Male-driven evolution

Wen-Hsiung Li*, Soojin Yi and Kateryna Makova

The strength of male-driven evolution – that is, the magnitude of the sex ratio of mutation rate – has been a controversial issue, particularly in primates. While earlier studies estimated the male-to-female ratio (α) of mutation rate to be about 4–6 in higher primates, two recent studies claimed that α is only about 2 in humans. However, a more recent comparison of mutation rates between a noncoding fragment on Y and a homologous region on chromosome 3 gave an estimate of $\alpha = 5.3$, reinstating strong male-driven evolution in hominoids. Several studies investigated variation in mutation rates among genomic regions that may not be related to sex differences and found strong evidence for such variation. The causes for regional variation in mutation rate are not clear but GC content and recombination are two possible causes. Thus, while the strong male-driven evolution in higher primates suggests that errors during DNA replication in the germ cells are the major source of mutation, the contribution of some replication-independent factors such as recombination may also be important.

Addresses
Department of Ecology and Evolution University of Chicago, 1101 East 57th Street Chicago, Illinois 60637, USA
*e-mail: whli@uchicago.edu

Current Opinion in Genetics & Development 2002, **12**:650–656

0959-437X/02/$ – see front matter

Published online 4 October 2002

Abbreviations

LINE	long interspersed elements
MEN2A/B	multiple endocrine neoplasia type 2A/B
ZFX	zinc finger protein, X-linked
ZFY	zinc finger protein, Y-linked

Introduction

Almost 70 years ago, Haldane [1] proposed that the male mutation rate in humans is much higher than the female mutation rate because the male germline goes through many more rounds of cell divisions (DNA replications) per generation than does the female germline. Under this hypothesis, mutations arise mainly in males, so that evolution is 'male-driven' [2]. Although a higher mutation rate in males than in females has been well accepted, the magnitude of the male-to-female ratio (α) of mutation rate remains a point of contention. Knowing the magnitude of α is important because it is related to the issue of whether DNA replication errors are the major source of mutation [3,4], which has been a subject of heated debate for the past several decades. Clearly, resolving these issues has implications for understanding the mechanism of mutagenesis and for the generation-time effect hypothesis, which postulates a faster molecular clock for organisms with a short generation time than for ones with a long generation time. In this article, we review studies on male-driven evolution in mammals and birds in the past decade and discuss factors that may affect the sex ratio of mutation rate. Note that mutation here refers to point (substitution) mutation; we are not concerned here with deletion or insertion mutation, which seem to have a mechanism of mutagenesis different from that of point mutation.

Estimating α from new or recently produced mutations

Dramatic advances in DNA technology have allowed the inference of the origin of a new or recently-produced mutation. When the origins of many mutations are inferred, α can be estimated as the ratio of the number of point mutations of paternal origin to that of maternal origin. This direct approach has replaced the indirect methods for estimating α from incidents of X-linked diseases [1]. Application of the direct approach to 119 families of haemophilia A (an X-linked recessive disease) led to an estimate of $\alpha = 15$ [5]. Two small datasets from X-linked dominant disorders are available (Table 1) and they give an average α value of $31/3 = 10.3$, not significantly different from the above estimate.

The direct approach has been applied to many autosomal dominant disorders and a compilation of such studies by Hurst and Ellegren [6] showed a male excess in most estimates. This remains true when more cases are added (Table 1). The estimates are highly variable. This may be due in part to small sample sizes. However, the samples for Apert syndrome and achondroplasia include 40 or more cases, but no female-derived mutation was found in either sample. On the other hand, there are two cases — neurofibromatosis type 2 and von Hippel–Lindau disease — where α is close to 1; but the sample size is small in both cases, so it is not clear what the true α is. Interestingly, despite the fact that α is infinite (∞) for many of the cases in Table 1, when all samples are pooled together the average α is only 10.8, similar to the above average α for X-linked disorders.

A serious problem with the direct methods of estimating α is that the majority of the mutations are only from a few specific sites that mutate at unusually high rates. About half of these sites are at CpG dinucleotide sites, which tend to be mutational hotspots because of methylation (see later). In four of the cases in Table 1 — achondroplasia, Apert syndrome, MEN2A and MEN2B — the mutations were recurrent at a few specific sites. As an extreme case, achondroplasia is mainly (>95% of the cases examined) caused by mutations at site 1138 of the coding sequence (mostly G→A mutations, or C→T mutations in the other strand). This site is at a CpG dinucleotide site. In the case of MEN2B, >98% of the patients have a specific T→C

Table 1

Estimates of the ratio (α) of the number of point mutations of paternal origin to those of maternal origin leading to X-linked dominant disorders or autosomal dominant disorders in humans.

Disease	Gene	No. of mutations	α	CpG dinucleotide involved	References
X-linked dominant diseases					
Pelizaeus–Merzbacher disease	*PLP*	5	4	No	[44]
Rett syndrome	*MECP2*	29	13.5	Yes (~70 %)	[45,46]
Subtotal		34	10.3		
Autosomal-dominant diseases					
Achondroplasia	*FGFR3*	40	∞	Yes (>97%)	[47]
Apert syndrome	*FGFR2*	57	∞	Yes (~62%)	[48]
Crouzon syndrome and Pfeiffer syndrome	*FGFR2*	22	∞	No	[49]
Denys–Drash syndrome	*WT1*	2	∞	No	[50]
Hirschsprung disease	*RET*	3	0	1/3	[51]
MEN2A	*RET*	10	∞	No	[52]
MEN2B	*RET*	25	∞	No	[53]
Neurofibromatosis type 2	*NF2*	23	1.3	Yes (~35%)	[54]
von Hippel–Lindau disease	*VHL*	7	1.3	2 paternal CpG transitions	[55]
Subtotal		189	10.8		
Total		223	10.7		

FGFR, fibroblast growth factor receptor; *MECP2*, methyl-CpG-binding protein 2; *NF2*, neurofibromatosis type 2; *PLP*, proteolipid protein; *RET*, Ret protooncogene; *VHL*, von Hippel-Lindau syndrome.

mutation at the second position of codon 918, which is not at a CpG site. Clearly, in these cases the inference of the exclusively paternal origins of the mutations do not reflect the general male to female mutation rate ratio. Thus, although the direct approach is an improvement over the indirect approach, there are difficulties in using it to estimate α.

Evolutionary approach

In addition to the drawbacks mentioned above, the direct method may not be applicable to non-human organisms. As an alternative, Miyata *et al.* [2] proposed to estimate α from the mutation rates of the two sex chromosomes or of a sex chromosome and an autosome (or autosomes). Let *Y*, *X*, and *A* be the mutation rates for a Y-linked sequence, an X-linked sequence, and an autosomal sequence, respectively. Noting that in a population all Y-linked sequences are derived from the fathers, whereas one-third of the X-linked sequences are derived from the fathers and two-thirds from the mothers, Miyata *et al.* [2] showed $Y/X = 3\alpha/(2 + \alpha)$. From this formula, one can estimate α if the ratio Y/X is known. To estimate the Y and X values, one can use a pair of homologous nonfunctional Y-linked and X-linked sequences from two or more species, because in a nonfunctional sequence, the rate of nucleotide substitution is equal to the rate of mutation. In the same manner, Miyata *et al.* [2] showed $Y/A = 2\alpha/(1 + \alpha)$ and $X/A = (2/3)(2 + \alpha)/(1 + \alpha)$, and so α can also be estimated from a pair of homologous Y-linked and autosomal sequences or a pair of X-linked and autosomal sequences from two or more species. This evolutionary approach benefits from the comparison of mutation rates over a large number of sites and from the accumulation of mutations over long evolutionary times.

In higher primates, the α value was estimated to range from 4.2 to 6.3 by obtaining the intron sequences of several pairs of homologous genes on chromosomes X and Y (Table 2). In mice and rats α was estimated to be ~2 from the intron sequences of two pairs of homologous genes on the X and Y chromosomes (Table 2). These studies all strongly supported male-driven evolution. However, McVean and Hurst [7] hypothesized that the high α value might be caused by a reduced mutation rate on chromosome X rather than an elevated mutation rate on chromosome Y. They reasoned that chromosome X is in a hemizygous state in males, so that all deleterious mutations on the X chromosome are exposed to natural selection every generation, and it is advantageous for the X chromosome to evolve a low mutation rate. This hypothesis was not supported by the finding of male-driven evolution in birds; that is, α was estimated to be 4–5, despite the fact that females are heterogametic (ZW) and males are homogametic (ZZ) [8]. The existence of male-driven evolution in birds was further supported by the analyses of intronic sequences of additional genes and in additional species [9,10] (Table 2). Male-driven evolution was also observed in felids and ovids [11,12] (Table 2). Thus, by the end of the last century, male-driven evolution was considered a well-established phenomenon in primates, rodents, and birds.

With the Human Genome Project approaching completion, it became tempting to use a large amount of sequence data to estimate α. Two studies claimed that α is only ~2 in humans, significantly lower than the earlier estimates. First, Bohossian *et al.* [13] studied a 38.6 kb segment that was transposed from X to Y in the human lineage after the human–chimpanzee split and estimated α to be only 1.55.

Table 2

Ratio of nucleotide substitution rates on different chromosomes (m) and estimated male-to-female ratio of mutation rate (α) in different organisms.

Taxa	Gene pair (length studied)	Rate ratio (m)	α (95% CI)	References
Higher primates	*AMELY/AMELX* (1.1 kb)	Y/X = 2.16	5.14 (2.42–16.6)	[56]
Higher primates	*ZFY/ZFX* (0.9 kb)	Y/X = 2.27	6.26 (2.63–32.4)	[57]
Higher primates	*SMCY/SMCX* (1.4 kb)	Y/X = 2.03	4.20 (2.20–10.0)	[17]
Human and apes	Noncoding (10.4 kb)	Y/A3 = 1.68	5.25 (2.44–∞)	[14••]
Cats	*ZFY/ZFX* (0.8 kb)	Y/X = 2.06	4.38 (3.76–5.14)	[11]
Sheep and goat	*ZFY/ZFX* (0.8 kb)	Y/X = 1.99	3.94 (1.25–32.29)	[12]
Mouse and rat	*ZFY/*ZFX (1–1.3 kb)	Y/X = 1.42	1.80 (1.0–3.2)	[3]
Mouse and rat	*Ube 1Y/Ube 1X* (0.9kb)	Y/X = 1.50	2.0 (1.0–3.9)	[58]
Birds	*CHD1Z/CHD1W* (0.3 kb)	Z/W = 4.65	6.5 (2.8–10.2)	[8]
Birds	*CHD1Z/CHD1W* (0.4 kb)	Z/W = 3.06	4.1 (3.1–5.1)	[9]
Birds	*ATP5A1Z/ATP5A1W* (0.8 kb)	Z/W = 1.53, 1.87, 3.67	1.8; 2.3; 5.0	[10]

However, this study has two problems [14••]: it used an erroneous phylogeny for the Y-linked sequence and it compared closely related sequences, so that α could have been underestimated because of the effect of pre-existing polymorphism (Figure 1). Second, α was estimated to be only 2.1 from a comparison of the substitution rates in young subfamilies of long interspersed nuclear elements on X and Y [15]. However, corrections for multiple substitutions were not made and it was assumed that the repetitive elements of the same subfamily were inserted into the genome at the same time, which is not true.

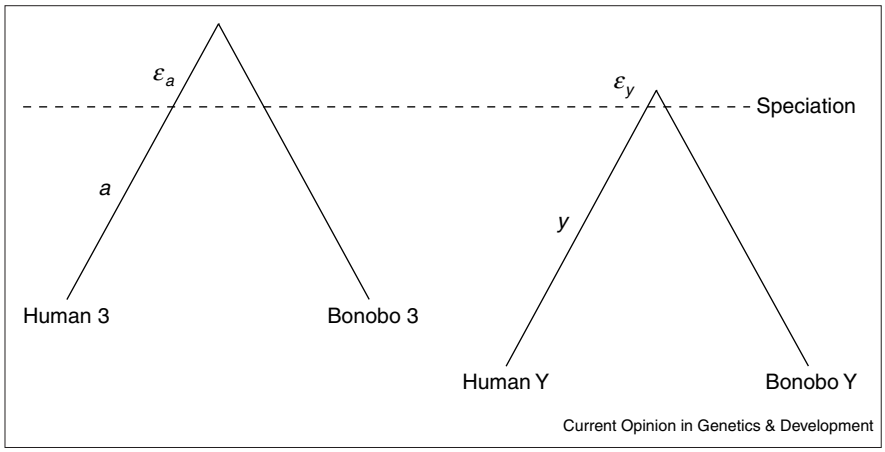
In an attempt to resolve the controversy, Makova and Li [14••] compared mutation rates in humans and apes in a pair of homologous noncoding regions (~11 kb) on chromosome Y and chromosome 3. This Y-linked locus was transposed from chromosome 3 after the New and Old World monkeys split. The α estimated from external (species-specific) branches was only 2.23, but it could have been underestimated because the species used are closely related so that the effect of pre-existing polymorphism in the ancestral population is not negligible (Figure 1). Indeed, the α value estimated from internal branches was 5.25. This supports the estimate of α being ~4–6 in higher primates.

Methylation effects

In mammalian cells, DNA methylation occurs mostly at the C residue of CpG dinucleotides and a methylated C residue is easily transformed to a T through deamination, which creates a C→T transitional mutation. If the C→T transition occurs on the antisense strand of DNA, it is reflected as a G→A transition on the sense strand. As methylation occurs at a considerably higher rate in sperm DNA than in oocyte DNA [16], it increases the frequency of the paternal origin of mutation. For example, in Rett syndrome and achondroplasia (Table 1), the majority of the mutations occurred at a specific CpG dinucleotide site, and it is likely that methylation was a major factor for the strong male bias in these two cases. However, methylation probably did not play a major role in the majority of the other disorders in Table 1. In the case of Apert syndrome (the most frequent case in Table 1), all mutations occurred at only two specific sites, one of which is a CpG dinucleotide site. However, because all mutations at this site were C→G transversions, they might not be caused by methylation. Note also that the majority of the other cases in Table 1 do not involve any CpG dinucleotide sites. Thus, the overall contribution of methylation to the paternal origin of genetic disorders may not be as important as commonly thought. In fact, when all the sites in Table 1 are considered together, the average of α is only 10.7. Nevertheless, as methylation can certainly produce an extreme bias in certain cases, it posts a serious problem for estimating α from new or recently-produced mutations.

By contrast, methylation causes no serious problem in the evolutionary approach for two reasons. First, this approach considers the mutation rates at a large number of nucleotide sites, so that it is less affected by biases at mutational hotspots. Second and more importantly, it uses nonfunctional sequences, which usually have very few CpG dinucleotide sites, as such sites tend to disappear quickly in a nonfunctional sequence. For example, in the primate ZFX (zinc finger protein, X-linked) and ZFY (zinc finger protein, Y-linked) introns studied by Chang *et al.* [17], no CpG site was found in the ZFY introns, and only five CpG sites were found in the ZFX introns, and only 2 of these were mutated in at least one of the lineages studied, representing only 2% of the total number of variable sites (97) in the sequence alignment (~1,140 bp long). This example shows that the effect of methylation on the estimation of α is minor and actually tends to reduce α because of an absence or near absence of CpG dinucleotide sites on Y-linked sequences. The same comment applies to the other introns studied (Table 2). In the case of a Y-linked sequence that has been derived recently from an autosomal or an X-linked sequence, there might initially be some CpG dinucleotide sites on the sequence, but

Figure 1



Effect of preexisting polymorphism on the estimation of α. In the figure, the divergence between two species is equal to $d_0 + 2\mu t$, where t is the divergence time between the two species, μ is the rate of substitution, and d_0 is the average divergence between two sequences (pre-existing polymorphism) in the ancestral population [60]. In practice, the ratio of mutation rates between chromosome Y and an autosome is estimated by $Y/A = (y + \varepsilon_y) / (a + \varepsilon_a)$, where *y* and *a* are the numbers of mutations per site on chromosome Y and on an autosome, respectively, because the divergence of the two species, and ε_y and ε_a are the preexisting polymorphisms at the point of speciation on chromosome Y and an autosome, respectively. The level of pre-existing polymorphism (ε_y) on chromosome Y is extremely low because there is no recombination and a small effective population size. On the other hand, ε_a – or ε_x, if X chromosome sequences are used – can be high enough to substantially underestimate the Y/A ratio (or the Y/X ratio) and, consequently, α. Thus, the effect of pre-existing polymorphism on the estimation of α is not negligible when the two species are closely related, or in other words, closely related species are not suitable for estimating α.

these sites would disappear quickly. Therefore, it can be concluded that in the evolutionary approach the effect of methylation on the estimation of α is minor.

Regional effects

The possibility that mutation rate may vary among genomic regions was first proposed, on the basis of very limited data, by Filipski [18] and Wolfe *et al.* [19]. Support for this hypothesis came from a study that detected substantial variation in both mutation rate and pattern among three primate arginosuccinate synthetase processed pseudogenes located in different regions of the genome [20]. Later, in a comparison of human and mouse genes, Matassi *et al.* [21] found that synonymous substitution rates are significantly more similar for neighboring genes than for genes located far apart. This study provided the first statistical evidence on a genomic scale that the rate of synonymous substitution varies among genomic regions. Further genomic-scale evidence came from the observation of a positive correlation between the GC content of old repetitive elements and their surrounding regions [22]. Comparing orthologous human and mouse genes, Lercher *et al.* [23••] detected a significant similarity in synonymous rates for linked genes; the rate similarities extended over whole chromosomes. By contrast, substantial rate differences were found between chromosomes. This finding was supported by a more extensive analysis of human and mouse genes [24] and an analysis of ~2 million base pairs between the chimpanzee and human genomes [25•]. Thus, there is now strong evidence for the regional mutation pressure hypothesis.

However, the causes for regional variation in mutation rate are not clear. One possible cause is the timing of replication and the abundance of free dNTPs during the cell cycle in the germ line [19]. Another possible cause is differences in the efficiency of DNA repair across the genome [18,21]. Two other possible factors, GC content and recombination rate, are discussed below.

There have been conflicting conclusions on whether the GC content of a region can affect the mutation rate. Earlier studies suggested a negative correlation between synonymous substitution rate and GC content (e.g. [26]) or a nonlinear relationship [19,24,27], or no correlation [21,28]. However, recent studies showed a positive correlation between GC content and synonymous rate [24,29,30,31•]. A positive correlation between GC content and substitution rate has also been found in primate introns [32]. The recent studies seem more reliable because they used more extensive data and employed a maximum likelihood method for synonymous rate estimation. Nevertheless, variation in GC content cannot completely account for the variation in synonymous rates [23••,24].

Recombination seems to be an even more important factor. There is evidence that mutation rate increases with local recombination rate. For instance, the pseudoautosomal region in mice has an unusually high recombination rate and a remarkably high synonymous rate [33]. Moreover, Lercher and Hurst [31•] found a strong positive correlation between nucleotide diversity (single nucleotide polymorphisms) sampled throughout the human genome and local recombination rate and showed that at least part of the correlation is mediated by mutation rate. Why is mutation rate correlated with recombination rate? Evidence from yeast [34] and from mammalian mutation hotspots [35] suggests that repair of double-stranded breaks during recombination is mutagenic. Moreover, GC content and local recombination rate are correlated [36,37•,38•]. However, the causality of this relationship remains uncertain. Some studies concluded that recombination might be a cause of positive correlation

between GC content and recombination rate [39,40], whereas others suggested that high levels of GC content enhance recombination (e.g. [41]). Although further studies are required to resolve this issue, GC content and recombination rate appear to be two important factors for regional variation in mutation rate.

Conclusions and future directions

The above survey of studies that used the method of Miyata *et al.* [2] to estimate α suggests that it is between 5 and 6 for hominoids (humans and apes) (Table 2). The α value (~11) estimated from the direct method (Table 1) was considerably higher for two reasons. First, it included the effect of methylation. Second, it referred to humans only; α is expected to be higher in humans than in apes because it increases with generation time. There is indeed a positive correlation between α and generation time (Table 2): the generation times are longest in hominoids, intermediate in cats and sheep, and shortest in rodents, and the α values for these organisms are ~5–6, 4, and 2, respectively.

Knowing the sex ratio of mutation rate (α) has important implications for the mechanisms of mutagenesis. Chang *et al.* [3] found that in rodents and higher primates the α value is approximately equal to the sex ratio (*c*) of the numbers of germ-cell divisions per generation in males and females, and concluded that errors during DNA replication in the germline are the major source of mutation (i.e. the germ-cell division hypothesis). For example, in mice α was estimated to be ~2 (Table 2) and *c* was estimated from gametogenesis data to be ~2 if the father is 5 months old at the time of fertilization. In humans, *c* was estimated to be ~6 if the father's age is 20 and ~10 if the father's age is 25. It has been claimed that in humans α is much smaller than *c* because the generation time in humans can be longer than 25 years and *c* can be larger than 10 [6]. However, it should be noted that the α values shown in Table 2 are not for humans but for higher primates — including Old World monkeys and New World monkeys in the majority of cases. In higher primates, an assumption of 20 years for the generation time seems reasonable. Further, it should be noted that data for estimating the number of germ-cell divisions in humans is scanty and estimates of *c* may not be reliable. This is a subject that deserves future attention. The estimate of *c* in mice was based on more data and should be more reliable.

It has been argued that the large variation in silent substitution rate among autosomes (i.e. regional effects) is incompatible with the germ-cell division hypothesis because rate heterogeneity among autosomes cannot be as a result of differing times spent in the male and female germlines [23••,25•]. However, it should be noted that a cause for regional variation in mutation rate may be compatible with the germ-cell division hypothesis. To illustrate this, let us consider the four possible factors for regional variation that have been proposed to date. The first factor — DNA replication time in the germ-cell cycle — assumes that the GC content in the nucleotide precursor pool (free dNTPs) changes with DNA replication time and that this change affects the rate of misincorporation of nucleotide bases during DNA replication. This model assumes that errors during DNA replication are the primary source of mutation. The second factor is variation in repair efficiency among genomic regions. Like the first model, this one also assumes that the ultimate source of mutation is misincorporation of bases during DNA replication. The third model assumes that the GC content of a region can affect the rate of misincorporation of bases during DNA replication, or the efficiency of error repair, or it can cause bias in gene conversion or can affect recombination rate. Except for the last two possibilities, this factor is also compatible with the germ-cell division hypothesis. The fourth factor is recombination. As recombination occurs at meiosis in both male and female germ lines, it may not contribute to sex differences in mutation rate except for the X and Y chromosomes. Therefore, if recombination is indeed mutagenic, it may be an important factor for mutation besides DNA replication errors.

It is worth noting that if recombination is mutagenic, then the α value can be underestimated from a comparison between a Y-linked sequence and an X-linked or an autosomal sequence because recombination is absent in Y unique regions. Note also that recombination rate is generally lower on the X chromosome than on autosomes because recombination on the X chromosome is absent in males. This may partly explain why silent substitution rates tend to be lower for X-linked sequences than for autosomal sequences [23••,24,25•,42]; the lower rates should be caused in part by sex differences in mutation rate. In view of this and the above issues, the potential effect of recombination on mutation should be eagerly pursued.

The germ-cell division hypothesis predicts a higher rate of nucleotide substitution in organisms with a short rather than a long generation time (i.e. the 'generation-time effect') because the number of male germ cell divisions per unit time is expected to be higher for short-lived organisms than for long-lived ones. This prediction has indeed been supported by the observation of a faster molecular clock in Old World monkeys than in hominoids [4,32] and a much faster molecular clock in rodents than in higher primates [4,43].

In conclusion, much progress has been made in the past decade on sex differences in mutation rate in mammals. The exact magnitude of sex differences and the causes for regional variation in mutation rate, however, still need much further investigation.

Acknowledgements

This study was supported by grants from the National Institutes of Health.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Haldane JBS: **The rate of spontaneous mutation of a human gene.** *J Genet* 1935, **31**:317-326.

2. Miyata T, Hayashida H, Kuma K, Mitsuyasu K, Yasunaga T: **Male-driven molecular evolution: a model and nucleotide sequence analysis.** *Cold Spring Harb Symp Quant Biol* 1987, **52**:863-867.

3. Chang BH, Shimmin LC, Shyue SK, Hewett-Emmett D, Li WH: **Weak male-driven molecular evolution in rodents.** *Proc Natl Acad Sci USA* 1994, **91**:827-831.

4. Li WH, Ellsworth DL, Krushkal J, Chang BHJ, Hewett-Emmett D: **Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis.** *Mol Phyl Evol* 1996, **5**:182-187.

5. Olderburg J, Schwaab R, Grimm T, Zerres K, Hakenberg P, Brackmann HH, Olek K: **Direct and indirect estimation of the sex ratio of mutation frequencies in hemophilia A.** *Am J Hum Genet* 1993, **53**:1229-1238.

6. Hurst LD, Ellegren H: **Sex biases in the mutation rate.** *Trends Genet* 1998, **14**:446-452.

7. McVean GT, Hurst LD: **Evidence for a selectively favourable reduction in the mutation rate of the X chromosome.** *Nature* 1997, **386**:388-392.

8. Ellegren H, Fridolfsson AK: **Male-driven evolution of DNA sequences in birds.** *Nat Genet* 1997, **17**:182-184.

9. Kahn NW, Quinn TW: **Male-driven evolution among Eoaves? A test of the replicative division hypothesis in a heterogametic female (ZW) system.** *J Mol Evol* 1999, **49**:750-759.

10. Carmichael AN, Fridolfsson AK, Halverson J, Ellegren H: **Male-driven evolution rates revealed from Z and W chromosome-linked ATP synthase α-subunit (ATP5A1) sequences in birds.** *J Mol Evol* 2000, **50**:443-447.

11. Slattery J, O'Brien SJ: **Patterns of Y and X chromosome DNA sequence divergence during the Felidae radiation.** *Genetics* 1998, **148**:1245-1255.

12. Lawson LJ, Hewitt GM: **Comparison of substitution rates in ZFX and ZFY introns of sheep and goat related species supports the hypothesis of male-biased mutation rates.** *J Mol Evol* 2002, **54**:54-61.

13. Bohossian HB, Skaletsky H, Page DC: **Unexpectedly similar rates of nucleotide substitution found in male and female hominids.** *Nature* 2000, **406**:622-625.

14. •• Makova KD, Li WH: **Strong male-driven evolution of DNA sequences in humans and apes.** *Nature* 2002, **416**:624-626.

The first study that used homologous noncoding regions in an autosome (chromosome 3) and chromosome Y to estimate the sex ratio of mutation rate. As this study does not involve any X-linked sequence, it avoids the possibility that the low mutation rate in X-linked sequences is caused by a reduced mutation rate favored by natural selection rather than as a result of sex differences. It also points out that closely related species are not suitable for estimating the sex ratio of mutation rate.

15. International Human Genome Sequencing Consortium: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.

16. Monk M: **Epigenetic programming of differential gene expression in development and evolution.** *Dev Genet* 1995, **17**:188-197.

17. Chang BH, Hewett-Emmett D, Li WH: **Male-to-female ratios of mutation rate in higher primates estimated from intron sequences.** *Zool Studies* 1996, **35**:36-48.

18. Filipski J: **Why the rate of silent codon substitutions is variable within a vertebrate genome.** *J Theor Biol* 1988, **34**:159-164.

19. Wolfe KH, Sharp PM, Li WH: **Mutation rates differ among regions of the mammalian genome.** *Nature* 1989, **337**:283-285.

20. Casane D, Boissinot S, Chang BHJ, Shimmin LC, Li WH: **Mutation pattern variation among regions of the primate genome.** *J Mol Evol* 1997, **45**:216-226.

21. Matassi G, Sharp PM, Gautier C: **Chromosomal location effects on gene sequences evolution in mammals.** *Curr Biol* 1999, **9**:786-791.

22. Gu Z, Wang H, Nekrutenko A, Li WH: **Densities, length proportions, and other distributional features of repetitive sequences in the human genome estimated from 430 megabases of genomic sequence.** *Gene* 2000, **259**:81-88.

23. •• Lercher MJ, Williams EJB, Hurst LD: **Local similarity in evolutionary rates extends over whole chromosomes in human-rodent and mouse-rat comparisons: implications for understanding the mechanistic basis of the male mutation bias.** *Mol Biol Evol* 2001, **18**:2032-2039.

The authors evaluate the relationship between evolutionary rates and genetic linkage, while minimizing statistical artefacts. They showed that in human–rodent comparisons and in mouse–rat comparisons, linked genes have similar K_a and K_s values. However, the correlation in the mouse–rat comparisons was weaker, suggesting that the human and rodent genomes differ in their local patterns of mutation or selection on synonymous sites. Further, the local similarity in evolutionary rates extends over whole mouse chomosomes.

24. Castresana J: **Genes on human chromosome 19 show extreme divergence from the mouse orthologs and a high GC content.** *Nucleic Acids Res* 2002, **8**:1751-1756.

25. • Ebersberger I, Metzler D, Schawarts C, Pääbo S: **Genomewide comparison of DNA sequences between humans and chimpanzees.** *Am J Hum Genet* 2002, **70**:1490-1497.

The authors obtained almost 2 million bases of nonrepetitive chimpanzee sequences by sequencing randomly sheared genomic DNA. Comparisons with their human counterparts revealed the patterns of substitutions that are dependent on both the genomic regions and the times spent in the two different germlines.

26. Ticher A, Graur D: **Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein-coding genes.** *J Mol Evol* 1989, **28**:286-298.

27. Wolfe KH, Sharp PM: **Mammalian gene evolution: nucleotide sequence divergence between mouse and rat.** *J Mol Evol* 1993, **37**:441-456.

28. Bernardi G, Mouchiroud D, Gautier C: **Silent substitutions in mammalian genomes and their evolutionary implications.** *J Mol Evol* 1993, **37**:583-589.

29. Bielawski JP, Dunn KA, Yang Z: **Rates of nucleotide substitution and mammalian nuclear gene evolution: approximate and maximum-likelihood methods lead to different conclusions.** *Genetics* 2000, **156**:1299-1350.

30. Hurst LD, Williams EJB: **Covariation of GC content and the silent site substitution rate in rodents: implications for methodology and for the evolution of isochores.** *Gene* 2000, **261**:107-114.

31. • Lercher MJ, Hurst LD: **Human SNP variability and mutation rate are higher in regions of high recombination.** *Trends Genet* 2002, **18**:337-340.

Using regression analyses, the authors demonstrate a positive correlation between recombination rate and nucleotide diversity in the human genome, and show that at least part of the correlation appears to be mediated by regional differences in mutation rate.

32. Yi S, Ellsworth DL, Li WH: **Slow molecular clocks in old world monkeys, apes and humans.** *Mol Biol Evol* 2002, in press.

33. Perry J, Ashworth A: **Evolutionary rate of a gene affected by chromosomal position.** *Curr Biol* 1999, **9**:987-989.

34. Strathern J, Shafer B, McGill C: **DNA synthesis errors associated with double-strand-break repair.** *Genetics* 1995, **140**:965-972.

35. Papavasiliou FN, Schatz DG: **Cell-cycle-regulated DNA double-strand breaks in somatic hypermutation of immunoglobulin genes.** *Nature* 2000, **408**:216-220.

36. Eyre-Walker A: **Recombination and mammalian genome evolution.** *Proc R Soc Lond Ser B* 1993, **252**:237-243.

37. • Fullerton SM, Bernardo Carvalho A, Clark AG: **Local rates of recombination are positively correlated with GC content in the human genome.** *Mol Biol Evol* 2001, **18**:1139-1142.

The first study to show that GC content and recombination are correlated within intron sequences. The authors suggest that local recombination rate is an important indicator of compositional heterogeneity in the human genome.

38. • Marais G, Mouchiroud D, Duret L: **Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes.** *Proc Natl Acad Sci USA* 2001, **98**:5688-5692.

The authors analyzed the pattern of codon usage in the *C. elegans* and *D. melanogaster* genomes and showed a positive correlation between recombination rate and the frequency of optimal codons. The authors claimed that this effect is as a result of a mutational bias toward G and C in regions of high recombination rate, possibly as a direct consequence of the recombination process. This is one of the first studies to contrast the

mutationist hypothesis and the selectionist hypothesis to explain the genomic scale codon usage. However, see [59] for additional analyses and critiques.

39. Brown TC, Jiricny J: **Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells.** *Cell* 1988, **54**:705-711.

40. Takano-Shimizu T: **Local changes in GC/AT substitution biases and in crossover frequencies on *Drosophila* chromosomes.** *Mol Biol Evol* 2001, **18**:606-619.

41. Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD: **Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*.** *Proc Natl Acad Sci USA* 2000, **97**:11383-11390.

42. Smith NGC, Hurst LD: **The causes of synonymous rate variation in the rodent genome: can substitution rates be used to estimate the sex bias in mutation rate?** *Genetics* 1999, **152**:661-673.

43. Gu X, Li WH: **Higher rates of amino acid substitution in rodents than in humans.** *Mol Phylogenet Evol* 1992, **1**:211-214.

44. Mimault C, Giraud G, Courtois V, Cailloux F, Boire JV, Dastugue B, Boespflug-Tanguy O: **Brain Dysmeylinating Disease: Proteolipoprotein gene analysis in 82 patients with sporadic Pelizaeus-Merzbacher Disease: duplications, the major cause of the disease, originate more frequently in male germ cells, but point mutations do not. The Clinical European Network on Brain Dysmyelinating Disease.** *Am J Hum Genet* 1999, **65**:360-369.

45. Girad M, Couvert P, Carrié A, Tardieu M, Chelly J, Beldjord C, Bienvenu T: **Parental origin of *de novo* MECP2 mutations in Rett syndrome.** *Eur J Hum Genet* 2001, **9**:231-236.

46. Trappe R, Laccone F, Cobilanschi J, Meins M, Huppke P, Hanefeld F, Engel W: **MECP2 mutations in sporadic cases of Rett syndrome are almost exclusively of paternal origin.** *Am J Hum Genet* 2001, **68**:1093-1101.

47. Wilkin D, Szabo JK, Cameron R, Henderson S, Bellus GA, Mack ML, Kaitila I, Loughlin J, Munnich A, Sykes B *et al.*: **Mutations in fibroblast growth-factor receptor 3 in sporadic cases of Achondroplasia occur exclusively on the paternally derived chromosome.** *Am J Hum Genet* 1998, **63**:711-716.

48. Moloney DM, Slaney SF, Oldridge M, Wall SA, Sahlin P, Stenman G, Wilkie AOM: **Exclusive paternal origin of new mutations in Apert syndrome.** *Nat Genet* 1996, **13**:48-53.

49. Glaser RL, Jiang W, Boyadjiev SA, Tran AK, Zachary AA, Maldergem LV, Johnson D, Walsh S, Oldridge M, Wall SA *et al.*: **Paternal origin of FGFR2 mutations in sporadic cases of Couzon syndrome and Pfeiffer syndrome.** *Am J Hum Genet* 2000, **66**:768-777.

50. Nordenskjöld A, Friedman E, Anvret M: **WT1 mutations in patients with Denys-Drash syndrome: a novel mutation in exon 8 and paternal allele origin.** *Hum Genet* 1994, **93**:115-120.

51. Yin L, Seri M, Barone V, Tocco T, Scaranari M, Romeo G: **Prevalence and parental origin of *de novo* RET mutations in Hirschsprung's disease.** *Eur J Hum Genet* 1996, **4**:356-358.

52. Shuffenecker I, Ginet N, Goldgar D, Eng C, Chambe B, Boneu A, Houdent C, Pallo D, Schlumberger M, Thivolet C, Lenoir GM: **Prevalence and parental origin of *de novo* RET mutations in multiple endocrine neoplasia type 2A and familial medullary thyroid carcinoma. Le Groupe D'Étude des Tumeur À Calcitonine.** *Am J Hum Genet* 1997, **60**:233-237.

53. Carson KM, Dou S, Chi D, Scavarda N, Toshima K, Jackson CE, Wells SA, Goodfellow PJ, Donis-Keller H: **Single missense mutation in the tyrosine kinase catalytic domain of the RET protooncogene is associated with multiple endocrine neoplasia type 2B.** *Proc Natl Acad Sci USA* 1994, **91**:1579-1583.

54. Kluwe L, Mauntner V, Parry DM, Jacoby LB, Baser M, Gusella J, Davis K, Stavrou D, MacCollin M: **The parental origin of new mutations in neurofibromatosis 2.** *Neurogenet* 2000, **3**:17-24.

55. Richards F, Payne SJ, Zbar B, Affara NA, Ferguson-Smith A, Maher ER: **Molecular analysis of *de novo* germline mutations in the von Hippel-Lindau disease gene.** *Hum Mol Genet* 1995, **4**:2139-2143.

56. Huang W, Chang BH, Gu X, Hewett-Emmett D, Li W: **Sex differences in mutation rate in higher primates estimated from AMG intron sequences.** *J Mol Evol* 1997, **44**:463-465.

57. Shimmin LC, Chang BH, Li WH: **Male-driven evolution of DNA sequences.** *Nature* 1993, **362**:745-747.

58. Chang BH, Li WH: **Estimating the intensity of male-driven evolution in rodents by using X-linked and Y-linked Ube 1 genes and pseudogenes.** *J Mol Evol* 1995, **40**:70-77.

60. Hey J, Kliman RM: **Interactions between natural selection, recombination and gene density in the genes of *Drosophila*.** *Genetics* 2002, **160**:595-608.

61. Li WH: **Distribution of nucleotide differences between two randomly chosen cistrons in a finite population.** *Genetics* 1977, **85**:331-337.