difference of ≥ 0.6 cycles between duplicate reactions; (ii) the reaction was not in log phase according to the fluorescence plot; (iii) melt curve analysis indicated an incorrect product or the presence of primer dimers. For all X-linked ampliconic genes with Y-linked counterparts, the qPCR analysis was repeated using female DNA as a template to confirm the absence of any cross-reaction with Y-linked family members (data not shown). For Y-linked genes, absence of cross-reaction with X copies was confirmed via endpoint non-quantitative PCR.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Mr Kenrick Ng Kien Hoe for assistance with quantitative PCR.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the Biotechnology and Biological Sciences Research Council (BB/F007434/1 to N.A.A. and P.J.I.E.).

REFERENCES

1. Fisher, R.A. (1931) The evolution of dominance. *Biol. Rev.*, **6**, 345–368.
2. Rice, W.R. (1984) Sex chromosomes and the evolution of sexual dimorphism. *Evolution*, **38**, 735–742.
3. Charlesworth, B. (1996) The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol.*, **6**, 149–162.
4. Vallender, E.J. and Lahn, B.T. (2004) How mammalian sex chromosomes acquired their peculiar gene content. *Bioessays*, **26**, 159–169.
5. Graves, J.A. (1995) The origin and function of the mammalian Y chromosome and Y-borne genes – an evolving understanding. *Bioessays*, **17**, 311–320.
6. Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T. *et al.* (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, **423**, 825–837.
7. Ross, M.T., Grafham, D.V., Coffey, A.J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G.R., Burrows, C., Bird, C.P. *et al.* (2005) The DNA sequence of the human X chromosome. *Nature*, **434**, 325–337.
8. Alföldi, J.E. (2008). Sequence of the mouse Y chromosome. Doctoral thesis, Massachusetts Institute of Technology. http://hdl.handle.net/1721.1/42398.
9. Rozen, S., Skaletsky, H., Marszalek, J.D., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K. and Page, D.C. (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature*, **423**, 873–876.
10. Warburton, P.E., Giordano, J., Cheung, F., Gelfand, Y. and Benson, G. (2004) Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res.*, **14**, 1861–1869.
11. Ellis, P.J., Clemente, E.J., Ball, P., Touré, A., Ferguson, L., Turner, J.M., Loveland, K.L., Affara, N.A. and Burgoyne, P.S. (2005) Deletions on

mouse Yq lead to upregulation of multiple X- and Y-linked transcripts in spermatids. *Hum. Mol. Genet.*, **14**, 2705–2715.
12. Mueller, J.L., Mahadevaiah, S.K., Park, P.J., Warburton, P.E., Page, D.C. and Turner, J.M. (2008) The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nat. Genet.*, **40**, 794–799.
13. Turner, J.M., Mahadevaiah, S.K., Ellis, P.J., Mitchell, M.J. and Burgoyne, P.S. (2006) Pachytene asynapsis drives meiotic sex chromosome inactivation and leads to substantial postmeiotic repression in spermatids. *Dev. Cell*, **10**, 521–529.
14. McKee, B.D. and Handel, M.A. (1993) Sex chromosomes, recombination, and chromatin conformation. *Chromosoma*, **102**, 71–80.
15. Touré, A., Clemente, E.J., Ellis, P.J., Mahadevaiah, S.K., Ojarikre, O.A., Ball, P.A., Reynard, L., Loveland, K.L., Burgoyne, P.S. and Affara, N.A. (2005) Identification of novel Y chromosome encoded transcripts by testis transcriptome analysis of mice with deletions of the Y chromosome long arm. *Genome Biol.*, **6**, R102.
16. Cocquet, J., Ellis, P.J., Yamauchi, Y., Mahadevaiah, S.K., Affara, N.A., Ward, M.A. and Burgoyne, P.S. (2009) The multicopy gene Sly represses the sex chromosomes in the male mouse germline after meiosis. *PLoS Biol.*, **7**, e1000244.
17. She, J.X., Bonhomme, F., Boursot, P., Thaler, L. and Catzeflis, F. (1990) Molecular phylogenies in the genus Mus: comparative analysis of electrophoretic, scnDNA hybridization, and mtDNA RFLP data. *Biol. J. Linn. Soc.*, **41**, 83–103.
18. DeBry, R. and Seshadri, S. (2001) Nuclear intron sequences for phylogenetics of closely related mammals: an example using the phylogeny of Mus. *J. Mammalogy*, **82**, 280–288.
19. Tucker, P.K. (2007) Systematics of the genus Mus. In Fox, J.G., Barthold, S.W., Davisson, M.T., Newcomer, C.E., Quimby, F.W. and Smith, A.L. (eds), *The Mouse in Biomedical Research: History, Wild mice, and Genetics*. Academic Press, Burlington, Massachusetts, Vol. 1, pp. 13–21.
20. Bishop, C.E. and Hatat, D. (1987) Molecular cloning and sequence analysis of a mouse Y chromosome RNA transcript expressed in the testis. *Nucleic Acids Res.*, **15**, 2959–2969.
21. Conway, S.J., Mahadevaiah, S.K., Darling, S.M., Capel, B., Rattigan, A.M. and Burgoyne, P.S. (1994) Y353/B: a candidate multiple-copy spermiogenesis gene on the mouse Y chromosome. *Mamm. Genome*, **5**, 203–210.
22. Ellis, P.J., Ferguson, L., Clemente, E.J. and Affara, N.A. (2007) Bidirectional transcription of a novel chimeric gene mapping to mouse chromosome Yq. *BMC Evol. Biol.*, **7**, 171.
23. Ferguson, L., Ellis, P.J. and Affara, N.A. (2009) Two novel mouse genes mapped to chromosome Yp are expressed specifically in spermatids. *Mamm. Genome*, **20**, 193–206.
24. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
25. Burt, A. and Trivers, R. (2006) *Genes in Conflict: The Biology of Selfish Genetic Elements*. Harvard University Press, Cambridge, Massachusetts.
26. Ellis, P.J. and Affara, N.A. (2006) Spermatogenesis and sex chromosome gene content: an evolutionary perspective. *Hum. Fertil. (Camb).*, **9**, 1–7.
27. Macholán, M., Baird, S.J., Munclinger, P., Dufková, P., Bímová, B. and Piálek, J. (2008) Genetic conflict outweighs heterogametic incompatibility in the mouse hybrid zone? *BMC Evol. Biol.*, **8**, 271.
28. Good, J.M., Handel, M.A. and Nachman, M.W. (2008) Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. *Evolution*, **62**, 50–65.
29. Good, J.M., Giger, T., Dean, M.D. and Nachman, M.W. (2010) Widespread over-expression of the X chromosome in sterile F_1 hybrid mice. *PLoS Genet.*, **6**(pii), e1001148.
30. Casellas, J., Farber, C.R., Verdugo, R.A. and Medrano, J.F. (2010) Segregation analysis of a sex ratio distortion locus in congenic mice. *J. Hered.*, **101**, 351–359.
31. Forejt, J. (1996) Hybrid sterility in the mouse. *Trends Genet.*, **12**, 412–417.
32. Storchová, R., Gregorová, S., Buckiová, D., Kyselová, V., Divina, P. and Forejt, J. (2004) Genetic analysis of X-linked hybrid sterility in the house mouse. *Mamm. Genome*, **15**, 515–524.
33. Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J.C. and Forejt, J. (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science*, **323**, 373–375.
34. Oka, A., Mita, A., Sakurai-Yamatani, N., Yamamoto, H., Takagi, N., Takano-Shimizu, T., Toshimori, K., Moriwaki, K. and Shiroishi, T. (2004) Hybrid breakdown caused by substitution of the X chromosome between two mouse subspecies. *Genetics*, **166**, 913–24.
35. Haldane, J.B.S. (1922) Sex ratio and unisexual sterility in animal hybrids. *J. Genet.*, **12**, 101–109.
36. Hurst, L.D. and Pomiankowski, A. (1991) Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics*, **128**, 841–858.
37. Reynard, L.N., Cocquet, J. and Burgoyne, P.S. (2009) The multi-copy mouse gene Sycp3-like Y-linked (Sly) encodes an abundant spermatid protein that interacts with a histone acetyltransferase and an acrosomal protein. *Biol. Reprod.*, **81**, 250–257.
38. Reynard, L.N. and Turner, J.M. (2009) Increased sex chromosome expression and epigenetic abnormalities in spermatids from male mice with Y chromosome deletions. *J. Cell Sci.*, **122**, 4239–4248.