Human mutation—blame (mostly) men

Hans Ellegren

Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden.
e-mail: Hans.Ellegren@ebc.uu.se

The relative contribution of mothers and fathers to mutation can be studied by evolutionary analysis of sex-linked DNA sequences. A new study shows that mutations occur in men at a rate five times of that in women, lending support to the idea of 'male-driven evolution'.

It is estimated that a human DNA sequence differs from that of one's parents at about 100 nucleotide positions[1]. These sites generally represent germline mutations that have arisen during the production of gametes in the parental generation. A classical question in human genetics is how many of these mutations come from mothers and fathers, respectively[2]. Is the paternal and maternal contribution to genetic novelty or erosion equal, or is there a male bias in the mutation rate? As the question has widespread implications for evolutionary, medical and molecular genetics, it has generated considerable interest over the years, without a consensus as to the extent of male bias. Some recent studies have indicated that the male bias may be much less than previously thought[3,4]. In a recent issue of *Nature*, however, Kateryna Makova and Wen-Hsiung Li[5] provide evidence that a significant proportion of human mutations is indeed of paternal origin.

Measuring mutation rates

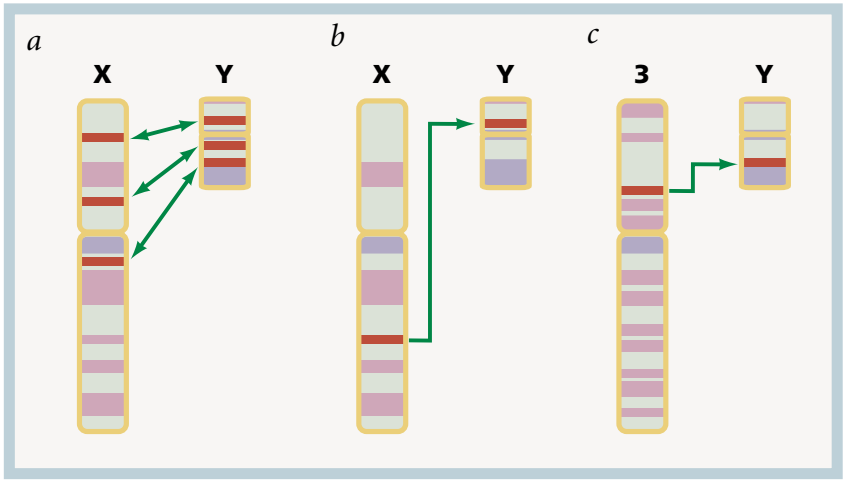
How to measure the rate of mutation in modern humans? One approach is to identify large numbers of random point mutations, which are transmitted from one generation to another in random families, thereby allowing one to calculate sex-specific mutation rates. This is a formidable task, however, and is hardly feasible with available technology. Rather than blindly grope for these random mutations, an alternative approach is to combine information from the very rare cases in which new mutations affect the phenotype of the offspring.

However, studies of the parental origin of novel mutations, like those causing achondroplasia and Apert syndrome, may be severely biased by recurrent mutations at individual CpG sites where methylation levels differ between the sexes. The observation that some diseases are caused exclusively by paternally derived mutations seems unlikely to apply to the genome as a whole[6]. The identification of *de novo* mutations in hypermutable markers such as microsatellites by pedigree analysis may offer another alternative, and has suggested a male:female mutation rate ratio (α_m) of 3–5 (refs 7,8). We are then considering replication slippage–induced mutations in repetitive DNA, however, rather than the point mutations that are most prevalent in unique DNA.

Given the great difficulty in identifying random mutations in pedigrees, the issue of sex-specific mutation rates is better approached by molecular evolutionary studies of sex chromosomes[9]. Whereas the X chromosome spends only one-third of the time in males, the Y chromosome is present in males only. For paralogous genes shared between X and Y (*ZFX*/*ZFY*, for example; Fig. 1*a*), the rate of presumably neutral substitutions in primate comparisons is higher on Y than on X. The observed rate differences correspond to an α_m of approximately 4–6, which suggests a rather distinct male bias in hominoids[10]. However, these data were recently challenged in a study by Bohossian *et al.*[3], who suggested that α_m is only 1.7, using data from a large region transposed from X to Y after the human lineage split with chimpanzees (Fig. 1*b*). Although perhaps surprising, this was a seemingly solid observation, given that the amount of sequence data included exceeded that of earlier studies by an order of magnitude.

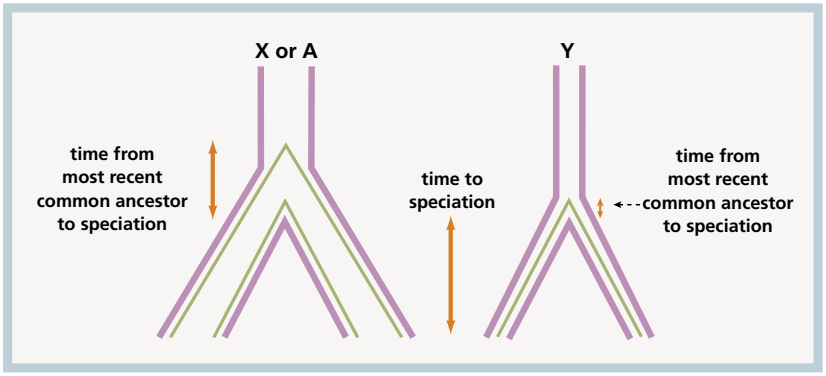
On ancestral polymorphism

Makova and Li[5] now argue that the low estimate of α_m was flawed, owing to the effect of ancestral polymorphism. Bohossian *et al.*[3] inferred the number of nucleotide substitutions having arisen in the X and Y chromosome lineages subsequent to transposition by comparison to chimpanzee X-chromosome outgroup sequence (note that as the X-to-Y transposition occurred after the human–chimpanzee split, chimpanzee has only the X-linked sequence). One possible scenario for the X-to-Y transposition event in the human lineage is that the particular X-chromosome allele being transposed already contained polymorphic sites that distinguished it from other X-chromosome sequences. This could



BOB CRIMI

Fig. 1 Chromosomal approaches used to address sex-specific mutation rates. ***a***, Comparison of substitution rates in homologous (gametologous) genes shared between X and Y. ***b***, Comparison of substitution rates in a region recently transposed from X to Y. ***c***, Comparison of substitution rates in a region transposed from chromosome 3 to Y. Although the comparison in *b* allows one to study mutations in the human lineage after the split with chimpanzees, the effect of ancestral polymorphism on estimates of divergence may be significant. Makova and Li[5] argue that the comparison depicted in *c* is advantageous compared with the one in *a*, as it involves homologous sequences of more recent ancestry and avoids the possible effects of a specific reduction in the X-chromosome mutation rate.



BOB CRIMI

Fig. 2 The effect of ancestral polymorphism (width of branches) on the possible times to the most recent common ancestor in relation to time to speciation. The left panel depicts a scenario with significant ancient nucleotide diversity, implying that the time to the most recent common ancestor may be longer than the time to speciation. The right panel shows a situation with limited ancestral polymorphism, where the time to the most recent common ancestor is about the same as the time to speciation.

imply that some of the mutations interpreted as being of Y-chromosome origin (sites where human X = chimpanzee X ≠ human Y) could actually have arisen on the X chromosome, and were polymorphic prior to transposition. The effect of this would be to underestimate α_m.

Makova and Li[5] also make a more general point about ancient polymorphism and estimates of α_m. Ancestral polymorphism can have a significant effect on the average divergence between two species—or two sequences—if they are closely related (Fig. 2). In other words, the time to the most recent common ancestor may significantly exceed the time to speciation. If ancestral polymorphism is low or effectively absent, however, the times to the most recent common ancestor and speciation would be more similar. This dichotomy may apply to X-and Y-chromosome sequences. The Y chromosome is typically very much lower in nucleotide diversity than the X[11], probably because the smaller effective population size and the absence of recombination allows selection to wipe out polymorphism. Therefore, when closely related species are compared, ancient nucleotide diversity may contribute to the estimates of divergence in X, but not in Y. The combined effect would be to lower the estimates of α_m, relative to that when more distantly related species are compared and where ancient nucleotide diversity has less influence on divergence.

DAZ to the rescue

To avoid these problems, and the potential pitfall of a suggested chromosome-specific reduction in the X-chromosome mutation-rate[12], Makova and Li[5] studied the *DAZ* (*d*eleted in *az*oospermia) locus that was transposed from chromosome 3 to Y after the split between New World and Old World monkeys (Fig. 1*c*). In phylogenetic trees based on sequences from a number of primate species, they noted that internal branch lengths were significantly longer in trees constructed using Y sequences than in trees based on autosomal sequences. In contrast, external branches generally showed less pronounced differences in length between the two trees. This is consistent with divergence of autosomal, but not Y, sequences being affected by ancestral polymorphism.

By using the formula $Y/A = 2\alpha_m/(1 + \alpha_m)$ on summed internal branch lengths of human–bonobo–gorilla–siamang–gibbon trees, they estimated α_m to be 5.2. Clearly, this estimate is based on mutations that arose during several million years of evolution and does not necessarily reflect the mutation processes in contemporary human populations. However, there are no obvious reasons to think that human α_m would be lower than in other primates. If anything, we should rather expect human α_m to be higher owing to a longer generation time and an older age at reproduction than our ancestors and the apes.

The new α_m estimate of 5.2 carries a large 95% confidence interval (2.44 to ∞), but it should be noted that it is similar to that obtained in previous comparisons of homologous X–Y genes[10]. If we accept a significant male bias in human mutation rate, it would suggest that a large proportion of germline mutations derive from replication-associated processes (replication errors, for example), a conclusion that can be drawn from the fact that the number of germline cell divisions in spermatogenesis vastly exceeds that in oogenesis[13]. The relative importance of replication-independent mutagenic factors on germline mutation would in this case be minor. As spermatogonial germ (stem) cells are continuously mitotically active in adult men, another consequence of these results is that we should expect the male mutation rate to increase with age[14].

Is the new study without caveats? Although the results seem convincing, it might be premature to consider the case entirely closed. One potentially confounding factor relates to a recent observation that there is significant heterogeneity in the local mutation rate[15]. One could argue that the use of homologous sequences on different chromosomes would imply that the only cause of a difference in their mutation rate is the varying length of time spent in the male and female germlines. However, if the local mutation rate varies over the genomic landscape in a way that it is primarily determined by the regional genomic context rather than the primary sequence context, then homologous sequences may not necessarily have similar intrinsic mutation rates. To judge whether this is the case, we need to learn more about the determinants of local mutation rate variation. □

1. Nachman, M.W. & Crowell, S.L. *Genetics* **156**, 297–304 (2000).
2. Haldane, J.B.S. *Annals Eug.* **13**, 262–271 (1947).
3. Bohossian, H.B., Skaletsky, H. & Page, D.C. *Nature* **406**, 622–625 (2000).
4. International Human Genome Sequencing Consortium. *Nature* **409**, 860–921 (2001).
5. Makova, K.D. & Li, W.H. *Nature* **416**, 624–626 (2002).
6. Hurst, L.D. & Ellegren, H. *Trends Genet.* **14**, 446–452 (1998).
7. Ellegren, H. *Nature Genet.* **24**, 400–402 (2000).
8. Xu, X., Peng, M. & Fang, Z. *Nature Genet.* **24**, 396–399 (2000).
9. Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K. & Yasunaga, T. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 863–867 (1987).
10. Shimmin, L.C., Chang, B.H.J. & Li, W.H. *Nature* **362**, 745–747 (1993).
11. The International SNP Map Working Group. *Nature* **409**, 928–933 (2001).
12. McVean, G.T. & Hurst, L.D. *Nature* **386**, 388–392 (1997).
13. Vogel, F. & Motulsky, A.G. *Human Genetics: Problems and Approaches* (Springer, Berlin, 1997).
14. Crow, J.F. *Naure Rev. Genet.* **1**, 40–47 (2000).
15. Lercher, M.J., Williams, E.J.B. & Hurst, L.D. *Mol. Biol. Evol.* **18**, 2032–2039 (2001).