3

Estimating Human Point Mutation Rates from Codon Substitution Rates

Kazuharu Misawa
Research Program for Computational Science,
Research and Development Group for Next-Generation
Integrated Living Matter Simulation,
Fusion of Data and Analysis Research and
Development Team, Yokohama City, Kanagawa,
Japan

1. Introduction

Estimation of point mutation rates is essential for studying molecular evolution and genetics. Point mutation rates are also important for developing tools for genome analyses, such as those used for homology searches (Altschul et al., 1990), sequence alignments (Katoh et al., 2002; Larkin et al., 2007), gene finding (Misawa and Kikuno, 2010), or detecting natural selection (Nei and Gojobori, 1986; Hughes and Nei, 1988; Yang, 2007; Yang and Nielsen, 2008), and for reconstructing phylogenetic trees (Felsenstein, 2004; Sullivan and Joyce, 2005). Patterns of mutations also affects the neutrality test for population genetics (Misawa and Tajima, 1997). According to the neutral theory (Kimura, 1968), new alleles may be produced at the same rate per individual as they are substituted in a population. On the basis of this theory, mutation rates were estimated from neutral substitution rates.

One of the causes of mutations is the error during DNA replication (Pray, 2008). Since cell division is tightly linked to DNA replication, mutation rates are expected to correlate the number of cell divisions so that they are higher in sperms than in egg (Haldane, 1956; Miyata et al., 1987). The phenomenon in which mutation rate are higher in males than in females is called 'male-driven evolution' (Miyata et al., 1987). The previous studies show large discrepancies with regard to the effect of male-driven evolution on mutation rates (Li et al., 2002). To investigate the effect of DNA replication on mutation rates, the effect of male male-driven evolution was also investigated in this study.

Recent studies on mutation rates in human non-coding regions have shown that mutation rates in the human genome are negatively correlated to local GC content (Fryxell and Moon, 2005; Taylor et al., 2006; Tyekucheva et al., 2008; Walser et al., 2008) and to the densities of functional elements (Hardison et al., 2003; Hellmann et al., 2005; Tyekucheva et al., 2008). Mutation rates have also been shown to correlate with the distance of a gene from the telomere (Hellmann et al., 2005; Tyekucheva et al., 2008). A study on mutation rates in human coding regions showed that mutation rates also depend on the chromosome sizes, and this is probably attributed to the distance between the genes and telomeres (Misawa, 2011).

Mutation rates are sometimes dependent of the local context, especially the adjacent nucleotides (Cooper and Youssoufian, 1988; Cooper and Krawczak, 1989; Hobolth et al., 2006; Misawa et al., 2008; Misawa and Kikuno, 2009; Misawa, 2011). CpG hypermutability is a major cause of codon substitution in mammalian genes (Huttley, 2004; Lunter and Hein, 2004; Misawa et al., 2008; Misawa and Kikuno, 2009). CpG hypermutation occurs approximately 10 times or more rapidly than other types of point mutations do (Scarano et al., 1967; Bird, 1980; Lunter and Hein, 2004). Figure 1 shows cytosine (C), methylcytosine (methyl-C), and thymine (T). The notation CpG is used to distinguish a C followed by a guanine (G) from a Watson-Crick pair of C and G. CpG dinucleotides are often methylated at C by DNA methyltransferase (DNMT) (Wu and Zhang, 2011), and methyl-C spontaneously undergoes deamination to generate T. The mutation pressure of CpG hypermutability is so high in organisms with DNMT, that they share a similar pattern of amino acid substitutions (Misawa et al., 2008). This pattern of amino acid substitutions used be called the 'universal' trend (Jordan et al., 2005). Misawa et al. (2008) also showed that organisms who lost DNMT, such as *Buchnera* and *Saccharomyces* do not share the 'universal' trend. In the mouse, the effect of CpG hypermutability on codon preference is stronger than that of tRNA abundance (Misawa and Kikuno, 2011). When mutation rates are estimated, the effect of adjacent nucleotides on mutation rates and the direction of mutation rates should be considered.
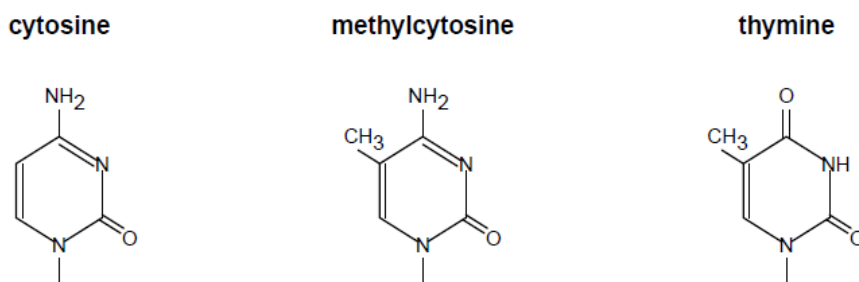

Fig. 1. Cytosine, methylcytosine, and thymine.

Some of previous studies used mutation models that do not consider the effect of adjacent nucleotides; these mutation models were the REV model (Tavare, 1986; Yang, 1994; Whelan et al., 2001) used by Hardison (Hardison et al., 2003), and the JC (Jukes and Cantor, 1969) and HKY models (Hasegawa et al., 1985) used by Tyekucheva (Tyekucheva et al., 2008). Tayler's (Taylor et al., 2006) study was based on two-species comparison, and therefore, the direction of mutation was unclear. Recently, I and my colleagues (Misawa and Kikuno, 2009; Misawa, 2011) estimated the mutation rates in humans by considering the effect of adjacent nucleotides on mutation rates and the direction of mutations. These studies, however, did not consider the effect of the distance of a gene from the telomere on mutation rates.

To understand the effect of DNA replication and the distance of a nucleotide site from the telomere, mutation rates were estimated using the codon substitution rates in the coding regions of thousands of human and chimpanzee genes by using autosomes and X chromosome; further, the ancestral gene sequences were inferred by assuming macaque

genes as the outgroup in this study. Regression analyses were conducted to evaluate the effect of GC content, gene density, and CpG island density on the rates of CpG-to-TpG mutations, TpG-to-CpG mutations, and non-CpG transitions and transversions.

2. Methods

2.1 Data set

To estimate the rates of mutation rates, we used 10,372 orthologous gene trios obtained from human, chimpanzee, and macaque genomes (Gibbs et al., 2007). Table 1 shows the sizes of human chromosomes of the NCBI human genome (Build 36) taken from the UCSC genome browser (http://genome.ucsc.edu). Table 1 also shows number of genes on each chromosome used in this study. These genes were binned into a series of 10-Mb windows of human DNA depending on the positions of the midpoint of genes.

Chromosome	Chromosome Size (base pair)	No. of genes used	No. of fourfold sites used
1	247249719	1123	236111
2	242951149	708	158041
3	199501827	644	140972
4	191273063	442	96920
5	180857866	521	120438
6	170899992	623	120260
7	158821424	472	97566
8	146274826	359	70764
9	140273252	420	89507
10	135374737	431	88630
11	134452384	593	121652
12	132349534	567	119251
13	114142980	198	44238
14	106368585	360	78504
15	100338915	321	75640
16	88827254	435	96218
17	78774742	549	119793
18	76117153	192	39089
19	63811651	525	100125
20	62435964	319	62578
21	46944323	127	25595
22	49691432	199	41400
X	154913754	244	41061

Table 1. Summary of data

2.2 Estimation of mutation rates from substitution rate on fourfold sites

To estimate mutation rates from nucleotide substitution rates, it is important to separate substitutions representing neutral evolutionary drift from those influenced by selection. Following Hardison et al.'s study (Hardison et al., 2003), fourfold sites were analyzed in this

study. Fourfold sites are sites marked "N" in the codons GCN (Ala), CCN (Pro), TCN (Ser), ACN (Thr), CGN (Arg), GGN (Gly), CTN (Leu), and GTN (Val). is worth noting that nucleotide substitutions at fourfold sites do not change amino acids so that they can be considered as neutral. In the case of mouse, the effect of codon preference on nucleotide substitutions would be smaller than that of mutations (Misawa and Kikuno, 2011). In this study, the mutation rates were estimated from the substitution rates at fourfold sites by assuming that the substitutions at fourfold sites are neutral.

2.3 Classification of sites and mutations

To evaluate the effect of CpG hypermutability on mutation rates, fourfold sites were classified into three categories depending on the adjacent nucleotides, namely, CpG sites, TpG sites, and usual sites. If the nucleotide at fourfold site is C and the first nucleotide of 3'-adjacent codon is G, the site is classified into CpG site. If the nucleotide at fourfold site is T and the first nucleotide of 3'-adjacent codon is G, the site is classified into TpG site. If the nucleotide at fourfold site is C and the first nucleotide of 3'-adjacent codon is A, the site is classified into TpG site, because on the complementary strand T is next to G. The sites that are neither CpG site nor TpG site are classified as usual site.

Mutations were classified into 4 categories: CpG to TpG mutations, TpG to CpG mutations, and non- CpG transitions and transversions. To distinguish CpG to TpG mutations and TpG to CpG mutations from non-CpG transitions, the adjacent nucleotides were also considered. If the observed nucleotide was T, its ancestral nucleotide was C, and the downstream nucleotide was G, the mutation was classified as a CpG to TpG substitution. For example, a mutation from CCC (Pro) to CCT (Pro) is classified into a CpG to TpG mutation only when 3' adjacent nucleotide of the codon is G. If the observed nucleotide was A, its ancestral nucleotide was G, and the upstream nucleotide was C, the mutation was again classified as a CpG to TpG mutation because a CpG to TpG mutation occurs on the complementary strand of DNA. For example, a mutation from CCG (Pro) to CCA (Pro) corresponds to a CpG to TpG mutation. If the observed nucleotide was C, its ancestral nucleotide was T, and the downstream nucleotide was G, the mutation was classified as a CpG to TpG mutation. If the observed nucleotide was G, its ancestral nucleotide was A, and the upstream nucleotide was C, the mutation was again classified as a CpG to TpG mutation. Other types of mutation are classified into two types: transitions and transversions (Misawa et al., 2008; Misawa and Kikuno, 2009; Misawa, 2011; Misawa and Kikuno, 2011).

2.4 Estimation of mutation rates by using the Maximum Parsimony (MP) method

I determined the codon sequences of the common ancestors of humans and chimpanzees by using the maximum parsimony (MP) method. Next, I counted the number of codon substitutions that had occurred along the human lineage. For some codon trios, the ancestral state between the human and chimpanzee codons appeared ambiguous when estimated by the MP method. In such cases, all possible ancestral states were treated equally. I also calculated the mutation rates by dividing the number of codon substitutions occurring annually by the number of ancestral codons. It was assumed that the human-chimpanzee divergence occurred 5 million years (MY) ago (Horai et al., 1995). Macaque genes were used as the outgroup. To calculate confidence intervals of the estimates, the binomial distribution was assumed. The awk program used in the analysis is available from the author upon request.

2.5 Comparison between mutation rate and the distance from the telomere by using regression analysis

I compared the mutation rates with the distance of a central position of 10-MB windows from the telomere. These values were used in regression analyses, which were performed using the statistical software R (R Development Core Team, 2008).

3. Result

Table 2 shows the mean value of the mutation rate estimates per BY per site and their 99% confidence intervals. The mutation rate of CpG to TpG is similar to that obtained by the previous study, but the rate of other types of mutation is lower than those obtained by the previous study (Misawa, 2011). The mutation rate of CpG to TpG is about 10 times higher than that of transitions and transversions on usual sites. This ratio is similar to previous studies (Scarano et al., 1967; Bird, 1980; Lunter and Hein, 2004). The mutation rates on autosomes were similar to those on X chromosome, except the rates of transversion on usual sites. The rate of transversion on usual sites on autosomes is significantly higher than that of X chromosomes after the Bonferroni correction ($P < 0.05$).

Mutation type	Site type	Autosomes				X chromosome				
CpG to TpG	CpG	3.91	(3.74	- 4.08	)	3.96	(2.07	- 4.58	)	
Transversion	CpG	0.79	(0.72	- 0.87	)	0.44	(0.15	- 1.18	)	
TpG to CpG	TpG	0.47	(0.44	- 0.49	)	0.47	(0.20	- 0.50	)	
non-CpG transition	TpG	0.27	(0.26	- 0.29	)	0.09	(0.11	- 0.35	)	
Transversion	TpG	0.29	(0.27	- 0.31	)	0.30	(0.04	- 0.22	)	
non-CpG transition	Usual	0.32	(0.29	- 0.34	)	0.41	(0.09	- 0.33	)	
Transversion	Usual	0.51	(0.49	- 0.54	)	0.24	(0.18	- 0.48	)	*

*Significantly lower than autosomes at 5% level

Table 2. Avarage mutation rates per BY per site and their 99% confidence intervals

Figures 2, 3, and 4 are a scatter plot of the mutation rates versus the distances to telomeres. Upper panel of figure 3 is for autosomes and the lower panel is for X chromosome. Figure 3 shows a scatter plot of the mutation rates on TpG sites. Figure 4 shows a scatter plot of the mutation rates on usual sites. These figures show that mutation rates are negatively correlated to the distances from telomeres. These figures also show that a large variation in the rates of mutations exist among genomic regions.

Table 3 shows Pearson's correlation coefficient between mutation rates and the distances from telomeres. Only the mutation rates of TpG to CpG and transversion on TpG sites were significantly correlated to the distance from telomeres ($P < 0.001$ and $P < 0.01$, respectively) after the Bonferroni correction. There were no significant differences of the correlation coefficients between autosomes and X chromosomes.

Mutation type	Site type	Autosomes		X chromosome
CpG to TpG	CpG	-0.06		-0.42
Transversion	CpG	0.18		-0.09
TpG to CpG	TpG	-0.20	***	-0.43
non-CpG transition	TpG	-0.10		-0.05
Transversion	TpG	-0.20	**	0.13
non-CpG transition	Usual	-0.13		-0.61
Transversion	Usual	-0.03		-0.34

*at 1% level of significance; ***at 0.1% level of significance

Table 3. Correlation coefficient between mutation rates and distances to telomeres by regression analysis
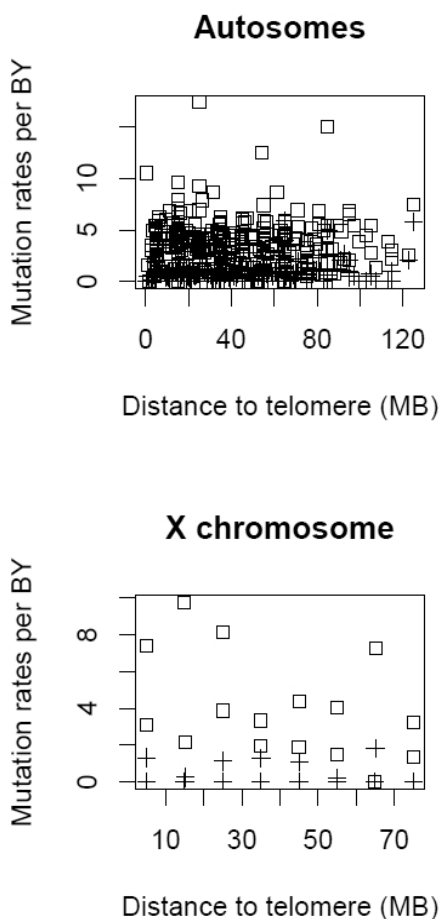
Fig. 2. A scatter plot of the mutation rates on CpG sites versus the distances to telomeres. The open circles (○) show the TpG to CpG mutation rates. The open triangles (Δ) show the rates of non-CpG transition. The crosses (+) show the rates of transition.
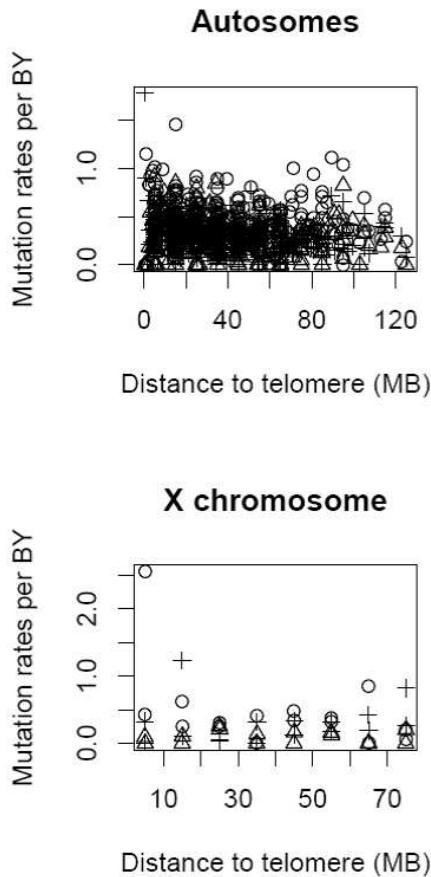
Fig. 3. A scatter plot of the mutation rates on TpG sites versus the distances to telomeres. The open squares (□) show the CpG to TpG mutation rates. The open circles (○) show the TpG to CpG mutation rates. The open triangles (∆) show the rates of non-CpG transition. The crosses (+) show the rates of transition.

4. Discussion

In this study, the rates of CpG to TpG mutations, TpG to CpG mutations, and non-CpG transitions and transversions were estimated by comparing the coding regions of thousands of human and chimpanzee genes from entire genome and inferring their ancestral sequences by assuming macaque genes as the outgroup. The mutation rate of CpG to TpG is about 10 times higher than that of transitions and transversions on usual sites. This ratio is similar to previous studies (Scarano et al., 1967; Bird, 1980; Lunter and Hein, 2004). The mutation rate of CpG to TpG is similar to that obtained by the previous study, but the rate of other types of mutation is lower than those obtained by the previous study (Misawa, 2011), probably because previous studies included nonsynonymous substitutions (Misawa, 2011) while only fourfold sites were analyzed in this study.
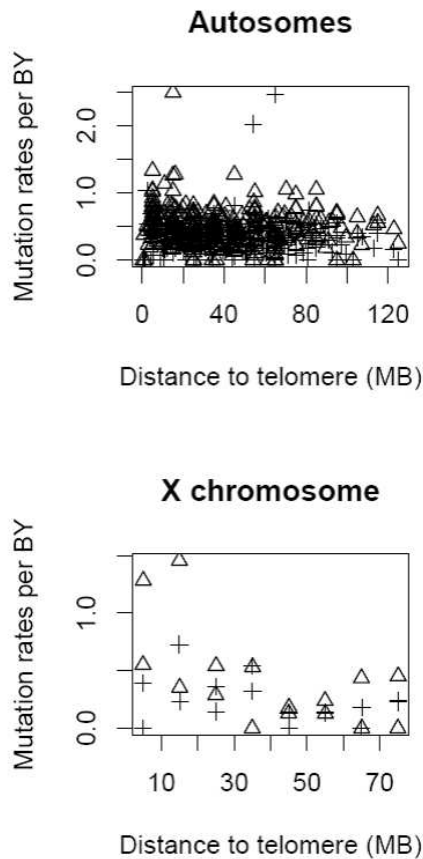
Autosomes

Distance to telomere (MB)

X chromosome

Distance to telomere (MB)

Fig. 4. A scatter plot of the mutation rates on usual sites versus the distances to telomeres. The open squares (□) show the CpG to TpG mutation rates. The open circles (○) show the TpG to CpG mutation rates. The open triangles (Δ) show the rates of non-CpG transition. The crosses (+) show the rates of transition.

As seen in table 2, a significant difference was not observed between the mutation rates of autosomes and X chromosome, except the rates of transversion on usual sites This result indicates that the effect of "male-driven evolution" (Miyata et al., 1987) is not strong. Figures 2, 3, and 4 show that a large variation in the rates of mutations among genomic regions. These results might be caused by the fact that mutation rates are affected by various factors, such as gene density, GC contents and the density of CpG islands (Misawa, 2011).

Vogel and Motulsky (Vogel and Motulsky, 1997) pointed out that since the deamination of methyl-C occurs spontaneously and is independent of DNA replication, the rate of CpG mutations should be scaled with time and not with the number of cell divisions. Recently, Taylor et al. (Taylor et al., 2006) investigated male mutation bias separately at non-CpG and CpG sites by using human-chimpanzee whole-genome alignments. They concluded that CpG hypermutation is weakly affected by the number of cell divisions. As pointed out by

Misawa (2011), the effect of male-driven evolution on CpG hypermutation is weaker than that of other chromosomal properties. Further study must be necessary.

Figures 2, 3, and 4 indicate that the CpG to TpG substitution rates were negatively correlated to the distances from telomeres. This is consistent with previous studies (Hellmann et al., 2005; Tyekucheva et al., 2008), although their methods for estimating mutation rates were different from this study. However, Table 3 shows that only the mutation rates of TpG to CpG and transversion on TpG sites were significantly correlated to the distance from telomeres after the Bonferroni correction. Tyekucheva et al. (Tyekucheva et al., 2008) suggested the existence of additional mutagenic mechanisms that increase neutral substitution rates in subtelomeric regions. Increased divergence near telomeres has been linked to direct and indirect effects of large-scale chromosomal structure. If the correlation coefficients between mutation rates and the distances from telomeres on X chromosome are different from that on autosomes, cell division and DNA replication might be a part of such mutation mechanisms. Unfortunately, no significant differences of the correlation coefficients between autosomes and X chromosomes were observed in this study.

The numbers data points are not very large; thus, dividing too many bins by the distance from the telomeres may yield weak results. As more data become available, incorporating these additional predictors in the regression analyses may be beneficial.

5. References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. Basic local alignment search tool. *J Mol Biol* 215 (1990), pp. 403-10.

Bird, A.P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* 8 (1980), pp. 1499-504.

Cooper, D.N. and Krawczak, M. Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Hum Genet* 83 (1989), pp. 181-8.

Cooper, D.N. and Youssoufian, H. The CpG dinucleotide and human genetic disease. *Hum Genet* 78 (1988), pp. 151-5.

Felsenstein, J., Inferring phylogenies. Sinauer Associates Sunderland, Mass, USA (2004).

Fryxell, K.J. and Moon, W.J. CpG mutation rates in the human genome are highly dependent on local GC content. *Mol Biol Evol* 22 (2005), pp. 650-8.

Gibbs, R.A., Rogers, J., Katze, M.G., Bumgarner, R., Weinstock, G.M., Mardis, E.R., Remington, K.A., Strausberg, R.L., Venter, J.C., Wilson, R.K., Batzer, M.A., Bustamante, C.D., Eichler, E.E., Hahn, M.W., Hardison, R.C., Makova, K.D., Miller, W., Milosavljevic, A., Palermo, R.E., Siepel, A., Sikela, J.M., Attaway, T., Bell, S., Bernard, K.E., Buhay, C.J., Chandrabose, M.N., Dao, M., Davis, C., Delehaunty, K.D., Ding, Y., Dinh, H.H., Dugan-Rocha, S., Fulton, L.A., Gabisi, R.A., Garner, T.T., Godfrey, J., Hawes, A.C., Hernandez, J., Hines, S., Holder, M., Hume, J., Jhangiani, S.N., Joshi, V., Khan, Z.M., Kirkness, E.F., Cree, A., Fowler, R.G., Lee, S., Lewis, L.R., Li, Z., Liu, Y.S., Moore, S.M., Muzny, D., Nazareth, L.V., Ngo, D.N., Okwuonu, G.O., Pai, G., Parker, D., Paul, H.A., Pfannkoch, C., Pohl, C.S., Rogers, Y.H., Ruiz, S.J., Sabo, A., Santibanez, J., Schneider, B.W., Smith, S.M., Sodergren, E.,

Svatek, A.F., Utterback, T.R., Vattathil, S., Warren, W., White, C.S., Chinwalla, A.T., Feng, Y., Halpern, A.L., Hillier, L.W., Huang, X., Minx, P., Nelson, J.O., Pepin, K.H., Qin, X., Sutton, G.G., Venter, E., Walenz, B.P., Wallis, J.W., Worley, K.C., Yang, S.P., Jones, S.M., Marra, M.A., Rocchi, M., Schein, J.E., Baertsch, R., Clarke, L., Csuros, M., Glasscock, J., Harris, R.A., Havlak, P., Jackson, A.R., Jiang, H., et al. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316 (2007), pp. 222-34.

Haldane, J.B.S. The estimation of viabilities. *J. Genet.* 54 (1956), pp. 294-296.

Hardison, R.C., Roskin, K.M., Yang, S., Diekhans, M., Kent, W.J., Weber, R., Elnitski, L., Li, J., O'Connor, M., Kolbe, D., Schwartz, S., Furey, T.S., Whelan, S., Goldman, N., Smit, A., Miller, W., Chiaromonte, F. and Haussler, D. Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. *Genome Res* 13 (2003), pp. 13-26.

Hasegawa, M., Kishino, H. and Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22 (1985), pp. 160-74.

Hellmann, I., Prufer, K., Ji, H., Zody, M.C., Paabo, S. and Ptak, S.E. Why do human diversity levels vary at a megabase scale? *Genome Res* 15 (2005), pp. 1222-31.

Hobolth, A., Nielsen, R., Wang, Y., Wu, F. and Tanksley, S.D. CpG + CpNpG analysis of protein-coding sequences from tomato. *Mol Biol Evol* 23 (2006), pp. 1318-23.

Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. and Takahata, N. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci U S A* 92 (1995), pp. 532-6.

Hughes, A.L. and Nei, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335 (1988), pp. 167-70.

Huttley, G.A. Modeling the impact of DNA methylation on the evolution of BRCA1 in mammals. *Mol Biol Evol* 21 (2004), pp. 1760-8.

Jordan, I.K., Kondrashov, F.A., Adzhubei, I.A., Wolf, Y.I., Koonin, E.V., Kondrashov, A.S. and Sunyaev, S. A universal trend of amino acid gain and loss in protein evolution. *Nature* 433 (2005), pp. 633-8.

Jukes, T.H. and Cantor, T.H.: Evolution of protein molecules. In: Munro, H.N. (Munro, H.N.)Munro, H.N.s), *Mammalian Protein Metabolism*. Academic Press, New York (1969).

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30 (2002), pp. 3059-66.

Kimura, M. Evolutionary rate at the molecular level. *Nature* 217 (1968), pp. 624-6.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. Clustal W and Clustal X version 2.0. *Bioinformatics* 23 (2007), pp. 2947-8.

Li, W.H., Yi, S. and Makova, K. Male-driven evolution. *Curr Opin Genet Dev* 12 (2002), pp. 650-6.

Lunter, G. and Hein, J. A nucleotide substitution model with nearest-neighbour interactions. *Bioinformatics* 20 Suppl 1 (2004), pp. I216-I223.

Misawa, K. A codon substitution model that incorporates the effect of the GC contents, the gene density and the density of CpG islands of human chromosomes. *BMC Genomics* 12 (2011), p. 397.

Misawa, K., Kamatani, N. and Kikuno, R.F. The universal trend of amino acid gain-loss is caused by CpG hypermutability. *J Mol Evol* 67 (2008), pp. 334-42.

Misawa, K. and Kikuno, R.F. Evaluation of the effect of CpG hypermutability on human codon substitution. *Gene* 431 (2009), pp. 18-22.

Misawa, K. and Kikuno, R.F. GeneWaltz--A new method for reducing the false positives of gene finding. *BioData Min* 3 (2010), p. 6.

Misawa, K. and Kikuno, R.F. Relationship between amino acid composition and gene expression in the mouse genome. *BMC Res Notes* 4 (2011), p. 20.

Misawa, K. and Tajima, F. Estimation of the amount of DNA polymorphism when the neutral mutation rate varies among sites. *Genetics* 147 (1997), pp. 1959-64.

Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K. and Yasunaga, T. Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb Symp Quant Biol* 52 (1987), pp. 863-7.

Nei, M. and Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3 (1986), pp. 418-26.

Pray, L. DNA replication and causes of mutation. *Nature Education* 1 (2008), p. 1.

R Development Core Team, R: A language and environment for statistical computing, Vienna, Austria (2008).

Scarano, E., Iaccarino, M., Grippo, P. and Parisi, E. The heterogeneity of thymine methyl group origin in DNA pyrimidine isostichs of developing sea urchin embryos. *Proc Natl Acad Sci U S A* 57 (1967), pp. 1394-400.

Sullivan, J. and Joyce, P. Model selection in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 36 (2005), pp. 445-466.

Tavare, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* 17 (1986), pp. 57-86.

Taylor, J., Tyekucheva, S., Zody, M., Chiaromonte, F. and Makova, K.D. Strong and weak male mutation bias at different sites in the primate genomes: insights from the human-chimpanzee comparison. *Mol Biol Evol* 23 (2006), pp. 565-73.

Tyekucheva, S., Makova, K.D., Karro, J.E., Hardison, R.C., Miller, W. and Chiaromonte, F. Human-macaque comparisons illuminate variation in neutral substitution rates. *Genome Biol* 9 (2008), p. R76.

Vogel, F. and Motulsky, A.G., Human genetics: problems and approaches. Springer-Verlag, Berlin (1997).

Walser, J.C., Ponger, L. and Furano, A.V. CpG dinucleotides and the mutation rate of non-CpG DNA. *Genome Res* 18 (2008), pp. 1403-14.

Whelan, S., Lio, P. and Goldman, N. Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends Genet* 17 (2001), pp. 262-72.

Wu, S.C. and Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* 11 (2011), pp. 607-20.

Yang, Z. Estimating the pattern of nucleotide substitution. *J Mol Evol* 39 (1994), pp. 105-11.

Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24 (2007), pp. 1586-91.

Yang, Z. and Nielsen, R. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Mol Biol Evol* 25 (2008), pp. 568-79.
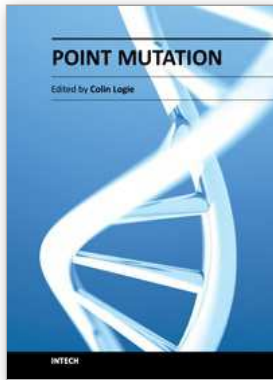
Point Mutation

Edited by Dr Colin Logie

This book concerns the signatures left behind in chromosomes by the forces that drive DNA code evolution in the form of DNA nucleotide substitutions. Since the genetic code predetermines the molecular basis of life, it could have been about any aspect of biology. As it happens, it is largely about recent adaptation of pathogens and their human host. Nine chapters are medically oriented, two are bioinformatics-oriented and one is technological, describing the state of the art in synthetic point mutagenesis. What stands out in this book is the increasing rate at which DNA data has been amassed in the course of the past decade and how knowledge in this vibrant research field is currently being translated in the medical world.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Kazuharu Misawa (2012). Estimating Human Point Mutation Rates from Codon Substitution Rates, Point Mutation, Dr Colin Logie (Ed.), ISBN: 978-953-51-0331-8, InTech, Available from: http://www.intechopen.com/books/point-mutation/estimating-human-point-mutation-rates-from-codon-substitution-rates-

INTECH
open science | open minds