How often does human DNA mutate?

Yali Xue

(The Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK)

Editor's comments: The human mutation rate – how often new changes appear in the DNA – is fundamental to understanding many aspects of medical genetics and human evolutionary genetics. But it is low, and has therefore been difficult to measure. In the past, scientists could only estimate it approximately, either by observing how often mutant phenotypes appeared, or by comparison of humans and closely related species, such as chimpanzee, where many mutations could accumulate but the time period was uncertain. Now, a new study supported by the NSFC in China and The Royal Society in the UK reports the first direct measurement of the human mutation rate at the individual letters (nucleotides or bases) of DNA. This was possible because new (next)-generation sequencing technology is much more powerful than the methods available previously. The work was published in the leading journal *Current Biology* on 15th September 2009. The results were reported in the news by Nature, Science and the BBC, as well as in more than 20 Chinese newspapers and radio stations after the work first appeared online on 27th August. It was also one of the research highlights in Nature on 3rd September, which commented "This direct measurement of the human mutation rate should help researchers to refine evolutionary dating and better understand the source of genetic disease". From the work, researchers could estimate that everyone has around 200 new mutations in their genome; as the authors said, "we are all mutants". The ability to reliably measure rates of DNA mutation means we can begin to ask how mutation rates vary between different regions of the genome and perhaps also between different individuals.

Key words: Mutation rate, deep-rooting pedigree, next-generation sequencing, human Y chromosome

Mutation is an inevitable and fundamental property of DNA: changes occur by chance when DNA is replicated and passed on to the next generation, and this process can be affected by factors in the environment. Fortunately, most changes have no effect whatever. A tiny fraction are harmful and can lead to cancer or genetic disease, and an even smaller proportion are actually beneficial, and provide the raw material for evolution. We therefore want to understand how often mutations do occur, so we can have a better understanding of both their bad and good effects.

In the past, it was just impossible to measure this important parameter directly because it was so low. Scientists only could estimate it approximately, and developed two ways of doing this. In one, they looked for new cases a specific medical phenotype in a family, such as the blood disease haemophilia. If they detected all

the cases, and could determine what proportion were due to new mutations and also knew how big the gene was, they could calculate a rate. Alternatively, they could compare the DNA between human and closely related species, such as chimpanzee. Here, if they knew how many years or generations separated the species, they could again calculate a rate. But there were a lot of uncertainties involved.

Now, with the new generation of sequencing technology with its lower cost and higher throughput, it has become possible to measure the rate directly for the first time. We therefore designed a study to do this as part of an international joint project supported by the National Natural Science Foundation, China and The Royal Society, UK. Our work has just been published in the 15th September issue of Current Biology[1], and we describe the results here. But even with the benefit of the latest technology, we had to design the project carefully.

We used the Y chromosome because it has several unique features which worked to our advantage. Most of it (almost 60,000,000 nucleotides) is transmitted simply from father to son each generation. This means that the son has exactly the same Y chromosome, apart from any mutations that have occurred. So does the son's son, although again apart from mutations that have occurred in the two generations, and so on for further generations. This would not be true of other parts of the genome, because they mix each generation.

In our study, we chose two men who were male-line relatives, but were separated by 13 generations. This meant that we would pick up mutations that had occurred in all these generations. We were particularly interested one particular family because it carries a Y-linked hearing impairment mutation[2], but this condition turned out to be irrelevant to our study. We began by purifying the Y chromosomes from the two men by flow sorting them. We then sequenced them by next-generation sequencing (in this case using the Illumina Genome Analyzer) to find the differences between the two Y chromosomes. Although the next-generation sequencing made this study possible, the high error rate of the technology gave us an enormous numbers of false positives (more than 30,000), far more than the number of real mutations we expected which was less than 10. But by using the known "gold standard" positions on the chromosome whose sequence was predicted from previous studies[3], we excluded the vast majority of these false positives and ended up with 23 candidate differences. This number was small enough to then use traditional Sanger sequencing to test all of them. In this way, we confirmed that 12 of the 23 were real differences.

But this was not the end. The chromosomes we sequenced came from cell lines and the 12 differences could have arisen in the family (real mutations) or in the cell lines (somatic mutations). Fortunately, we could distinguish between these possibilities by sequencing blood DNAs from the two men as well as from other individuals in the family. In the end, we found that eight of the differences were somatic mutations, and only four were real mutations that happened during transmission in the family. For technical reasons, we restricted the analysis to around 10 million bases from the two chromosomes, but this was enough to calculate the rate for the whole genome. Four mutations in 10 million bases of DNA over 13 generations gave us about one mutation

in every 30 million nucleotides per generation. In the six billion nucleotides in the complete genome, everyone has about 200 mutations.

The mutation rates are thought to vary in different parts of the genome, and this kind of variation can be followed up in future studies. Our study, for the first time ever, has shown that one can use next-generation sequencing technology to measure the very low mutation rate of human nuclear DNA reliably. Reassuringly, the mutation rate we observed is consistent with that inferred from evolutionary comparisons but can potentially be measured more precisely and provide new insights into human mutation processes.

In addition to demonstrating the power of the technology and paving the way for future mutation studies, this study also highlighted the potential advantages of using SNPs in forensic genetics in the future[4]. Currently, forensic scientists use markers called Y-STRs when they need to distinguish between different Y chromosomes. However, as we have seen, sons carry the same Y chromosome as their fathers, and these cannot be distinguished unless a mutation has occurred. In practice, male line relatives less than 20 generations apart are likely to carry the same Y-STR type, so are not distinguished by current methods. This was the case for the two men tested in our study: they showed exactly the same pattern with Y-STRs[5]. But they have four Y-SNP differences, so this means that SNPs are better than STRs at telling Y chromosomes apart. Indeed, if more of the Y chromosome was used, almost every Y chromosome could be distinguished. Some technical advances and cost reductions are needed before this can come into practice, but the possibility illustrates the surprising ways in which science advances.

Selected related links (comments on the work):

http://www.sanger.ac.uk/Info/Press/2009/090827.shtml

http://www.nature.com/nature/journal/v461/n7260/full/461015b.html

http://www.nature.com/news/2009/090827/full/news.2009.864.html

http://esciencenews.com/articles/2009/08/27/we.are.all.mutants

http://news.xinhuanet.com/tech/2009-08/29/content_11962256.htm

References:

1 Xue, Y. *et al*., Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Current Biology* (in press) (2009).

2 Wang, Q. J. *et al*., Y-linked inheritance of non-syndromic hearing impairment in a large Chinese family. *Journal of Medical Genetics* **41**, e80 (2004).

3 Karafet, T. M. *et al*., New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res*earch **18**, 830-838 (2008).

4 Xue, Y. and Tyler-Smith, C. The hare and the tortoise: one small step for four SNPs, one giant leap for SNP-kind. *Forensic Science International: Genetics* (in press) (2009).

5 Vermeulen, M. *et al*., Improving global and regional resolution of male lineage differentiation by simple single-copy Y-chromosomal short tandem repeat polymorphisms. *Forensic Science*

International: Genetics **3**, 205-213 (2009).