PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Nature of Gene Evolution on the Mammalian Y Chromosome: Lessons from Sry [and Discussion]

Priscilla K. Tucker, Barbara L. Lundrigan and M. A. Ferguson-Smith

Phil. Trans. R. Soc. Lond. B 1995 **350**, 221-227
doi: 10.1098/rstb.1995.0155

The nature of gene evolution on the mammalian Y chromosome: lessons from *Sry*

PRISCILLA K. TUCKER AND BARBARA L. LUNDRIGAN

The Department of Biology and Museum of Zoology, University of Michigan, Ann Arbor, Michigan 48019, U.S.A.

SUMMARY

With the exception of a small region, heteromorphic sex chromosomes of mammals do not undergo recombination in male meiosis. As a result, the majority of the Y chromosome is clonally transmitted through paternal lineages. Numerous phenomena, including the Hill-Robertson effect, Muller's ratchet, genetic hitch-hiking, and male-driven molecular evolution, are associated with the special transmission properties of the Y chromosome, and can potentially explain the tempo and pattern of gene evolution on the mammalian Y. We explore these phenomena in light of comparative data from the Y-linked sex-determining locus, *Sry*. *Sry* exhibits rapid amino acid divergence between species and little to no variation within species. We find no evidence for directional selection acting on this locus. The pattern of evolution between species is consistent with the Hill-Robertson effect and Muller's ratchet. Lack of variation in *Sry* within species may reflect genetic hitch-hiking, however, we cannot exclude the confounding effects of small effective population size of Y chromosomes. We find no support for male-driven molecular evolution for *Sry* in Old World mice and rats. However, a more appropriate test of this hypothesis would be to compare the evolution of *Sry* to the X-linked *Sox3* gene in these same species. Clearly, more comparative studies of *Sry* and other Y-linked loci are needed to characterize the effects of Y chromosome transmission on the evolution of Y-linked sequences.

1. INTRODUCTION

Molecular evolutionary studies of gene sequences are increasingly utilized to identify functionally significant regions of genetic loci. With some exceptions, highly conserved regions signify functional domains, whereas less conserved regions indicate domains of relaxed functional constraint. The basic assumption behind these studies is that different regions of the genome evolve independently through the processes of mutation, selection and drift. However, theoretical and empirical studies (see below) have shown that linked genetic loci may not evolve independently of one another. These results are especially relevant to comparative studies of genes found on the non-recombining region of the mammalian Y chromosome, where linkage is complete.

In this paper, we discuss various phenomena that have been proposed to explain the evolution of Y-linked genes. These include the Hill-Robertson effect (Hill & Robertson 1966; Felsenstein 1974, 1988), Muller's ratchet (Muller 1918; Felsenstein 1974; Charlesworth 1978; Haigh 1978), genetic hitch-hiking (Maynard-Smith & Haigh 1974; Kaplan *et al.* 1989; Begun & Aquadro 1992), and male-driven molecular evolution (Haldane 1947; Miyata *et al.* 1987). We also consider the possible impact of hemizygosity on the evolution of Y-linked genes, and the consequences of small effective population size of Y chromosomes relative to autosomes. Finally, we examine within- and between-species comparative data from the Y chromosome-linked sex-determining locus, *Sry*, in light of these theoretical considerations. In so doing, we provide a framework in which to interpret the pattern and rate of sequence evolution for Y chromosome-linked genes in mammals.

2. THE EVOLUTION OF HETEROMORPHIC SEX CHROMOSOMES

Muller (1914) hypothesized that the heteromorphic sex chromosomes of *Drosophila* evolved from a homologous pair of chromosomes by suppression of recombination and the subsequent loss of gene function on the chromosome that became the Y. This hypothesis was later recalled to explain the evolution of heteromorphic sex chromosomes in vertebrates, including the evolution of the mammalian X and Y chromosomes (Ohno 1967). Two observations from recent genetic mapping data provide support for this hypothesis. First, all the known functional Y-linked genes isolated from human or mouse are related to genes on the X chromosome (Stevanovic *et al.* 1993; Foster & Graves 1994; Collignon *et al.* 1996; reviewed in Disteche 1995), giving support to the idea that the X and Y chromosomes of mammals evolved from a homologous pair of chromosomes. These loci are effectively non-allelic and divergence in structure and function are possible. Second, in contrast to the X chromosome,

Phil. Trans. R. Soc. Lond. B (1995) **350**, 221–227
Printed in Great Britain

only a few functional loci have been mapped to the mouse Y chromosome (see, for example, Bishop 1993; Brown *et al.* 1993), suggesting that loss of gene function on the Y has taken place. Numerous studies of human and mouse sex chromosomes have shown that, with the exception of a small region that recombines with the X, the Y chromosome does not undergo recombination in male meiosis. The differential segment of the Y chromosome is thus clonally transmitted from father to son.

3. THE EVOLUTION OF THE Y CHROMOSOME

Several phenomena (see, for example, Muller 1918; Fisher 1935; Hamilton 1967; Nei 1970; Charlesworth 1978, 1991; Lucchesi 1978; Rice 1987) have been proposed to explain the evolution of the Y chromosome. Noteworthy among these are three that especially affect chromosome regions where there is no recombination: the Hill-Robertson (H-R) effect (Hill & Robertson 1966; Felsenstein 1974, 1988), Muller's ratchet (Muller 1964; Felsenstein 1974; Charlesworth 1978; Haigh 1978), and genetic hitch-hiking (Maynard-Smith & Haigh 1974; Rice 1987; Birky & Walsh 1988; Kaplan *et al.* 1989).

The Hill-Robertson effect is a phenomena whereby linked genetic loci interfere with each other's response to selection. Birky & Walsh (1988) have shown that the H-R effect results in a decrease in the accumulation of advantageous mutations, but a more rapid accumulation of slightly deleterious mutations, at linked sites over time. Because Y-specific genes are by definition completely linked, they are especially susceptible to the H-R effect and one expects a more rapid accumulation of replacement substitutions, but not silent ones, as the latter are probably neutral and thus not affected by recombination rate (Birky & Walsh 1988; Charlesworth *et al.* 1993*b*). The overall effect at the nucleotide level would be a higher ratio of replacement to silent substitutions for Y-linked loci relative to X-linked and autosomal loci.

Muller's ratchet is a strictly stochastic process whereby the class of non-recombining chromosomes (for example, the mammalian Y) with the fewest number of deleterious mutations is lost from the population (reviewed in Charlesworth *et al.* 1993*b*). Because there is no recombination, and if back mutations are rare, the chromosome class with the fewest number of mutations cannot be recovered. This results in an increase in genetic load on the non-recombining chromosome from one generation to the next. Felsenstein (1988) argues that both the H-R effect and Muller's ratchet result from 'linkage disequilibria randomly generated by genetic drift' in finite populations.

The interaction of linkage and selection at the population level is referred to as genetic hitch-hiking. Hitch-hiking occurs when an advantageous mutation is fixed in a population by selection and carries with it a linked neutral or slightly deleterious mutation. Hitch-hiking is likely to occur in genome regions with little or no recombination (such as the mammalian Y) and can result, especially when the population size is small, in reduced levels of within population variation at linked neutral or slightly deleterious sites, relative to genome regions that undergo recombination. Alternatively, lack of variation could result from background selection against deleterious mutations elsewhere on the Y chromosome (Charlesworth *et al.* 1993*a*).

The transmission of the Y chromosome strictly through paternal lineages can have additional consequences for its evolution. Because Y-linked genes are effectively hemizygous, the expression of a rare allele on the Y chromosome can not be masked by a dominant allele on a homologous chromosome. In contrast to the H-R effect, this should increase the rate at which rare beneficial alleles accumulate and decrease the rate at which deleterious alleles accumulate. However, like genetic hitch-hiking, it has the effect of lowering the production of allelic diversity at the population level (Charlesworth *et al.* 1987; reviewed in Rice 1988).

In addition, because of its' paternal transmission, the effective population size of Y chromosomes relative to X chromosomes and autosomes is significantly reduced. When the male-to-female breeding sex ratio is 1, Y-linked genes are only one-quarter as numerous as autosomes and one-third as numerous as X chromosomes. If the male-to-female breeding sex ratio is greater than one, as is the case for polygynous species, then the effective population size of Y chromosomes relative to autosomes is reduced even further. Given the potentially small population size of Y chromosomes, reduced variation at Y-linked loci relative to X-linked and autosomal loci could result from stochastic bottleneck effects (Nei *et al.* 1975).

The transmission of Y-linked loci through paternal lineages may also produce an increase in mutation rate relative to autosomal and X-linked loci. This hypothesis, referred to as male-driven molecular evolution and first proposed by Haldane (1947), is based on the observation that the number of germ-cell divisions per generation is much higher in the male germ-line than in the female germ-line. Several studies (Miyata *et al.* 1987; Pamilo & Bianchi 1993; Shimmin *et al.* 1993, 1994; Chang *et al.* 1994) in which Y-linked loci were compared with X-linked or autosomal loci, provide evidence to support this hypothesis. Specifically, these studies suggest that the neutral rate of substitution for Y-linked genes is significantly higher than the neutral rate of substitution for autosomal or X-linked genes, and that this effect is more pronounced in species with long generation times (for example, primates), than in species with short generation times (for example, rodents).

4. EMPIRICAL DATA FROM *Sry*

The Y chromosome-linked sex determining locus, identified in 1990 (Sinclair *et al.* 1990; Gubbay *et al.* 1990), has been sequenced in over 20 mammalian species (Sinclair *et al.* 1990; Gubbay *et al.* 1990; Foster *et al.* 1992; Bianchi *et al.* 1993; Tucker & Lundrigan

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

1993; Whitfield *et al.* 1993; Coward *et al.* 1994; Lundrigan & Tucker 1994; Payen & Cotinot 1994). A fewer number of population level surveys of *Sry* have also been conducted (Lundrigan & Tucker 1994; Nachman & Aquadro 1994; Miller *et al.* 1995). Taken together, these comparative data permit description of *Sry* evolution in mammals, and allow us to examine whether the observed patterns are consistent with the H-R effect, Muller's ratchet, genetic hitch-hiking, and male-driven molecular evolution.

(*a*) *Species level comparisons*

Sry contains a single exon consisting of a central DNA-binding motif (HMG box), 79 amino acids in length, and flanking sequences. The DNA sequence of the HMG box is similar across species of marsupial (infraclass Metatheria) and placental (infraclass Eutheria) mammals (Gubbay *et al.* 1990; Sinclair *et al.* 1990; Foster *et al.* 1992; Bianchi *et al.* 1993; Tucker & Lundrigan 1993; Whitfield *et al.* 1993; Coward *et al.* 1994; Hacker *et al* 1995; Lundrigan & Tucker 1994; Payen & Cotinot 1994). However, there is little to no sequence similarity in flanking regions between these two infraclasses, or among orders within each infraclass. Flanking sequence varies in both length and composition. For example, although the N-terminal region of *Sry* in primates is 58 amino acids long, it is only two amino acids long in species belonging to the rodent family Muridae, subfamily Murinae. The C-terminal region in primates varies in length from 70 amino acids in humans, to 94 amino acids in marmosets (Whitfield *et al.* 1993), and in Old World mice and rats (subfamily Murinae), from 92 amino acids in *Hylomyscus alleni*, to 313 amino acids in *Mus musculus* (Tucker & Lundrigan 1993; Hacker *et al.* 1995). Significant length variation in the C-terminal region is evident even between sibling species of house mice, *M. musculus* and *M. domesticus*: the C-terminal region of *Sry* in laboratory mice carrying a *M. musculus* Y chromosome is 313 amino acids in length (Hacker *et al.* 1995). This same region in laboratory mice carrying a *M. domesticus* Y chromosome ranges in length from 153 to 155 amino acids (Coward *et al.* 1994). The difference in size between *M. musculus* and *M. domesticus Sry* sequences is due to a C to T transitional substitution at nucleotide position 9906 (numbers refer to the base pair position of *Sry* in GenBank entry X67204) resulting in the replacement of a glutamine in *M. musculus* with a stop codon in *M. domesticus*.

The flanking regions also show a markedly different pattern of substitution than the HMG box domain. For example, a comparison of the pattern of substitutions between human and marmoset indicates a much higher ratio of replacement to silent substitutions (dn/ds) in flanking regions (dn/ds = 0.44 for N-terminus, dn/ds = 0.79 for C-terminus) than in the HMG box (dn/ds = 0.12) (Whitfield *et al.* 1993). A similar pattern is observed in comparisons of other primate species, and in comparisons of Old World mice and rats. For example, in a comparison of the house mouse species, *M. musculus*, with the African wood mouse, *Hylomyscus alleni*, the ratio of replacement to silent substitutions (dn/ds) in the C-terminal region is 0.41, while in the HMG box it is only 0.21 (Tucker & Lundrigan 1993). It is interesting to note that although the dn/ds ratios in the HMG box are lower than those in the flanking regions, they are high relative to many other protein encoding genes. For example, we noted from a comparative study of 28 loci from laboratory mouse and rat (O'hUigin & Li 1992) that only two loci had a higher dn/ds than the HMG box. These data suggest that the amino acid sequence of *Sry* is diverging rapidly relative to most other protein encoding genes.

There are several evolutionary scenarios that could explain this result. First, it is possible that *Sry* is undergoing adaptive change. For example, it has been hypothesized that amino acid changes in *Sry* act as a reproductive isolating mechanism between species because if these changes cause *Sry* to malfunction on a hybrid genetic background, the result would be an all female hybrid population in which XY females are likely to be sterile. There is evidence at the molecular level that directional selection in abalone sperm lysins may have evolved as a 'pre-mating' reproductive isolating mechanism (Lee & Vacquier 1992; Lee *et al.* 1995). However, it is difficult to imagine how directional selection could operate as a 'post-mating' reproductive isolating mechanism, as would have to be the case for *Sry*. Another adaptive explanation is that rapid sequence evolution of *Sry* may result from maternal/foetal conflict in which an arms race develops between foetally expressed selfish Y-linked genes and maternally derived suppressors (Hurst 1994*a*,*b*). However, no mechanism for this has, as yet, been identified.

Following Hughes *et al.* (1990), we tested the hypothesis that selection is acting on certain characteristics of the *Sry* amino acid sequence (charge, polarity, and functional characteristics as defined by Miyata *et al.* 1979) by examining whether replacement substitutions resulted in conservative or radical amino acid changes with respect to these characteristics. The underlying assumption of this analysis is the following: if amino acid substitutions take place at random with respect to a particular characteristic (e.g. charge), we would expect the proportion of conservative (i.e. same charge) and radical (i.e. different charge) amino acid replacements to be equal. If there are significantly more conservative than radical changes, we could argue that a particular amino acid characteristic is adaptively constrained, and if there are significantly more radical than conservative, then it is under directional selection. We focused our analysis on a comparison between *M. musculus* and *H. alleni*, two species known to carry only a single copy of *Sry* (Nagamine 1994; Tucker & Lundrigan, personal observation). The results (see table 1) indicate some conservation of charge in the HMG box ($p < 0.05$), but no other evidence for selection, at least when comparing broadly defined regions such as the HMG box and C-terminus.

A second possible explanation for the relatively high number of amino acid replacement substitutions in *Sry* is the H-R effect. Recall that the H-R effect is likely to be most pronounced in genome regions with little or no recombination. Slightly deleterious mutations are

Table 1. *Conservative and radical amino acid substitutions with respect to charge, polarity and functional group* (*Miyata* et al. *1979*) *in* Sry *between two species of murine rodent*, Mus musculus *and* Hylomyscus alleni

(Calculations were made using the program SCR provided by A. Hughes. Separate comparisons are made for the N-terminus/HMG box and C-terminus (to amino acid position 137). The total number of silent substitutions per potential number of silent sites is 6.00/48.42 for the N-terminus/HMG box and 6.67/34.92 for the C-terminus. The total number of replacement substitutions per potential number of replacement sites is 5.00/188.58 for the N-terminus/HMG box and 11.33/118.08 for the C-terminus. Standard errors are given in parentheses. Probability values (p) were determined by Fisher's Exact Test.)

	type of amino acid change	no. of sites	no. of hits	proportion of differences	significance
charge					
N-terminus/HMG box					
	conservative	91.25	5.00	0.0548 (0.0238)	
	radical	97.33	0.00	0.000 (0.0000)	$p = 0.03$
C-terminus					
	conservative	66.67	9.33	0.1400 (0.0425)	
	radical	51.42	2.00	0.0389 (0.0270)	$p = 0.06$
polarity					
N-terminus/HMG box					
	conservative	129.25	4.00	0.0309 (0.0152)	
	radical	59.33	1.00	0.0169 (0.0167)	$p = 0.34$
C-terminus					
	conservative	75.50	5.33	0.0706 (0.0295)	
	radical	42.58	6.00	0.1409 (0.0533)	$p = 0.11$
functional group					
N-terminus/HMG box					
	conservative	53.00	2.00	0.0377 (0.0262)	
	radical	135.58	3.00	0.0221 (0.0126)	$p = 0.30$
C-terminus					
	conservative	35.00	6.33	0.1810 (0.0651)	
	radical	83.08	5.00	0.0602 (0.0261)	$p = 0.05$

expected to accumulate as a result of being linked to an adaptive mutation. This will result in a faster rate of accumulation of slightly deleterious mutations over time than would be expected for recombining sequences (Birky & Walsh 1988). As Charlesworth *et al.* (1993*b*) noted, these slightly deleterious mutations are likely to be replacement substitutions. Presumably, the accumulation of slightly deleterious mutations will be most pronounced in regions of relaxed functional constraint. Indeed, the higher dn/ds ratio in the C-terminus than in the HMG box (Tucker & Lundrigan 1993; Whitfield *et al.* 1993) may reflect the fact that the former is under relaxed functional constraints relative to the latter. We know from in vitro studies that the HMG box can bind specific DNA sequences (Nasrin *et al.* 1991; Ferrari *et al.* 1992; King & Weiss 1993; Natesan & Gilman 1993; Harley *et al.* 1994), and mutations in the HMG box have been associated with gonadal dysgenesis (reviewed in Hawkins 1993; Poulat *et al.* 1994). We know less about the action of the N-terminal and C-terminal regions. However, given their rapid evolution, it is difficult to imagine that these regions are under strict functional constraint.

Expression studies of mouse and human have shown that, in *M. musculus musculus*, *Sry* has multiple start codons, and a TCTG microsatellite in the 5′ untranslated region that could form secondary structure (Hacker *et al.* 1995) and may result in inefficient translation. In human, there is only one start codon, but its non-optimal position may similarly result in inefficient translation (Vilain *et al.* 1992; Behlke *et al.* 1993; Clepet *et al.* 1993; Su & Lau 1993). This may reflect increased genetic load on the Y chromosome and is consistent with the process of Muller's ratchet.

If the pattern of evolution on the Y chromosome is shaped by the H-R effect, the relatively rapid rate of replacement substitutions will not be unique to *Sry*. Other genes in the non-recombining region of the Y should show the same pattern. Two kinds of studies are needed to determine whether the pattern of evolution for *Sry* is gene-specific or characteristic of the entire non-recombining portion of the Y chromosome. First, to minimize gene-specific effects, an examination of *Sox3*, the putative ancestral homologue of *Sry* (Stevanovic *et al.* 1993; Foster & Graves 1994; Collignon *et al.* 1996), from the same species used in the *Sry* studies, is needed. Some data are already available. Comparisons across broad taxonomic groups (marsupials, primates, and rodents) indicate that, in contrast to *Sry*, *Sox3* shows a highly conserved pattern of sequence evolution (Stevanovic *et al.* 1993; Foster & Graves 1994; Collignon *et al.* 1996). Second, comparative studies of other Y-linked genes are needed to test the generality of the results from *Sry*. In a comparative study of exon sequences from the zinc finger genes (*Zfx* and *Zfy*) in species of the genus *Mus*, we found that the dn/ds ratio is higher for *Zfy* than for *Zfx* (R. Adkins & P. Tucker, unpublished data).

Finally, determination of the neutral rate of mutation in *Sry* and *Sox3* for the same pair of species would provide a test for the theory of male-driven molecular evolution. The absence of intron sequences in *Sry* and

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Sox3 limits this analysis to synonymous sites which may or may not be under selection. These data have not been collected for rodents or primates. We did find, however, that the synonymous rate of change in *Sry* between *M. musculus* and *Rattus exulans* fell well within the range of synonymous rates for 28 X-linked or autosomal loci compared for the same genera (Tucker & Lundrigan 1993). This suggests that male-driven molecular evolution has not markedly increased the rate of neutral change on the Y chromosome. However, this hypothesis needs more rigorous testing. Our comparison suffers from two weaknesses: first, we compared *Sry* to non-homologous loci with different nucleotide compositions (Wolfe & Sharp 1993); and second, our comparison of *Sry* between *M. musculus* and *R. exulans* may not have been between orthologous copies of *Sry* as multiple copies of *Sry* are found on the Y chromosome of *R. exulans* (P. Tucker & B. Lundrigan, personal observations).

(*b*) *Population-level studies*

Population level variation in *Sry* has been less extensively studied than between species variation. We examined 515 base pairs of unique sequence DNA from *Sry*, including 86 base pairs of the 5′ untranslated region, the N-terminus, the HMG box, and 186 base pairs of the C-terminus in several species of *Mus* (Lundrigan & Tucker 1994). Our samples included representatives from four populations of *M. musculus* and six populations of *M. domesticus*. We found only two base pair substitutions (one replacement and one silent substitution) among the four geographically disjunct populations of *M. musculus*. No variation was found among the six geographically disjunct populations of *M. domesticus* (Lundrigan & Tucker 1994). However, in a subsequent comparison that included the C-terminal repeat region of *M. domesticus*, we did find variation among populations in the number of CAG repeats (Miller *et al.* 1995). Greater variation in repeat regions than in regions of unique sequence has been documented in other genes, and probably reflects their more rapid rate of mutation (Richards & Sutherland 1992, 1994). There are many possible interpretations for the low level of polymorphism in our sequence data. For example, the low level of variation in *M. musculus* and *M. domesticus* may reflect recent bottlenecks. This, however, seems an unlikely explanation for the absence of polymorphism in unique sequence among our six populations of *M. domesticus* as variation in allozymes and mitochondrial DNA occurs throughout the range of this species (Britton-Davidian 1990; Prager *et al.* 1993). Low polymorphism may also indicate the presence of functional constraints, however, the absence of substitutions at synonymous sites, coupled with the rapid evolution of this gene between species, makes this interpretation also unlikely. A more plausible explanation may be found in the special transmission properties of Y chromosomes. Both the small effective population size of Y chromosomes and their clonal inheritance (and consequent susceptibility to genetic hitch-hiking) will reduce polymorphism of Y-linked loci relative to autosomal and X-linked loci. To test for genetic hitch-hiking on the Y chromosome in *M. domesticus* and *Mus spretus*, Nachman & Aquadro (1994) compared levels of polymorphism to divergence for 1036 b.p. of non-coding DNA flanking *Sry* and 887 b.p. of the mitochondrial D-loop. The neutral expectation is that the ratio of polymorphism to divergence will be the same for different gene regions unless selection is acting on the gene or a closely linked locus (Hudson *et al.* 1987). There was no evidence for genetic hitch-hiking on the Y chromosome in *M. domesticus* (i.e. the proportion of polymorphism to divergence was the same for *Sry* and the D-loop). Their test was biased in favour of rejection of the neutral hypothesis because the effective population size of males is probably much smaller than females in *M. domesticus*. Nevertheless, their test may be a conservative estimate for two reasons. First, both the levels of polymorphism and divergence are small as the species used in their study are recently derived. Second, neither of the gene regions compared (Sry flanking sequence and the mitochondrial DNA control region) undergo recombination. They are, thus, equally susceptible to the effects of hitch-hiking.

5. CONCLUSION

Comparative data from population level and between species surveys of the sex-determining locus, *Sry*, show rapid gene evolution between species and little to no sequence variation within species. These data are consistent with predictions of the H-R effect, Muller's ratchet, and genetic hitch-hiking. One conclusion of our between-species study is that, because Y-specific genes are completley linked, selection can not operate efficiently on the non-recombining portion of the Y chromosome. However, to rigorously test whether the observed between-, as well as within-, species patterns of variation in *Sry* result from the unique transmission properties of the Y chromosome, more Y-linked genes, along with autosomal sequences, should be surveyed from a variety of species. For example, a study comparing species of ancient origin is likely to provide more nucleotide variation, and thus more information on the interaction between linkage and selection on the evolution of the Y, than a study of recently derived species such as *M. musculus* and *M. domesticus*. We must also come to better understand the effects of population size of Y chromosomes within species as this may confound any interpretation of population level variation. To do this, population level variation of Y-linked and autosomal sequences should be sampled from species with distinct mating systems (e.g. polygynous versus monogamous). These data should then be juxtaposed with information on between species divergence. In sum, these kinds of studies are essential to understanding evolutionary pattern and process for clonally inherited nuclear genes. They also provide an essential framework in which to interpret comparative data for the purposes of understanding gene structure and function on the mammalian Y.

We thank R. Lovell-Badge for providing papers in preparation and in press before the writing of this manuscript. We are especially grateful to C.W. Birky who provided many

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

helpful comments on an earlier version of this manuscript. The paper was supported by an NSF grant to P.K.T.

REFERENCES

Begun, D. J. & Aquadro, C. F. 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature*, *Lond*. **356**, 519–520.

Behlke, M. A., Bogan, J. S., Beer-Romero, P. & Page, D. 1993 Evidence that the SRY protein is encoded by a single exon on the human Y chromosome. *Genomics* **17**, 736–739.

Bianchi, N. O., Bianchi, M. S., Bailliet, G. & de la Chapelle, A. 1993 Characterization and sequencing of the sex determining region Y gene (*Sry*) in *Akodon* (Cricetidae) species with sex reversed females. *Chromosoma* **102**, 389–395.

Birky, C. W. Jr & Walsh, J. B. 1988 Effects of linkage on rates of molecular evolution. *Proc. natn Acad. Sci. U.S.A.* **85**, 6414–6418.

Bishop, C. E. 1993 Mouse Y chromosome. *Mamm. Genome* **4**, 282–283. (Suppl.)

Britton-Davidian, J. 1990 Genic differentiation in M. m. domesticus populations from Europe, the Middle East, and North Africa: geographic patterns and colonization events. *Biol. J. Linn. Soc.* **41**, 27–45.

Brown, S. D. M., Avner, P., Boyd, Y., Chapman, V., Rastan, S., Sefton, L. Thomas, J. D. & Herman, G. E. 1993 Mouse X chromosome. *Mamm. Genome* **4**, 269–281. (Suppl.)

Chang, B. H.-J., Shimmin, L., Shyue, S.-K., Hewett-Emmett, D. & Li, W.-H. 1994 Weak male-driven molecular evolution in rodents. *Proc. natn Acad. Sci. U.S.A.* **91**, 827–831.

Charlesworth, B. 1978 Model for evolution of Y chromosomes and dosage compensation. *Proc. natn Acad. Sci. U.S.A.* **75**, 5618–5622.

Charlesworth, B. 1991 The evolution of sex chromosomes. *Science*, *Wash.* **251**, 1030–1033.

Charlesworth, B., Coyne, J. A. & Barton, N. H. 1987 The relative rates of evolution of sex chromosomes and autosomes. *Am. Nat.* **130**, 113–146.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. 1993*a* The effect of deleterious mutations on neutral molecular evolution. *Genetics* **134**, 1289–1303.

Charlesworth, D., Morgan, M. T. & Charlesworth,B. 1993*b* Mutation accumulation in finite outbreeding and inbreeding populations. *J. Hered.* **84**, 321–325.

Clepet, C., Schafer, A. J., Sinclair, A. H. *et al.* 1993 The human *SRY* transcript. *Hum. molec. Genet.* **2**, 2007–2012.

Collignon, J., Hacker, A. & Sockanathan, S. *et al.* 1996 *Sox-3*, an X-linked gene closely related to *Sry*, shows genital ridge and CNS expression. *Development*. (In the press.)

Coward, P., Nagai, K., Chen, D., Thomas, H. D., Nagamine, C. M. & Lau, Y.- F. C. 1994 Polymorphism of a CAG trinucleotide repeat within *Sry* correlates with B6.^{Ydom} sex reversal. *Nature Genet.* **6**, 245–250.

Disteche, C. M. 1995 Escape from X inactivation in human and mouse. *Trends Genet.* **11**, 17–22.

Felsenstein, J. 1974 The evolutionary advantage of recombination. *Genetics* **78**, 737–756.

Felsenstein, J. 1988 Sex and the evolution of recombination. In *The evolution of sex* (ed. R. E. Michod & B. R. Levin), pp. 74–86. Sunderland, Massachusetts: Sinauer Associates Inc.

Ferrari, S., Harley, V. R., Pontiggia, A., Goodfellow, P. N., Lovell-Badge, R. & Bianchi, M. E. 1992 SRY, like HMG1, recognizes sharp angles in DNA. *EMBO J.* **11**, 4497–4506.

Fisher, R. A. 1935 The sheltering of lethals. *Am. Nat.* **69**, 446–455.

Foster, J. W., Brennan, F. E., Hampikian, G. K. *et al.* 1992 Evolution of sex determination and the Y chromosome: *SRY*-related sequences in marsupials. *Nature*, *Lond.* **359**, 531–533.

Foster, J. W. & Marshall Graves, J. A. 1994 An *SRY*-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene. *Proc. natn Acad. Sci U.S.A.* **91** 1927–1931.

Gubbay, J., Collignon, P., Koopman, P. *et al.* 1990 A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature*, *Lond.* **346**, 245–250.

Hacker, A., Capel, B., Goodfellow, P. & Lovell-Badge, R. 1995 Expression of *Sry*, the mouse sex determining gene. *Development* **121**, 1603–1614.

Haigh, J. 1978 The accumulation of deleterious genes in a population- Muller's ratchet. *Theor. Popul. Biol.* **14**, 251–267.

Haldane, J. B. S. 1947. The mutation rate of the gene for haemophilia and its segregation ratios in males and females. *Ann. Eugen.* **13**, 262–271.

Hamilton, W. D. 1967 Extraordinary sex ratios. *Science*, *Wash.* **156**, 477–488.

Harley, V. R., Lovell-Badge, R. & Goodfellow, P. N. 1994 Definition of a consensus DNA binding site for SRY *Nucl. Acids Res.* **22**, 1500–1501.

Hawkins, J. R. 1993 Mutational analysis of *SRY* in XY females. *Hum. Mut.* **2**, 347–350.

Hill, W. G. & Robertson, A. 1966 Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**, 226–231.

Hudson, R. R., Kreitman, M. & Aguade, M. 1987 A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Hughes, A. L., Ota, T. & Nei, M. 1990 Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Molec. Biol. Evol.* **7**, 515–524.

Hurst, L. D. 1994*a* Embryonic growth and the evolution of the mammalian Y chromosome. I. The Y as an attractor for selfish growth factors. *Heredity* **73**, 223–232.

Hurst, L. D. 1994*b* Embryonic growth and the evolution of the mammalian Y chromosome. II. Suppression of the selfish Y-linked growth factors may explain escape from X-inactivation and rapid evolution of *Sry*. *Heredity* **73**, 233–243.

Kaplan, N. L., Hudson, R. R. & Langley, C. H. 1989 The 'Hitchhiking Effect' revisited. *Genetics* **123**, 887–899.

King, C.-Y. & Weiss, M. A. 1993 The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: a topological mechanism of sequence specificity. *Proc. natn Acad. Sci. U.S.A.* **90**, 11990–11994.

Lee, Y.-H., Ota, T. & Vacquier, V.D. 1995 Positive selection is a general phenomenon in the evolution of abalone sperm lysin. *Molec. Biol. Evol.* **12**, 231–238.

Lee, Y.-H. & Vacquier, V. D. 1992 The divergence of species-specific abalone sperm lysins is promoted by positive Darwinian selection. *Biol. Bull.* **182**, 97–104.

Lucchesi, J. C. 1978. Gene dosage compensation and the evolution of sex chromosomes. *Science*, *Wash.* **202**, 711–716.

Lundrigan, B. L. & Tucker, P. K. 1994 Tracing paternal ancestry in mice, using the Y-linked sex-determining locus, *Sry*. *Molec. Biol. Evol.* **11**, 483–492.

Maynard-Smith, J. & Haigh, J. 1974 The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35.

Miller, K. E., Lundrigan, B. L. & Tucker, P. K. 1995 Length variation of CAG repeats in *Sry* across populations of *Mus domesticus*. *Mamm. Genome* **6**, 206–208.

Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K. & Yasunaga, T. 1987 Male-driven molecular evolution: a

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

model and nucleotide sequence analysis *Cold Spring Harb. Symp. quant. Biol.* **52**, 863–867.
Miyata, Y., Miyazawa, S. & Yasunaga, T. 1979 Two types of amino acid substitutions in protein evolution. *J. molec. Evol.* **12**, 219–236.
Muller, H. J. 1914 A gene for the fourth chromosome of Drosophila. *J. exp Zool.* **17**, 325–336.
Muller, H. J. 1918 Genetic variability, twin hybrids and constant hybrids, in a case of balanced lethal factors. *Genetics* **3**, 422–499.
Muller, H. J. 1964 The relation of recombination to mutational advance. *Mut. Res.* **1**, 2–9.
Nachman, M. W. & Aquadro, C. F. 1994 Polymorphism and divergence at 5′ flanking region of the sex determining locus, *Sry*, in mice. *Molec. Biol. Evol.* **11**, 539–547.
Nagamine, C. M. 1994 The testis-determining gene, *SRY*, exists in multiple copies in Old World rodents. *Genet. Res.* **64**, 151–159.
Nasrin, N., Buggs, C., Kong, X. F., Carnazza, J., Goebl, M. & Alexander-Bridges, M. 1991 DNA-binding properties of the product of the testis-determining gene and a related protein. *Nature, Lond.* **354**, 317–320.
Natesan, S. & Gilman, M. Z. 1993 DNA bending and orientation-dependent function of YY1 in the c-*fos* promoter. *Genes Dev.* **7**, 2497–2509.
Nei, M. 1970 Accumulation of nonfunctional genes on sheltered chromosomes. *Am. Nat.* **104**, 311–322.
Nei, M., Maruyama, T. & Chakraborty, R. 1975 The bottleneck and genetic variability in populations. *Evolution* **29**, 377–384.
Ohno, S. 1967 *Sex chromosomes and sex-linked genes*. pp.1–192. New York: Springer-Verlag.
O'hUigin, C. & Li, W.-H. 1992 The molecular clock ticks regularly in muroid rodents and hamsters. *J. molec. Evol.* **35**, 377–384.
Pamilo, P. & Bianchi, N. O. 1993 Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. *Molec. Biol. Evol.* **10**, 271–281.
Payen, E. J. & Cotinot, C. Y. 1994 Sequence evolution of SRY gene within *Bovidae* family. *Mamm. Genome* **5**, 723–725.
Poulat, F., Soullier, S., Goze, C., Heitz, F., Calas, B. & Berta, P. 1994 Description and functional implications of a novel mutation in the sex-determining gene SRY. *Hum. Mut.* **3**, 200–204.
Prager, E. M., Sage, R. D., Gyllensten, U. *et al.* 1993 Mitochondrial DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. *Biol. J. Linn. Soc.* **50**, 85–122.
Rice, W. R. 1987 Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics* **116**, 161–167.
Rice, W. R. 1988 The effect of sex chromosomes on the rate of evolution. *Trends Evol. Ecol.* **3**, 2–3.
Richards, R. I. & Sutherland, G. R. 1992 Heritable unstable DNA sequences. *Nature Genet.* **1**, 7–9.
Richards, R. I. & Sutherland, G. R. 1994 Simple repeat DNA is not replicated simply. *Nature Genet.* **6**, 114–116.
Shimmin, L. C., Chang, B. H.-J. & Li, W.-H. 1993 Male-driven evolution of DNA sequences. *Nature, Lond.* **362**, 745–747.
Shimmin, L. C., Chang, B. H.-J. & Li, W.-H. 1994 Contrasting rates of nucleotide substitution in the X-linked and Y-linked zinc finger genes. *J. molec. Evol.* **39**, 569–578.
Sinclair, A. H., Berta, P., Palmer, M. S. *et al.* 1990 A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature, Lond.* **346**, 240–244.
Stevanovic, M., Lovell-Badge, R., Collignon, J. & Goodfellow, P. N. 1993 *SOX3* is an X-linked gene related to *SRY*. *Hum. molec. Genet.* **2**, 2013–2018.
Su, H. & Lau, Y.-H. C. 1993 Identification of the transcriptional unit, structural organization, and promoter sequence of the human sex-determining region Y (SRY) gene, using a reverse genetic approach. *Am. J. Hum. Genet.* **52**, 24–38.
Tucker, P. K. & Lundrigan, B. L. 1993 Rapid evolution of the sex determining locus in Old World mice and rats. *Nature, Lond.* **364**, 715–717.
Vilain, E., Fellous, M. & McElreavey, K. 1992 Characterization and sequence of the 5′ flanking region of the human testis-determining factor *Sry*. *Methods molec. cell. Biol.* **3**, 128–134.
Whitfield, L. S., Lovell-Badge, R. & Goodfellow, P. N. 1993 Rapid sequence evolution of the mammalian sex-determining gene *SRY*. *Nature, Lond.* **364**, 713–715.
Wolfe, K. H. & Sharp, P. M. 1993 Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. *J. molec. Evol.* **37**, 441–456.

Discussion

M. A. Ferguson-Smith (*Cambridge University, Department of Pathology, U.K.*). It is possible that sex reversal in inbred C57BL/6J which also carry a *Mus domesticus poschiavinus* Y chromosome is related to variation in the number of repeats in the C-terminal region of the *Sry* gene. Does Dr Tucker know of any other *Sry* variants of similar type known to cause sex reversal when placed in a *Mus musculus* background, perhaps through insufficient binding of the HMG domain to target sites?

P. K. Tucker. The inbred mouse strain, C57Bl/6J, like many inbred strains, is a mixture of both *Mus musculus* and *Mus domesticus* genomes, but carries a *M. musculus* Y chomosome. Sex reversal in C57Bl/6J mice carrying a *Mus domesticus poschiavinus* Y chromosome has been correlated with the number of repeats in a specific region of the C-terminus of *Sry* in *Mus domesticus poschiavinus*. It has not yet been determined whether sex reversal in C57Bl/6J mice carrying other *Mus domesticus* Y chromosomes is also correlated with variation in the number of repeats in this same region of *Sry*. As stated in my paper, the correlation between repeat length and sex reversal may be coincidental reflecting allelic variation elsewhere in the gene.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES