THE ROYAL SOCIETY

An apparent excess of sex- and reproduction-related genes on the human X chromosome

G. Mustafa Saifi[1] **and H. Sharat Chandra**[1,2*]

[1]*Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560 012, India*
[2]*Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur Post, Bangalore 560 064, India*

We describe here the results of a search of *Mendelian inheritance in man*, GENDIAG and other sources which suggest that, in comparison with autosomes 1, 2, 3, 4 and 11, the X chromosome may contain a significantly higher number of sex- and reproduction-related (SRR) genes. A similar comparison between X-linked entries and a subset of randomly chosen entries from the remaining autosomes also indicates an excess of genes on the X chromosome with one or more mutations affecting sex determination (e.g. *DAX1*), sexual differentiation (e.g. *androgen receptor*) or reproduction (e.g. *POF1*). A possible reason for disproportionate occurrence of such genes on the X chromosome could be that, during evolution, the 'choice' of a particular pair of homomorphic chromosomes for specialization as sex chromosomes may be related to the number of such genes initially present in it or, since sex determination and sexual dimorphism are often gene dose-dependent processes, the number of such genes necessary to be regulated in a dose-dependent manner. Further analysis of these data shows that XAR, the region which has been added on to the short arm of the X chromosome subsequent to eutherian-marsupial divergence, has nearly as high a proportion of SRR genes as XCR, the conserved region of the X chromosome. These observations are consistent with current hypotheses on the evolution of sexually antagonistic traits on sex chromosomes and suggest that both XCR and XAR may have accumulated SRR traits relatively rapidly because of X linkage.

Keywords: sex chromosome evolution; sexual dimorphism; X inactivation; XY heteromorphism; gene dosage; X-chromosome conservation

1. INTRODUCTION

In *Drosophila melanogaster*, the X chromosome forms slightly less than one-third of the haploid genome, and the autosomes, over two-thirds. Sex lethal (*Sxl*), the master gene whose activity state determines fly sex, is located on the X chromosome. Among the known regulators of *Sxl*, five, or possibly six, are on the X and seven are on the autosomes (Flybase, http://morgan.harvard.edu; Ryner & Swain 1995; Cline 1993; Baker 1989). In butterflies (ZZ, male; ZW, female), a disproportionate number of genes related to sexuality, reproduction and speciation are located on the Z chromosome, which forms approximately one-sixtieth of the genome in females. Female mate-selection behaviour, male courtship signals, female limitation of colour polymorphism and mimicry are thought to result largely from interactions between autosomal genes and uncompensated Z-linked regulatory genes (Stehr 1959; Sheppard 1961; Cock 1964; Grula & Taylor 1980). In humans, it has been known for some time that *TFM*, the gene for testicular feminization (Meyer *et al*. 1975) and several genes for ovarian development (see Ogata & Matsuo 1995) are X-linked. These findings, as well as the discovery that the closest relative of *SRY* is X-linked (Stevanovic *et al*. 1993) and reports of apparent linkage between the Xq28 region and sexual orientation in human males (Hamer *et al*. 1993; Hu *et al*. 1995), prompted us to ask whether there might be a similar concentration of genes related to sexuality and reproduction on the human X chromosome.

*Author for correspondence (sharat@mcbl.iisc.ernet.in).

2. MATERIALS AND METHODS

The scope of the terms sex determination, sexual differentiation and reproduction as used in this study is as follows: those genes in which a mutation results in gonadal sex reversal (presence of ovaries or atrophic gonads in an XY individual or testis or ovotestis in an XX individual) are termed as sex determining genes (e.g. *DAX1*), those that result in abnormalities of the genitalia (without gonadal sex reversal) as sexual differentiation genes (e.g. *androgen receptor*) and those that lead to infertility as genes involved in reproduction (e.g. *POF1*). We have also included under 'sexual differentiation genes' mutations which result in gynaecomastia or other types of abnormal breast development. All other entries were treated as not being related to sex or reproduction ('non-SRR') for the purposes of this study. A small number of genes contributing to certain aspects of sexual dimorphism, such as hair distribution, teeth development, etc., were also included in the latter category because they do not affect sexuality or reproduction in any direct manner. For the same reason, genes showing sex-limited expression but no direct connection to sex determination, sexual differentiation or reproduction were included in the non-SRR category. Needless to say, in classifying these genes as sex- or reproduction-related

© 1999 The Royal Society
Received 15 September 1998 *Accepted* 30 September 1998

or otherwise, the same criteria were applied to both X-linked and autosomal entries in the databases (see below). It may also be noted that in many of the entries considered, the effect of the relevant gene on sexual development or reproduction was only one of several pleiotropic effects of the gene and not necessarily the 'primary' effect.

(a) *Database search*

Entries in *Mendelian inheritance in man* (MIM) (McKusick 1992) were examined to identify loci with one or more mutations affecting sex determination, sexual differentiation or reproduction. Information from MIM on each of these entries was updated using *Online mendelian inheritance in man* (OMIM, (http://www.ncbi.nlm.nih.gov/repository/OMIM/omim.txt.Z), October 1996). Information on the chromosomal location of OMIM entries given in genmap was also downloaded (http://www.ncbi.nlm.nih.gov/repository/OMIM/genmap) and a program was written to identify from this file entry numbers pertaining to each chromosome. OMIM information given against entries from chromosomes 1, 2, 3 and X was individually examined for an abnormality in sex determination, sexual differentiation or reproduction. Those entries that contained at least one such abnormality were segregated. These data were supplemented from GENDIAG (a database similar to OMIM, (http://www.infobiogen.fr/services/gendiag/)) and other sources (Schinzel 1984; Ogata & Matsuo 1995; Tar *et al*. 1995; Thode *et al*. 1988; Sala *et al*. 1997). Among the entries so segregated, each of those identified by an asterisk in MIM or OMIM was considered as representing a distinct locus and was included for statistical analysis. The non-asterisked entries were excluded for reasons given under § 2c and also in § 3 and 4. A list of such validated entries (marked with an asterisk) is given in Appendix A. The incompletely validated entries (non-asterisked) are given in Appendix B. Further details of this search and classification are given in footnotes to Appendices A and B. (These can be found on the Royal Society Web site at (http://www.pubs.royalsoc.ac.uk/publish/pro_bs/jan99pb2.htm).)

(b) *Disease-associated loci (OMIM)*

A list of disease-associated loci on autosomes 1, 2, 4 and 11 was compiled from international workshops on these chromosomes (Weith *et al*. 1996; Spur *et al*. 1996; Riess *et al*. 1996; Shows *et al*. 1996). A similar list was compiled for chromosomes 1, 3 and X using information in the genmap file. The two lists were merged, and the information in OMIM and GENDIAG on each of these entries was individually examined. If at least one mutation in the locus identified with an entry was associated with an anomalous phenotype, that entry was segregated into a separate category. From among such disease-associated loci, those in which there was an abnormality in sex determination, sexual differentiation or reproduction were identified.

(c) *Disease-associated loci (GENDIAG)*

Unlike OMIM, which contains a description of both genes and genetic disorders, GENDIAG is a database only of genetic disorders. It employs specific keywords to describe a particular trait and a combination of these traits is used to define a syndrome. Keywords that specifically relate to entries with SRR anomalies can thus be obtained from GENDIAG. By individually examining each GENDIAG entry pertaining to chromosomes 1, 2, 3, 4, 11 and X, a list of SRR keywords was prepared. We have assumed that this list represents nearly all SRR keywords. However, it is possible that a few SRR keywords described only in entries pertaining to one or more of the remaining autosomes may have been missed out.

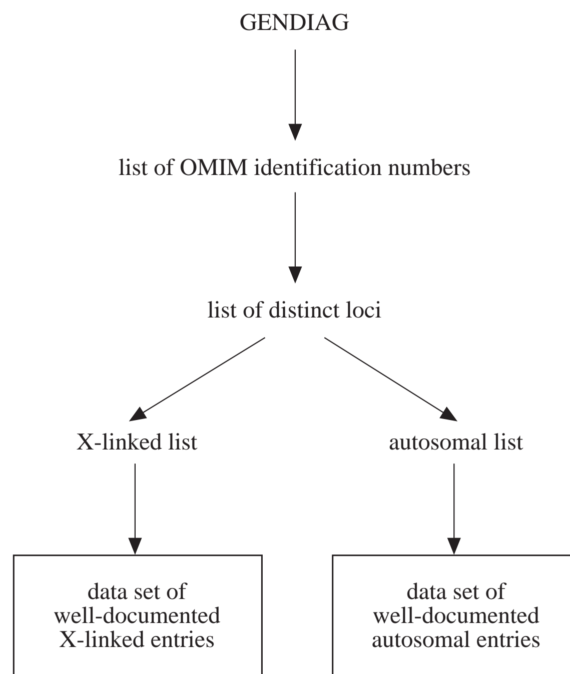

Figure 1. Procedure employed for obtaining X-linked and autosomal datasets of phenotypically well-documented distinct loci (additional details are given in § 2c).

GENDIAG does not give information on the uniqueness of an entry, that is, whether it represents a distinct locus or not. However, this information is given in OMIM, in which an asterisk preceding the entry identification number signifies that it represents a distinct locus. Since most GENDIAG entries include the identification number of the corresponding OMIM entries, we made a list of such identification numbers from GENDIAG. From this list we identified that subset which represented distinct loci (figure 1). Two sublists were then prepared—one X-linked and the other autosomal—using the fact that the first digit of the OMIM identification numbers is '3' for nearly all X-linked entries and '1' or '2' for nearly all autosomal entries.

Autosomal recessives are likely to become well documented more slowly than X-linked recessives (see § 3 and 4). Additionally, there is the likelihood that the subtle effects of a mutation may be recognized more slowly than the primary effects. One might thus underestimate the number of autosomal SRR entries in which the SRR phenotype is subtle or whose penetrance is low. However, such bias can be addressed by taking into account only well-documented entries. Information on the extent of documentation of each entry was also obtained from OMIM: the presence of a clinical synopsis section (which describes the features of a disorder) or a section on allelic variants (mutations within a gene associated with particular phenotypes) or both was taken as an indication of adequate documentation. GENDIAG information corresponding to each of these well-documented entries was retrieved and two data sets prepared—one X-linked and the other autosomal. Sex- and reproduction-related keywords were then made use of in a program written to identify SRR entries from these data sets. The proportion of X-linked SRR traits from among the total number of X-linked traits was calculated. Similarly, the proportion of autosomal SRR traits

Table 1. *Numbers of sex- and reproduction-related loci on human chromosomes**

(*a*) Data on chromosomes 1, 2, 3 and X

chromosome	number of entries	loci related to sex or reproduction
1	412	7
2	222	5
3	188	7
total (1–3)	822	19
X	296	46

Difference between X and the three autosomes: ($p<0.00001$); odds ratio = 7.8 (95% confidence limits: 4.5, 13.5); propensity = 6.7 (95% confidence limits: 4.1, 12.1)

(*b*) Data on disease-associated loci on chromosomes 1, 2, 3, 4, 11 and X

chromosome	number of disease-associated loci	number of loci related to sex or reproduction
1	90	7
2	17	1
3	64	7
4	33	5
11	60	6
total (1–4, 11)	264	26
X	141	46

Difference between X and the five autosomes: ($p<0.00001$); odds ratio = 4.4 (95% confidence limits: 2.6, 7.6); propensity = 3.3 (95% confidence limits: 2.2, 5.3)

(*c*) Data on disease-associated loci on autosomes and the X chromosome

chromosome	number of disease-associated loci	number of loci related to sex or reproduction
autosomes	674	165
X	93	27

(*d*) Data on disease-associated loci assigned to XCR and XAR

gene type	number on XCR	number on XAR
SRR	29	11
others	57	33

Difference between XCR and the five autosomes: ($p<0.00001$); odds ratio = 4.7 (95% confidence limits: 2.6, 8.5); propensity = 3.4 (95% confidence limits:2.1, 5.7).

Difference between XAR and the five autosomes: ($p<0.0005$); odds ratio = 3.1 (95% confidence limits: 1.4, 3.5); propensity = 2.5 (95% confidence limits:1.2, 4.7)

* See § 2d for details of the statistical methods.

was also calculated. Since GENDIAG also indicates the proportion of individuals affected by each trait, this value was taken as an indication of the penetrance of the trait in that particular syndrome. These numbers were used to calculate the overall average penetrance values of X-linked and autosomal SRR and non-SRR traits.

(d) *Statistical analysis*

The number of SRR and non-SRR entries listed against chromosomes 1, 2, 3 and X are given in table 1*a* and those obtained by a similar scrutiny of disease-related loci on chromosomes 1, 2, 3, 4, 11 and X are given in table 1*b*. The numbers obtained by examining well-documented GENDIAG entries on the X and all the autosomes are given in table 1*c*. Chi square, odds ratio, propensity and confidence limits were calculated as follows: (i) $\chi^2=[(a+b+c+d)\ (ad-bc)^2]/[(a+b)\ (c+d)\ (a+c)\ (b+d)]$; (ii) odds ratio $=(a/b)/(c/d)=ad/bc$; (iii) standard deviation (s.d.) $=[(1/a)+(1/b)+(1/c)+(1/d)]^{0.5}$; (iv) 95% confidence intervals for odds ratio: lower limit = antilog [$\log_e$ (odds ratio) $1.96\times$ s.d.], upper limit = antilog [$\log_e$ (odds ratio) $+1.96\times$ s.d.]; (v) propensity $=[a/(a+b)]/[c/(c+d)]$, where a is the number of X-linked sex- or reproduction-related loci; b, the number of remaining X-linked loci, i.e. other than those related to sex or reproduction; c, the number of autosomal loci related to sex or reproduction; and d, the number of remaining autosomal loci. The 95% confidence limits for propensity were calculated by simulations using a program provided by Dr Niranjan Joshi.

3. RESULTS AND DISCUSSION

As noted in § 2, care was taken to exclude from this analysis entries which, because of insufficient information, might introduce a bias of one kind or another. For instance, an autosomal gene the expression of which is limited to males may initially be classified as an X-linked gene. Similarly, an X-linked dominant trait may initially be labelled as autosomal dominant, and an X-linked recessive as autosomal recessive (McKusick 1962). A phenotype resulting from an X-linked gene, but one located in the pseudoautosomal regions (Burgoyne 1982), may be classified as being due to an autosomal gene. Hence, misclassifications can occur when the pattern of inheritance of a newly described disorder is not well defined, and they can occur in both directions. For this reason, only those entries for which there is no such ambiguity were included in the analysis.

The phenotype of intersexual marsupials suggests that homeobox genes concerned with mammary gland development are present on the X chromosome (Cooper *et al.* 1995). Since the mammalian X chromosome is conserved to a high degree (Ohno 1967), it is likely that homologues of these genes may be X-linked in humans. However, in the absence of direct evidence that these genes are located on the human X chromosome, they were not included. *Alpha-inhibin* and *beta-2 inhibin* appear to function in murine sexual development. Although their human homologues (MIM 147380 and 147390) are located on chromosome 2, these genes were not included because it is not known whether they perform similar functions in humans.

Three types of search were done to estimate the number of SRR genes on the X chromosome and the autosomes: (i) all OMIM and GENDIAG entries pertaining to chromosomes 1, 2, 3 and X; (ii) all OMIM and GENDIAG disease-associated entries pertaining to chromosomes 1, 2, 3, 4, 11 and X; and (iii) all GENDIAG entries that represent distinct loci and are well-documented in OMIM. In (i) and (ii) most of the data were from OMIM, whereas in (iii) the data were entirely from GENDIAG.

Anomalies of sexual development or reproduction occur in over 80 X-linked entries (details given in Appendices A and B). However, only 39 of them, representing 46 distinct loci, were considered (Appendix A); the remaining (Appendix B) were excluded either because: (i) it was not clear whether they represent loci distinct from those listed in Appendix A; (ii) their mode of inheritance—X-linked or autosomal—is not certain; or (iii) the same locus may be represented by more than one entry and, therefore, the relevant gene has already been taken into account in Appendix A. The last follows from the recognition that two or more clinically separate syndromes are sometimes associated with mutations in the same gene. For example, mutations in *XH2* (300032) may give rise to alpha-thalassaemia and mental retardation (301040) or the Juberg–Marsidi syndrome (309590). The gene *XH2* was placed in Appendix A, whereas the two syndromes were included in Appendix B; thus the locus was counted only once for this analysis. Five entries, although marked with an asterisk in MIM or OMIM, are included in Appendix B because, in our opinion, they are insufficiently validated to be included in Appendix A (see footnotes to Appendix B). For instance, precocious puberty has been reported in about 80% of individuals carrying a particular *G6PD* allele (entry 6629 in GENDIAG), but since no such reproduction-related effect has been observed in over 200 other alleles at this locus, this entry (305900), although asterisked in MIM, was excluded from Appendix A. The outcome of a similar search of autosomes 1, 2 and 3 was that only seven syndromes involving anomalies of sexual differentiation were traceable to seven genes on chromosome 1, six syndromes resulting from five different genes to chromosome 2 and seven such syndromes (seven genes) to chromosome 3 (Appendix A). Although the total number of entries on these three autosomes is nearly three times the number on the X chromosome, the X chromosome has more than twice the number of SRR entries than the autosomes (table 1*a*). Thus, there appears to be a disproportionately large number of genes related to sexuality and reproduction on the X chromosome ($p < 0.00001$). Maximum-likelihood analysis (Fienberg 1977) suggests that the odds for a locus on the X chromosome to be sex- or reproduction-related are 7.8 times greater than those for a locus on chromosomes 1, 2 or 3 (95% confidence limits: 4.5, 13.5). Propensity (also known as preference; Williams *et al.* 1987), a value similar to the likelihood ratio, was also calculated for an X-linked gene to be related to sex or reproduction. This value is about 6.7 times higher for the X chromosome (95% confidence limits: 4.1, 12.1) than for the three autosomes. An analysis of 300 randomly chosen OMIM entries from the remaining autosomes also suggests that there are fewer (seven) SRR genes among them than among the 296 X-linked entries. Again this difference is highly significant ($p < 0.00001$). The odds ratio (7.7; 95% confidence limits: 3.4, 17.4) and propensity (6.7; 95% confidence limits: 3.5, 20.3) values are also similar to those obtained from the comparison of X with the first three autosomes (see above).

In general, recessive genes on the X chromosome would be more easily uncovered than those on the autosomes. X inactivation may further facilitate such uncovering of recessives because it would make the mutation effectively hemizygous in many cells in females as well. These factors are expected to result in better documentation of the X chromosome in respect of its functions, at least in so far as recessive genes are concerned. This expectation is supported by the observation that 54% of the X-linked loci listed in OMIM are associated with abnormal phenotypes, whereas for the five autosomes (1, 2, 3, 4 and 11) that were similarly scanned, the proportion of such loci is 22%, 10%, 41%, 24% and 24%, respectively (data not shown). Chromosome 11 was chosen because three well-studied loci—*WT1*, *IGFII* and *H19*—related to sex determination and sexual differentiation are known to be located on it (Coppes *et al.* 1993; Rainier *et al.* 1993). As noted above, the number of entries for which little or no information on the clinical consequences of mutations in the relevant gene(s) is available appears to be higher on the autosomes than on the X. This may create a bias in favour of the hypothesis that there is a higher number of SRR genes on the X chromosome. To avoid such a possibility, comparisons were made between the proportion of genes that are sex- and reproduction-related among the total number of loci for which some information about function is available.

As shown in Appendix A and table 1*b*, for the X chromosome there are 46 loci that are sex- or reproduction-related out of a total of 141 loci with some information on function. The five autosomes together have 26 such loci out of a total of 264. This difference is highly significant ($p < 0.00001$). Maximum-likelihood analysis of these data suggests that the odds for a locus on the X chromosome to be sex- or reproduction-related are 4.4 times greater than for a locus on chromosomes 1, 2, 3, 4 or 11 (95% confidence limits: 2.6, 7.6). Propensity was also calculated for an X-linked gene to be related to sex or reproduction. This value is about 3.3 times higher for the X chromosome than for the five autosomes (95% confidence limits: 2.2, 5.3). The difference between the X chromosome and the autosomes in the number of such

genes is thus highly significant even when as many as 40 X-chromosomal entries were excluded from consideration as being insufficiently validated (Appendix B). In comparison, the sum total of such insufficiently validated entries for autosomes 1, 2, 3, 4 and 11 is ten. Similarly, an analysis of 76 randomly chosen disease-associated OMIM entries from the remaining autosomes also suggests that there are only seven SRR genes among them than among X-linked entries (46 out of 141 disease-associated loci). These differences also are statistically significant ($p < 0.0005$) and the odds ratio (4.8; 95% confidence limits: 2, 11.2) and propensity (3.5; 95% confidence limits: 1.9, 10) values are similar to those obtained from the comparison of X with the autosomes 1, 2, 3, 4 and 11.

GENDIAG enables the identification of particular traits in an entry through specific keywords. From among such keywords we prepared a subset related to sex and reproduction. This subset of keywords was then used to search all GENDIAG entries that are well documented in terms of the phenotype and are also thought to represent distinct loci (see § 2c for details). This search encompassed the X chromosome and all autosomes. The results were that 27 out of the 93 entries on the X chromosome were classified as being sex- or reproduction-related (29%) whereas 165 out of 674 autosomal entries were similarly classified (24%). Thus, in this search of well-characterized GENDIAG entries also, the number of SRR loci appears to be higher on the X chromosome than on the autosomes, although the difference is not statistically significant ($p > 0.25$). The reasons for this apparent difference between the results of the first two methods of analysis and those of the third are not clear. One possibility is that, since the last approach involved a global search of all chromosomes, those autosomes that were not searched for in the first two approaches (i.e. autosomes other than 1, 2, 3, 4 and 11) may contain a relatively high proportion of SRR genes. However, this is unlikely because a search of randomly chosen OMIM entries from autosomes other than 1, 2, 3, 4 and 11 also gives a low proportion of SRR genes (see above). Another possibility is that in the third approach the method followed to obtain well-documented entries from GENDIAG may have introduced a bias favouring autosomes. The manner in which the two databases are structured, or differences in the criteria used for creating and updating their entries, may have made a difference. For instance, OMIM is descriptive whereas GENDIAG relies on the usage of specific keywords to describe the different aspects of a phenotype and to identify particular syndromes.

Mutations with high penetrance and clear effects are likely to come to the attention of clinicians and geneticists earlier than those with poor penetrance and subtle phenotypes. Among the latter, it is likely that X-linked recessives would be identified and characterized earlier than autosomal recessives. As a result, in general, autosomal entries, including those with minor effects, might become validated more slowly than X-linked entries. Since GENDIAG gives information on the proportion of individuals carrying the mutation and who show a particular feature or defect attributed to the mutation (indicated in square brackets in Appendices A and B against those entries for which such information was available), we calculated the average number of SRR traits per SRR entry as well as the average penetrance of SRR and non-SRR traits for the GENDIAG datasets (§ 2c). For example, the average penetrance of an X-linked SRR trait was estimated by summing the penetrance values of each X-linked SRR trait and dividing the total by the number of X-linked SRR traits. Our observations can be summarized as follows: (i) the overall proportion of SRR traits is higher for X-linked entries than for autosomal entries; (ii) although the average penetrance of a trait in an autosomal non-SRR entry is greater than that for a similar X-linked entry, X-linked SRR traits show greater penetrance than autosomal SRR traits: in 27 out of 29 X-linked entries in Appendix A more than 50% of the affected individuals show sex- or reproduction-related anomalies, whereas the relevant figure for the autosomal entries is 8 out of 13. Thus, the X chromosome appears to have not only a higher proportion of SRR genes, but it also appears that such genes show greater pleiotropy and higher levels of penetrance of SRR traits than similar autosomal genes.

Another point to note is that mutations in many SRR genes, both X-linked and autosomal, show a recognizable phenotypic effect in males but not in females. In only a few entries is a phenotypic anomaly observed in both males and females or only in females (Appendix A). In general, X-linked mutations would be more easily uncovered in males than in females. Hemizygosity may therefore be a sufficient explanation for the observed excess of X-linked SRR entries in which more males are affected than females. Does the excess of SRR entries in which more males are affected indicate a possible bias in our data sets? *A priori* there is no reason to believe that hemizygosity would uncover more SRR genes than other (non-SRR) genes on the X chromosome. However, some explanation other than hemizygosity is necessary to account for the observations that: (i) the proportion of SRR entries in which more males are affected than females is high among the autosomes 1, 2, 3, 4 and 11 as well; and (ii) SRR traits show greater phenotypic deviation from normality in males than in females. One possibility is that in mammals the male pathway may be more vulnerable to developmental perturbations brought about by mutant genes because, in effect, the default pathway is femaleness.

Among the X-linked genes listed in Appendix A, a few require additional comment: *DSS* (300018) affects sex determination, and several genes affect sexual differentiation in the male (e.g. *androgen receptor*; 313700) or reproduction in the female (e.g. *premature ovarian failure, type I*; 311360) without causing severe dysmorphic features. However, two genes, one thought of as 'major' and another as 'minor' on the basis of the severity of phenotypic effects, may have similar fitness coefficients. For instance, if there are mutations influencing sexual orientation, their effect on fitness may be as severe as that of mutations causing major dysmorphic features. *DSS* is of particular interest because it appears to function in a dose-dependent manner and cause sex reversal when present in two active copies in XY individuals. Mutations in *DAX1*, a gene located in the critical region of *DSS*, are associated with congenital adrenal hypoplasia and hypogonadotropic hypogonadism (300200) in XY individuals.

Transgenic XY mice in which an extra copy of *Dax1* is expressed at high levels in the presence of weak *Sry* alleles show sex reversal (Swain *et al*. 1998). Since only one copy of *DSS* is normally present in XY individuals, an understanding of the basis of its dose-dependent action on sex determination may have to be sought in the interrelationships, during evolution, among sex determination, X inactivation and the emergence of XY heteromorphism (Chandra 1985, 1986).

Why might there be a disproportionate representation of sex- and reproduction-related genes on the X chromosome? One possibility is that during evolution, the 'choice' of a particular pair of homomorphic chromosomes for specialization as sex chromosomes may be related to the number of such genes initially present in it or, since sex determination and sexual dimorphism are often gene dose-dependent processes, the number of such genes necessary to be regulated in a dose-dependent manner. Alternatively, the X chromosome may come to consist of a higher number of SRR genes because accumulation of such traits on this chromosome is promoted by X linkage (Fisher 1931; Rice 1989). Rice (1989) has argued on theoretical grounds that if genetic variation is present throughout the genome for sexually antagonistic traits (i.e. phenotypes that are selectively favoured in one sex but disfavoured in the other), then selection favours an increase in the frequency of such genes on the X chromosome under a much wider range of conditions than on autosomes. In a *D. melanogaster* model system Rice (1992) has shown that, on artificial selection for as few as 29 generations, sexually antagonistic traits accumulate in linkage with a sex-determining gene.

It has been known for some time that the mammalian X chromosome is conserved to a remarkable extent (Ohno 1967). Comparative gene-mapping studies among the extant taxa of mammals (Graves & Watson 1991; Wakefield & Graves 1996) have begun to yield important information on the evolution of their sex chromosomes. Graves and co-workers have shown that the long arm and pericentric region of the human X chromosome are conserved among the X chromosomes of monotremes, marsupials and other mammals. This conserved region of the human X (referred to as XCR, the X conserved region) must therefore have been present in the common ancestral lineage before these three groups diverged from each other. Most of the X short arm other than the pseudoautosomal region appears to have been recently added to the X from autosomes. This region has therefore been called the X-added region, or XAR (Graves & Watson 1991). If the hypothesis that the original homomorphic pair of chromosomes already contained a significant number of SRR genes is correct, then the proportion of such genes would be expected to be particularly high in XCR. To verify this possibility we classified X-linked entries as belonging to XCR or XAR. Out of the 40 X-linked SRR loci that could be classified in this manner, 29 are located in the XCR (table 1*d*). The difference between the number of SRR entries on autosomes and those in the conserved region of the X chromosome is greater than when the entire X chromosome is taken into account (table 1). However, the proportion of SRR genes in XAR is not significantly lower than that in XCR ($p > 0.25$). One possible explanation—for which there is support from population-genetic arguments (Charlesworth & Charlesworth 1980)—is that X-autosomal translocations in which the autosomal segment was rich in SRR functions prior to the translocation were favoured during evolution. These results may also mean that X linkage promotes accumulation of SRR functions relatively rapidly (Rice 1989, 1992) and, as a consequence, both XAR and XCR have come to contain more or less the same proportion of such genes.

We thank Dr Niranjan Joshi and Dr Vidyanand Nanjundiah for suggestions on statistical methods and critical review of the manuscript, and Dr Jennifer Graves (LaTrobe University, Australia), Dr Lucio Luzzatto (Memorial Sloan-Kettering Cancer Center, New York) and two anonymous referees for useful suggestions. This work was supported by the Indian Council of Medical Research. G.M.S. is a Senior Research Fellow of the Council of Scientific and Industrial Research, New Delhi.

REFERENCES

Baker, B. S. 1989 Sex in flies: the splice of life. *Nature* **340**, 521–524.

Burgoyne, P. S. 1982 Genetic homology and crossing over in the X and Y chromosomes of mammals. *Hum. Genet.* **61**, 85–90.

Chandra, H. S. 1985 Is human X chromosome inactivation a sex-determining device? *Proc. Natn. Acad. Sci. USA*. **82**, 6947–6949.

Chandra, H. S. 1986 X chromosomes and dosage compensation. *Nature* **319**, 18.

Charlesworth, D. & Charlesworth, B. 1980 Sex-differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genet. Res.* **35**, 205–214.

Cline, T. 1993 How do flies count two? *Trends Genet.* **10**, 376–380.

Cock, A. G. 1964 Dosage compensation and sex chromatin in non-mammals. *Genet. Res.* **5**, 354–365.

Cooper, D. W., Johnston, P. G., Hughes, R. L., Gemmell, R., Smith, M. & Watson, C. M. 1995 X-chromosome involvement in marsupial sex determination. *Cytogenet. Cell Genet.* **71**, 307–342. (Abstract P5.)

Coppes, M. J., Cambell, C. E. & Williams, B. R. G. 1993 The role of *WT1* in Wilms tumorigenesis. *FASEB J.* **7**, 886–895.

Fienberg, S. E. 1977 *The analysis of cross-classified categorical data.* Cambridge, MA: MIT Press.

Fisher, R. A. 1931 The evolution of dominance. *Biol. Rev.* **6**, 345–368.

Graves, J. A. M. & Watson, J. M. 1991 Mammalian sex chromosomes: evolution of organisation and function. *Chromosoma* **101**, 63–68.

Grula, J. W. & Taylor, O. R. Jr. 1980 The effect of X-chromosome inheritance on mate selection behaviour in the sulfur butterflies, *Colias eurytheme* and *Colias philodice*. *Evolution* **34**, 688–695.

Hamer, D. H., Hu, S., Magnuson, V. L., Hu, N. & Pattatucci, A. M. L. 1993 A linkage between DNA markers on the X chromosome and male sexual orientation. *Science* **261**, 321–327.

Hu, S., Pattatucci, A. M. L., Patterson, C., Li, L., Fulker, D. W., Cherny, S. S., Kruglyak, L. & Hamer, D. H. 1995 Linkage between sexual orientation and chromosome Xq28 in males but not in females. *Nature Genet.* **11**, 248–256.

Kanaan, C., Habecker-Green, J. & Cohn, G. 1995 Prenatal diagnosis of a unilateral pelvic multicystic dysplastic kidney in a fetus with a 47,XXY karyotype and congenital megacolon. *Am. J. Hum. Genet.* **57**. (Abstract 1776.)

McKusick, V. 1962 On the X chromosome of man. *Q. Rev. Biol.* **37**, 69–175.

McKusick, V. 1992 *Mendelian inheritance in man. Catalog of human genes and genetic disorders*, 10th edn. Baltimore, MD: Johns Hopkins University Press.

Meyer, W. J. III, Migeon, B. R. & Migeon, C. J. 1975 Locus on human X chromosome for dihydrotestosterone receptor and androgen insensitivity. *Proc. Natn. Acad. Sci. USA* **72**, 1469–1472.

Ogata, T. & Matsuo, N. 1995 Turner syndrome and female sex chromosome aberrations: deduction of the principal factors involved in the development of clinical features. *Hum. Genet.* **95**, 607–629.

Ohno, S. 1967 *Sex chromosomes and sex-linked genes.* Berlin and New York: Springer-Verlag.

Rainier, S., Johnson, L. A., Dobry, C. J., Ping, A. J., Grundy, P. E. & Feinberg, A. P. 1993 Relaxation of imprinted genes in human cancer. *Nature* **362**, 747–749.

Rice, W. R. 1989 Sex chromosomes and the evolution of sexual dimorphism. *Evolution* **38**, 735–742.

Rice, W. R. 1992 Sexually antagonistic genes: experimental evidence. *Science* **256**, 1436–1439.

Riess, O., Kozak, C. & Van Ommen, G.-J. 1996 Report of the fourth international workshop on human chromosome 4 mapping 1996. *Cytogenet. Cell Genet.* **74**, 57–69.

Ryner, L. C. & Swain, A. 1995 Sex in the '90s. *Cell* **81**, 483–493.

Sala, C. (and 10 others) 1997 Eleven X chromosome breakpoints associated with premature ovarian failure (POF) map to a 15-Mb YAC contig spanning Xq21. *Genomics* **40**, 123–131.

Schinzel, A. 1984 *Catalogue of unbalanced chromosomal aberrations*. Berlin and New York: Walter de Gruyter.

Sheppard, P. M. 1961 Some contributions to population genetics resulting from the study of the Lepidoptera. *Adv. Genet.* **10**, 165–216.

Shows, T. B. (and 25 others) 1996 Report of the fifth international workshop on human chromosome 11 mapping 1996. *Cytogenet. Cell Genet.* **74**, 1–56.

Spur, N. K. (and 16 others) 1996 Report of the fourth international workshop on human chromosome 2 mapping 1996. *Cytogenet. Cell Genet.* **73**, 255–273.

Stehr, G. 1959 Hemolymph polymorphism in a moth and the nature of sex-controlled inheritance. *Evolution* **13**, 537–560.

Stevanovic, M., Lovell-Badge, R., Collignon, J. & Goodfellow, P. N. 1993 *SOX3* is an X-linked gene related to *SRY*. *Hum. Mol. Genet.* **2**, 2013–2018.

Swain, A., Narvaez, V., Burgoyne, P., Camerino, G. & Lovell-Badge, R. 1998 *Dax1* antagonizes *Sry* action in mammalian sex determination. *Nature* **391**, 761–767.

Tar, A., Solyom, J., Gyorvari, B., Ion, A., Telvi, L., Barbaux, S., Souleyreau, N., Vilain, E., Fellows, M. & McElreavey, K. 1995 Testicular development in an *SRY*-negative 46,XX individual harbouring a distal Xp deletion. *Hum. Genet.* **96**, 464–468.

Thode, A., Parrington, M. W., Yip, M.-Y., Chapman, C., Richardson, V. F. & Turner, G. 1988 A new syndrome with mental retardation, short stature and an Xq duplication. *Am. J. Med. Genet.* **30**, 239–250.

Wakefield, M. J. & Graves, J. A. M. 1996 Comparative maps of vertebrates. *Mammalian Genome* **7**, 715–716.

Weith, A., Brodeur, G. M., Bruns, G. A. P., Matise, T. C., Mischke, D., Nizetic, D., Seldin, M. F., Roy, N. van & Vance, J. 1996 Report of the second international workshop on human chromosome 1 mapping 1995. *Cytogenet. Cell Genet.* **72**, 113–154.

Williams, R. W., Chang, A., Juretic, D. & Loughran, S. 1987 Secondary structure predictions and medium range interactions. *Biochim. Biophys. Acta* **916**, 200–204.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.